Limited Feedback Schemes for Downlink OFDMA Based on Sub-Channel Groups

Jieying Chen, Randall A. Berry, Michael L. Honig

***Abstract*—In a downlink Orthogonal Frequency Division Multiple Access (OFDMA) system, optimally allocating sub-channels across mobile users can require excessive feedback of channel state information (CSI). We consider an OFDMA model in which the feedback overhead is explicitly taken into account, given a fixed feedback rate and finite coherence time. The tradeoff between feedback rate and sum capacity is studied for two limited feedback schemes: a sequential scheme in which the users send compressed feedback bits over consecutive time slots, and a contention scheme in which users send their feedback via a random access protocol. For both schemes each feedback bit indicates a request for a group containing multiple sub-channels. We show that the sum capacity for both schemes with optimized sub-channel groups grows linearly with the number of sub-channels N, and that the associated constant increases as the log of the normalized feedback rate measured in bits per coherence time per sub-channel. We also compare the asymptotic (large N) performance of the two limited feedback schemes as a function of the feedback rate and load (users per sub-channel). The sequential scheme performs best with moderate to large feedback rates, or small loads, whereas the contention scheme performs best with small feedback rates or large loads.**

I. INTRODUCTION

Orthogonal Frequency Division Multiple Access (OFDMA) can exploit both frequency and multiuser diversity through an appropriate assignment of users to sub-channels. Given perfect Channel State Information (CSI) at the Base Station (BST), 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]–[3]. Although those schemes can achieve substantial capacity gains, relative to having 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 BST. Schemes for limiting the feedback in single-user OFDM and downlink OFDMA have been studied recently, including [4]–[12], [15]. In many of these schemes (e.g., [5]–[7], [9], [12]) each user provides one bit of feedback per sub-channel to indicate whether or not the particular channel gain exceeds a pre-determined threshold.

Even one-bit feedback per sub-channel can be excessive as the system size scales. Namely, with perfect CSI at the BST the sum capacity grows as $N \log \log K$, assuming Rayleigh fading sub-channels, where N is the number of sub-channels and K is the number of users [7], [9], [13]. This optimal order of growth can also be achieved with one-bit feedback per user per sub-channel. Yet the total feedback rate scales as NK. Hence given a fixed coherence time T, during which the feedback occurs, as N and K increase, the feedback eventually dominates the coherence time, so that the capacity growth is unsustainable.[1] This problem motivates the feedback model in this paper. Namely, here we assume that both the feedback rate per sub-channel R_F (in bits per second) and the coherence time T are fixed, i.e., do not scale with the number of users K. Also, the duration of the feedback is explicitly modeled as part of the coherence time T. This corresponds to a time-division duplex (TDD) system in which the BST waits for CSI feedback on the feedback link before transmitting.[2] 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 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. The receiver requests a sub-channel group only if all sub-channel gains in the group exceed a threshold. The total feedback therefore 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 BST 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

[1]Throughput the paper, we assume that each sub-channel corresponds to a coherence band. Hence increasing N corresponds to increasing the system bandwidth.

[2]In a TDD system, the BST may be able to estimate some CSI from the uplink traffic, assuming reciprocity. We do not model this possibility. One reason for this is that when each user transmits on a relatively small fraction of the uplink channels, the CSI gained through such an approach will be small. Also, in practice channel reciprocity may not hold even when the uplink and downlink share the same sub-channels so that explicit CSI feedback is still needed.

The authors are with the Dept. of EECS, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208 USA, e-mail: j-chenm@northwestern.edu, {rberry,mh}@eecs.northwestern.edu. This work was presented in part at the 2007 IEEE International Symposium on Information Theory, Nice, France, and at Milcom 2007. This work was supported in part by the Motorola-Northwestern Center for Seamless Communications and by ARO under grant W911NF-06-1-0339.

identification bits over the assigned sub-channel group, provided that all sub-channel gains exceed the threshold, where the number of bits in the identification depends on the number of users assigned to a group. If multiple users request the same sub-channel group, a collision occurs, and 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.

With perfect CSI at the receivers 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 $\Theta(N)$ as $N \to \infty$ and $K \to \infty$ with a fixed ratio $\rho = K/N$.[3] Hence the feedback constraint eliminates the multiuser diversity term $\log \log K$, which is present with unlimited feedback (perfect CSI at the BST). However, the constants associated with the $\Theta(N)$ growth for both schemes have the form $\log(R_F T)$, where $R_F T$ is the *normalized* feedback rate per sub-channel measured in bits per coherence time. Consequently, if $R_F T$ is allowed to grow faster than $\log K$, we obtain the order-optimal sum capacity growth of $N \log \log K$.

We also compare the performance of the two feedback schemes for large N and K. When $R_F T$ is small or the load $\rho = K/N$ is large, the contention feedback scheme achieves a higher sum capacity than the sequential scheme, whereas the reverse is true when $R_F T$ is large or ρ is small. Numerical results are presented, which show that the sequential scheme generally achieves a higher sum capacity over a range of parameters of interest (i.e., $\rho < 1$ and $R_F T > 10$ bits). However, the sequential scheme requires additional synchronization overhead for the user feedback. Additional numerical results show that the asymptotic (large N) analysis is accurate for moderate system sizes (e.g., $K > 50$).

Related work on limited feedback schemes for OFDM and downlink OFDMA has been presented in [4]–[12], [15]. A key distinguishing feature of our model is the ability to vary the tradeoff between feedback rate and downlink performance (achievable sum rate) by varying the size of the sub-channel groups along with the activation threshold. (Sub-channel groups are used to reduce feedback for an uplink Multi-Input/Multi-Output (MIMO) OFDM model in [22].) Furthermore, the associated performance analysis explicitly accounts for feedback overhead. Other related work on limited feedback for the narrowband MIMO downlink has been presented in [14], [16]–[18]. Analogous asymptotic scaling results are presented in those references, although again the total feedback across users (including user identification) is not explicitly modeled as part of the downlink overhead.

In the next section we present the system model and specify the two limited feedback schemes to be analyzed. The main asymptotic scaling results are presented in Section III, and numerical performance results, which compare the performance of the two schemes and illustrate the accuracy of the asymptotic analysis for finite-size systems, are presented in Section IV. Proofs of the main results are given in the appendices.

[3] We use the following notation: for any $\overline{y} \in [0, \infty]$, as $y \to \overline{y}$, $f(y) = o(g(y))$ if $\lim_{y \to \overline{y}} \frac{|f(y)|}{|g(y)|} = 0$, $f(y) = \Theta(g(y))$ if $\lim_{y \to \overline{y}} \frac{|f(y)|}{|g(y)|} = M$, $0 < M < \infty$, and $f(y) \asymp g(y)$ if $\lim_{y \to \overline{y}} \frac{|f(y)|}{|g(y)|} = 1$.

II. System Model

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

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

$1 \leq k \leq K$, $1 \leq n \leq N$, 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, $1 \leq n \leq N$, are known at receiver k. The *i.i.d.* Rayleigh sub-channels can be interpreted as a set of coherence bands, corresponding to a block fading model in the frequency domain. In particular, these need not be single OFDMA tones, but could be represent groups of tones within a coherence band.

