Performance of Limited Feedback Schemes for Downlink OFDMA with Finite Coherence Time

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

***Abstract*—We consider the capacity of a downlink Orthogonal Frequency Division Multiple Access (OFDMA) system with limited feedback rate R_F per sub-channel and finite coherence time T. The feedback is used to relay channel state information (CSI) from K users to the base station. The order-optimal capacity growth with Rayleigh fading sub-channels is $\Theta(N \log\log K)$ as N and K increase with fixed ratio, where N is the number of sub-channels. However, to achieve this, previous work requires a feedback rate per subchannel that scales linearly with the system size. Here we explicitly include the feedback overhead when calculating the sum capacity, and study the tradeoff between feedback rate and sum capacity. We propose two limited feedback schemes, one based on sequential transmissions across users and the other based on random access, in which the each feedback bit requests the use of a sub-channel group containing multiple sub-channels. With fixed $R_F T$, the sum capacity for both schemes with optimized sub-channel groups increases as $\Theta(N)$. If $R_F T$ grows faster than $\log K$, then both schemes can achieve the order-optimal capacity growth. We also show that when $R_F T$ is small, the random access scheme performs better than the sequential transmission scheme, whereas the reverse is true for large $R_F T$.**

I. Introduction

Orthogonal Frequency Division Multiple Access (OFDMA) can exploit both frequency and multiuser diversity through an appropriate assignment of users to sub-channels. We consider downlink OFDMA system in which the base station assigns at most a single user to each sub-channel. Given perfect channel state information (CSI), i.e., knowledge of all sub-channel gains across all users, the sum capacity is achieved by assigning the user with the best channel gain to each sub-channel and water-filling the power over the sub-channels. Related optimized power and rate allocations are discussed in [1], [2]. Although those schemes can achieve substantial capacity gains, relative to that with no CSI at the transmitter, the associated feedback required in a mobile environment is likely to be excessive in practice.

The feedback overhead for downlink OFDMA can be substantially reduced by coarsely quantizing the CSI at the receivers before sending it back to the base station. Feedback schemes in which each user feeds back one bit per sub-channel have been proposed and studied in [3]–[5]. Each feedback bit indicates whether or not the particular channel gain exceeds a pre-determined threshold. It shown in [6] that with Rayleigh fading sub-channels the corresponding weighted sum capacity grows as $N \log(\log(K))$, where N is the number of sub-channels, which is same the capacity growth achieved with perfect CSI at the base station. Furthermore, the gap between the capacity with perfect CSI and the one bit feedback scheme is bounded by a constant [6]. Other related work on limited feedback schemes for the downlink *narrowband* Multiple Input Multiple Output (MIMO) channel has appeared in [7]–[10].

Even one-bit feedback per sub-channel is likely to be excessive in many applications. Namely, the total amount of feedback grows as NK, which is much faster than the rate at which the downlink capacity grows. Hence, given a fixed coherence time T, during which the feedback occurs, as N and K scale up, the feedback eventually dominates the coherence time, so that the optimal capacity growth is unsustainable. This problem motivates the work in this paper. Namely, here we assume that both the feedback rate per sub-channel R_F and the coherence time T are fixed, i.e., do not scale up with the number of users K. Also, the duration of the feedback is explicitly modeled as part of the coherence time T. Our objective is then to maximize the sum capacity, accounting for the loss in channel uses due to feedback.

We consider two feedback schemes, which can be used to reduce the feedback rate below one bit per sub-channel. In both schemes, non-overlapping groups of sub-channels are formed, where each group contains the same number of sub-channels. Each feedback bit then requests the use of all sub-channels in that group. Here we assume that a sub-channel group is requested only if all sub-channel gains exceed a threshold. Clearly, the total feedback decreases with the size of the sub-channel groups.

