An Adaptive Limited Feedback Scheme for MIMO-OFDM Based on Optimal Stopping

Jieying Chen, Randall A. Berry, Michael L. Honig
Department of Electrical Engineering and Computer Science
Northwestern University
Evanston, Illinois 60208
j-chenm@northwestern.edu, {rberry,mh}@ece.northwestern.edu

Abstract

We propose and analyze a limited feedback scheme for downlink MIMO-OFDMA. Our analysis explicitly accounts for the feedback overhead by assuming a time division duplex system in which all feedback and data transmission must occur with a coherence time. As the fraction of coherence time devoted to feedback increases, the base station can allocate resources more efficiently, but has less time available for data transmission. In the proposed scheme, the base-station sequentially receives feedback from the users and decides when to stop receiving additional feedback and begin data transmission. Each user feeds back their best codeword (beam) selected from a beam-forming codebook on each group of OFDM sub-channels, provided that the channel gain exceeds a given threshold. For a given feedback threshold, the *optimal stopping rule* used by the base station is derived. With this rule we show that the total throughput of this scheme scales linearly with the number of users, provided that the number of OFDM sub-channels also scales with fixed ratio. The effect of varying the coherence time and feedback rate is also characterized.

I. INTRODUCTION

Orthogonal Frequency Division Multiple Access (OFDMA) combined with Multiple Input Multiple Output (MIMO) techniques provide numerous degrees of freedom in space and frequency. To efficiently exploit these in a cellular downlink, the base station needs sufficient channel state information (CSI) to allocate resources and schedule users. In a multiuser system, the feedback overhead for acquiring CSI at the base stations can be prohibitive. This has motivated the study of several limited feedback schemes for single antenna OFDMA systems [1], [3]–[5], single user MIMO-OFDM systems [11], [12] and MIMO-OFDMA systems [13].

In [1], [4], [5] a threshold-based feedback scheme has been studied for single antenna OFDMA systems. There, each user sends the base station one bit per sub-channel to indicate whether the channel gain is above or below a given threshold. Limited feedback schemes have also been considered in the setting of narrow band downlink MIMO system (e.g., see [7]–[9]). In MIMO systems, the feedback information usually contains the channel magnitude as well as the channel direction in order to exploit spatial diversity. In [11]–[13], the authors exploit correlation in the frequency or time domain to reduce the feedback for MIMO-OFDM. However, a limitation of these schemes is that as the system size scales (i.e., the number of sub-channels and number of users becomes large), the required feedback also increases. Given a finite coherence time T and feedback rate R_F per sub-channel, the time for feedback will eventually dominate the entire coherence time T.

In [2] we introduced two limited feedback schemes for the downlink of a single-antenna OFDMA system. Both schemes were shown to achieve positive throughput as the system size scales when the time for feedback is explicitly taken into account. Here we study an extension of the *sequential scheme* from [2] for a MIMO-OFDMA system. The feedback in this scheme is limited by two techniques: *i*) Channels are grouped. One bit is used to indicate whether the magnitude of all the channels within one group is above a threshold. *ii*) Each user compresses the binary feedback vector before sending it to the

This work was supported by the Motorola Center for Seamless Communications and the Army Research Office under grant W911NF-07-1-0028.

base station. The users feed back sequentially. Hence once the channel group size, threshold and number of users are set, the fraction of time devoted to feedback becomes fixed. That is due to the assumption that every user must feed back its CSI to the base station. Provided that the load $\beta = K/N$ is less than $R_F T$ (number of feedback bits per coherence time), the total data throughput scales linearly with the number of sub-channels N as $N \to \infty$ and $K \to \infty$ (with fixed β). In contrast, with unlimited CSI feedback the total throughput scales like $N \log \log K$, where the additional growth is due to multiuser diversity. With finite $R_F T$ this additional terms disappears. Furthermore, for high enough loads the sequential scheme in [2] achieves zero throughput due to the requirement that all users feed back. This motivates us to consider an *adaptive* version of this scheme in which only a subset of users feed back within each coherence time.

Given a fixed coherence time, we develop an adaptive sequential feedback scheme for MIMO-OFDM in which, based on the feedback received so far, the base station decides whether to request CSI feedback from an additional user or start data transmission. This decision balances the multiuser diversity gain associated with additional feedback with the required overhead. We formulate this decision as an optimal stopping time problem [10] and characterize the optimal stopping rule. We then optimize the parameters of our scheme, which include the sub-channel group size and the channel gain threshold. The optimized scheme is shown to have a throughput that scales linearly with the system size for all loads. We also study the impact of $R_F T$ on the data throughput, and show that if this quantity scales faster than $\log K$ as the system scales, the multi-user diversity gain of $\log \log K$ is recovered.

