The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'07)

ACHIEVABLE THROUGHPUT OF MIMO DOWNLINK BEAMFORMING WITH LIMITED CHANNEL INFORMATION

Giuseppe Caire
University of Southern California
Los Angeles, CA, USA

Nihar Jindal
University of Minnesota
Minneapolis, MN, USA

Mari Kobayashi
Supelec
Paris, France

Niranjay Ravindran
University of Minnesota
Minneapolis, MN, USA

ABSTRACT

We consider a MIMO fading broadcast channel and study the achievable throughput of zero-forcing downlink beamforming when the channel state information (CSI) available to the transmitter and/or receivers is imperfect. Each receiver (mobile) acquires imperfect CSI via downlink training pilots, and the transmitter acquires CSI through explicit feedback from each mobile. We analyze both analog and digital (i.e., quantized) channel feedback techniques. Our analysis quantifies the throughput degradation due to limited training, limited feedback resources (measured in feedback channel symbols rather than bits), and errors on the feedback channel. Using these results we are able to quantify scenarios in which digital feedback outperforms analog and also provide guidelines for the optimal allocation of resources to training and channel feedback.

I MODEL SETUP AND BACKGROUND

We consider a multi-input multi-output (MIMO) Gaussian broadcast channel modeling the downlink of a system where the base station (transmitter) has M antennas and K user terminals (receivers) have one antenna each. A channel use of such channel is described by

$$y_k = \mathbf{h}_k^{\mathsf{H}}\mathbf{x} + z_k, \quad k = 1, \dots, K \tag{1}$$

where y_k is the channel output at receiver k, $z_k \sim \mathcal{CN}(0, N_0)$ is the corresponding AWGN, $\mathbf{h}_k \in \mathbb{C}^M$ is the vector of channel coefficients from the k-th receiver to the transmitter antenna array and $\mathbf{x}$ is the channel input vector. The channel input is subject to the average power constraint $\mathbb{E}[|\mathbf{x}|^2] \leq P$.

We assume that the channel *state*, given by the collection of all channel vectors $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_K] \in \mathbb{C}^{M\times K}$, varies in time according to a block fading model where $\mathbf{H}$ is constant over each frame of length T channel uses, and evolves from frame to frame according to an ergodic stationary jointly Gaussian process; i.i.d. block-fading channel, where the entries of $\mathbf{H}$ are Gaussian i.i.d. with elements $\sim \mathcal{CN}(0, 1)$ is a special case of this.

A Capacity results

If $\mathbf{H}$ is perfectly and instantaneously known to all terminals (perfect CSIT and CSIR), the capacity region of the channel (1) is obtained by MMSE-DFE beamforming and Gaussian dirty-paper coding (see [1, 2] and references therein). Because of simplicity and robustness to non-perfect CSIT, simpler *linear precoding* schemes with standard Gaussian coding have been extensively considered. A particularly simple scheme consists of zero-forcing (ZF) beamforming, where the transmit signal is formed as $\mathbf{x} = \mathbf{V}\mathbf{u}$, such that $\mathbf{V} \in \mathbb{C}^{M\times K}$ is a zero-forcing beamforming matrix and $\mathbf{u} \in \mathbb{C}^K$ contains the symbols from K independently generated Gaussian codewords. For $K \leq M$, the k-th column $\mathbf{v}_k$ of $\mathbf{V}$ is chosen to be a unit vector orthogonal to the subspace $\mathcal{S}_k = \mathrm{span}\{\mathbf{h}_j : j \neq k\}$. In this case, the achievable sum rate is given by

$$R^{\mathrm{ZF}} = \max_{\sum_k \mathbb{E}[\mathcal{P}_k(\mathbf{H})] \leq P} \sum_{k=1}^{K} \mathbb{E}\left[\log\left(1 + \frac{|\mathbf{h}_k^{\mathsf{H}}\mathbf{v}_k|^2 \mathcal{P}_k(\mathbf{H})}{N_0}\right)\right]. \tag{2}$$

We consider the situation where $K = M$, and thus do not consider user selection. Furthermore, we are mainly interested in the high-spectral efficiency regime, where we can characterize the achievable sum rate as $\kappa \log P/N_0 + O(1)$, and κ is the "system multiplexing gain" or "pre-log factor" of the ergodic sum rate. Hence, it is well-known that using uniform power $\mathcal{P}_k = P/M$ for all $k = 1, \dots, M$, rather than performing optimal water-filling, incurs a loss only in the $O(1)$ term, and we shall restrict to this choice in the rest of this paper.