During each coherence time T, the BST 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 BST receives from the mobiles during the start of the coherence time T (e.g. in a TDD system). 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 sequential scheme, in which the users sequentially transmit feedback, and a contention, or random access scheme. In both schemes, the feedback is reduced by grouping sub-channels. Namely, each sub-channel (or 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 $\mathcal{H}_m$, $1 \leq m \leq 1/\alpha$, provided that $h_k^n \geq t_o$ for all $n \in \mathcal{H}_m$.

A. Feedback Protocols

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

1) Sequential scheme:

- Each user can request any 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 $h_k^n \geq t_o$ 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 using a fixed-to-variable length source code.
- The BST decodes the compressed feedback bits from all users. If a channel group is requested by more than two users, then the BST randomly assigns one of those users

to that channel group. If a channel group is not requested by any user, then it is not used.[4]

The probability that a user requests a particular channel group is the probability that all channels in that group exceed t_o given by $p_o = e^{-\alpha N t_o/\sigma^2}$. 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 [19, Thm. 5.4.1], there exists a source coding scheme such that the expected codeword length L satisfies $\frac{1}{\alpha}H(p_o) \leq L \leq \frac{1}{\alpha}H(p_o)+1$. This corresponds to a variable-length coding scheme so that the feedback varies across users (depending on the channel conditions). 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 $\frac{1}{\alpha}H(p_o)+1$.[5]

Suppose that the total feedback rate is NR_F, i.e., it scales linearly with the number of sub-channels N. (We implicitly assume that the feedback is coded across all coherence bands.) Also, we scale the number of users K in proportion with N, with a fixed load of $\rho = K/N$. 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 (mean) 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 (i.e., K and N both tend to infinity with fixed ρ). Hence asymptotically the fraction of the coherence time devoted to feedback is

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

The BST 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.[6] The average received Signal-to-Noise Ratio on a particular sub-channel n assigned to user k is then $KPh_k^n/[(p_s/\alpha)\times(\alpha N)] = \rho P h_k^n/p_s$. We assume that the code rate is matched to the channel threshold t_o, so that for large N using an optimal code the achievable rate per active sub-channel is

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

where we assume that the transmitter codes across multiple coherence blocks in frequency and/or time (so that there are enough degrees of freedom available to approach this rate).[7]

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

$$\tilde{R}_{seq} = Np_s r_{seq}(1-f_{seq})^+ \tag{5}$$

where Np_s is the average number of active sub-channels. Here, $(1-f_{seq})^+ = \max(0, 1-f_{seq}$ which denotes the fact that if $f_{seq} > 1$ then there is no time remaining to send data. Namely, as K and N both tend to infinity with fixed ρ, the fraction of time devoted to the feedback converges to the mean f_{seq} in (3). In what follows, we maximize $\tilde{R}_{seq}$ over the parameters $\alpha \in [0,1]$ and $t_o \geq 0$, giving the optimized objective

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

We will compare this rate with the corresponding rate for the contention scheme, defined next.

2) Contention scheme:

- For each channel group, βK users are allowed to contend for that group, where $0 < \beta < 1$. Each user can be assigned to multiple groups, and can request only those groups to which she has been assigned.
- To request a 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 BST allocates the channel group to the user who successfully contends for the group. If multiple users contend for the same group, a collision occurs, and the group remains idle.

A minimum of $\log(\beta K)$ feedback bits are needed to identify a user within the user group assigned to a particular channel group. The additional bit assumed here 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 channel group is assigned to a user if and only if one out of βK users requests that channel group.

In analogy with (5), the average sum capacity objective for large N is

$$\tilde{R}_{con} = Np_t r_{con}(1-f_{con})^+, \tag{7}$$

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,

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

and

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

is the fraction of the coherence time used for feedback. Note that in this case f_{con} is a deterministic constant for all K.

[4]Alternatively, each sub-channel group requested by more than one users could be divided among the users. That would not change the total sum rate, which is determined by the threshold, although it may improve fairness.

[5]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.

[6]For the asymptotic results with fixed ρ the total power scales linearly with the number of users K, or equivalently, with the number of sub-channels N.

[7]We assume that the rate of a user is matched to t_0 instead of the actual channel gain since given the feedback the BST does not know the actual gain. If the BST codes over many coherence times, it could transmit at a higher rate given by the expected rate conditional on the channel exceeding t_0. However this does not change the asymptotic analysis that follows.

We can again maximize $\tilde{R}_{con}$ over the parameters α and t_o, and the additional parameter $\beta \in [0, 1]$ giving the optimized objective

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

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

III. Analytical Performance Results

In this section we state our main analytical results, which include a characterization of the sum rate of the limited feedback schemes, and optimized group size and threshold for the sequential scheme. These results are asymptotic as N and K tend to infinity with fixed ratio $\rho = K/N$.

A. Capacity Growth Order

If there is no limit on the feedback rate and/or the coherence time, then the sum-capacity grows at the rate $\Theta(N \log \log K)$ [7], [9], [13]. In this section, we specify the corresponding growth rate for the two schemes described in the preceding section assuming that the normalized feedback rate $R_F T$ is fixed.

Proposition 1: For any $R_F T > \rho$, $R_{seq} = \Theta(N)$ as $N \to \infty$. For any $R_F T > 0$, $R_{con} = \Theta(N)$, as $N \to \infty$.

The proof is given in Appendix A.[8]

The proposition states that accounting for feedback overhead in the downlink model reduces the order growth in capacity of both schemes from $N \log \log K$ to N. Of course, this result applies only to the two feedback schemes considered. Proving a converse, which states that no feedback scheme can achieve a capacity growth faster than $\Theta(N)$, appears to be difficult. This is mainly due to the many different types of feedback schemes, which could be constructed. However, the following observation suggests that the converse may be true. Suppose that there is a single sub-channel shared by K users. Given a limited total feedback rate, the number of users that can feed back any information (e.g., one bit) is also limited. Hence as $K \to \infty$, only a finite number of users can contend for the sub-channel, eliminating the asymptotic growth in K due to multiuser diversity. Extending this argument to the limit in which both $N \to \infty$ and $K \to \infty$ with fixed ratio does not appear to be straightforward.

According to Prop. 1, the order-growth with limited feedback is the same as with no feedback. Namely, for the channel model considered, in the absence of any feedback the ergodic sum capacity does not depend on how the N sub-channels are allocated among users. The maximum order growth in that scenario is also $\Theta(N)$. Feedback does, however, affect the associated first-order constant. This dependence will be specified in the next subsection.

The asymptotic behavior of the optimal channel group size, threshold, and probability that a channel group is requested for the sequential scheme is summarized in the next proposition. Here "optimal" means that the group size α and threshold t_o are optimized for each K and N.

[8]The behavior of R_{seq} when $R_F T \leq \rho$ is discussed in Section III-D.

Proposition 2: In the optimal sequential scheme, the probability that a user requests a channel group decreases as $\Theta(1/K)$ as $K \to \infty$. The optimal group size increases as $\Theta(\log K)$ and the average number of channel groups requested by one user decreases as $\Theta(1/ \log K)$.