II. System Model

We consider a downlink MIMO-OFDMA system with K users and N OFDM sub-channels (each with bandwidth normalized to one). Each user has a single antenna and the base station has M transmit antennas. The received signal on the l^{th} $(1 \leq l \leq N)$ sub-channel of the k^{th} user is

$$y_{k,l} = h_{k,l}^H V_l S + n_{k,l} \tag{1}$$

where $h_{k,l}$ is an $M \times 1$ channel vector for user k and $n_{k,l}$ is complex Gaussian noise with unit variance, which is independent across sub-channels and users. The channel vectors $h_{k,l}$ are modeled as a block-fading process with block-length T, which we will refer to as the channel coherence time. During each block, each component of $h_{k,l}$ for all k and l is generated according to an *i.i.d.* complex Gaussian distribution with zero mean and unit variance, corresponding to a rich scattering environment. The realization of $h_{k,l}$ is assumed to be known perfectly at receiver k at the start of each coherence block, but not at the transmitter or any other receiver. All beamforming vectors (codewords) are assumed to be selected from a single codebook with M unit norm vectors $[v_1, \ldots, v_M]$. The matrix $V_l = [v_{l,1}, \ldots, v_{l,M_l}]$ in (1) is $M \times M_l$, where M_l is the number of data streams scheduled on the l^{th} sub-channel. The columns of V_l correspond to the codewords assigned to the scheduled users on sub-channel l. The matrix $S = [s_1, s_2, \ldots, s_{M_l}]^T$ contains the data symbols of the scheduled users. The base station is assumed to use an on-off power control policy, so that if a sub-channel is requested by at least one user, a constant power P is allocated to this sub-channel. If multiple streams are scheduled on a sub-channel, this power is divided equally across the streams.

As in [2], we consider a feedback scheme in which the users send limited feedback at the start of each coherence block. The system is time-division duplex, so that when feedback is being sent, data can not be transmitted. All feedback is assumed to be sent at a fixed rate R_F bits/sub-channel, so that if there are N subchannels the total feedback rate is $R_F N$. Hence, if L_F bits of feedback are sent, then $T(1 - \frac{L_F}{R_F N})$ seconds remain for data transmission. During each coherence block, the users feed back CSI information sequentially in a given order, which may change from block to block according to a pre-determined schedule. The base station adaptively decides how many users send feedback before it begins transmitting data.[1]

[1] This requires the base-station to send a control signal to the users to indicate when to stop sending feedback.

Each user k, which sends feedback in a given coherence block, first quantizes its channel gain $h_{k,l}$ on each subchannel l to the closest codeword defined by its inner product as in [8], [9],

$$d_{k,l} = \arg \max_{1 \leq m \leq M} \mid h_{k,l}^H v_m \mid . \tag{2}$$

As in [7], [9], the user then estimates the received signal-to-interference-and-noise ratio (SINR) it would see on sub-channel l if it were allocated the codeword $v_{d_{k,l}}$. This estimate assumes that the maximum of M users are scheduled simultaneously on this sub-channel. The resulting SINR estimate is given by

$$\begin{aligned} \tilde{\gamma}_{k,l} &= \frac{\frac{P}{M}|h_{k,l}^H v_{d_{k,l}}|^2}{1 + \sum_{j \neq d_{k,l}} \frac{P}{M}|h_{k,l}^H v_j|^2} \\ &= \frac{|h_{k,l}^H v_{d_{k,l}}|^2}{1/\rho + \sum_{j \neq d_{k,l}} |h_{k,l}^H v_j|^2} \end{aligned} \tag{3}$$

where $\rho = P/M$. User k then compares the estimated SINR, $\tilde{\gamma}_{k,l}$, with a given threshold t_0. User k only requests sub-channels for which $\tilde{\gamma}_{k,l}$ exceeds the threshold. Let

$$p_0 = \Pr(\tilde{\gamma}_{k,l} \geq t_0)$$

denote the probability the estimated SINR on a particular sub-channel exceeds t_0.

Lemma 1: If $t_0 \geq 1$, then $p_0 = \frac{M}{e^{t_0/\rho}(1+t_0)^{M-1}}$.

The condition $t_0 \geq 1$ ensures that the estimated SINR associated with at most one of the codewords for each user will exceed t_0 on a given sub-channel. The result follows from this assumption and the assumed fading distribution. To simplify our analysis, we assume that $t_0 \geq 1$ for the remainder of the paper.