In the first feedback scheme, each user forms a binary vector, which indicates the set of requested sub-channel groups. That vector is losslessly compressed, and the users then transmit their compressed vectors to the base station sequentially. In the second scheme, a group of users is assigned to each sub-channel group. (The user groups may overlap.) Users assigned to a particular sub-channel group then contend for the use of that group via random access. That is, each user transmits an ID over the assigned sub-channel group, provided that all sub-channel gains exceed the threshold, where the number of bits in the ID depends on the number of users assigned to a group. If multiple users request the same sub-

This work was supported in part by the Motorola-Northwestern Center for Seamless Communications and by ARO under grant W911NF-06-1-0339.

channel group, we assume a collision, so that the sub-channel group remains idle. For both schemes, we optimize the size of the sub-channel group and the channel gain threshold, and for the random access scheme, we also optimize the size of the user groups.

As in [4], [6], we assume perfect CSI at the receiver and *i.i.d.* Rayleigh fading sub-channels. We show that for both feedback schemes with fixed R_F and T, the sum capacity grows as $\Theta(N)$[1]. Hence the feedback constraint eliminates the multiuser diversity term $\log(\log(K))$, which is present with unlimited feedback. However, the constants associated with the $\Theta(N)$ growth for both schemes have the form $\log(R_F T)$. Consequently, if $R_F T$ grows faster than $\log(K)$, we obtain the order-optimal growth of $N\log(\log(K))$. We also show that when $R_F T$ is small, the random access feedback scheme outperforms the first feedback scheme, whereas the reverse is true when $R_F T$ is large.

II. System Model

For the downlink OFDMA system considered, the ith received sample for user k, assigned to sub-channel n, is given by

$$y_k^n(i) = \sqrt{h_k^n} e^{j\theta_k^n} x_k^n(i) + w_k^n(i) \tag{1}$$

$1 \le k \le K$, $1 \le n \le N$, where x_k^n is the transmitted symbol, h_k^n is the squared channel gain, θ_k^n is the random phase uniformly distributed in $[0, 2\pi]$, and w_k^n is additive white Gaussian noise with zero mean and unit variance. The channel gains are assumed to be Rayleigh distributed with variance σ^2, and are independent across users and sub-channels. Also, we assume that all channel gains remain constant during a coherence time of T seconds, and that each receiver has perfect CSI, i.e., the gains h_k^n, $1 \le n \le N$, are known at receiver k.

During each coherence time T, the base station assigns users to sub-channels to maximize the sum rate over all users. At most one user can be assigned to any sub-channel. This assignment is based on feedback, which the base station receives from the mobiles during the start of the coherence time T (e.g. in a time-division duplex(TDD) system). [2] We assume a fixed coherence time T, and a limited feedback rate per sub-channel R_F. To reduce the total feedback from all users, we consider two limited feedback protocols: a random access, or contention-based scheme, and a non-contention (sequential feedback) scheme. In both schemes, the feedback is reduced by grouping sub-channels. Namely, each sub-channel group contains αN sub-channels, where $0 < \alpha < 1$. The sub-channel groups do not overlap, so that there are $1/\alpha$ groups. A user k can request a particular sub-channel group $\mathcal{H}_m$, $1 \le m \le 1/\alpha$, provided that $h_k^n \ge t_o$ for all $n \in \mathcal{H}_m$.

[1]We use the notation: $x_K = O(y_K)$ if $\lim_{K\to\infty} \frac{|x_K|}{|y_K|} \le M$; $x_K = \Omega(y_K)$ if $y_K = O(x_K)$; $x_K = \Theta(y_K)$ if $x_K = O(y_K)$ and $x_K = \Omega(y_K)$; $x_K \asymp y_K$ if $\lim_{K\to\infty} \frac{x_K}{y_K} = 1$.

[2]Of course in a TDD system, the base station could also use the uplink traffic to estimate some the channel conditions, provided that the user was transmitting on the uplink. We do not model this source of channel information here.

A. Feedback Protocols

Here we specify the feedback protocols along with the corresponding total feedback and sum rate objectives.