The proof is given in Appendix B. The proposition implies that for large K and N, the optimal number of channel groups is proportional to $N/(\log K)$. Hence each user requests on the order of $N/(K \log K) = 1/(\rho \log K)$ sub-channel groups. The optimized parameters reflect the tradeoff between the benefit and cost of reducing the amount of feedback. Namely, decreasing the probability that a channel group is requested and increasing the channel group size reduces the feedback, which allows more of the coherence time to be used to transmit information symbols. However, this also decreases the achievable rate per transmitted information symbol. The proof of the proposition relies on results from extreme order statistics [20], which are used to characterize the asymptotic probability that a channel is requested.[9]

Characterizing the asymptotic growth of α, t_o, and β for the contention scheme is more difficult than for the sequential scheme. This is due to the more complicated interactions among the three parameters, instead of the two for the sequential scheme.[10] We are therefore unable to characterize this behavior explicitly, although we have the following condition on the threshold.

Proposition 3: In the contention scheme, if $t_o \to \infty$ as $K \to \infty$, then the asymptotic rate can be positive only if $\beta K \to \infty$ and $\alpha N \to \infty$.

In other words, if the threshold approaches infinity, then so must the channel group size and the number of users per group.[11] The proposition follows from the proof of Prop. 1. Essentially, letting $t_o \to \infty$ increases the transmission rate on a successful channel, but also drives the probability of success p_t to zero. To achieve a positive asymptotic rate, p_t cannot converge to zero too quickly, so that βK must increase. If βK increases, then αN must also increase to keep the feedback time bounded.

B. Relative Performance versus Feedback Rate

Proposition 1 shows that the sum capacity of both schemes increases as $\Theta(N)$ (provided that $R_F T > \rho$). Next we compare the performance of the two schemes in terms of their asymptotic first-order constants. Let γ_{seq} (γ_{con}) denote this constant for the optimal sequential (contention) scheme, e.g., $R_{seq} \asymp \gamma_{seq} N$ as $N \to \infty$.

To determine the first-order constants, we must optimize each scheme over the relevant parameters. However for each scheme, it appears difficult to find closed-form expressions for these constants. Indeed as we indicated in the discussion before Prop. 3, it even appears difficult to characterize the

[9]See also [6], in which related techniques are used to derive asymptotic results for the one-bit feedback scheme.

[10]Moreover, because β can change with K, the number of users on a channel group may not grow without bound, which precludes us from applying the results from extreme order statistics used to analyze the sequential scheme.

[11]Note that this does not imply that the optimized threshold $t_o \to \infty$.

asymptotic behavior of the optimized parameters for the contention scheme (though this is possible for the sequential case). However, we are able to compare the performance of these two schemes for large $R_F T$ and small $R_F T$.

Proposition 4: There exist constants $b_1^* \geq b_2^* \geq \rho$, such that $\gamma_{seq} > \gamma_{con}$ when $R_F T > b_1^*$, and $\gamma_{seq} < \gamma_{con}$ when $R_F T < b_2^*$.

A sketch of the proof is given in Appendix C. The numerical results in the next section suggest that the first-order constants for both schemes are increasing and concave functions of $R_F T$, and that they cross at a single point (i.e. $b_1^* = b_2^*$).

C. Effect of Feedback on Capacity

Prop. 1 states that the asymptotic order-growth in sum capacity depends on neither the feedback rate R_F nor the coherence time T. In contrast, previous work [6], [9] has shown that feedback of one bit per sub-channel can achieve the optimal growth rate in sum capacity of $\Theta(N \log\log K)$. With that scheme the total amount of feedback per sub-channel is K bits. In our model with finite R_F and T, that would result in a feedback time of $\frac{K}{R_F}$, which exceeds T for large enough K. Hence as K and N become large, the sum capacity for the one-bit feedback scheme tends to zero unless $R_F T$ increases (at least) linearly with K. The next proposition states that for the two schemes considered here $R_F T$ only needs to increase as $\log K$ to recover the increase in sum capacity due to multi-user diversity.

Proposition 5: If $R_F T$ increases as $o(\log K)$ as $K \to \infty$, then R_{seq} and R_{con} both increase as $\Theta(N \log(R_F T))$ If $R_F T$ increases at least as fast as $\Theta(\log K)$, then R_{seq} and R_{con} both increase as $\Theta(N \log\log K)$.

The proof is given in Appendix D. Prop. 5 states that the first-order constant associated with the sum capacity growth in Prop. 1 is proportional to $\log(R_F T)$.

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

$$R_{nf} = N \int_{x=0}^{\infty} \log(1 + \rho P x) dF(x), \tag{11}$$

where $F(x)$ is the cumulative distribution function (c.d.f.) of the channel gains. This quantity is also increasing as $\Theta(N)$. However, from Prop. 5, 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 perform better than those without feedback; however, the improvement does not increase the first-order growth rate.

D. Effect of Load on Capacity

In this section, we characterize the capacity growth per sub-channel achieved by both feedback schemes for small and large loads ρ. Namely, referring to the asymptotic order-growth in Prop. 1, we specify the behavior of the first-order constants associated with the sequential and contention feedback schemes as $\rho \to 0$ and as ρ becomes large.

As $\rho \to 0$, the feedback overhead for both schemes diminishes. In particular, for the sequential scheme (3) implies that the feedback overhead $f_{seq} \to 0$. This is true asymptotically even if each user feeds back any fixed number of B bits. (For example, if $B = 1$ and the group size $\alpha N = 1$, then $f_{seq} = \rho \frac{NH(p_o)+1}{R_F T}$.) Hence as $\rho \to 0$, the base station can obtain one feedback bit per sub-channel from each user without decreasing the asymptotic sum throughput. In contrast, for the contention scheme, the CSI obtained for each user corresponds to only the assigned sub-channels. Hence we expect the system capacity for the sequential scheme to be larger.

As ρ increases and approaches $R_F T$, the achievable rate for the sequential scheme $R_{seq}/N \to 0$. This is because each user is assumed to transmit at least one feedback bit, so that if $\rho \geq R_F T$, then the entire coherence time is needed for feedback. In contrast, for the contention scheme the sizes of the sub-channel and user groups, and hence feedback, can be adjusted to optimize the capacity per sub-channel R_{con}/N as the system scales. As ρ increases, the optimal feedback fraction $f_{con} < 1$, allowing the achievable rate to increase with K.

Proposition 6: For the sequential scheme, as $\rho \to 0$, the asymptotic capacity per sub-channel $R_{seq}/N \to 0$ as $\Theta(\sqrt{\rho})$. Furthermore, as $\rho \to R_F T$ from below, $R_{seq}/N \to 0$.

Proposition 7: For the contention scheme, as $\rho \to 0$, the asymptotic capacity per sub-channel $R_{con}/N \to 0$ as $\Theta(\rho \log(1/\rho))$. Furthermore, as $\rho \to \infty$, R_{con}/N increases as $\Theta(\log \rho)$.