To reduce the required feedback, we divide the N sub-channels into non-overlapping *sub-channel groups*, each containing αN sub-channels, where $0 \leq \alpha \leq 1$. A user who sends feedback in a coherence block requests a particular sub-channel group if and only if the estimated SINR for all the αN sub-channels within the group are above the threshold. Therefore, the probability a user requests a particular group is $p_1 := p_0^{\alpha N}$. The amount of feedback can be reduced by increasing αN or t_0. If user k requests a group, we assume it feeds back the beam indices, $d_{k,l}$, corresponding to the αN sub-channels within the group. Hence, the CSI at the receiver is a $M^{\alpha N}$-ary sequence with length $N/(\alpha N)$. In order to reduce the feedback overhead further, the feedback bits of each user can be losslessly compressed before transmission. The length of the compressed vector of feedback bits depends on the particular compression scheme used. We will specify a particular scheme in Section IV. However, we emphasize that the stopping rule developed in the next section does not depend on this particular scheme.

Our main performance objective is the sum-rate, which can be written as

$$\tilde{R} = (1-f)N_a r \tag{4}$$

where f indicates the fraction of a coherence-block used for feedback, N_a denotes the number of data streams scheduled and r is the achievable rate per scheduled sub-channel. We assume that r is matched to the feedback threshold t_0 and is given by[2]

$$r = \log(1 + t_0).$$

The value of N_a and f depend on both the channel group size, the feedback threshold and the number of users, which send feedback during each coherence band. Exact expressions for these will be studied in the following sections.

[2] This is reasonable assuming that the users are only assigned a few sub-channels and do not code over multiple coherence blocks. If user's could code over many sub-channels then they could achieve the larger rate of $E(\log(1 + h_{k,l}^H v_m||h_{k,l}^H v_m|^2 > t_0)$.

III. Optimal Stopping Rule

In this section, we characterize the optimal stopping rule that the base station should use to decide when to stop receiving additional feedback. For this we assume that the total coherence time T is divided into $K^{'}$ slots, where each slot is used either for CSI feedback from one user or data transmission. If the slot is used for CSI feedback, we assume that the scheduled user feeds back its CSI using all subchannels to the base station within the given slot.[3] Given the cumulative CSI at the end of each slot, the base station must decide to either request CSI feedback from another user in the next time-slot or to allocate resources according to the current available CSI and start transmitting data. Once it starts transmitting data, it continues doing so for all of the remaining slots in the current coherence block.

Let $s_n = [x_{1,n}, x_{2,n}, \ldots, x_{M,n}, x_{0,n}]$ be a vector summarizing the feedback information after the n^{th} slot, where $x_{i,n}$ $(0 \leq i \leq M)$ denotes the number of sub-channels on which i distinct codewords have been requested by the n^{th} time-slot, i.e., i data streams can be scheduled simultaneously on $x_{i,n}$ sub-channels. The sum of all the elements in s_n, $\sum_{i=0}^{M} x_{i,n}$, is equal to the total number of sub-channels N in the system. We refer to s_n as the state of the system at time n. If the base station stops receiving CSI feedback after slot n, when the system is in state s_n, then the corresponding sum-rate is given by

$$R(n, s_n) = \left(1 - \frac{n}{K'}\right) \sum_{i=1}^{M} (ix_{i,n}) \log(1 + t_0) \tag{5}$$

Comparing with (4), the fraction of the coherence time used to transmit data is $f = 1 - \frac{n}{K'}$ and the total number of data streams scheduled is $N_a = \sum_{i=1}^{M}(ix_{i,n})$.

Given the previously defined state, we can view the decision faced by the base station as an optimal stopping problem. Namely, the base station's decision after each slot is to either stop requesting CSI feedback, in which case it receives a pay-off given by (5), or request additional feedback, in which case it can determine an expected future pay-off given the current state. Furthermore, the sequence of states is a Markov process. Hence, we can use results from optimal stopping theory to design a stopping rule. Such a rule is given in the following Proposition.

Proposition 1: The optimal stopping time j^* for a system with $K^{'}$ slots exists and is given by

$$j^* = \min \left\{ j \geq 1 : \sum_{i=1}^{M} (ix_{i,j}) \geq \frac{(K' - (j+1))Np_1}{1 + \frac{p_1}{M}(K' - (j+1))} \right\}. \tag{6}$$

Proposition 1 gives a simple threshold policy that the base station can follow to determine the stopping time j^*. Namely, after each time-slot the base-station stops if and only if the total number of streams it can schedule, given by $\sum_{i=1}^{M}(ix_{i,j})$, exceeds the time-varying threshold on the right-hand side of (6). Given that each coherence time T is divided into a finite number of time slots, determining the optimal stopping time is a finite horizon dynamic programming problem, which can in general be solved using backward induction. However, for this problem backward induction is not needed, i.e., it can be shown using the rate expression in (5) that a one-stage look-ahead policy (i.e. stopping if $R_n \geq E(R_{n+1}|s_n)$) is optimal. This follows from showing that the problem is a *monotone stopping problem* [10]. The detailed proof is omitted due to space considerations.