1) Sequential scheme: The sequential feedback scheme is specified as follows:

- Each user can request any sub-channel group. For a particular user k the set of requests is represented by a $(1/\alpha)$-bit feedback vector, where the m^{th} entry is '1' if $h_k^n \ge t_o$ for all $n \in \mathcal{H}_m$, and is '0' otherwise.
- The users transmit their binary feedback vectors sequentially. Each binary vector is losslessly compressed before transmission.
- The base station decodes the compressed feedback bits from all users. If a channel group is requested by more than two users, then the base station randomly assigns one of them to that channel group. If the channel group is not requested by any user, then the group is not used during the coherence time.

The probability that a user requests a particular channel group is $p_o = e^{-\alpha N t_o/\sigma^2}$. Since the feedback bit sequence is *i.i.d.*, its entropy is given by $\frac{1}{\alpha} \times H(p_o)$ where

$$H(p_o) = -p_o \log(p_o) - (1-p_o)\log(1-p_o) \tag{2}$$

is the binary entropy function. According to [11, Thm. 5.4.1], we can find a coding scheme such that the expected codeword length L satisfies $\frac{1}{\alpha}H(p_o) \le L \le \frac{1}{\alpha}H(p_o) + 1$. Since the feedback time-slot allocated to each user should contain at least one bit, we assume that the average number of feedback bits per user is $\frac{1}{\alpha}H(p_o) + 1$.[3]

Suppose that the total feedback rate is NR_F, i.e., it scales linearly with the number of sub-channels N. Also, we scale the number of users K in proportion with N, i.e., $K/N = \rho$. The average duration of the feedback time slot allocated to a particular user is then $\frac{\frac{1}{\alpha}H(p_o)+1}{NR_F}$ channel uses, and the average total feedback time within a coherence time T is $K \times \frac{\frac{1}{\alpha}H(p_o)+1}{NR_F}$. It can be shown that for the optimal system parameters discussed in Section III-A, the total feedback time converges to its mean with probability one, as the system scales. Hence, asymptotically the fraction of the coherence time devoted to feedback is then

$$f_{seq} = \rho \frac{\frac{1}{\alpha}H(p_o) + 1}{R_F T}. \tag{3}$$

We assume that the base station allocates power uniformly over the active sub-channels. Given power P per user, the average power per sub-channel group is then the total power divided by the average number of active channel groups, or $KP/(p_s/\alpha)$, where $p_s = 1-(1-p_o)^K$ is the probability that a channel group is requested by at least one user. The average received Signal-to-Noise Ratio on a particular subchannel n assigned to user k is then $KPh_k^n/[(p_s/\alpha)\times(\alpha N)] = \rho P h_k^n/p_s$.

[3]Here we ignore additional feedback, which may be required to demarcate the user transmissions. For example, this may be required if the users are unable to decode the feedback transmissions from all other users. Accounting for this additional feedback does not change the main results presented in Section III-A.

We assume that the code rate is matched to the channel threshold t_o, so that with an optimal code the achievable rate per active sub-channel is

$$r_{seq} = \log\left(1 + \frac{\rho P t_o}{p_s}\right). \quad (4)$$

Accounting for the feedback in (3) as part of the coherence time enables us to write the average sum rate as

$$\tilde{R}_{seq} = N p_s r_{seq}(1 - f_{seq}) \quad (5)$$

where Np_s is the average number of active sub-channels. In what follows, we maximize $\tilde{R}_{seq}$ over the parameters α and t_o, giving the optimized objective

$$R_{seq} = \max_{\alpha, t_o} \tilde{R}_{seq}. \quad (6)$$

2) Contention scheme: The contention, or random access scheme is defined as follows:

- For each sub-channel group, βK users are allowed to contend for that group, where $0 < \beta < 1$. Each user can therefore request only a subset of available sub-channel groups.
- To request a sub-channel group (i.e., if the channel gains are above the threshold), user k transmits $\log(\beta K) + 1$ identification bits over the associated αN sub-channels.
- The base station allocates the group to the user whose feedback bits are successfully received by the base station. If multiple users contend for the same group, then a collision occurs, and the group remains idle.