It is well-known that, under perfect CSIT and CSIR, both the optimal "Dirty-Paper" sum-rate C and the zero-forcing sum-rate R^{ZF} are equal to $M \log P/N_0 + O(1)$. On the contrary, under non-perfect CSIT the rate sum may behave in a radically different way; for example, if there is perfect CSIR and no CSIT when $\mathbf{H}$ has i.i.d. Gaussian entries, the sum rate is equal to $\log P/N_0 + O(1)$ [1]

B Channel state information model

We consider a model that includes the effect of downlink training, channel feedback, and secondary downlink training, as shown in Fig. 1. The initial training and feedback stages are standard, but note that an additional round of training is required because terminals do not know the channels of other terminals and thus are not aware of the chosen beamforming vectors.. Furthermore, we consider the explicit transmission of channel feedback symbols over an AWGN (unfaded) channel with SNR $\frac{P}{N_0}$.

1. **Initial Training**: In order to allow for channel estimation, $\beta_1 M$ shared pilots ($\beta_1 \geq 1$ symbols per antenna) are transmitted on the downlink. If the true channel state is $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_K]$, each receiver estimates its channel on the basis of the received signal s_k:

$$\mathbf{s}_k = \sqrt{\beta_1 P}\, \mathbf{h}_k + \mathbf{z}_k \tag{3}$$

where the entries of $\mathbf{z}_k$ are i.i.d. . $\sim \mathcal{CN}(0, N_0)$. The MMSE estimate ($\tilde{\mathbf{h}}_k$) is $\tilde{\mathbf{h}}_k = \frac{\sqrt{\beta_1 P}}{N_0 + \beta_1 P}\mathbf{s}_k$, and the true

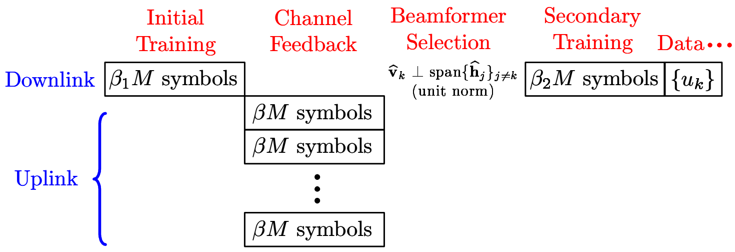

Figure 1: Training & Feedback Model

channel $\mathbf{h}_k$ can be written as:

$$\mathbf{h}_k = \tilde{\mathbf{h}}_k + \mathbf{n}_k \tag{4}$$

where the entries of $\mathbf{n}_k$ are i.i.d. $\sim \mathcal{CN}(0, \sigma_n^2)$ with $\sigma_n^2 = (1 + \beta_1 P/N_0)^{-1}$, and $\mathbf{n}_k$ and $\tilde{\mathbf{h}}_k$ are independent.

2. **Channel Feedback**: Each mobile feeds back the estimate $\tilde{\mathbf{h}}_k$ to the transmitter immediately after the initial training phase using βM feedback channel symbols (per mobile). A simplifying assumption we make is that we consider no fading and orthogonal access in the CSIT feedback link (uplink), and we assume that the SNR on the feedback channel is equivalent to the un-faded downlink SNR $\frac{P}{N_0}$. We use $\widehat{\mathbf{H}} = [\widehat{\mathbf{h}}_1, \ldots, \widehat{\mathbf{h}}_K] \in \mathbb{C}^{M \times K}$ to denote the (imperfect) channel information available at the transmitter corresponding to true channel state $\mathbf{H}$ (which is constant for each frame), and represent the feedback mechanism as a probabilistic mapping from the mobile estimate $\tilde{\mathbf{h}}_k$ to the transmitter estimate $\widehat{\mathbf{h}}_k$. The choice of feedback mechanism (i.e., digital or analog) determines the parameters of this mapping.

3. **Beamformer Selection**: The transmitter selects beamforming vectors using the zero-forcing criterion applied to its channel estimate $\widehat{\mathbf{H}}$. Following the ZF procedure, $\widehat{\mathbf{v}}_k$ is a unit vector orthogonal to the subspace $\mathcal{S}_k = \text{span}\{\widehat{\mathbf{h}}_j : j \neq k\}$, with $\widehat{\mathbf{V}} \triangleq [\widehat{\mathbf{v}}_1, \ldots, \widehat{\mathbf{v}}_K]$.

