Multiuser Diversity with Limited Feedback

Mario H. Castañeda and Josef A. Nossek
Institute for Circuit Theory and Signal Processing
Technische Universität München
80290 Munich, Germany
Email: {maca,jono}@nws.ei.tum.de

***Abstract*—Multiple antennas at the base station can be employed to serve multiple single-antenna users in the downlink of a system. When the number of active users K is larger than the number of transmit antennas M, the base station can exploit multiuser diversity by cleverly scheduling a set of M users. For this purpose, the base station needs to know the downlink channel state information of all the users in the system. In a *frequency division duplex* (FDD) system, this is attained with limited feedback from the users. Albeit most of the literature assumes a per-user limited feedback, a *system* limited feedback is actually more appropriate. Given a constraint on the total amount of *airtime* available for feedback, in this work we are interested in determining the optimum number of users that should feed back. We observe a tradeoff between the attainable degree of multiuser diversity and the feedback quality. The users are scheduled based on minimizing the sum of the *mean square error* (MSE) for the selected set of users.**

I. Introduction

Consider the downlink of a system with M transmit antennas at the base station and K single-antenna active users. In such a scenario, the M antennas at the base station can be employed to simultaneously transmit to M users. When $K > M$, the base station can perform user selection, i.e., *channel-aware* scheduling, in order to cleverly select a set of M users to serve out of all the K users. For this purpose, the base station needs to know the downlink *channel state information* (CSI) of all the K users. In an FDD system, however, the transmit CSI for the downlink is obtained through limited feedback in the uplink [1].

In the literature, numerous works dealing on limited feedback for multiuser systems with $K \leq M$ assume a limited feedback of B bits per user. With a limited feedback of B bits per user, each of the K users can feed back a quantized version of its *channel direction information* (CDI), i.e. of its normalized *estimated* channel vector, to the base station. Subsequently, the base station can perform zero-forcing beamforming [2], [3] or MMSE beamforming [4] based on the fed back CDI of the K users.

Nonetheless, the previous works focus on the case $K \leq M$. i.e., no user selection is required or no channel-aware scheduling is involved. User selection in order to maximize the sum rate with zero-forcing beamforming based on limited feedback is considered in [5], [6], where $K > M$ users must report their *channel quality indicator* (CQI) in addition to their quantized CDI. The CQI includes information about the *channel magnitude information* (CMI) and/or about the quantization error, which aids the user selection performed at the base station. However, in [5], [6] it is assumed that the base station has access to an unquantized version of the CQI, i.e., the B feedback bits are used solely to quantize the CDI.

The prior works assume that the overall required feedback load of the system increases linearly with the number of users, i.e., the total number of feedback bits for K users is KB bits. Albeit most of the current work considers a per-user limited feedback, an overall *system* limited feedback is actually more appropriate. This issue has recently been addressed in [7], where given a total number of feedback bits, Ravindran et. al discuss whether it is more benefecial to acquire coarsely quantized channel feedback from many users or instead to obtain more accurate channel feedback from a small number of users in order to maximize the sum rate under zero-forcing beamforming. This leads to a tradeoff between the attainable degree of multiuser diversity and the users' feedback quality.

In the work at hand, we investigate a system limited feedback load, but not by considering a constraint on the total number of feedback *bits* as in [7]. Instead we consider a constraint on the total amount of *airtime* T_{F} available for feedback, i.e., given a number of T_{F} time instances, we are interested in determining the optimum number of users K_{opt} that optimize the system's performance. In our case we will consider the sum MSE of the selected users as the figure of merit. For a given number of users, we assume the users are scheduled based on minimizing the sum MSE for the selected set of users. As in [3], [4], we assume the feedback bits are sent using QPSK symbols in the uplink. In addition, we make use of an available degree of freedom in the feedback link: the M *receive* antennas at the base station in the uplink, i.e. the feedback link, can be used to detect the feedback of at most M users simultaneously with *receive* beamforming. Summarizing, in contrast to [7] we consider that:

- the system feedback load consists of a total amount of airtime T_{F}, where at each feedback instance, the base station can receive the feedback of at most M users,
- the feedback of the users consists of solely the CDI and
- the users are scheduled based on minimizing the sum MSE for the selected set of users.

To this end, this paper is orgnaized as follows. The system model is presented in Section II, while Section III discusses the user selection. How the users relay their feedback to the base station given a system feedback load T_{F} is presented in Section IV. In Section V, the numerical results are shown. The paper is concluded with Section VI.

II. System Model

Consider an FDD downlink in a single-cell with M antennas at the base station and $K > M$ single-antenna users, i.e., a *multiuser multiple-input single-output* (MU-MISO) system. The available transmit power at the BS is taken to be P_{DL}. The signals to be transmitted to the users are assumed to be independent zero-mean with variance σ_s^2. The MISO downlink channel of user k is denoted as $\boldsymbol{h}_k \in \mathbb{C}^M$ where the elements of $\boldsymbol{h}_k \; \forall\, k$ are i.i.d. zero-mean unit-variance complex Gaussian random variables. We assume that during the transmission the users experience independent zero-mean AWGN at their receivers with variance σ_n^2.