Here $\log(\beta K)$ feedback bits are needed to identify a user within the user group assigned to a particular sub-channel group. The additional bit ensures that at least one feedback bit is sent. Instead of allocating one dedicated time slot for each user, as in the sequential scheme, all βK users simultaneously access the same bandwidth to transmit their feedback bits. A sub-channel group is assigned to a user if and only if one out of βK users requests that sub-channel group.

In analogy with (5), the sum capacity objective is

$$\tilde{R}_{con} = N p_t r_{con}(1 - f_{con}), \quad (7)$$

where $p_t = \beta K e^{-\alpha N t_o/\sigma^2}(1 - e^{-\alpha N t_o/\sigma^2})^{\beta K - 1}$ is the probability that a single user requests a sub-channel group,

$$r_{con} = \log\left(1 + \frac{\rho p}{p_t} t_o\right), \quad (8)$$

and

$$f_{con} = \frac{\log(\beta K) + 1}{\alpha N R_F T} \quad (9)$$

is the fraction of the coherence time used for feedback.

We can again maximize $\tilde{R}_{con}$ over the parameters α and t_o, and the additional parameter β giving the optimized objective

$$R_{con} = \max_{\alpha, \beta, t_o} \tilde{R}_{con}. \quad (10)$$

In what follows, we will compare the performance of the sequential and contention schemes as a function of the feedback $R_F T$.

III. Main Results

A. Capacity Growth Order

If there is no limit on the feedback rate and/or the coherence time, then the sum-capacity grows at rate $\Theta(N \log(\log(K)))$ as N and K increase with fixed ratio. In this section, we characterize this growth rate for the two schemes in the previous section, assuming that the feedback rate and coherence time are fixed.

Lemma 1: In the optimal sequential scheme, the probability that a user requests a channel group is decreasing as $\Theta(1/K)$ as $K \to \infty$. The optimal grouping size is increasing as $\Theta(\log(K))$. The average number of channel groups requested by one user is decreasing as $\Theta(1/\log(K))$.

Here, by "optimal" we mean that the parameters α and t_o are optimized for each K and N. The key idea behind this lemma is that under the optimal scheme, the total feedback time must be bounded. Both decreasing the probability that a channel is requested and increasing the grouping size help to limit this quantity. However, these also decrease the sum capacity if the feedback overhead is not taken into account. Hence, the optimal scheme must decrease these "fast enough", but not too fast. To determine the optimal rate, as in [4], we use results from extreme order statistics [12] to characterize the asymptotic probability that a channel is requested. We omit the detailed proof due to space considerations.

Lemma 2: In the contention scheme, if $t_o \to \infty$ as $K \to \infty$, then to have a non-zero asymptotic rate, it must be that $\beta K \to \infty$ and $\alpha N \to \infty$.

In other words, in the contention scheme, if the threshold approaches infinity, the channel group size and the number of users per group must also.[4] This follows since increasing t_o increases the transmission rate on a successful channel, but also effects the probability of success p_t. To keep p_t from going to zero too quickly, βK must increase. If βK increases, then αN must also increase to keep the feedback time bounded.

Using these two lemmas, we have the following proposition.

Proposition 1: Given a fixed value of $R_F T$, both R_{seq} and R_{con} grow as $\Theta(N)$, as $N \to \infty$.

B. Performance Comparison

Proposition 1 shows that the sum capacity of both schemes increases as $\Theta(N)$. Next we compare the performance of the two schemes in terms of their asymptotic first order constant. Let γ denote this constant for the optimal sequential scheme, i.e., $R_{seq} \asymp \gamma N$ as $N \to \infty$. Using asymptotic order statistics, it can be shown that this constant has the form

$$\gamma = \nu\left(1 - \frac{-\log(1-\mu)\mu(e^{\frac{\nu}{\mu}} - 1)}{\rho P \sigma^2 R_F T} - \frac{\rho}{R_F T}\right), \quad (11)$$

where $\mu \triangleq p_s$ is the asymptotic probability that a group is requested by more than one user and $\nu \triangleq \mu \log(1 + \frac{\rho P}{\mu} t_o)$ represents the asymptotic rate per sub-channel carrying the