4. **Secondary Training**: An additional round of downlink training is performed to allow each terminal to estimate its useful signal coefficient denoted by $a_k = \mathbf{h}_\mathbf{k}^\mathbf{H} \hat{\mathbf{v}}_\mathbf{k}$, which then enables coherent detection. The training is accomplished in $\beta_2 M$ symbols by transmitting along each of the beamforming vectors for $\beta_2 \geq 1$ symbols on the downlink. MMSE estimation is performed on received signal $r_k = \sqrt{\beta_2 P}\, a_k + z_k$, which results in estimate $\hat{a}_k$ which satisfies:

$$a_k = \hat{a}_k + f_k, \tag{5}$$

where f_k and $\hat{a}_k$ are independent complex Gaussian's with variance $\sigma_f^2 = \frac{N_0}{N_0 + \beta_2 P}$ and $1 - \sigma_f^2$, respectively.

5. **Data Transmission**: The duration of the frame is dedicated to transmission of data symbols. If $\mathbf{u}$ is a vector of data symbols intended for the M terminals, then setting $\mathbf{x} = \widehat{\mathbf{V}}\mathbf{u}$ in (1) yields:

$$\begin{aligned} y_k &= (\mathbf{h}_k^\mathsf{H} \widehat{\mathbf{v}}_k)\, u_k + \sum_{j \neq k} (\mathbf{h}_k^\mathsf{H} \widehat{\mathbf{v}}_j)\, u_j + z_k & (6) \\ &= a_k u_k + I_k + z_k & (7) \end{aligned}$$

where $I_k = \sum_{j \neq k} (\mathbf{h}_k^\mathsf{H} \widehat{\mathbf{v}}_j) u_j$ is the interference at receiver k and $a_k = \mathbf{h}_k^\mathsf{H} \widehat{\mathbf{v}}_k$ is the useful signal coefficient. The effective output of the channel is the pair $(y_k, \hat{a}_k)$.

II Rate Gap Bound with CSIT Training and Feedback

We now provide a general lower bound on the throughput achievable in terms of the CSI model described in the previous section. This bound directly captures the effect of the training/feedback parameters β_1, β, β_2, and allows for different feedback mechanisms that are captured by the interference term. Although we do not have room for the proof, note that we use techniques similar to those in [3, 4]:

Theorem II.1 *The achievable rate for ZF beamforming with CSIT training and feedback can be bounded from below by:*

$$R_k(P) \geq \mathbb{E}\left[\log\left(1 + \frac{|\hat{a}_k|^2 P}{\sigma_f^2 P + N_0 M + \mathbb{E}\left[|I_k|^2 | \hat{a}_k\right] M}\right)\right],$$

where $R_k(P) = \sup_{u_k} I(u_k; y_k, \hat{a}_k)$ is the rate achieved with the optimal input distribution.

A useful measure of the performance is the difference between the rate achieved with perfect CSIT and the rate achieved with channel estimation/feedback. If we let $R_k^{\text{ZF}}(P)$ denote the per-user rate achieved with zero-forcing and uniform (in time and across users) power allocation $\mathcal{P}_k(\mathbf{H}) = \frac{P}{M}$ in (2), the rate gap defined as follows:

$$\Delta R(P) \triangleq R_k^{\text{ZF}}(P) - R_k(P). \tag{8}$$

We now provide a bound on the rate gap incurred under channel estimation and feedback with respect to ideal CSIT under ZF beamforming.

Theorem II.2 *The rate gap $\Delta R(P)$ is upper bounded by:*

$$\Delta R(P) \leq \log\left(1 + \sigma_f^2 \frac{P}{N_0 M} + \frac{\mathit{Var}\,[I_k]}{N_0}\right) \tag{9}$$

The proof is similar to the proof of Theorem 1 in [5] with the addition of a step to handle the $|\hat{a}_k|^2 P$ in the numerator of the achievable rate expression.

Interestingly, the bound (9) only depends on Var $[I_k]$ rather than on the conditional variance Var $[I_k | \hat{a}_k]$. In general, Var $[I_k]$ depends on the channel feedback method used as well as the errors introduced in the channel estimation phase (the initial training phase). In the following sections, we shall compute Var $[I_k]$ and particularize the above results for specific feedback schemes, i.e., specific schemes that map $\tilde{\mathbf{h}}_k$ to $\hat{\mathbf{h}}_k$.