The proof of Prop. 6 is given in Appendix E. The proof of Prop. 7 is omitted due to the space limitation.[12]

For small ρ, $\rho \log(1/\rho) < \sqrt{\rho}$, hence as expected, the asymptotic capacity for the sequential scheme is larger than that for the contention scheme. Also, the capacity for the contention scheme grows as $\log \rho$. In contrast, as K and N increase with full CSI, the sum capacity increases to infinity as $N \log\log K$ for any ρ. These results are illustrated in Fig. 1, which shows the asymptotic capacity per sub-channel for both schemes as a function of system load ρ. The curves are obtained by evaluating (5) numerically with channel variance $\sigma^2 = 10$ and power per user $P = 10$ dB. Fig. 1 shows a crossover load ρ^* for which the sequential scheme performs better than the contention scheme for $\rho < \rho^*$, and vice versa for $\rho > \rho^*$. For this example, $\rho^* > 1$, which indicates that the sequential scheme performs best for (practical) loads $\rho < 1$.

IV. Numerical Results

A. Asymptotic Comparisons

In this section, we provide some numerical examples, which illustrate the asymptotic performance of both feedback schemes. For the examples in this section the channel gains are *i.i.d.* Rayleigh with variance $\sigma^2 = 1$, and the power per user $P = 10$.

Fig. 2 shows the asymptotic sum capacity per sub-channel versus $R_F T$ for different loads $\rho = K/N$. For each scheme,

[12] A sketch of the proof can be found in [21].

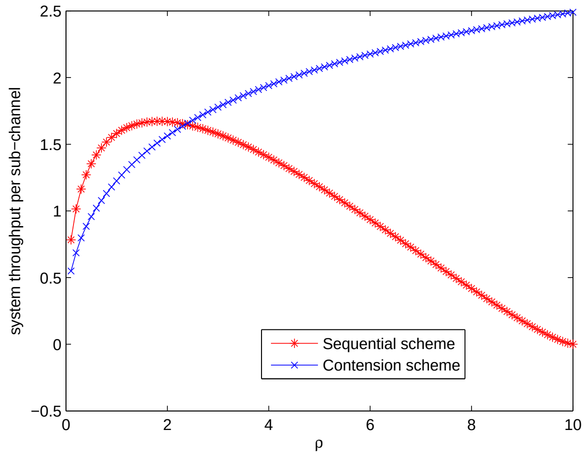

Fig. 1. Capacity versus load for the two feedback schemes.

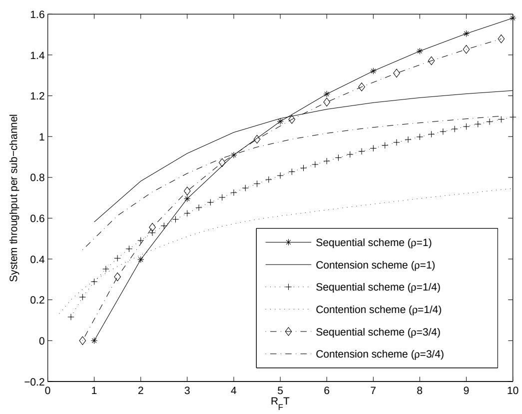

Fig. 2. Asymptotic capacity per sub-channel versus feedback $R_F T$ for the sequential and contention feedback schemes with different loads ρ.

the capacity, given by (6) or (10), is optimized numerically over the relevant parameters (i.e., t_o, α and β.) As stated in Prop. 4, for a given ρ, when $R_F T$ is small the contention scheme achieves a larger sum capacity, whereas when $R_F T$ is large the sequential scheme performs better. (Note that the sum capacity of the sequential scheme approaches zero as $R_F T$ approaches ρ.) In each case there is a single crossing point ρ^*, which shifts to the right as ρ increases.

The relative performance of the sequential and contention schemes is further illustrated in Fig. 3, which shows how the crossing point in Fig. 2 depends on $R_F T$ and P. Namely, the load ρ^* at which the relative performance of the two feedback schemes switches is shown as a function of $R_F T$ for different values of P. (For example, in Fig. 1, $\rho^* \approx 2.4$). For the range of parameters shown the crossing-point increases almost linearly with $R_F T$ for each P. For values of ρ and $R_F T$ above (below) the line in this figure, the contention (sequential) scheme performs best.

Fig. 3 also shows that given $R_F T$, the crossing point ρ^* decreases as P increases, so that the region in which the contention scheme performs better becomes larger. This is because increasing the power P leads to a decrease in the threshold (i.e., the minimum received power per active

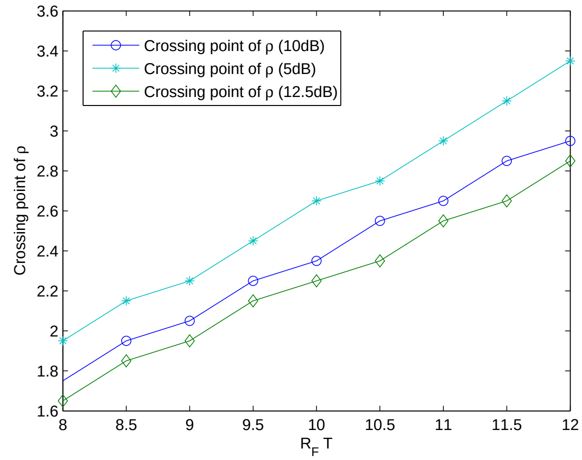

Fig. 3. The crossing point ρ^* as a function of $R_F T$.

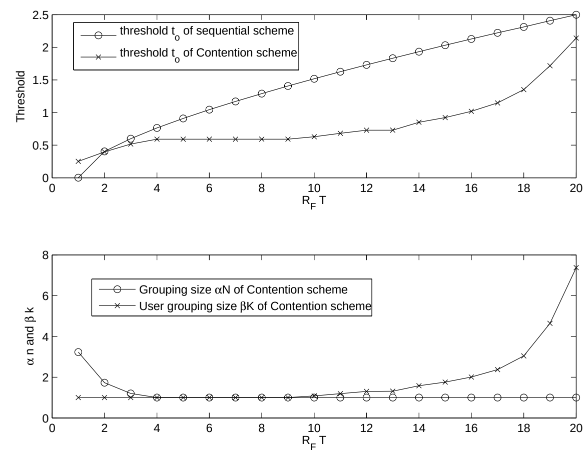

Fig. 4. Optimized parameters for each feedback scheme versus $R_F T$.

group stays approximately constant). The feedback overhead in the sequential scheme then increases, since the probability of requesting a sub-channel group increases. In contrast, the feedback overhead in the contention scheme is less sensitive to this change, due to the additional flexibility of being able to adjust the user and channel group sizes.

Fig. 4 shows the optimized parameters for both schemes as a function of $R_F T$ with $\rho = 1$. 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 $R_F T$. In contrast, for the one-bit feedback scheme in [6], the optimal threshold approaches infinity as the system scales. For that scheme, the asymptotic rate at which the threshold increases (namely $\log K$) ensures that the SNR per sub-channel increases as $\log K$, so that the capacity increases as $N \log \log K$. Here the threshold must increase at a slower rate in order to bound the fraction of coherence time used for feedback. The lower part of the figure shows the optimal group size versus $R_F T$ for the contention scheme. (For the sequential scheme Prop. 2 states that the optimal group size approaches infinity.) As $R_F T$ increases, the number of sub-channels in each group decreases to one, while the number of users per group increases.