IV. Performance

A. Feedback constraint

We first express the feedback constraint in terms of the system parameters. Since the total bandwidth is divided into sub-channel groups with size αN, the CSI at the receiver is a vector of length $N/(\alpha N)$

[3] Of course, the validity of this assumption depends on the scheme used for compressing the feedback and the feedback threshold. This relationship will be explored in the following section.

with elements that can take one of $M^{\alpha N}+1$ values, where one of these values indicates that a sub-channel group is not requested. Conditioned on a sub-channel group being requested, the corresponding feedback symbol takes on one of $M^{\alpha N}$ values with uniform probability. We assume that each user compresses this feedback vector to within one bit of the entropy. Since the channel follows an *i.i.d.* distribution across the subchannels, the $(M^{\alpha N}+1)$-ary feedback symbols are also *i.i.d.*, and the entropy of each feedback symbol is[4]

$$L_{ent} = \left(M^{\alpha N} \frac{p_1}{M^{\alpha N}} \log\left(\frac{M^{\alpha N}}{p_1}\right) + (1-p_1)\log\left(\frac{1}{1-p_1}\right)\right) = H(p_1) + \alpha N p_1 \log M \tag{7}$$

where $H(p_1) = p_1 \log(1/p_1) + (1-p_1)\log(1/(1-p_1))$.

As the system size scales (N becomes large), we can assume that each user's feedback can be done within the given time duration $T/K^{'}$ at the rate NR_F. Namely, with a variable length coding scheme the actual feedback L_i for each user i is random; however, as the number of sub-channels N increases, the time to send user i's feedback satisfies

$$\frac{L_i}{NR_F} \rightarrow \frac{L_{ent}}{\alpha N R_F}$$

by the law of large numbers.[5] Based on this, we model the relationship between the system parameters and the number of slots by assuming that

$$\frac{L_{ent}+1}{\alpha N R_F} = \frac{T}{K'} \tag{8}$$

where we have added an extra one bit to the entropy to ensure that each user must send back at least one bit. Substituting for L_{ent}, this is equivalent to the following

$$\frac{1}{\alpha N} H(p_1) + p_1 \log M + \frac{1}{\alpha N} \quad = \quad \frac{R_F T}{K'}. \tag{9}$$

Using that $p_1 = p_0^{\alpha N}$ and the value for p_0 in Lemma 1 gives

$$K^{'} p_1 \left(y + (y - \log M) \frac{(1-p_1)\log(1-p_1)}{p_1 \log(p_1)} \right) + \gamma\beta = R_F T \tag{10}$$

where

$$y \triangleq \log(\frac{M}{p_0}) = \frac{t_0}{\rho} + (M-1)\log(1+t_0) \tag{11}$$

and the group size

$$\alpha N = \frac{\log(p_1)}{\log(M) - y}. \tag{12}$$

[4] It would be more precise to denote the entropy as $L_{ent}(N, \alpha, p_1)$, but to simplify our notation we suppress the dependence on these parameters.

[5] Some care is required in showing this since α and p_1 may vary with the system size; in particular, here we are assuming that $\alpha \rightarrow 0$ so that each user is compressing an infinite number of symbols, as the following results show this is the case in an optimized system.

B. Asymptotic Results

As shown in Proposition 1, the stopping time j^* is a random variable in each coherence time T. It appears to be difficult to determine the probability distribution of j^*; however, we can determine the asymptotic behavior of j^* as K and N both tend to infinity with fixed ratio $\beta = K/N$. Namely, referring to the stopping criterion in Proposition 1, dividing the left-hand side by N, we note that almost surely

$$\frac{\sum_{i=1}^{M}(ix_{i,j})}{N} \to E(M_j) \tag{13}$$

by the law of large numbers. The right-hand side of (13) depends only on the stopping time j, the request probability p_1, and the total number of time slots $K^{'}$. An expression for $E(M_j)$ as a function of j and p_1 can be derived by induction on j, and is given by the following Lemma.

Lemma 2: As K and N goes to infinity with fixed ratio, j^* asymptotically converges to a constant satisfying

$$\left(1-\frac{p_1}{M}\right)^{j^*}\left(1+\frac{p_1}{M}(K^{'}-(j^*+1))\right)=1. \tag{14}$$

Lemma 2 indicates the average fraction of time devoted to feedback for a large system. Based on this result, we can derive the corresponding average data throughput per sub-channel. Namely, the capacity objective with fixed parameters α, t_0, and $K^{'}$ is given by

$$C(\alpha,t_0,K')=M\left(1-\frac{j^*}{K'}\right)\left(1-\left(1-\frac{p_1}{M}\right)^{j^*}\right)\log(1+t_0). \tag{15}$$

We wish to maximize this expression over α, t_0 and $K^{'}$ subject to the feedback constraint (10), where j^* is determined by (14) and y is given by (11). Also, we have the additional constraints $\alpha N \geq 1$ and $t_0 \geq 1$ (by assumption), which lead to

$$y > \log M \tag{16}$$

$$y \geq \frac{1}{\rho} + (M-1)\log 2. \tag{17}$$

The preceding optimization problem is difficult to solve analytically, although we are able to derive certain scaling properties. These are summarized in the following proposition.