We consider a packet length of T time instances, where the channels of the K users are assumed to be constant. The transmit CSI for MMSE beamforming in the downlink is obtained at the base station as follows. First, each of the K users estimate their downlink channel $\boldsymbol{h}_k$ with a *common* downlink pilot of length T_{DL} time instances (symbols) as shown in Figure 1. To this end, we assume MMSE channel estimation and that $T_{\text{DL}} \geq M$ symbols, in order to have a meaningful estimate. We denote the estimated channel of user k as $\hat{\boldsymbol{h}}_k$ $\forall k$, which is a zero-mean complex Gaussian random variable with variance $1 - \sigma^2_{e_{\text{DL}}}$, where $\sigma^2_{e_{\text{DL}}}$ is given by [9]

$$\sigma^2_{e_{\text{DL}}} = \frac{1}{1 + \frac{P_{\text{DL}}}{M\sigma_n^2} T_{\text{DL}}}, \tag{1}$$

which under the given assumptions is the same for all users. Each user k afterwards normalizes its estimated channel to obtain $\hat{\boldsymbol{h}}_{\text{n},k}$ which is an estimate of its CDI. The estimated CDI is thereupon quantized with B bits by employing *random vector quantization* (RVQ) [8] with a codebook consisting of 2^B unit-norm codewords. The quantized CDI of user k is denoted as $\hat{\boldsymbol{h}}_{\text{q},k}$ and is obtained based on the minimum chordal distance as follows

$$\hat{\boldsymbol{h}}_{\text{q},k} = \underset{\boldsymbol{t}_{k,j} \in \mathcal{C}_k}{\text{argmax}} \; |\boldsymbol{t}_{k,j}^{\text{H}} \hat{\boldsymbol{h}}_{\text{n},k}|. \tag{2}$$

Note that each user k has a different codebook $\mathcal{C}_k$ consisting of 2^B unit-norm random beamforming vectors $\boldsymbol{t}_{k,j} \in \mathcal{C}_k$, which is also available at the BS. We assume a different codebook for each user, since otherwise there exists a non-zero probability that two or more users feed back the same channel vector.

Subsequently each user k feeds back B bits corresponding to the index of $\hat{\boldsymbol{h}}_{\text{q},k}$ in its codebook, such that the base station knows all the K users' quantized CDI. The feedback of the B bits per user for all the K users takes places during the system limited feedback, which consists of T_{F} time instances in the uplink as shown in Figure 1. We will discuss later in Section IV how the feedback of the K users takes place over the T_{F} time instances. In addition, we will consider no feedback errors as in [7].

As a figure of merit of the system we consider the sum MSE of the scheduled users. We will assume that the base station always serves M users out of the $K > M$ active users during the remaining $T - T_{\text{DL}}$ time instances of a packet as shown in Figure 1. Let us define the set of M scheduled users as $\mathcal{S}$ and let us denote the users in the set as follows

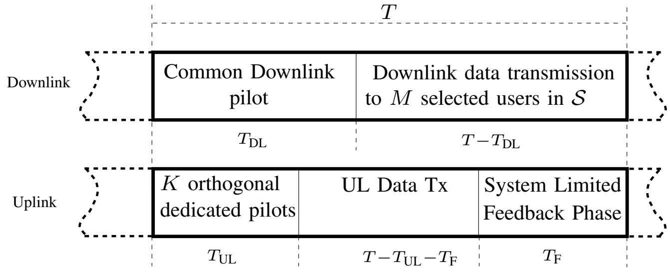

Fig. 1. Resource Allocation

$$\mathcal{S} = \{\pi(1), \pi(2), \cdots, \pi(M)\},$$

where $\pi(m) \in \{1, 2, \cdots, K\}$ for $m = 1, \ldots, M$. We can collect the CDI of the M users in this set in the following matrix

$$\hat{\boldsymbol{H}}_{\text{q}} = \left[\hat{\boldsymbol{h}}_{\text{q},\pi(1)}, \hat{\boldsymbol{h}}_{\text{q},\pi(2)}, \cdots, \hat{\boldsymbol{h}}_{\text{q},\pi(M)}\right]^{\text{T}} \in \mathbb{C}^{M \times M}. \tag{3}$$

Let us assume the M users in the set $\mathcal{S}$ are served by the base station with MMSE beamforming based on the M selected users' CDI, i.e. based on $\hat{\boldsymbol{H}}_{\text{q}}$. To this end, let us collect the transmit signals of the M users in the set $\mathcal{S}$ in the vector $\boldsymbol{s} \in \mathbb{C}^M$. In addition, we denote the channel matrix $\boldsymbol{H}$ of the M users in the set $\mathcal{S}$ as

$$\boldsymbol{H} = \left[\boldsymbol{h}_{\pi(1)}, \boldsymbol{h}_{\pi(2)}, \ldots, \boldsymbol{h}_{\pi(M)}\right]^{\text{T}} \in \mathbb{C}^{M \times M}. \tag{4}$$

After the estimation (a), normalization (b) and quantization (c), we can rewrite $\boldsymbol{H}$ as follows [4]

$$\begin{aligned} \boldsymbol{H} &\overset{(a)}{=} \hat{\boldsymbol{H}} + \boldsymbol{E} \\ &\overset{(b)}{=} \hat{\boldsymbol{B}} \hat{\boldsymbol{H}}_{\text{n}} + \boldsymbol{E} \\ &\overset{(c)}{=} \hat{\boldsymbol{B}} \left(\boldsymbol{C} \hat{\boldsymbol{H}}_{\text{q}} + \boldsymbol{E}_q\right) + \boldsymbol{E} \\ &= \hat{\boldsymbol{B}} \boldsymbol{C} \hat{\boldsymbol{H}}_{\text{q}} + \hat{\boldsymbol{B}} \boldsymbol{E}_q + \boldsymbol{E}, \end{aligned} \tag{5}$$