B. Performance of Finite-Size Systems

In this section we present numerical results for finite-size systems. Specifically, we select the parameters for each scheme to maximize the asymptotic sum capacity per sub-channel and then simulate the performance in a finite system. For each of the following scenarios, results are shown for 1000 different channel realizations, where each realization consists of NK independent sub-channel gains generated according to an exponential distribution with variance $\sigma^2 = 10$.

Fig. 5 shows the throughput per sub-channel for the contention scheme as a function of the number of users K with $R_F T = 10$ and $\rho = 1$. Even for a relatively small system size (e.g., $K = 10$), averaging the throughput per sub-channel over all channel realizations gives a value, which is very close to the asymptotic throughput per sub-channel (2.49 nats/sec). The variance in throughput across realizations decreases with the system size due to increased averaging across users and channels.

Fig. 6 shows analogous results for the sequential scheme with $\rho = 1$ and $R_F T = 200$. (The larger value of $R_F T$ is chosen in this case, because as discussed in Section III-B, this is the regime in which the sequential scheme performs better than the contention scheme.) For the sequential scheme a fixed-to-variable lossless source code, which has average codeword length less than or equal to $\frac{1}{\alpha}H(p_0) + 1$, is used to compress the feedback for each user. The results are shown in Fig. 6 along with the asymptotic throughput per sub-channel (5.60 nats/sec). Compared to the contention scheme, the averaged throughput per sub-channel for the sequential scheme converges to its asymptotic value at a much slower rate. This is due to the convergence rate of the feedback per sub-channel to its asymptotic value. Specifically, in the optimal sequential scheme the probability that a user requests a sub-channel p_0 converges to zero as the system scales, so that $H(p_0) \to 0$. The average number of compressed feedback bits per sub-channel therefore converges to a constant, which is upper bounded by one. However, for the coding scheme used in our simulations, this convergence is slow, leading to the slower convergence in throughput per channel. Once again, we note that the variance in the throughput per sub-channel across realizations decreases as the system scales.

Also shown in Fig. 6 for comparison is the sum capacity of a one-bit feedback scheme in which each user indicates whether or not each sub-channel gain exceeds a threshold (e.g., see [6]). The threshold is selected to maximize the sum rate. The maximum number of users that can be supported with the one-bit feedback scheme is $R_F T$. Fig. 6 shows that with $K = 10$ users the one-bit scheme gives a slightly larger average throughput than the sequential scheme. Initially, the sum capacity of the one-bit scheme increases, then it decreases almost linearly to zero. On the other hand, the throughput per sub-channel achieved by the sequential scheme converges to a positive constant as the system size scales.

V. Conclusions

We have presented two feedback schemes for downlink OFDMA for which the feedback overhead remains bounded as the system size scales. Given a finite coherence time T and limited feedback link capacity R_F, the sum capacity growth for both schemes is $\Theta(N)$ as the number of users and sub-channels increase, as opposed to the optimal order growth of $\Theta(N \log\log K)$ with perfect CSI. The first-order constant for the limited feedback schemes increases as $\log(R_F T)$, so that the multi-user diversity term $\log\log K$ can be recovered if the feedback $R_F T$ is allowed to scale with the system size as $\log K$.

An asymptotic comparison of the two schemes shows that the sequential scheme performs best when $R_F T$ is sufficiently large, or when the load ρ is sufficiently small. Numerical results indicate that this is the case when $R_F T$ takes on moderate values (e.g., 100 bits per sub-channel per coherence time) and $\rho < 1$. Hence, for the model considered the sequential scheme achieves a higher sum capacity than the contention scheme for a practical range of values. However, the sequential scheme may require additional overhead to synchronize the user feedback, which is not taken into account

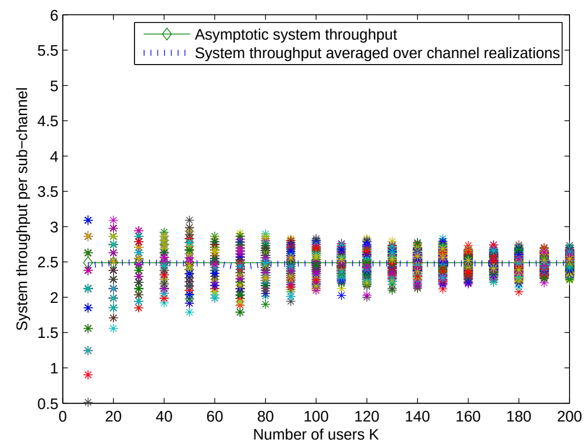

Fig. 5. Achievable rate per sub-channel for the contention scheme as a function of K for $\rho = 1$. Each point corresponds to a different channel realization.

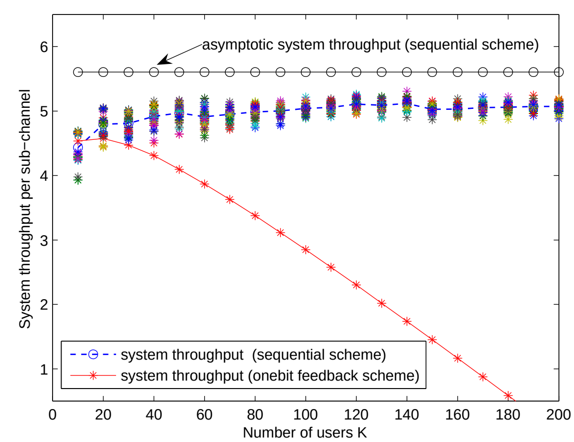

Fig. 6. Achievable rate per sub-channel for the sequential scheme as a function of K for $\rho = 1$. Also shown is the achievable rate for the one-bit feedback scheme presented in [6]

in our model. This synchronization overhead will, of course, compromise the relative advantage of the sequential scheme. (Some numerical examples, which illustrate this point are presented in [21].)

The model and results presented here can be extended in a few different ways. 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. In practice, this criterion may be replaced by other selection criteria, which give different performance. It may also be possible to reduce feedback further by exploiting correlation among neighboring groups of sub-channels. Specifically, correlation should reduce the required feedback per user in the sequential scheme. It is more difficult to predict how correlation will affect the contention scheme since feedback collisions become correlated across sub-channels. We do not expect that correlation will affect the asymptotic scaling results, since correlation can be decreased by merging sub-channel groups into larger groups, although correlation will affect the associated constant. Finally, another direction is to extend the current model to OFDMA systems with multiple antennas, which may require much larger amounts of feedback and overhead to exploit the available spatial dimensions.

Appendix A

Proof of Proposition 1: We prove this proposition via the following four steps:

Step 1: For a fixed R_FT, $R_{seq} = O(N)$ as $N \to \infty$. To establish a contradiction, suppose this is not true, i.e., $R_{seq}/N \to \infty$. Since $f_{seq} \geq 0$, this can only occur if $p_s r_{seq} = p_s \log(1 + \frac{\rho P t_o}{p_s}) \to \infty$. We show that the required feedback needed for this must eventually exceed the available coherence time and so this cannot occur.