Proposition 2: As $K \to \infty$ and $N \to \infty$ with fixed ratio, to maximize the throughput $C(\alpha, t_0, K')$, the system parameters must scale as follows:

- The number of slots $K' \log K' = \theta(K)$.
- The optimal stopping time j^* grows as $\theta(K')$.
- The group size αN grows as $\theta(K')$.
- The threshold t_0 is bounded.

Furthermore, given such parameters, $C(\alpha, t_0, K') \to C^*$, for some non-zero, finite C^*.

The last part of this proposition states that the sum-capacity of an optimized system scales linearly with the number of sub-channels, i.e. like C^*N. The proof consists of proving the following lemmas. Namely, let $Kp_1 \to \mu_1$ as K and N tend to infinity, where $0 \leq \mu_1 \leq \infty$. Suppose also that $K^{'}p_1 \to \gamma\mu_1$, where $0 \leq \gamma \leq 1$.

Lemma 3: If $\gamma\mu_1 = 0$ or ∞, the average capacity converges to zero per sub-channel.

This lemma indicates that the optimal $\gamma\mu_1$ has to converge to a finite positive value. The proof consists of noting that if $\gamma\mu_1 = \infty$, (10) implies that the threshold t_0 has to be zero, and if $\gamma\mu_1 = 0$, Lemma 3 implies that the average number of data streams scheduled on a particular sub-channel approaches zero. In both cases, the capacity must then converge to zero.

Lemma 4: Given any finite positive value for $\gamma\mu_1$, to maximize $C(\alpha, t_0, K')$ the optimal $p_1 \to 0$ and the optimal $K^{'} \to \infty$.

For any $0 < \gamma\mu_1 < \infty$, if we fix the fraction of time devoted to feedback, i.e., $\varphi = j^*/K^{'}$ is a constant, we can show that $C(\alpha, t_0, K')$ is a decreasing function of p_1. Therefore, the optimal asymptotic limit of p_1 should be 0.

From Lemma 4 we conclude that the sum throughput C can be expressed as a function of $\gamma\mu_1$. The proposition then follows by optimizing C over this value.

C. System Comparison

We now compare the performance of the sequential scheme proposed in [2] with the adaptive sequential scheme. In the scheme in [2], all users compress their feedback and send it at the start of each coherence block. Such a scheme can be treated as an extreme case of the adaptive scheme in which we allocate $K^{'} > K$ slots within one coherence time T. The base station waits for the feedback bits from all the users in each coherence time T, which implies that K slots are used for feedback and $K^{'} - K$ slots for data transmission. As in (15), the average system throughput per sub-channel can be expressed as

$$C_{seq}(\alpha, t_0, K') = M\left(1 - \frac{K}{K'}\right)\left(1 - \left(1 - \frac{p_1}{M}\right)^K\right)\log(1 + t_0). \tag{18}$$

We can again optimize this over the system parameters t_0, α and $K^{'}$, subject to (10),(12),(16),(17) as well as the constraint that $K' \geq K$. The following lemma characterizes some scaling properties of the solution to this optimization.

Lemma 5: As $K \to \infty$ and $N \to \infty$ with fixed ratio, to maximize the throughput $C_{seq}(\alpha, t_0, K_1)$ the system parameters must scale as follows:

- Kp_1 converges to a positive finite value.
- K'/K converges to a positive finite value.

Furthermore, given such parameters, $C_{seq}(\alpha, t_0, K') \to C^*_{seq}$, for some non-zero, finite C^*_{seq}.

The proof of lemma 5 is similar to lemma 3. Namely we show that if the parameters do not scale in this way we show that the throughput must go to zero.

The last part of Lemma 5 implies that the sequential scheme has a capacity that scales like $C^*_{seq}N$.

From the above discussion it is clear that $C^*_{seq} \leq C^*$, since the adaptive scheme optimizes over a larger range of K'. Next we show that even if we restrict the sequential scheme to using $K' = K$ time-slots, it will still perform better than the sequential scheme proposed in [2]. Specifically, let $C_f(\alpha, t_0)$ denote the system throughput per sub-channel achieved by the adaptive sequential scheme when K' is set to K. We can again consider optimizing the throughput of this restricted scheme over α and t_0. By following a similar argument the asymptotic behavior of the optimal parameters can be characterized and it can again be shown that under the optimal parameters $C_f(\alpha, t_0) \to C^*_f$. The next lemma compares the first order growth of this scheme with that of the sequential scheme from [2].

Lemma 6: For all loads ρ, $C^*_{seq} \leq C^*_f$.