where $\hat{\boldsymbol{H}}_{\text{q}}$ is given in (3) and additionally

$$\begin{aligned} \hat{\boldsymbol{H}} &= \left[\hat{\boldsymbol{h}}_{\pi(1)}, \hat{\boldsymbol{h}}_{\pi(2)}, \ldots, \hat{\boldsymbol{h}}_{\pi(M)}\right]^{\text{H}} \in \mathbb{C}^{M \times M} \\ \hat{\boldsymbol{H}}_{\text{n}} &= \left[\hat{\boldsymbol{h}}_{\text{n},\pi(1)}, \hat{\boldsymbol{h}}_{\text{n},\pi(2)}, \ldots, \hat{\boldsymbol{h}}_{\text{n},\pi(M)}\right]^{\text{H}} \in \mathbb{C}^{M \times M} \\ \boldsymbol{E} &= \left[\boldsymbol{e}_{\pi(1)}, \boldsymbol{e}_{\pi(2)}, \cdots, \boldsymbol{e}_{\pi(M)}\right]^{\text{T}} \in \mathbb{C}^{M \times M} \\ \boldsymbol{E}_{\text{q}} &= \left[\boldsymbol{e}_{\text{q},\pi(1)}, \boldsymbol{e}_{\text{q},\pi(2)}, \cdots, \boldsymbol{e}_{\text{q},\pi(M)}\right]^{\text{T}} \in \mathbb{C}^{M \times M} \\ \hat{\boldsymbol{B}} &= \text{diag}\left(\left[\|\hat{\boldsymbol{h}}_{\pi(1)}\|_2, \cdots, \|\hat{\boldsymbol{h}}_{\pi(M)}\|_2\right]\right) \in \mathbb{R}_+^{M \times M} \\ \boldsymbol{C} &= \text{diag}\left(\left[c_{\pi(1)}, c_{\pi(2)}, \cdots, c_{\pi(M)}\right]\right) \in \mathbb{C}^{M \times M}, \end{aligned}$$

where for user $k \in \mathcal{S}$, $\boldsymbol{e}_k$ is the estimation error and $\boldsymbol{e}_{\text{q},k}$ is the quantization error, such that $\hat{\boldsymbol{h}}_{\text{q},k}$ and $\boldsymbol{e}_{\text{q},k}$ are orthogonal. In addition, θ_k is the angle between the normalized channel vector $\hat{\boldsymbol{h}}_{\text{n},k}$ and the quantized channel vector $\hat{\boldsymbol{h}}_{\text{q},k}$. The angle

θ_k has the same distribution among all the users $k = 1, \ldots, K$. We also have that $c_k = \hat{\boldsymbol{h}}_{\mathrm{q},k}^{\mathrm{H}} \hat{\boldsymbol{h}}_{\mathrm{n},k} \in \mathbb{C}$ with $|c_k| \leq 1$.

In order to match the channel observed at the BS, each user $k \in \mathcal{S}$ multiplies its received signal with $\frac{1}{c_k \|\hat{\boldsymbol{h}}_k\|_2}$ besides the scalar filter g_{q}. Multiplying with $\frac{1}{c_k}$ removes the phase uncertainty introduced by the quantization. The estimated signal of the M users in $\mathcal{S}$ is given by

$$\begin{aligned} \hat{\boldsymbol{s}}_{\mathrm{q}} &= g_{\mathrm{q}}\, \boldsymbol{C}^{-1} \hat{\boldsymbol{B}}^{-1} \left(\boldsymbol{H} \boldsymbol{P}_{\mathrm{q}} \boldsymbol{s} + \boldsymbol{n} \right) \\ &= g_{\mathrm{q}} \left(\hat{\boldsymbol{H}}_{\mathrm{q}} \boldsymbol{P}_{\mathrm{q}} \boldsymbol{s} + \boldsymbol{n}_{\mathrm{q}} \right), \end{aligned} \tag{6}$$

where in the second step we have used (5) and we have the effective noise

$$\boldsymbol{n}_{\mathrm{q}} = \boldsymbol{C}^{-1} \boldsymbol{E}_q \boldsymbol{P}_{\mathrm{q}} \boldsymbol{s} + \boldsymbol{C}^{-1} \hat{\boldsymbol{B}}^{-1} \boldsymbol{E} \boldsymbol{P}_{\mathrm{q}} \boldsymbol{s} + \boldsymbol{C}^{-1} \hat{\boldsymbol{B}}^{-1} \boldsymbol{n}. \tag{7}$$

The scalar receiver at user k, for $k \in \mathcal{S}$, is actually $\frac{g_{\mathrm{q}}}{c_k \|\hat{\boldsymbol{h}}_k\|_2}$ which is based not only on the estimated CDI but also on the estimated CMI and c_k.

We are now interested in computing $\boldsymbol{P}_{\mathrm{q}}$ and g_{q} which minimize the MSE based on $\hat{\boldsymbol{H}}_{\mathrm{q}}$ and on the statistics of the estimated CMI, i.e. $\hat{\boldsymbol{B}}$, which is the information assumed to be available at the BS. Recall that we have assumed that the statistics of the users' CMI to be i.i.d. The optimizaton problem reads as

$$\{\boldsymbol{P}_{\mathrm{q}}, g_{\mathrm{q}}\} = \underset{\{\boldsymbol{P}_{\mathrm{q}}, g_{\mathrm{q}}\}}{\operatorname{argmin}}\ \mathrm{E}\left[\|\boldsymbol{s} - \hat{\boldsymbol{s}}_{\mathrm{q}}\|_2^2 \right] \ \text{ s.t. } \sigma_{\mathrm{s}}^2 \operatorname{tr}\left(\boldsymbol{P}_{\mathrm{q}} \boldsymbol{P}_{\mathrm{q}}^{\mathrm{H}} \right) \leq P_{\mathrm{DL}},$$