A Analog CSIT feedback

If analog feedback is used, the (downlink) channel coefficients are explicitly transmitted on the feedback channel (modeled as AWGN with SNR $\frac{P}{N_0}$) by *each mobile* using unquantized quadrature-amplitude modulation [6–9]. Recall that each mobile receives $\mathbf{s}_k = \sqrt{\beta_1 P}\ \mathbf{h}_k + \mathbf{z}_k$ during the initial training phase. If each mobile transmits a scaled version of $\mathbf{s}_k$ during the channel feedback phase, the transmitter observation for a particular frame is given by (using (4)):

$$\begin{aligned} \mathbf{g}_k &= \frac{\sqrt{\beta P}}{\sqrt{\beta_1 P + N_0}}\,\mathbf{s}_k + \tilde{\mathbf{w}}_k && (10)\\ &= \frac{\sqrt{\beta\beta_1}P}{\sqrt{\beta_1 P + N_0}}\,\mathbf{h}_k + \frac{\sqrt{\beta P}}{\sqrt{\beta_1 P + N_0}}\,\mathbf{z}_k + \tilde{\mathbf{w}}_k && (11)\\ &= \frac{\sqrt{\beta\beta_1}P}{\sqrt{\beta_1 P + N_0}}\,\mathbf{h}_k + \mathbf{w}_k && (12) \end{aligned}$$

where $\tilde{\mathbf{w}}_k$ represents the AWGN noise on the uplink and (11) is obtained from (3). $\mathbf{w}_k$ and $\tilde{\mathbf{w}}_k$ are spatially and spectrally white Gaussian processes with components distributed as $\mathcal{CN}(0, \sigma_w^2)$, $\mathcal{CN}(0, N_0)$ respectively. From (10) and the expression for $\mathbf{s}_k$, σ_w^2 is computed to be $\frac{\beta P N_0}{\beta_1 P + N_0} + N_0$. The power scaling β corresponds to the number of channel uses per channel coefficient, assuming that transmission in the feedback channel has fixed peak power P and that the channel state vector is modulated by a $\beta M \times M$ unitary spreading matrix [6]. The transmitter then computes the MMSE estimate $\widehat{\mathbf{h}}_k$ and we can write

$$\mathbf{h}_k = \widehat{\mathbf{h}}_k + \mathbf{e}_k \qquad (13)$$

where $\widehat{\mathbf{h}}_k$ and $\mathbf{e}_k$ are mutually independent and the components of $\mathbf{e}_k$ are i.i.d. . $\sim \mathcal{CN}(0, \sigma_e^2)$ with $\sigma_e^2 = \frac{N_0^2 + N_0(\beta+\beta_1)P}{(N_0+\beta P)(N_0+\beta_1 P)}$. Using the representation in (13) and some other basic properties we are able to compute the variance of the interference term:

$$\begin{aligned} \mathrm{Var}\,[I_k] &= \sum_{j \neq k} \frac{P}{M}\,\mathbb{E}\left[|\mathbf{h}_k^{\mathsf{H}} \widehat{\mathbf{v}}_j|^2\right] \\ &= \sum_{j \neq k} \frac{P}{M}\,\mathbb{E}\left[|\mathbf{e}_k^{\mathsf{H}} \widehat{\mathbf{v}}_j|^2\right] \\ &= (M-1)\,\frac{P}{M}\,\sigma_e^2 \end{aligned}$$

By plugging this into (9) and simplifying σ_e^2 we have:

$$\begin{aligned} \Delta R^{\mathrm{A}}(P) &\le \log\left(1 + \sigma_f^2\,\frac{P}{N_0 M} + \frac{M-1}{M}\,\frac{P}{N_0}\,\sigma_e^2\right) \\ &\le \log\left(1 + \frac{1}{\beta_2 M} + \frac{M-1}{M}\left(\frac{1}{\beta} + \frac{1}{\beta_1}\right)\right). \end{aligned}$$

It is interesting to note that the rate gap is bounded, 0 implying that *multiplexing gain is preserved in spite of imperfect CSIR*.

B Rate gap bound for digital CSIT feedback

We now compute the rate gap bound when digital feedback is used, in which case the channel coefficients are quantized at each mobile and represented by B bits. The packet of B bits is fed back by each receiver through orthogonal feedback channels on the uplink. For the time being we consider error-free transmission of the feedback bits; later we relate feedback 'bits' to feedback channel "symbols", and we also incorporate the effect of errors on the feedback link.