This shows that the asymptotic performance of the sequential scheme is no better than that of the adaptive scheme with $K' = K$. In other words, fixing the number of feedback slots at K but adaptively stopping, is better than having $K' > K$ slots but requiring all users to feedback.

D. Effect of R_FT on Capacity

Previously we have assumed that the the number of potential feedback bits per sub-channel R_FT was fixed as the system scales. In this section, we study the asymptotic performance of the adaptive feedback scheme when R_FT also scales.

Proposition 3: When $R_F T$ increases slower than $\log K$, the parameters of the optimized system satisfy the following:

- The asymptotic limit of $K^{'}p_1$ scales with $R_F T$ in the form of $M\log(\rho R_F T/M)$.
- The fraction of time devoted for feedback goes to zero at the order $\frac{\log(\log(\rho R_F T/M))}{\log(\rho R_F T/M)}$.
- The threshold t_1 scales with $R_F T$ at the order $\frac{\rho R_F T/M}{\log(\rho R_F T/M)}$.
- The capacity per sub-channel grows at the order $M\log(\rho R_F T/M) - 2M\log(\log(\rho R_F T/M))$.

When $R_F T$ increases faster than $\log K$, the capacity per subchannel of an optimized system grows at the order $M\log(\log K)$.

Proposition 3 states that the throughput scaling is proportional to $\log(R_F T)$ if $R_F T$ increases slower than $\log K$. If $R_F T$ scales faster than $\log K$, the full multiuser diversity gain can be exploited.

V. Numerical Results

In this section, numerical results are shown to illustrate the performance of the proposed scheme. We assume in the simulation that the channel from each transmit antenna to each user is i.i.d. Rayleigh with unit variance. The base station has $M = 4$ transmit antennas. Each user has one receive antenna. The system load is $\beta = 2$. The maximum number of feedback bits per sub-channel per coherence time is set to be $R_F T = 10$. The power assigned to each active data stream is normalized to 1.5 Watts.

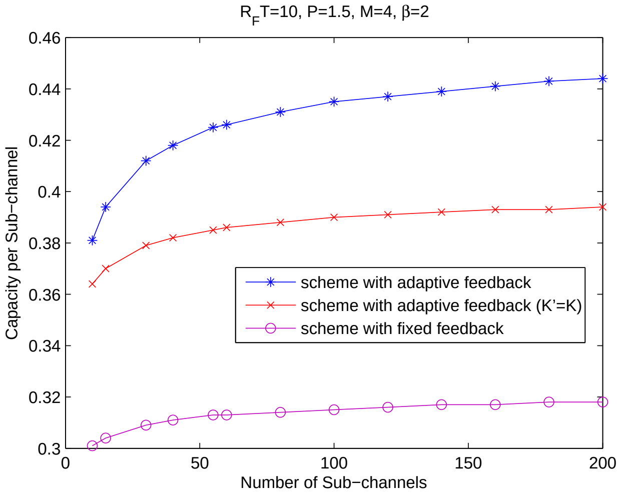

Fig. 1. Performance Comparison

Figure 1 shows the performance the three schemes considered above: the adaptive scheme, the sequential feedback scheme from [2], and the adaptive scheme with $K' = K$. In this case, both adaptive scheme perform significantly better than the non-adaptive scheme for all system sizes considered.

Figure 2 shows the performance of both schemes as a function of $R_F T$. As shown in [2], when $\beta \geq R_F T$, the throughput of the sequential scheme goes to zero. In contrast, the adaptive feedback scheme can work for any value of system load and $R_F T$. The plot on the right illustrates this. From Proposition 3, the performance of the adaptive scheme should increase like $\log(R_F T)$ and a similar scaling holds for the sequential scheme; this can be seen in the curve on the left, which shows shows these curves over a larger range of values of $R_F T$.

Figure 3 shows the fraction of time allocated for feedback for different values of $R_F T$. The fraction of time will converge to a constant as the system size increases. For the values of $R_F T$ used in the simulation, the fraction of feedback converges to values between 0.4 and 0.5, i.e. roughly, half of the coherence time T is devoted for feedback. Also, note that as $R_F T$ increases, this fraction decreases. Figure 4 shows the average number of subchannel groups requested by one user and the optimal threshold