where the expected value is taken over $\boldsymbol{s}$, $\boldsymbol{n}$, $\boldsymbol{E}$, $\hat{\boldsymbol{B}}$, $\boldsymbol{C}$ and $\boldsymbol{E}_{\mathrm{q}}$ since only $\hat{\boldsymbol{H}}_{\mathrm{q}}$ is known at the BS. As shown in [4], g_{q} and $\boldsymbol{P}_{\mathrm{q}}$ read as:

$$g_{\mathrm{q}} = \sqrt{\frac{\operatorname{tr}\left(\left((1-\kappa) \hat{\boldsymbol{H}}_{\mathrm{q}}^{\mathrm{H}} \hat{\boldsymbol{H}}_{\mathrm{q}} + \xi_{\mathrm{q}} \mathbf{1}_M \right)^{-2} \sigma_{\mathrm{s}}^2 \hat{\boldsymbol{H}}_{\mathrm{q}}^{\mathrm{H}} \hat{\boldsymbol{H}}_{\mathrm{q}} \right)}{P_{\mathrm{DL}}}} \tag{8}$$

$$\boldsymbol{P}_{\mathrm{q}} = \frac{1}{g_{\mathrm{q}}} \left((1-\kappa) \hat{\boldsymbol{H}}_{\mathrm{q}}^{\mathrm{H}} \hat{\boldsymbol{H}}_{\mathrm{q}} + \xi_{\mathrm{q}} \mathbf{1}_M \right)^{-1} \hat{\boldsymbol{H}}_{\mathrm{q}}^{\mathrm{H}}, \tag{9}$$

where

$$\kappa = \frac{\mathrm{E}\left[\tan^2 \theta_k \right]}{M-1} \tag{10}$$

$$\xi_{\mathrm{q}} = K\kappa + \frac{K\, \mathrm{E}\left[\cos^{-2} \theta_k \right]}{(M-1)(1-\sigma_{e_{\mathrm{DL}}}^2)} \left(\sigma_{e_{\mathrm{DL}}}^2 + \frac{\sigma_n^2}{P_{\mathrm{DL}}} \right), \tag{11}$$

where we have $\mathrm{E}\left[\tan^2 \theta_k \right] \approx \frac{\mathrm{E}\left[\sin^2 \theta_k \right]}{1 - \mathrm{E}\left[\sin^2 \theta_k \right]}$ and $\mathrm{E}\left[\sin^2 \theta_k \right]$ given as [2]

$$\mathrm{E}\left[\sin^2 \theta_k \right] = 2^B \,\mathrm{Beta}\left(2^B, \frac{M}{M-1} \right). \tag{12}$$

In addition we can also approximate $\mathrm{E}\left[\cos^{-2} \theta_k \right] \approx \frac{1}{\mathrm{E}\left[\cos^2 \theta_k \right]}$, where $\mathrm{E}\left[\cos^2 \theta_k \right] = 1 - \mathrm{E}\left[\sin^2 \theta_k \right]$.

The resulting sum MSE as a function of $\hat{\boldsymbol{H}}_{\mathrm{q},i}$ has been derived in [4]: $\mathrm{MSE}_{\mathrm{q}} =$

$$\sigma_s^2 \operatorname{tr}\left(\left(\xi_{\mathrm{q}} \mathbf{1}_M - \kappa \hat{\boldsymbol{H}}_{\mathrm{q}} \hat{\boldsymbol{H}}_{\mathrm{q}}^{\mathrm{H}} \right) \left((1-\kappa)\, \hat{\boldsymbol{H}}_{\mathrm{q}} \hat{\boldsymbol{H}}_{\mathrm{q}}^{\mathrm{H}} + \xi_{\mathrm{q}} \mathbf{1}_M \right)^{-1} \right). \tag{13}$$

III. User Selection

As stated before, the figure of merit of the system is the sum MSE of the scheduled users in the set $\mathcal{S}$ as given in (13). We will assume that the base station always serves M users out of the $K > M$ active users. To find the set with the smallest $\mathrm{MSE}_{\mathrm{q}}$, requires a brute search over all the $\frac{K!}{M!(K-M)!}$ possible sets with M users. For moderate values of K this is not practical or feasible. Therefore, we propose to employ a suboptimal user selection algorithm which still captures the benefit from multiuser diversity. The suboptimal set of selected users is chosen as follows:

Step 1) *Initialization:* We select the first user in the set by finding the user which is on average the most orthogonal to the other users. Recall that the base station has access only to the quantized CDI of the users and to the statistics of the users' CMI which is the same for all users. We define a measure for the average orthogonality of user k as

$$\nu_k = \frac{1}{K-1} \sum_{x=1, x\neq k}^{K} \left| \hat{\boldsymbol{h}}_{\mathrm{q},k}^{\mathrm{H}} \hat{\boldsymbol{h}}_{\mathrm{q},x} \right|, \tag{14}$$

for $k = 1, \ldots, K$. Recall that $\hat{\boldsymbol{h}}_{\mathrm{q},k}$ are unit-norm vectors. In addition, we initialize the set of available users $\mathcal{T}_1$ for the selection. Hence, the first step of the user selection consists of:

$$\begin{aligned} \mathcal{T}_1 &= \{1, 2, \ldots, K\} && (15)\\ i &= 1 && (16)\\ \pi(i) &= \underset{k \in \mathcal{T}_i}{\operatorname{argmin}}\ \nu_k && (17)\\ \mathcal{S} &= \{\pi(i)\} && (18)\\ \mathcal{T}_{i+1} &= \{k \in \mathcal{T}_i, k \neq \pi(i)\} && (19)\\ i &\leftarrow i+1 && (20) \end{aligned}$$