The quantization codebook $\mathcal{C}$ used to quantize the channel coefficients at each terminal is assumed to be known at the transmitter and consists of 2^B unit-norm vectors in $\mathbb{C}^M$ i.e. $(\mathbf{p}_1, \ldots, \mathbf{p}_{2^B})$. The quantization $\widehat{\mathbf{h}}_k$ of $\tilde{\mathbf{h}}_k$ is selected from the $\mathcal{C}$ according to:

$$\widehat{\mathbf{h}}_k = \arg\max_{\mathbf{p} \in \mathcal{C}} \ \cos^2\left(\angle(\tilde{\mathbf{h}}_k, \mathbf{p})\right), \qquad (14)$$

where $\cos^2\left(\angle(\tilde{\mathbf{h}}_k, \mathbf{p})\right) \triangleq |\tilde{\mathbf{h}}_k^{\mathsf{H}}\mathbf{p}|^2/||\tilde{\mathbf{h}}_k||^2$. The index of the chosen vector is represented by B bits and fed back to the transmitter. Note that $\widehat{\mathbf{h}}_k$ is of unit-norm and hence no channel magnitude information is fed back in this model.

In [5], it is shown that if a *Random Vector Quantizer* (see [5] and references therein) is used, the relationship between the beamforming vectors and $\tilde{\mathbf{h}}_k$ satisfies:

$$\mathbb{E}\left[\cos^2\left(\angle\left(\tilde{\mathbf{h}}_k, \widehat{\mathbf{v}}_j\right)\right)\right] \le \frac{1}{M-1} 2^{-\frac{B}{M-1}}\ (j \neq k) \qquad (15)$$

We can use this to bound the variance of the interference term:

$$\begin{aligned} \mathrm{Var}\,[I_k] &= \sum_{j \neq k} \frac{P}{M}\,\mathbb{E}\left[|\mathbf{h}_k^{\mathsf{H}}\widehat{\mathbf{v}}_j|^2\right] \\ &\overset{(a)}{=} \sum_{j \neq k} \frac{P}{M}\left(\mathbb{E}\left[||\tilde{\mathbf{h}}_k||^2\right]\mathbb{E}\left[\frac{|\tilde{\mathbf{h}}_k^{\mathsf{H}}\widehat{\mathbf{v}}_j|^2}{||\tilde{\mathbf{h}}_k||^2}\right] + \mathbb{E}\left[|\mathbf{n}_k^{\mathsf{H}}\widehat{\mathbf{v}}_j|^2\right]\right) \\ &\overset{(b)}{\le} \frac{P}{M}\,\mathbb{E}\left[||\tilde{\mathbf{h}}_k||^2\right]\ 2^{-\frac{B}{M-1}} + \sum_{j \neq k}\frac{P}{M}\,\mathbb{E}\left[\widehat{\mathbf{v}}_j^{\mathsf{H}}\mathbb{E}[\mathbf{n}_k\mathbf{n}_k^{\mathsf{H}}]\widehat{\mathbf{v}}_j\right] \\ &\overset{(c)}{=} \frac{\beta_1 P^2}{N_0 + \beta_1 P}\,2^{-\frac{B}{M-1}} + (M-1)\,\frac{P}{M}\,\sigma_n^2 \\ &\le P 2^{-\frac{B}{M-1}} + (M-1)\,\frac{P}{M}\,\sigma_n^2 \end{aligned}$$

where (a) is obtained from the representation (4), (b) from (15) and (c) by computing the expected norm of $\tilde{\mathbf{h}}_k = \frac{\sqrt{\beta_1 P}}{N_0 + \beta_1 P}\mathbf{s}_k$ using $\mathbf{s}_k = \sqrt{\beta_1 P}\ \mathbf{h}_k + \mathbf{z}_k$. Plugging into (9), we have:

$$\begin{aligned} \Delta R^{\mathrm{D}}(P) &\le \log\left(1 + \sigma_f^2\,\frac{P}{N_0 M} + P 2^{-\frac{B}{M-1}} + \frac{M-1}{M}\,\frac{P}{N_0}\,\sigma_n^2\right) \\ &\le \log\left(1 + \frac{1}{M\beta_2} + \frac{P}{N_0}2^{-\frac{B}{M-1}} + \frac{M-1}{M}\,\frac{1}{\beta_1}\right) \end{aligned}$$