For $p_s \log(1 + \frac{\rho P t_o}{p_s}) \to \infty$, it must be that $t_0 \to \infty$ and $p_s \log(t_0) \to \infty$ [13] Next, note that p_s is equal to the probability that the maximum of K *i.i.d.* random variables with *c.d.f.* $F(x) = 1 - e^{-\alpha N x/\sigma^2}$ exceeds t_0. Using the theory of extreme order statistics [20], it can be shown that (see e.g. [6, Lemma 6])

$$p_s \asymp 1 - \exp(-Kp_0), \tag{12}$$

where $p_0 = e^{-\frac{\alpha N t_o}{\sigma^2}}$ is the probability one user's sub-channel gains all exceed t_0 in a given group. Let $\mu = \lim_{N\to\infty} p_s$. We consider two cases.

Case 1: $\mu > 0$. From (12) it follows that

$$Kp_o \to \begin{cases} -\ln(1-\mu), & \text{if } \mu \in (0,1), \\ \infty, & \text{if } \mu = 1. \end{cases} \tag{13}$$

Also, since $t_0 \to \infty$ and $\alpha N \geq 1$, it must be that $p_0 \to 0$. As $p_0 \to 0$, $H(p_0) \asymp -p_0 \log(p_0)$ and so, $\frac{KH(p_0)}{\alpha N} \asymp Kp_o(\frac{t_0}{\sigma^2})$. Furthermore, (13) and $t_0 \to \infty$ imply $\frac{KH(p_0)}{\alpha N} \to \infty$. From (3) it follows that $f_{seq} \to \infty$, i.e., the feedback eventually exceeds the coherence time.

Case 2: $\mu = 0$. From (12) it follows that $p_s \asymp Kp_0$, so that again $p_0 \to 0$ and $\frac{KH(p_0)}{\alpha N} \asymp p_s \frac{t_0}{\sigma^2}$. Since $p_s \log(t_0) \to \infty$, it again follows that $\frac{KH(p_0)}{\alpha N} \to \infty$, so that $f_{seq} \to \infty$.

[13] When $p_s \to 0$, this follows since in that case $p_s \log(p_s) \to 0$.

Step 2: For any fixed $R_FT > \rho$, $R_{seq} \asymp \gamma_{seq} N$ for some $\gamma_{seq} > 0$. Equivalently, we show there exists a sequence of $(t_0,\, \alpha)$ so that

$$p_s r_{seq}(1 - f_{seq}) \to \gamma_{seq}. \tag{14}$$

We showed in Step 1 that if $p_s r_{seq} \to \infty$, then $f_{seq} \to \infty$, and so for (14) to hold it must be that $p_s r_{seq}$ is asymptotically bounded. Also, since $1 - f_{seq} \leq 1$, (14) cannot hold if $p_s r_{seq} \to 0$. Therefore, the only possibility is that $p_s r_{seq} \to \nu$ and $1 - f_{seq} \to \gamma_{seq}/\nu$, for $\nu > 0$ and $\gamma_{seq}/\nu \in (0, 1]$. We next show that there exists a sequence (t_0, α) for which this is true.

Consider a sequence (t_0, α) for which $p_s \to \mu$ for some $\mu \in (0, 1)$. Then $p_s r_{seq} \asymp \mu \log(1 + \frac{\rho P t_0}{\mu})$, and so if $p_s r_{seq} \to \nu$, it must be that

$$t_0 \to \frac{\mu}{\rho P}(e^{\frac{\nu}{\mu}} - 1). \tag{15}$$

As in Step 1, we have $p_s \asymp 1 - e^{-Kp_0}$ and so

$$Ke^{-\frac{\alpha N t_o}{\sigma^2}} \to -\log(1 - \mu). \tag{16}$$

Solving for t_0 and comparing with (15), it follows that the group size αN must satisfy

$$\alpha N \asymp \frac{\rho p \sigma^2}{\mu(e^{\frac{\nu}{\mu}} - 1)}(\log K - \log(-\log(1-u))). \tag{17}$$

Hence $\alpha N = \Theta(\log K)$, which implies $p_0 \to 0$. Thus as in Step 1, $H(p_0) \asymp -p_0 \log p_0$, and from (16) we have $\frac{KH(p_0)}{\alpha N} \asymp -\log(1-\mu)t_0/\sigma^2$. Substituting this into the definition of f_{seq} and substituting the limiting value of t_0 in (15) gives

$$f_{seq} \to \frac{-\log(1-\mu)\mu(e^{\frac{\nu}{\mu}} - 1)}{\rho P \sigma^2 R_F T} + \frac{\rho}{R_F T}. \tag{18}$$

With an appropriate choice of the constants μ and ν (equivalently, t_0 and α), the limiting value of f_{seq} can be set to any value greater than $\frac{\rho}{R_F T}$. Therefore, for $R_F T > \rho$ we can find a sequence of parameters so that f_{seq} is bounded away from one, as desired.

Step 3: Given fixed $R_F T$, $R_{con} = O(N)$ as $N \to \infty$. To establish a contradiction, assume that $R_{con}/N \to \infty$. As in Step 1, since $f_{con} \geq 0$, this can only occur if $p_t r_{con} \to \infty$. We again show that the associated feedback must exceed the coherence time.

If $p_t r_{con} \to \infty$, it must be that $t_0 \to \infty$ and $p_t \log(t_0) \to \infty$. Recall that $p_t = \beta K e^{-\alpha N t_o/\sigma^2}(1 - e^{-\alpha N t_o/\sigma^2})^{\beta K - 1}$ and $\alpha N \geq 1$, so that $p_t \leq \beta K e^{-t_o/\sigma^2}$. Suppose that $\beta K \leq L$ (a constant) for all K. Then it follows that $p_t \leq Le^{-t_o/\sigma^2}$, which implies $p_t \log(t_o) \to 0$. This contradicts the assumption that $p_t r_{con} \to \infty$, and so $\beta K \to \infty$ as K increases.

If $\beta K \to \infty$, then to keep $f_{con} = \frac{log(\beta K)+1}{\alpha N R_F T}$ bounded, it must be that $\alpha N \to \infty$. In particular, there must exist a $W > 0$ such that $\frac{\log(\beta K)}{\alpha N} \leq W$. Since $\beta K \to \infty$, we apply extreme value theory to get[14] $(1 - e^{-\alpha N t_o/\sigma^2})^{\beta K} \asymp e^{-\kappa}$, where $\kappa = \beta K e^{-\alpha N t_o/\sigma^2}$. It follows that

$$p_t \asymp \kappa e^{-\kappa}. \tag{19}$$

[14] Note the left-hand side is simply the probability that the maximum of βK i.i.d. random variables exceeds t_0.

Next we show that $\kappa \to 0$ as $K \to \infty$. Note that

$$\log(\kappa) = \alpha N(\frac{\log(\beta K)}{\alpha N} - t_o/\sigma^2) \asymp -\alpha N t_o/\sigma^2 \to -\infty, \quad (20)$$

since $t_o \to \infty$ and $\frac{\log(\beta K)}{\alpha N} \le W$, so that $\kappa \to 0$. Combining (19) and (20) yields

$$p_t \log(t_o) \asymp \kappa \log(t_0) \le e^{\alpha N(W - t_0/\sigma^2)} \log(t_0) \to 0. \quad (21)$$

The last step follows from the fact that when t_0 is large enough, $W - t_0/\sigma^2 < 0$, and so $e^{\alpha N(W-t_0/\sigma^2)} \le e^{W-t_0/\sigma^2}$. Therefore $p_t \log(t_o)$ remains bounded, which implies that R_{con} cannot increase faster than N.