Step 2) *Partial channel matrix:* In order to add a user, we need to construct a partial channel matrix for each of the remaining users in the set $\mathcal{T}_i$. To this end, we denote at the i-th iteration of the algorithm, the partial channel matrix $\hat{\boldsymbol{H}}'_{\mathrm{q},k}(i)$ for $k \in \mathcal{T}_i$ as follows

$$\hat{\boldsymbol{H}}'_{\mathrm{q},k}(i) = \left[\hat{\boldsymbol{h}}_{\mathrm{q},\pi(1)}, \cdots, \hat{\boldsymbol{h}}_{\mathrm{q},\pi(i-1)}, \hat{\boldsymbol{h}}_{\mathrm{q},k} \right]^{\mathrm{T}} \in \mathbb{C}^{(|\mathcal{S}|+1) \times M}, \tag{21}$$

which includes the channels from the already selected users which are in the set $\mathcal{S} = \{\pi(1), \cdots, \pi(i-1)\}$ and the channel from user k, where $k \in \mathcal{T}_i$.

Step 3) *Adding a user:* The i-th user is selected as follows

$$\begin{aligned} \pi(i) &= \underset{k \in \mathcal{T}_i}{\operatorname{argmax}}\ \det\left(\hat{\boldsymbol{H}}'^{,\mathrm{H}}_{\mathrm{q},k}(i)\, \hat{\boldsymbol{H}}'_{\mathrm{q},k}(i) \right) && (22)\\ \mathcal{S} &\leftarrow \mathcal{S} \cup \{\pi(i)\}, && (23) \end{aligned}$$

where $\det(\boldsymbol{A})$ is the determinant of the matrix $\boldsymbol{A}$. The motivation behind (22) is that we try to choose the best user from the remaining users in $\mathcal{T}_i$ which is most orthogonal to the already selected users.

Step 4) *Next iteration or stop:* If $|\mathcal{S}| < M$, then we update the set $\mathcal{T}_i$ of available users for the next iteration.

$$\mathcal{T}_{i+1} = \{k \in \mathcal{T}_i, k \neq \pi(i)\} \tag{24}$$
$$i \leftarrow i + 1. \tag{25}$$

If $|\mathcal{S}| < M$ and if $\mathcal{T}_{i+1}$ is non-empty (although we have assumed that $K > M$) then we can still add users and hence, we go to Step 2). If this is not the case, the algorithm is terminated.

Here we have described how the user selection takes places, i.e. how the set $\mathcal{S}$ of M users is determined. In the next section, we will discuss how the feedback of the K users takes place over the T_F time instances.

IV. System Limited Feedback Design

As mentioned in the introduction, we assume a constraint on the total amount of airtime T_{F} available for the feedback of the system. We believe that such a constraint is more appropriate than a constraint on the number of feedback bits as presented in [7]. This is due to the fact that in the feedback link, i.e. the uplink of the system, the base station has M *receive* antennas such that it can detect the feedback of at most M users simultaneously using *receive* beamforming. To this end, the base station must nonetheless know the *uplink* channels of the K users! This can be achieved with K orthogonal dedicated pilots (one for each user) in the uplink, each consisting of $T_{\mathrm{UL}} \geq K$ time instances (symbols) as shown in Figure 2, such that the base station can obtain an estimate of the K users' uplink channels. However, we point out that the base station would need to estimate the uplink channels of the K users *anyhow* in order to perform user selection in the uplink, i.e. to benefit from multiuser diversity in the uplink data transmission. Hence, this does not really imply an additional headover for the feedback detection.

The receive antennas at the base station represent an additional degree of freedom for the feedback of the users in the uplink. Let us assume, for instance, that the base station performs MMSE receive beamforming in order to receive the feedback from at most M users. With MMSE receive beamforming the base station is able to spatially separate the feedback transmission from M distinct users. Our contribution to the the feedback design is the utilization of the additional degree of freedom presented by the multiple receive antennas at the base station in the uplink for the feedback detection. More precisely put, the fact that multiple users can feed back simultaneously in the uplink. This comes at the cost of an increase in the errors in the feedback link due to the presence of multiuser interference. Nonetheless, recalling that the feedback is sent using QPSK symbols we will assume for simplicity that base station receives the feedback of the K users without errors. Due to this assumption, we do not go into details about the estimation of the K uplink channels, where every user employs a different dedicated pilot sequence. In addition, we also do not discuss the MMSE receive beamforming based on the estimated uplink channels, which is used for the feedback detection. We simply assume that the feedback is received without errors, without affecting our main contribution to the feedback design, i.e. that at most M users can each simultaneously feed back one QPSK symbol. We will comment on the feedback errors in Section VI.

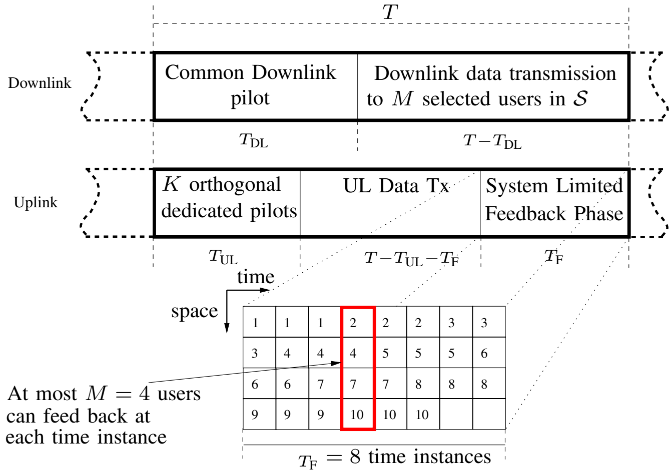

Fig. 2. System Limited Feedback Design.