[4]Note that this does imply that in the optimal contention scheme, the threshold approaches infinity.

downlink data. To find the first order constant, we must maximize (11) over μ and ν. The Karush-Kuhn-Tucker (KKT) conditions for this optimization problem result in the following two equations that the optimal μ and ν must satisfy:

$$\log(1-\mu)(e^t - 1 - te^t) = \frac{\nu}{1-\mu}(e^t - 1), \quad (12)$$

$$\rho P\sigma^2 R_F T - \rho^2 P\sigma^2 = -\frac{\log(1-\mu)}{\mu}((t+1)e^t - 1), \quad (13)$$

where $t \triangleq \nu/\mu$.

For the contention scheme, there are three parameters to optimize over. Furthermore, we can't necessarily apply asymptotic order statistics in this case, because the optimal number of channel groups may not approach infinity as K increases. However, we can still compare the schemes in two extreme cases: large $R_F T$ and small $R_F T$.

Theorem 1: There exists constants $b_1^* \geq b_2^* \geq \rho$, such that when $R_F T > b_1^*$ ($R_F T < b_2^*$) the first order constant of the sequential scheme is greater than (less than) that of the contention scheme.

We give a brief sketch of the proof of this next. We consider the following two cases: (a.) $R_F T$ is large, and (b.) $R_F T$ is small. For case (a.), it can be shown that the $\frac{\rho}{R_F T}$ term in (11) can be neglected. We then compare the first order constants of the two schemes by assuming that the fraction of time devoted to feedback and the channel threshold for both schemes lie within one of three given sets. Within each set, the sequential scheme performs better than the contention scheme. For case (b.), we show that as $R_F T$ decreases to ρ, the capacity of the sequential scheme approaches 0 while the capacity of collision scheme is bounded away from 0.

We conjecture that the first order constant of both schemes is an increasing and concave function of $R_F T$ and that they cross at one point (i.e. $b_1^* = b_2^*$). The numerical results below support this statement.

C. Impact of $R_F T$ on Capacity

In [4], the capacity achieved by a one bit feedback scheme is shown to increase as $\Theta(N \log(\log(K)))$, which is the same as optimal growth rate with full CSI at the base station. However, in Prop. 1, the capacity only scales like $\Theta(N)$; i.e. there is no longer the multiuser diversity gain of $\log(\log(K))$. As we will show next, this is due to the constraint that $R_F T$ is fixed. Indeed, for the one bit feedback scheme in [4], each user sends back one bit per sub-channel. Thus the total amount of feedback per sub-channel is K bits. In our model, this would result in a feedback time of $\frac{K}{R_F}$, which would eventually exceed T. To prevent this from happening, in the one-bit feedback scheme, $R_F T$ would need to increase linearly with K as the system scales. The next proposition shows that for the two schemes considered here $R_F T$ only needs to increase faster than $\log(K)$ to recover the multi-user diversity gains.

Proposition 2: If $R_F T$ increases slower than $\Theta(\log(K))$ as $K \to \infty$, then R_{seq} and R_{con} both increase as $\Theta(N \log(R_F T))$. If $R_F T$ increases faster than $\Theta(\log(K))$, R_{seq} and R_{con} both increase as $\Theta(N \log(\log(K)))$.

We give a brief sketch of the proof. If $R_F T$ increases faster than $\log(K)$, then the fraction of time used for feedback in the contention scheme satisfies

$$\frac{\log(\beta K)+1}{\alpha N R_F T} \leq \frac{\log(K)+1}{R_F T} \to 0.$$

Therefore, $R_{con}/N \asymp P_t \log(1 + \frac{\rho p}{P_t} t_o)$. We construct a lower bound on R_{con}/N by setting $\alpha N = 1$, $\beta K = K$ and $t_o = \sigma^2 \log(K)$. For these parameters, $P_t = K e^{-\log(K)}(1 - e^{-\log(K)})^{K-1} \to e^{-1}$, and the throughput per sub-channel grows as $\Theta(N \log(\log(K)))$. Therefore, R_{con} is increasing at least as fast as $\Theta(N \log(\log(K)))$. From Theorem 1, R_{con} is upper bounded by R_{seq} when $R_F T$ is large. Both R_{con} and R_{seq} are upper bounded by the optimal sum capacity with full CSI, which also increases as $\Theta(N \log(\log(K)))$. Combining these observations it follows that R_{con} and R_{seq} both increase as $\Theta(N \log(\log(K)))$.