If B is scaled linearly with $\log_2(P)$ as $B = \alpha(M-1)\log_2(P)$ then we have:

$$\Delta R^{\mathrm{D}}(P) \le \log\left(1 + \frac{1}{M\beta_2} + \left(\frac{P}{N_0}\right)^{1-\alpha} + \frac{M-1}{M}\,\frac{1}{\beta_1}\right).$$

If $\alpha < 1$ this bound goes to infinity and indeed full multiplexing gain is not achieved [5, Theorem 4]. If $\alpha >= 1$ this quantity is bounded and full multiplexing is achieved. Furthermore, note that if $\alpha > 1$ we have $P^{1-\alpha} \to 0$ which implies that the effect of limited rate feedback vanishes at high SNR.

III Comparison between analog and digital CSIT feedback

In this section we compare analog and digital feedback under the assumption that it is possible to transmit digital feedback error-free at rate equal to the capacity of the feedback channel.

A Perfect CSIR

For the sake of simplicity, we first restrict ourselves to the case of perfect CSIR, i.e., $\beta_1, \beta_2 \to \infty$. From (14) and using $\frac{M-1}{M} \leq 1$ we have:

$$\Delta R^{\mathrm{A}}_{\mathrm{CSIR}}(P) \leq \log\left(1 + \frac{1}{\beta}\right), \tag{16}$$

which corresponds to βM channel symbols of feedback per mobile during the feedback phase.

Let us now consider digital feedback for the same situation. Note that analog feedback has the additional advantage of providing a noisy version of the channel norm information (which, although irrelevant for ZF beamforming, could be made use of by user selection algorithms), while digital feedback provides no norm information. Thus, for fair comparison, we assume that βM feedback symbols in the analog feedback scheme corresponds to $\beta(M-1)$ feedback symbols for the digital feedback scheme. Let us assume (very unrealistically) that the digital feedback link can operate error-free at capacity, i.e., it can reliably transmit $\log(1 + P/N_0)$ bits per feedback symbol. Thus, the number of feedback bits per mobile is $B = \beta(M-1)\log_2(1 + P/N_0)$. Replacing this into the rate gap bound (16), we obtain:

$$\Delta R^{\mathrm{D}}_{\mathrm{CSIR}} \leq \log\left(1 + \frac{P/N_0}{(1+P/N_0)^\beta}\right). \tag{17}$$

If $\beta = 1$ digital and analog feedback achieve essentially the same rate gap of at most 1 b/s/Hz. However, if $\beta > 1$, unlike the analog feedback case, the rate gap of the quantized feedback vanishes for $P/N_0 \to \infty$. For example, for $\beta = 2$ the rate gap is upperbounded by $\log(1 + \frac{N_0}{P})$, which quite small for even moderate values of P/N_0 (e.g., $P/N_0 = 10$ dB gives 0.13 b/s/Hz). We conclude that digital is far superior to analog for $\beta > 1$.

This conclusion finds an appealing interpretation in the context of rate-distortion theory. It is well-known (see [10] and references therein) that analog transmission (which consists of scaling the source symbols and sending them uncoded and unquantized through the channel) is an optimal strategy to send a Gaussian source over a Gaussian channel with minimal end-to-end quadratic distortion. In our case, the source is the Gaussian channel vector $\mathbf{h}_k$ and the noisy channel is the feedback AWGN channel with SNR P/N_0. Hence, the fact that analog feedback cannot be essentially outperformed for $\beta = 1$ is expected. However, it is also well-known that if the channel rate is larger than the source rate (i.e., less than one Gaussian source symbol arrives per channel symbol, which corresponds to $\beta > 1$ in our case), then analog is strictly suboptimal as compared to separate source and channel coding because the distortion with analog transmission scales as $1/\beta$ whereas it decreases exponentially with β (i.e., along the vector quantizer R-D curve) for digital transmission.