The feedback load per user consists of B bits and we have assumed the feedback bits are sent using QPSK symbols. Considering without loss of generality only even values of B, each user can feed back its B bits using $\frac{B}{2}$ QPSK symbols. Let us now assume that the K users are somehow ordered into T_{F} groups with at most M users per group. Let us denote the ℓ-th group as $\mathcal{G}_\ell$ for $\ell = 1, \ldots, T_{\mathrm{F}}$ where $|\mathcal{G}_\ell| \leq M$. The users in the group $\mathcal{G}_\ell$ are the users which feed back one QPSK symbol from its feedback load during the ℓ-th time instance of the feedback phase. As discussed before, we assume that the base station can detect error-free one QPSK feedback symbol from each user in group $\mathcal{G}_\ell$ by performing for instance MMSE receive beamforming at time instance ℓ based on the estimated uplink channels of the users in group $\mathcal{G}_\ell$ for $\ell = 1, \ldots, T_{\mathrm{F}}$. The base station can inform the users via control channels in the downlink during which time instances in the feedback phase should each user relay its feedback to the base station, i.e. at which time instances should each user send one QPSK symbol from its feedback load.

Given a total system feedback load consisting of an airtime of T_{F} time instances, where at each time instance at most M users can each feed back one QPSK feedback symbol, we have that at least $K = M$ users can each feed back $B = 2T_{\mathrm{F}}$ bits or at most $K = T_{\mathrm{F}}M$ users can each feed back $B = 2$ bits. In the former case, each of the groups $\mathcal{G}_\ell$ for $\ell = 1, \ldots, T_{\mathrm{F}}$ would contain exactly the same set of M users, while in the latter each user appears only in one of the T_{F} groups. In general, we have that for a given airtime T_{F} and number of users K, where $M \leq K \leq T_{\mathrm{F}}M$, each user can feedback

$$B = 2\left\lfloor \frac{T_{\mathrm{F}}M}{K} \right\rfloor \tag{26}$$

feedback bits. To better explain this, let us consider an example. For

$$\begin{aligned} M &= 4 \quad \text{antennas} \\ K &= 10 \quad \text{users and} \\ T_{\mathrm{F}} &= 8 \quad \text{time instances,} \end{aligned}$$

each user could feedback

$$B = 2\left\lfloor \frac{8 \cdot 4}{10} \right\rfloor = 6 \tag{27}$$

bits using 3 QPSK symbols. Without loss of generality, we assume that the each of the $K = 10$ users are assigned randomly to 3 distinct groups out of the T_{F} groups in the feedback phase, where each group has at most 4 users, i.e. $|\mathcal{G}_\ell| \leq 4$. For this case, one possible allocation of the user to the groups in the feedback phase could be

$$\begin{aligned} \mathcal{G}_1 &= \{1, 3, 6, 9\} \\ \mathcal{G}_2 &= \{1, 4, 6, 9\} \\ \mathcal{G}_3 &= \{1, 4, 7, 9\} \\ \mathcal{G}_4 &= \{2, 4, 7, 10\} \\ \mathcal{G}_5 &= \{2, 5, 7, 10\} \\ \mathcal{G}_6 &= \{2, 5, 8, 10\} \\ \mathcal{G}_7 &= \{3, 5, 8\} \\ \mathcal{G}_8 &= \{3, 6, 8\} \end{aligned}$$

This feedback allocation is also depicted in Fig. 2 in the system limited feedback phase. The numbers represent the user indices, i.e. $k = 1, 2, \ldots, K$. The users in group $\mathcal{G}_\ell$ indicate the users that have to feedback at the ℓ-th time instance of the feedback phase one QPSK symbol with 2 bits from its $B = 6$ feedback bits. For example, at the time instance $\ell = 4$ in the feedback phase, users $2, 4, 7$ and 10 would each feed back one QPSK symbol. As illustrated in Figure 2, at the time instance $\ell = 4$ user 2 sends its first 2 bits from its feedback load, while user 4 would send its last 2 bits from its feedback load. Note that each user appears only in $\frac{B}{2} = \frac{6}{2} = 3$ different groups. We also point out that not at all time instances $M = 4$ users have to feed back, since $|\mathcal{G}_7| = |\mathcal{G}_8| = 3$.

V. Numerical Results

With the minimum and maximum value for K for a given T_{F}, i.e., $K = M$ and $K = T_{\mathrm{F}}M$, we can observe the tradeoff between multiuser diversity and feedback quality. In the former case we attempt to have the smallest quantization error per user with the largest possible $B = 2T_{\mathrm{F}}$, while in the latter case we rather try to have the largest degree of multiuser diversity with the largest possible $K = T_{\mathrm{F}}M$ with $B = 2$. The former case is achieved with the smallest required value of K, while the latter case is achieved with the smallest possible number of feedback bits per user. However, the optimum operating point for the system experiences a tradeoff between the degree of multiuser diversity and the feedback quality.

For a given limited feedback of T_{F} time instances, the tradeoff between multiuser diversity and the feedback quality can be observed in Figure 3, where we have plotted for $M = 4$ the average MSE based on quantized CDI, i.e. $\mathrm{E}[\mathrm{MSE}_q]$, which has been obtained by averaging over 10000 channels. Recall that MSE_q represents the sum MSE of the selected users which is given in (13) based on the quantized CDI of the selected users. We have assumed that $T_{\mathrm{DL}} = 4$ and in the following we refer to SNR as $\mathrm{SNR} = \frac{P_{\mathrm{DL}}}{M\sigma_n^2}$, which represents the ratio of the transmit power per antenna and the noise variance, which is assumed to be the same for all the users. We have employed the same settings for the rest of the results in this section.