If $R_F T$ increases slower than $\Theta(\log(K))$, it can be shown that R_{seq} increases like $\Theta(N \log(R_F T))$. For the contention scheme, we first lower bound R_{con} by setting $\alpha N = 1$ and $t_o = \sigma^2 \log(\beta K)$. With these settings, $p_t = (1 - \frac{1}{\beta K})^{\beta K - 1}$. As βK increases, p_t is lower bounded by e^{-1}. Therefore, we have

$$\begin{aligned} &R_{con}/N \\ &\geq e^{-1} \log\left(1 + \frac{\rho P \sigma^2}{e^{-1}} \log(\beta K)\right)\left(1 - \frac{\log(\beta K)+1}{R_F T}\right) \\ &= \Theta(\log(R_F T)). \end{aligned}$$

Theorem 1 shows that R_{con} is upper bounded by R_{seq} when $R_F T$ is large. Therefore, R_{con} also scales $\Theta(N \log(R_F T))$.

Note that if the base-station does not have any CSI (zero feedback bits) and codes over many channel realizations, it can achieve an average sum capacity of

$$R_{nf} = N \int_{x=0}^{\infty} \log(1 + \rho P x) dF(x) x, \quad (14)$$

where $F(x)$ is the cumulative distribution function of the channel gains. This quantity is also increasing as $\Theta(N)$. However, Proposition 2 implies that the first order constants of the two limited feedback schemes increase with $R_F T$, while the constant for R_{nf} does not. This implies that for large enough values of $R_F T$, these schemes will perform better than a no feedback scheme; however, the improvement does not increase the first order growth rate.

IV. Numerical Results

In this section, we provide some numerical examples which illustrate the asymptotic performance of both feedback schemes. For all the users, the channel gains are modeled as Rayleigh with variance $\sigma^2 = 1$, and we set the power per user, $P = 10$ (10 *dB*). Figure 1 shows the optimized first order constant for each scheme as a function of different values of $R_F T$ and for different loads, $\rho = K/N$. For each scheme, we optimized this constant numerically over the relevant parameters (i.e. t_o, α and β). As stated in Theorem 1, for a given ρ, when $R_F T$ is small the contention scheme has

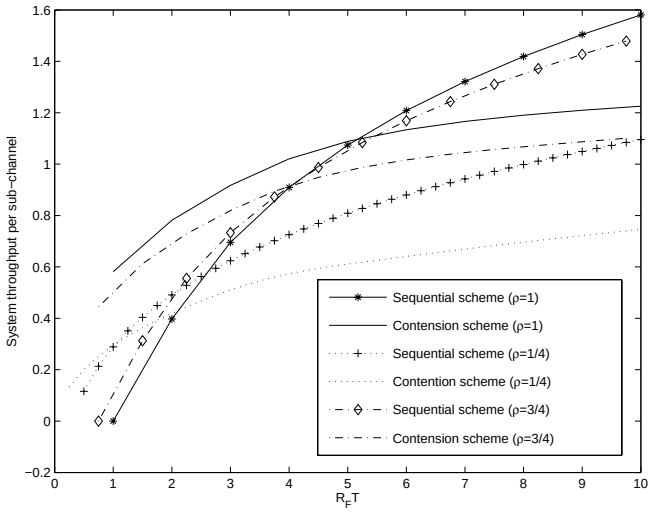

Fig. 1. Comparison of the first order constants of each scheme for different values of $R_F T$ and loads ρ.