Step 4: For any fixed $R_F T$, $R_{con} \asymp \gamma_{con} N$ for some $\gamma_{con} >$ 0. We need only provide a sequence of (t_0, α, β) such that $\tilde{R}_{con}/N \to \gamma_{con}$. Let $\alpha = \frac{1}{N}$ and $\beta = \frac{1}{K}$, so that there is one user and one channel per group. In this case $p_t = e^{-t_0/\sigma^2}$ and so $\tilde{R}_{con}/N = e^{-t_0/\sigma^2} \log(1 + \frac{\rho P t_0}{e^{-t_0/\sigma^2}})$ for all N. Hence for any $t_0 > 0$ $\tilde{R}_{con}/N$ converges to a positive constant. ■

APPENDIX B

Proof of Proposition 2: Step 2 of the proof of Proposition 1 shows that there exist sequences of parameters (t_0, α) that achieve the optimal growth rate with $p_s \to \mu$ for $\mu \in (0, 1)$. Furthermore, (17) shows that any such sequence must have $\alpha N = \Theta(\log K)$ and (15) shows that $p_0 = \Theta(1/K)$. The average number of channel groups requested by one user is given by $p_0 N/(\alpha N)$, which then decreases as $\Theta(1/\log K)$. Though such a sequence achieves the optimal growth rate, it is possible that the optimal sequence is not in this class (i.e., it could achieve the same order growth with a larger constant). To complete the proof, we show that this is not the case. Specifically, we show that if the optimal $p_s \to 0$ or $p_s \to 1$, then the feedback must be unbounded.

Case 1: $p_s \to 0$. As in the proof of Prop. 1, it must be that $p_s r_{seq} \to \nu$ for this scheme to be order-optimal. If $p_2 \to 0$, then it must be that $p_s \log(t_0) \to \nu$. Since $p_s \asymp K p_0$, it follows that $K p_0 \log(t_0) \to \nu$ and $p_0 \to 0$. Hence $\frac{KH(p_0)}{\alpha N} \asymp \frac{\nu t_0}{\sigma^2 \log(t_0)}$, which implies $f_{seq} \to \infty$.

Case 2: $p_s \to 1$. Again it must be that $p_s r_{seq} \to \nu$, which now implies that $t_0 \to \frac{e^\nu - 1}{\rho P}$. Since $p_s \asymp 1 - \exp(-K p_0)$, it must be that $K p_0 \to \infty$. Hence $\frac{KH(p_0)}{\alpha N} \asymp K p_0 \frac{t_o}{\sigma^2}$ and again $f_{seq} \to \infty$. ■

APPENDIX C

Proof (sketch) of Proposition 4: We consider two cases, when $R_F T$ is small and $R_F T$ is large.

Case 1: Large $R_F T$. The proofs of Prop. 1 and Prop. 2 show that with the optimal parameters for the sequential scheme, there exists a constant $\mu \in (0, 1)$ such that $p_s \to \mu$, and a constant $t_0^* > 0$ such that $t_0 \to t_0^*$. Furthermore, given any $t_0^* > 0$, we can find a sequence of α's such that p_s converges to any $\mu \in (0, 1)$. Therefore the first-order constant for the sequential scheme is

$$\gamma_{seq} = \max_{t_0 > 0, \mu \in (0,1)} \mu \log\left(1 + \frac{\rho P t_0}{\mu}\right)\left(1 - \frac{-\log(1-\mu) t_0}{\sigma^2 R_F T} - \frac{\rho}{R_F T}\right). \quad (22)$$

As $R_F T$ increases, $\frac{\rho}{R_F T} \to 0$; hence we drop this term and rewrite the preceding maximization as

$$\max_{c^*, t_o} \mu \log(1 + \frac{\rho P}{\mu} t_o)(1 - c^*). \quad (23)$$

where $c^* = \frac{-\log(1-\mu) t_0}{\sigma^2 R_F T}$ is the fraction of the coherence time dedicated to feedback, and $\mu = 1 - e^{-\frac{c^* R_F T}{t_o/\sigma^2}}$.

From (7)-(10) it can be seen that γ_{con} can be expressed as

$$\begin{aligned} \max_{\alpha, c^*, t_o} \quad & p_t \log(1 + \frac{\rho P}{p_t} t_o)(1 - c^*) \\ \text{s.t.} \quad & \beta K = e^{\alpha N c^* R_F T - 1}. \end{aligned} \quad (24)$$

where p_t is a function of α, c^* and t_0. The function $x \log(1 + \frac{a}{x})$ is increasing in $x > 0$ for any $a > 0$. Hence, if c^* and t_o are the same for both schemes, then we can compare the first-order constants by comparing the values of μ and p_t. Here μ is determined by c^* and t_o while p_t is given by optimizing over α. It can be shown that for fixed c^* and t_0, $\mu > p_t$, which implies $\gamma_{seq} > \gamma_{con}$. (We omit the details due to the space limitation.)

Case 2: Small $R_F T$. In the sequential scheme because each user must send back at least one bit, $f_{seq} \ge \frac{\rho}{R_F T}$. As $R_F T \to \rho$, $1 - f_{seq} \to 0$, so that $\gamma_{seq} \to 0$. For the contention scheme γ_{con} does not go to zero as $R_F T \to \rho$ since the channel and user group sizes αN and βK can adjusted to reduce the feedback while keeping a positive first-order constant. ■

APPENDIX D

Proof of Prop. 5: If $R_F T$ increases faster than $\log K$, then the fraction of time used for feedback in the contention scheme satisfies

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

so that $R_{con}/N \asymp p_t \log(1 + \frac{\rho P}{p_t} t_o)$. Setting $\alpha N = 1$, $\beta K = K$, and $t_o = \sigma^2 \log K$ gives $p_t = K e^{-\log K}(1 - e^{-\log K})^{K-1} \to e^{-1}$, and the throughput per sub-channel grows as $\Theta(N \log\log K)$. Furthermore, from Prop. 4 $R_{con} \le R_{seq}$ when $R_F T$ is large, and both R_{con} and R_{seq} are upper bounded by the sum capacity with full CSI, which also increases as $\Theta(N \log\log K)$. It follows that R_{con} and R_{seq} both increase as $\Theta(N \log\log K)$.

If $R_F T$ scales slower than $\Theta(\log K)$, then the proof of Prop. 1 shows that R_{seq} increases as $\Theta(N \log(R_F T))$. For the contention scheme, we first lower bound R_{con} by setting $\alpha N = 1$ and $t_o = \sigma^2 \log(\beta K)$, so that $p_t = (1 - \frac{1}{\beta K})^{\beta K - 1}$. As βK increases, p_t is lower bounded by e^{-1}, so that

$$\begin{aligned} & R_{con}/N \\ & \ge e^{-1} \log\left(1 + \frac{\rho P \sigma^2}{e^{-1}} \log(\beta K)\right)\left(1 - \frac{\log(\beta K) + 1}{R_F T}\right) \\ & = \Theta(\log(R_F T)). \end{aligned} \quad (25)$$

Prop. 4 states that $R_{con} \le R_{seq}$ when $R_F T$ is sufficiently large. Therefore R_{con} also scales as $\Theta(N \log(R_F T))$. ■

Appendix E

Proof of Prop. 6: As K and N increase with fixed ratio ρ, $R_{seq}/N = p_s r_{seq}(1 - f_{seq})$ converges to a constant, which we denote by $\gamma_{seq}(\rho)$. We first show that $\gamma_{sec}(\rho) = \Theta(\sqrt{\rho})$ as $\rho \to 0$.