The tradeoff in Figure 3 is shown for different SNR and values of T_{F}. For small K, the quantization error is small but there is little degree of multiuser diversity, while for large K there is potentially larger multiuser diversity but the quantization error is large. This leads, as observed, to an optimum number of users K_{opt} which should feedback given an SNR and a feedback load consisting of T_{F} time instances. The optimum number of users K_{opt} achieves the minimum average MSE for a given curve. For a given T_{F}, the K_{opt} users then feedback the resulting optimum number of feedback bits per user

$$B_{\mathrm{opt}} = 2\left\lfloor \frac{T_{\mathrm{F}}M}{K_{\mathrm{opt}}} \right\rfloor. \tag{28}$$

The users feedback their B_{opt} bits as described in Section IV. Note that for a given T_{F}, the optimum number of users which should feedback is independent of the packet length T. We only require that $T > T_{\mathrm{F}}$.

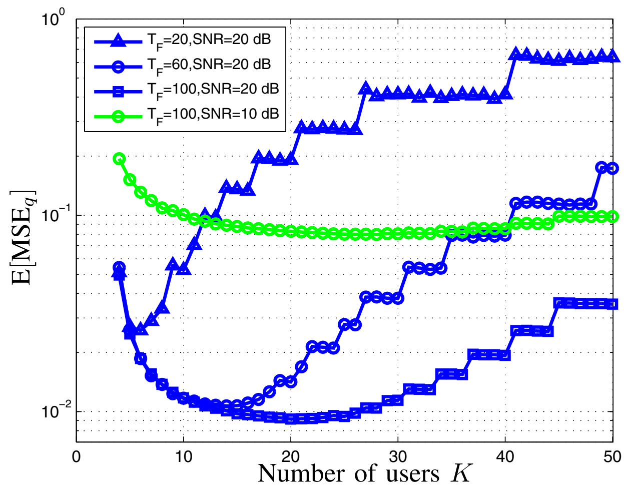

Fig. 3. Average User MSE vs. K.

In addition, Figure 4 shows K_{opt} as a function of T_{F} for $M = 4$ and different SNR. K_{opt} is obtained as explained in the discussion for Figure 3. Not surprisingly, the number of users who should feedback increases with T_{F}, since more users can be accomodated without increasing B and without increasing the MSE, as there is larger degree of multiuser diversity with the same feedback quality. Observe also that for a given T_{F}, K_{opt} decreases with increasing SNR which means that the optimum number of feedback bits B_{opt} increases, thus increasing the feedback quality. This is due to the fact that the quantization error has a larger effect as SNR increases, i.e. it is the main source of error instead of the noise. Hence, the number of feedback bits have to increase with SNR in order to avoid the saturation due to quantization [2], [4].

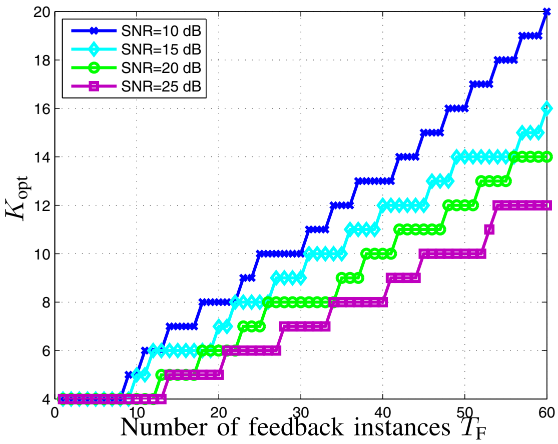

Fig. 4. Optimum K vs. T_{F}, $M = 4$.

As T_{F} increases, the optimum number of users K_{opt} increases such that the optimum number of feedback bits B_{opt} also increase in general. This can be observed in Figure 5. This means that in general as T_{F} increases the optimum operating point can profit from a larger degree of multiuser diversity (increase of K_{opt}) and a better feedback quality (increase of B_{opt}). The general step form of the curves is due to the fact that for a given T_{F}, the optimum number of users that feedback at each time instance is not always equal to M. This can be elaborated considering the following example. Note for instance in Figure 4 and Figure 5 that the optimum number of users and the optimum number of feedback bits remain constant ($K_{\text{opt}} = 10$ and $B_{\text{opt}} = 20$) for $25 \leq T_{\text{F}} \leq 27$ when SNR $= 10$ dB. For $T_{\text{F}} = 25$, we have that $M = 4$ users feed back at each time instance during the feedback phase, since $T_{\text{F}} M = K_{\text{opt}} \frac{B_{\text{opt}}}{2}$. However, for $T_{\text{F}} = 27$, we have that $M = 4$ users feed back during 19 time instances while only 3 users feedback during the remaining $T_{\text{F}} - 19 = 8$ time instances of the feedback phase, since $(19 \times 4) + (8 \times 3) = K_{\text{opt}} \frac{B_{\text{opt}}}{2}$. However, note that for $T_{\text{F}} = 28$, $K_{\text{opt}} = 10$ users can now each feed back $B_{\text{opt}} = 22$, i.e. one QPSK feedback symbol more than for $T_{\text{F}} = 27$. In this case, $M = 4$ users feed back during 26 time instances while only 3 users feedback during the remaining $T_{\text{F}} - 26 = 2$ time instances of the feedback phase.