B Imperfect CSIR

We now compare analog and digital feedback under the assumption of imperfect CSIR. From (14) for analog feedback (again using $\frac{M-1}{M} \leq 1$) and (16) with $B = \beta(M-1)\log_2(1 + P/N_0)$ we have:

$$\begin{aligned} \Delta R^{\mathrm{A}}(P) &\leq \log\left(1 + \frac{1}{M\beta_2} + \frac{1}{\beta} + \frac{1}{\beta_1}\right) \\ \Delta R^{\mathrm{D}}(P) &\leq \log\left(1 + \frac{1}{M\beta_2} + \frac{P/N_0}{(1+P/N_0)^\beta} + \frac{1}{\beta_1}\right) \end{aligned}$$

The $\frac{1}{\beta_1}$ term reflects the effect of initial training error ($\beta_1 M$ downlink pilots) on the CSI available to the transmitter: this is due to the fact that the CSIT is a corrupted version of the CSIR obtained during initial training (phase 1). The leading $\frac{1}{M\beta_2}$ term is due to the inaccurate signal coefficient estimate during secondary training, i.e., the fact that the channel is not quite coherent. Finally, the middle term in both expressions is the further CSIT imperfection incurred during the feedback process.

Comparing the equations we come to the same general conclusions as in Section A: if $\beta = 1$ then digital and analog are roughly equivalent, but if $\beta > 1$ digital is superior to analog because the effect of feedback noise vanishes at high SNR for digital but not for analog.

However, note that there are some important differences with the perfect CSIR scenario. Firstly, imperfect CSIR leads to residual interference that does not vanish with SNR. As a result, the rate gap is not driven to 0 even when $\beta > 1$ for digital feedback when β_1 and β_2 are fixed. For instance, even when there is perfect feedback (i.e., $\beta = \infty$) with imperfect CSIR (β_1, β_2 finite), we have the following bound which applies to either analog or digital feedback

$$\Delta R^{\text{Perf. FB}}_{\mathrm{CSIR}}(P) \leq \log\left(1 + \frac{1}{\beta_2 M} + \frac{1}{\beta_1}\right). \tag{18}$$

Although the feedback channel is perfect, the CSIT is still imperfect because it is based upon the imperfect terminal channel estimate. Another effect of imperfect CSIR is that the advantage of digital feedback relative to analog feedback is somewhat reduced because of the residual rate loss (suffered by either analog or digital) due to the imperfect CSIR.

Based on the above rate gap equations, we can come up with some rough guidelines about allocation of resources for training and feedback. For either digital or analog feedback, secondary training is clearly less important than initial training and therefore it makes sense to choose $\beta_1 > \beta_2$. Furthermore, for digital feedback it is sufficient (at high SNR) to choose β slightly larger than one and to put the remaining resources into training. For analog, on the other hand, initial training and feedback have essentially the same effect, although note that the feedback stage requires βM symbols *per mobile*.

IV Effects of CSIT feedback errors

We now investigate the impact of removing the optimistic assumption that the quantized feedback channel can operate error-free at capacity. To concentrate on the effect of feedback channel errors, we assume perfect CSIR, by which we mean the SINR (after selection of the beamforming vectors) is known perfectly at each terminal.

We consider a very simple CSIT feedback scheme that certainly represents a lower bound on the best quantized feedback strategy. The user terminals perform quantization using RVQ and transmit the feedback bits using simple uncoded QAM. No intelligent mapping of the quantization bits onto the QAM symbols is used, and therefore even a single erroneous feedback bit from user k results in CSIT that is completely independent (due to the properties of RVQ) of the actual k-th channel vector. Since uncoded QAM is used, error detection is not possible and the base station computes beamforming vectors based on the possibly erroneous feedback.

We again use βM symbol periods to transmit the feedback bits. There is a non-trivial tradeoff between quantization and channel errors. In order to maintain a bounded gap, feedback must be scaled at least as $(M-1)\log_2(1+P/N_0) \approx M\log_2 P/N_0$. Therefore, we consider sending $B = \alpha M \log_2 P/N_0$ for $1 \leq \alpha \leq \beta$ bits in βM symbol periods, which corresponds to $\frac{\alpha}{\beta}\log_2(P/N_0)$ bits per QAM symbol.

From [11], using the fact that the QAM constellation size is equal to $L = (P/N_0)^{\frac{\alpha}{\beta}}$, we have the following upper bound to the symbol error probability for QAM modulation:

$$P_s \quad \leq \quad 2\exp\left(-\frac{3}{2}\left(\frac{P}{N_0}\right)^{1-\alpha/\beta}\right) \tag{19}$$