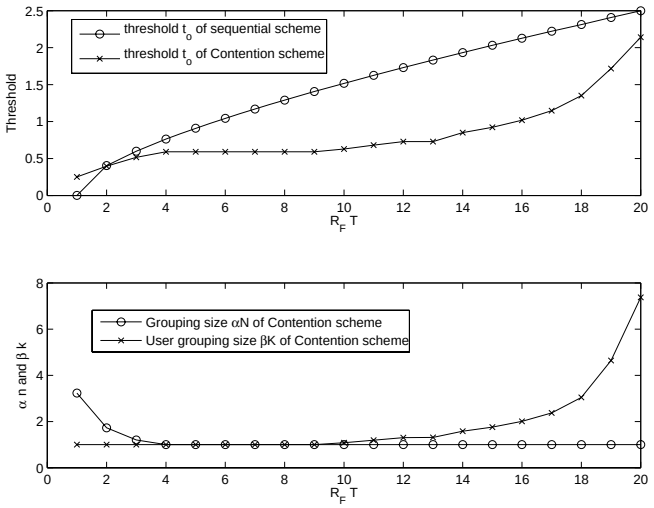

Fig. 2. Optimized parameters for each scheme versus $R_F T$.

the larger constant, while when $R_F T$ is large the sequential scheme performs better. For each of these cases, there is a single crossing point as conjectured after Theorem 1. The crossing point is shifted to the right as ρ increases. The sum capacity of the sequential scheme goes to 0 when $R_F T$ approaches ρ. This is because the entire coherence time T is used for feedback.

Figure 2 show the optimized parameters for both schemes as a function of $R_F T$, when $\rho = 1$. The top part shows the optimal asymptotic thresholds for the two schemes. For both schemes, the optimal thresholds converge to a finite value that increases with $R_F T$. This can be contrasted with the one-bit feedback scheme in [4], in which the optimal thresholds approach infinity as the system scales. The lower part of the figure shows the optimal group sizes for the contention scheme.[5] As $R_F T$ increases, the number of channels in each group decreases to 1, while the number of users in each group increases.

V. Conclusion

We have presented two feedback schemes for downlink OFDMA with finite coherence time T and limited feedback link capacity R_F. The capacity growth for both schemes is $\Theta(N)$ as the number of users and sub-channels increase, although the multi-user diversity $\log \log K$ term can be recovered if the feedback $R_F T$ is allowed to increase as $\log K$.

For purposes of analysis, we have assumed that a user requests a sub-channel group only if all sub-channel gains in the group exceed a pre-determined threshold. This criterion is rather severe, and it is of interest to re-evaluate performance with other selection criteria, e.g., based on associated rates. Also, we have assumed perfect CSI at the receiver, and have not accounted for the additional overhead associated with channel estimation. Finally, we have assumed a single antenna at the base station. Extending the feedback schemes considered here to MIMO OFDMA, and examining the associated tradeoff between downlink capacity and feedback is an interesting direction for future work.

References

[1] J. Huang, V. Subramanian, R. Agrawal, and R. Berry, "Downlink Scheduling and Resource Allocation for OFDM Systems," *Proc. of 40th Annual Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, March 2006.

[2] L. Hoo, B. Halder, J. Tellado, and J. Cioffi, "Multiuser Transmit Optimization for Multicarrier Broadcast Channels: Asymptotic FDMA Capacity Region and Algorithms," *IEEE Trans. on Communications*, vol. 52, no. 6, June 2004.

[3] Y. Sun, "Asymptotic Capacity of Multi-Carrier Transmission with Frequency-Selective Fading and Limited Feedback," *submitted to IEEE Trans. on Information Theory*.

[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 Trans. 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] S. Sanghavi, B. Hajek, "Adaptive Induced Fluctuations for Multiuser Diversity," *IEEE International Symposium on Information Theory*, Lausanne, Switzerland, July 2002.

[9] 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.

[10] 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.

[11] T. Cover, J. Thomas, *Elements of Information Theory*, John Wiley and Sons, New York, 1991.

[12] J. Galambos, *The Asymptotic Theory of Extreme Order Statistics*, John Wiley and Sons, New York, 1978.

[5]From Lemma 2, we know that the optimal group size in the sequential scheme approaches infinity.