Finally, we plot the optimum average sum MSE as a function of T_{F} for different SNR in Figure 6. The MSE decreases monotonically with T_{F}. As discussed before, as T_{F} increases the optimum number of user K_{opt} increases and/or the optimum number of feedback bits B_{opt} increases, such that the data transmission profits from an increase in the multiuser diversity and/or an increase in the feedback quality which lead

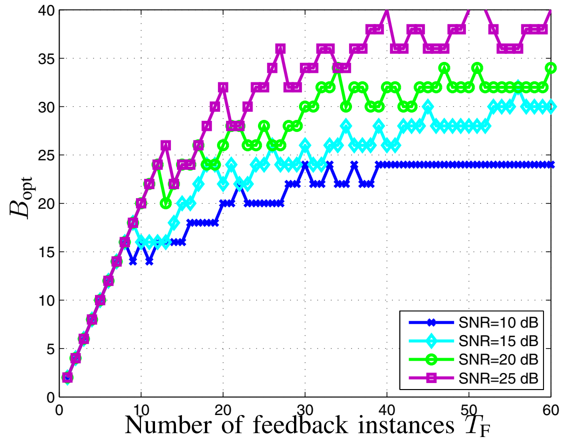

Fig. 5. Optimum B vs. T_{F}, $M = 4$.

to a decrease in the average MSE.

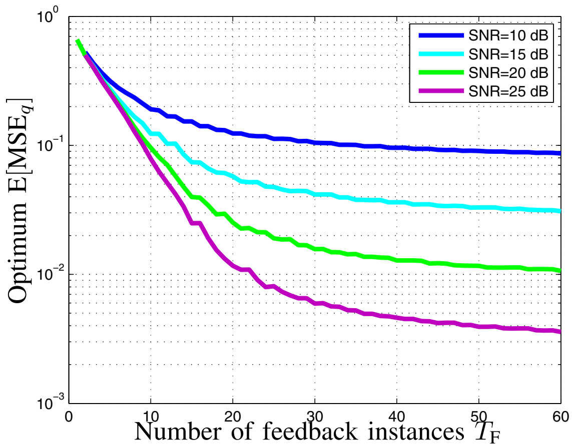

Fig. 6. Optimum average sum MSE vs. T_{F}, $M = 4$.

VI. Conclusion and Future Work

In this paper we have addressed the tradeoff between the attainable degree of multiuser diversity and the feedback quality given a constraint on the total amount of airtime T_{F} available for feedback in a system. For a given T_{F}, the question arises whether it is preferable to have more users relay coarsely quantized feedback or to have less users relay higher quality feedback. As a figure of merit we have considered the sum MSE of the selected users based on quantized CDI. We have presented a suboptimal user selection scheme which still profits from the multiuser diversity. In addition, we have taken into account for the feedback design, that multiple users (at most M) can each relay back to the base station one QPSK feedback symbol at each time instance during the feedback phase. To this end, the uplink channels of the users have to be known, but we point out that for the uplink transmissions to profit from multiuser diversity, the base station needs to estimate anyhow the uplink channels of the users.

For a given T_{F}, the optimum number of users K_{opt} which should feed back decreases with increasing SNR allowing for higher quality feedback from those users who feed back. This is due to the fact that the quantization is the main source of errors at higher SNR. In addition, we have observed how K_{opt} and B_{opt} could both increase with T_{F}, which means that we can benefit from an increase in multiuser diversity and an increase in the feedback quality due to the fact that multiple users can feed back at the same time instance as discussed in the feedback design.

In this work we have not considered the fact that the multiuser interference during the feedback detection can lead to feedback errors. However, the feedback detection can make use of the inherent multiuser diversity in the feedback link! Instead of allocating the users randomly, we could allocate the users in the groups $\mathcal{G}_\ell$ for $\ell = 1, \ldots, T_{\mathrm{F}}$ during the feedback phase, such that the users in each group are almost orthogonal to one another. If the number of users which feed back is small, then we can reduce the number of users which feed back at each time instance. The main idea is basically to perform scheduling in the feedback link to benefit from the inherent multiuser diversity in order to reduce the feedback error probability. This matter is the subject of our current work.

REFERENCES

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

[2] N. Jindal, "MIMO Broadcast Channels with Finite-Rate Feedback", *IEEE Transactions on Information Theory*, Vol. 52, No. 11, Nov. 2006.

[3] G. Caire, N. Jindal, M. Kobayashi and N. Ravindran, "Multiuser MIMO Downlink Made Practical: Achievable Rates with Simple Channel State Estimation and Feedback Schemes", *IEEE Trans. Information Theory*, Vol. 56, No. 6, pp. 2845-2866, June 2010.

[4] M. Castañeda, Israa Slim, A. Mezghani and J. A. Nossek, "Transceiver Design in Multiuser MISO Systems with Imperfect Transmit CSI", in *ITG Workshop in Smart Antennas (WSA)*, Feb. 2010.

[5] T. Yoo, N. Jindal and A. Goldsmith, "Multi-Antenna Broadcast Channels with Limited Feedback and User Selection", in *IEEE Journal on Selected Areas in Communications*, Vol. 25, No. 7, pp. 1478-1491, Sep. 2007

[6] M. Trivellato, F. Boccardi, and F. Tosato, "User Selection Schemes for MIMO Broadcast Channels with Limited Feedback," in *Proc. Vehicular Technology Conference* (VTC'07), Apr. 2007.

[7] N. Ravindran and N. Jindal, "Multi-User Diversity vs. Accurate Channel State Information in MIMO Downlink Channels", *Arxiv preprint cs.IT/0907.1099v1*, July 2009.

[8] C. K. Au-Yeung and D. J. Love, "On the Performance of Random Vector Quantization Limited Feedback Beamforming in a MISO System", *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, Feb. 2007.

[9] B. Hassibi and B. M. Hochwald, "How much Training is needed in a Multiple-antenna Wireless Link?", *IEEE Trans. Inform. Theory*, vol. 49, pp. 951-964, Apr. 2003.