For $\alpha = \beta$ (which means trying to signal at capacity with uncoded modulation!) P_s does not decreases with SNR and the system performance is very poor. However, for $\alpha/\beta < 1$, which corresponds to transmitting at a constant fraction of capacity, $P_s \to 0$ as $P/N_0 \to \infty$. The upper bound on the error probability of the whole quantized vector (transmitted in βM symbols) is given by $P_{e,fb} = 1-(1-P_s)^{\beta M}$. A lower bound on the achievable ergodic rate is obtained by assuming that when a feedback error occurs for user k its SINR is zero while if no feedback error occurs its rate is given $R_k^{\rm ZF} - \Delta R_{\rm quant.}$, that is, the rate of ideal ZF decreased by the (upper bound to) the rate gap. It follows that the ergodic rate of user k is upper-bounded by

$$R_k \geq (1-P_s)^{\beta M}\left(R_k^{\rm ZF} - \log\left(1+(P/N_0)^{1-\alpha}\right)\right) \tag{20}$$

Choosing $1 < \alpha < \beta$ we achieve both vanishing P_s and vanishing $\Delta R_{\rm quant.}$ as $P/N_0 \to \infty$. Thus, even under this very simple CSIT feedback scheme the optimal ZF performance can be eventually approached for sufficiently high SNR.

It should be noted that the assumption of perfect SINR at each mobile is important in this analysis of the effect of feedback errors. It is not sufficient to only know the useful signal coefficient a_k, but it necessary to know the SINR which incorporates the interference terms as well. By learning the SINR, the terminal implicitly learns whether a feedback error occurred or not because the SINR is likely to be extremely low whenever an error occurs; this fact is what makes it possible to consider the rate conditioned on no feedback error as in the above expression. Without SINR information, feedback errors could lead to considerably more degradation because each terminal would not be able to determine when feedback errors have occurred.

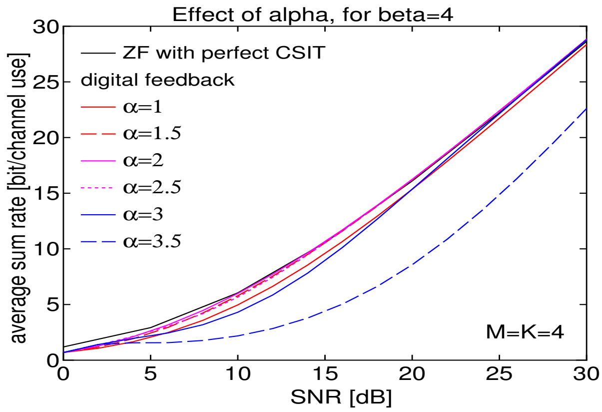

Figure 2: Quantized feedback with QAM modulation.

Fig. 2 shows the ergodic rate achieved by ZF beamforming with quantized CSIT and QAM feedback transmission for $M = K = 4$, independent Rayleigh fading, $\beta = 4$ and different values of α. It is noticed that by proper design of the feedback parameters the performance can be made very close to the ideal CSIT case.

References

[1] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel", *IEEE Trans. on Inform. Theory*, vol. 49, no. 7, 2003.

[2] H. Weingarten and Y. Steinberg and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel", *ISIT. Proceedings.*, 2004.

[3] M. Medard, "Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel", *IEEE Trans. on Inform. Theory*, vol. 46, no. 3, 2000.

[4] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?", *IEEE Trans. on Inform. Theory*, vol. 49, no. 4, 2003.

[5] N. Jindal, "MIMO broadcast channels with finite rate feedback", *IEEE Trans. on Inform. Theory*, vol. 52, no. 11, 2006.

[6] T. L. Marzetta and B. M. Hochwald, "Fast Transfer of Channel State Information in Wireless Systems", *IEEE Trans. on Signal Proc.*, vol. 54, no. 4, 2006.

[7] M. Kobayashi and G. Caire, "Joint Beamforming and Scheduling for a Multi-Antenna Downlink with Imperfect Transmitter Channel Knowledge", *to appear in IEEE J. Select. Areas Commun.*, 2007.

[8] T. A. Thomas and K. L. Baum and P. Sartori, "Obtaining channel knowledge for closed-loop multi-stream broadband MIMO-OFDM communications using direct channel feedback", *IEEE GLOBECOM*, vol. 6, 2005.

[9] D. Samardzija and N. Mandayam, "Unquantized and Uncoded Channel State Information Feedback on Wireless Channels", *IEEE WCNC. Proceedings.*, 2005.

[10] M. Gastpar and B. Rimoldi and M. Vetterli, "To Code, or not to code: Lossy source-channel communication revisited", *IEEE Trans. on Inform. Theory*, vol. 49, 2003.

[11] A. Goldsmith, "Wireless Communications", *Cambridge University Press*, 2005.