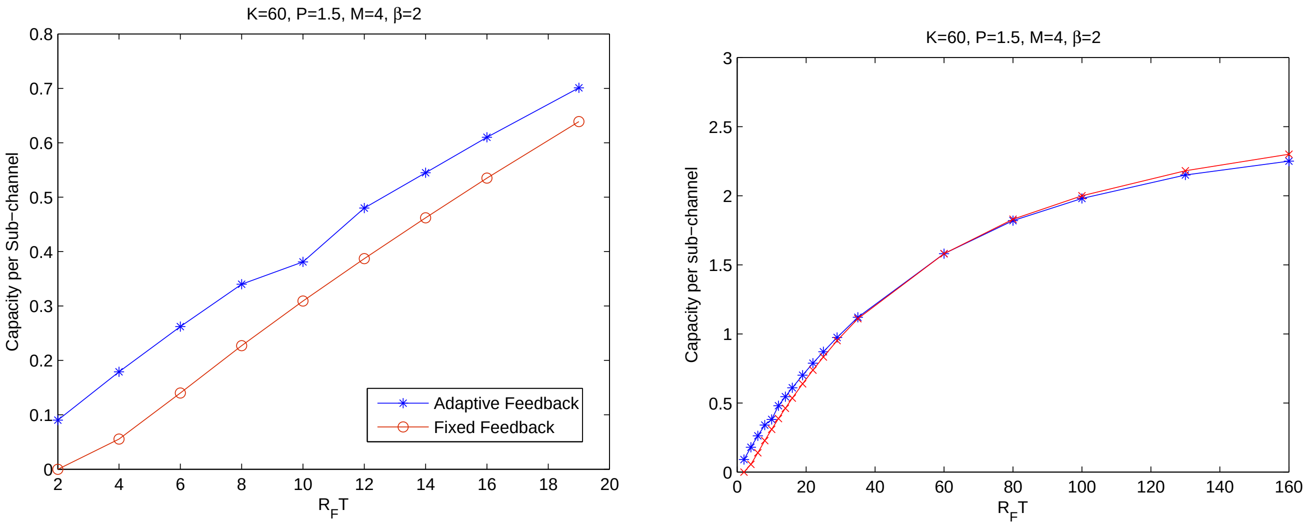

Fig. 2. Performance versus $R_F T$ Small values of $R_F T$ are shown on the left; larger values on the right.

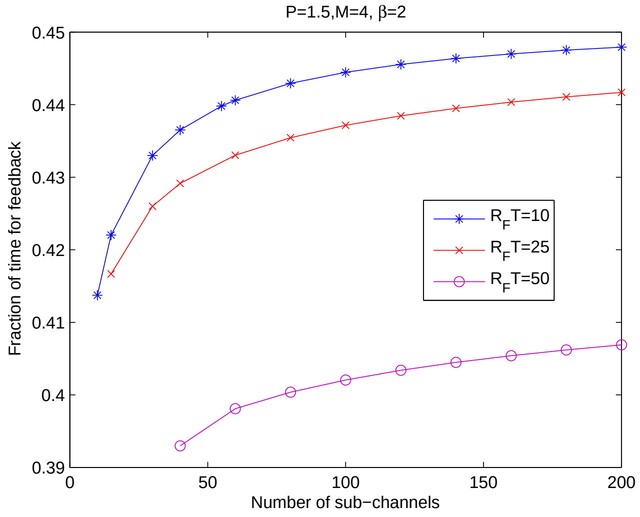

Fig. 3. Average fraction of time for feedback

growth for different values of $R_F T$. In Figure 3, the fraction of time for feedback is around 0.5 for both $R_F T = 25$ and $R_F T = 50$. However, in the case of $R_F T = 50$, the threshold is higher than $R_F T = 25$ and the average number of requested subchannels are lower than $R_F T = 25$. This implies that when $R_F T$ increases, the base station must probe more users to exploit the multiuser diversity.

VI. Conclusion

We have proposed an adaptive limited feedback scheme for downlink MIMO-OFDMA with a finite coherence time T and limited feedback link capacity R_F. This scheme is based on using an optimal stopping rule to determine the feedback time adaptively within one coherence time interval. We characterized the optimal stopping rule and studied the scaling behavior of the system parameters and the sum-throughput as the number of users and number of sub-channels scale with fixed ratio. In this limiting regime the optimized system was shown to have a sum-throughput that scales linearly with the system size for any load. In this paper, we have assumed that the codebook size is equal to the number of transmit antennas. Potential future directions include considering larger codebooks as well as models

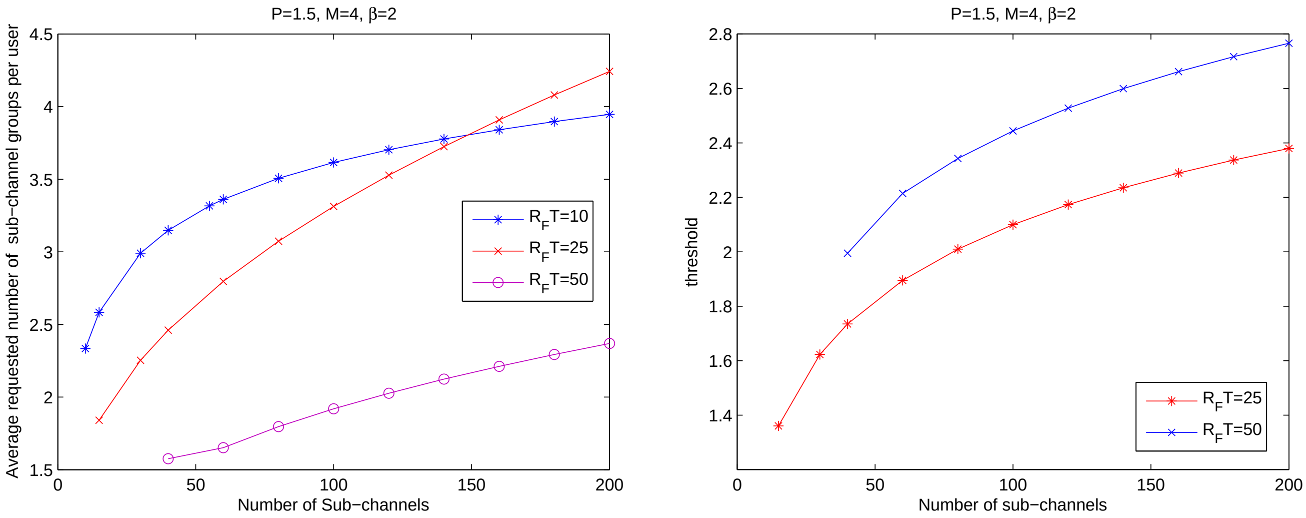

Fig. 4. The average number of sub-channel groups requested (left) and the optimal threshold (right) for the adaptive feedback scheme as a function of the total number of sub-channels.

with correlation in either frequency or time.

REFERENCES

[1] Y. Sun, “Asymptotic Capacity of Multi-Carrier Transmission with Frequency-Selective Fading and Limited Feedback,” *submitted to IEEE Transactions Information Theory*.

[2] J. Chen, R. Berry, M. Honig, “Limited Feedback Schemes for Downlink OFDMA” to appear IEEE Journal on Selected Areas in Communication special issue on ”Exploiting Limited Feedback in Tomorrow’s Wireless Communication Networks”.

[3] G. Wunder, T. Michel, “A (Not So) Many User Sum Capacity Analysis of the MIMO-OFDM Broadcast Channel,” *International ITG/IEEE Workshop on Smart Antennas (WSA)*, Duisburg, Germany, Apr. 2005.

[4] J. Chen, R. Berry, M. Honig, “Large System Performance of Downlink OFDMA with Limited Feedback,”*IEEE International Symposium on Information Theory*, Seattle, Wa, July 2006.

[5] S. Sanayei and A. Nosratinia, “Opportunistic downlink transmission with limited feedback,” *submitted to IEEE Transactions on Information Theory*, Aug. 2005.

[6] J. Chen, R. Berry, M. Honig, “Asymptotic Analysis of Downlink OFDMA Capacity,”*Annual Allerton Conference on Communication, Control and Computing*, Monticello, IL, USA, September, 2006.

[7] M. Sharif, B. Hassibi, “On the Capacity of MIMO Broadcast Channels with Partial Side Information,” *IEEE Trans. on Information Theory*. vol.51, no.2, Feb. 2005.

[8] C. Swannack, G. W. Wornell, E. Uysal-Biyikoglu, “MIMO Broadcast Scheduling with Quantized Channel State Information,” *IEEE International Symposium on Information Theory*, Seattle, Wa, July 2006.

[9] T. Yoo, N. Jindal, A. Goldsmith, “Finite-Rate Feedback MIMO Broadcast Channels with a Large Number of Users,”*IEEE International Symposium on Information Theory*, Seattle, Wa, July 2006. *International ITG/IEEE Workshop on Smart Antennas (WSA)*, Duisburg, Germany, Apr. 2005.

[10] T. Ferguson, Optimal Stopping and Applications. [Online]. Available: http://www.math.ucla.edu/ tom/Stopping/Contents.html, 2006.

[11] K. Nadia, M. Bishwarup, L. Geert, R. W. Heath, P. Frederik, “ Interpolation-based multi-mode precoding for MIMO-OFDM systems with limited feedback,” *IEEE Trans. on Wireless Communications*. vol. 6, pp. 1003-1013, 2007.

[12] Jae Yeun Yun, Sae-Young Chung, Jihoon Choi, Yong-Up Jang, Yong Hoon Lee, ” Predictive transmit beamforming for MIMO-OFDM in time-varying channels with limited feedback”. *International Conference on Wireless Communications and Mobile Computing* (IWCMC), 2007.

[13] Hooman Shirani-Mehr, Giuseppe Caire, “Channel State Feedback Schemes for Multiuser MIMO-OFDM Downlink,” *submitted to IEEE Trans. on Communications.*, April, 2008.