From Prop. 1 and Prop. 2, we know that for any $\rho \in (0, R_F T)$ with the optimal parameters: (*i*) there exists a $\mu \in (0,1)$ such that as the system scales $p_s \to \mu$, and (*ii*) there exists a constant $t_0^* > 0$ such that $t_0 \to t_0^*$ where

$$f_{seq} \to \frac{-\log(1-\mu)t_0^*}{\sigma^2 R_F T} + \frac{\rho}{R_F T} \leq 1; \tag{26}$$

Let f_{seq}^* be the limiting value of f_{seq} in (26). It follows that $\gamma_{sec}(\rho) = \mu \log(1 + \frac{\rho P t_0}{\mu})(1 - f_{seq}^*)$. As ρ is varied, the optimal values of μ and t_0 vary along with f_{seq}^*. If as $\rho \to 0$, μ and t_0 converge to any non-zero constant and f_{seq}^* converges to a value less than one, then it can be easily seen that the first-order constant decreases as $\Theta(\rho)$. If μ and t_0 converge to any non-zero constant and $f_{seq}^* \to 1$, then the first-order constant decreases faster than $\Theta(\rho)$.

Finally, suppose that $\mu \to 0$ when $\rho \to 0$. In this case,

$$\frac{-\log(1-\mu)t_o}{\sigma^2 R_F T} \asymp \frac{\mu t_o}{\sigma^2 R_F T}. \tag{27}$$

From (26) and (27), it follows that $\frac{\mu t_0}{\sigma^2 R_F T} \to f_{seq}^*$, as $\rho \to 0$. Furthermore, since $f_{seq}^* \leq 1$, it must be that for any $\epsilon > 0$, and ρ small enough, $\frac{\mu t_0}{\sigma^2 R_F T} \leq 1 + \epsilon$. Hence, for ρ small enough we have

$$\gamma_{seq}(\rho) \leq \mu \log(1 + \frac{\rho P (1+\epsilon)\sigma^2 R_F T}{\mu^2}). \tag{28}$$

The right-hand side of (28) is concave in $\mu \geq 0$ and from the first-order optimality conditions is maximized by choosing μ to satisfy $\frac{\rho P(1+\epsilon)\sigma^2 R_F T}{\mu^2} = x^*$, where x^* is the unique positive solution to $(1 + x^*)\log(1 + x^*) = 2x^*$. This implies that the γ which optimizes (28) decreases as $\Theta(\sqrt{\rho})$ for ρ small enough. Substituting this into (28) shows that the upper bound on $\gamma_{seq}(\rho)$ decreases as $\Theta(\sqrt{\rho})$.

Next we give a lower bound for $\gamma_{seq}(\rho)$, which also grows as $\Theta(\sqrt{\rho})$. Letting $\mu = \sqrt{\rho}$ and $t_o = 1/\sqrt{\rho}$, $\gamma_{seq}(\rho)$ can be lowered bounded by

$$\sqrt{\rho}\log(1 + P\sigma^2 R_F T)(1 - \frac{1}{\sigma^2 R_F T} - \frac{\rho}{R_F T}) = \Theta(\sqrt{\rho}). \tag{29}$$

Setting the group size $\alpha N \asymp \frac{\sigma^2}{\sqrt{\rho}}(\log(K) - \log(\sqrt{\rho}))$, it follows that $p_{seq} \to \mu$. Therefore as $\rho \to 0$, $\gamma_{sec}(\rho) = \Theta(\sqrt{\rho})$, as desired.

As $\rho \to R_F T$, because each user feeds back at least one bit per sub-channel, $f_{seq} \to 1$ and $R_{seq}/N \to 0$. Hence in this limit $\gamma_{sec}(\rho) \to 0$. ■

References

[1] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Select. Areas Commun*, vol. 17, no. 10, pp. 1747–1758, 1999.

[2] 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.

[3] 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.

[4] D. J. Love, R. W. Heath, "OFDM Power Loading Using Limited Feedback," *IEEE Trans. on Vehicular Technology*. vol. 54, No.5, Sept. 2005.

[5] Y. Rong, S. A. Vorobyov, A. B. Gershman, "Adaptive OFDM Techniques With One-Bit-Per-Subcarrier Channel-State Feedback," *IEEE Trans. on Communications.* vol. 54, no. 11, Nov. 2006.

[6] 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.

[7] 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.

[8] Y. Sun, W. Xiao, R. Love, K. Stewart, A. Ghosh, R. Ratasuk, B. Classon, "Multi-User Scheduling for OFDM Downlink with Limited Feedback for Evolved UTRA," *Vehicular Technology Conference* Sept. 2006.

[9] S. Sanayei and A. Nosratinia, "Opportunistic downlink transmission with limited feedback," *IEEE Trans. on Information Theory*, vol. 53, no.11, Nov. 2007.

[10] R. Agarwal, V. Majjigi, Z. Han, R. Vannithamby, J. Cioffi, "Low Complexity Resource Allocation with Opportunistic Feedback over Downlink OFDMA Networks," *IEEE JSAC Special Issue on Limited Feedback*, this issue.

[11] G. Wunder, C. Zhou, H.E. Bakker, S. Kaminski, " Throughput maximization under rate requirements for the OFDMA downlink channel with limited feedback," *EURASIP Journal on Wireless Communications and Networking*, vol.2008, article ID 437921, 2008.

[12] Y. Sun and M. Honig, "Asymptotic Capacity of Multi-Carrier Transmission with Frequency-Selective Fading and Limited Feedback," *IEEE Trans. on Information Theory*, to appear July 2008.

[13] 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.

[14] 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.

[15] J. Choi, B. Mondal, R. W. Heath, "Interpolation Based Unitary Precoding for Spatial Multiplexing MIMO-OFDM with Limited Feedback," *IEEE Trans. on Signal Processing.* vol. 54, no. 12, December 2006.

[16] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-Antenna Downlink Channels with Limited Feedback and User Selection," *IEEE Journal Sel. Areas in Communications.* vol. 25, No. 7, pp. 1478-1491, Sept. 2007.

[17] 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.

[18] K. Huang, R. Heath, and J. Andrews, "Multiuser limited feedback for wireless multi-antenna communication,"*IEEE International Symposium on Information Theory*, Nice, France, June 2007.

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

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

[21] J. Chen, R. Berry, M. Honig, "Performance of Scalable Feedback Schemes for Downlink OFDMA,"*2007 Military Communications Conference*, Orlando, Florida, Oct. 2007.

[22] E. Jorswieck, A. Sezgin, B. Ottersten, A. Paulraj, "Feedback Reduction in Uplink MIMO OFDM Systems by Chunk Optimization," *EURASIP Journal on Advances in Signal Processing*, vol.2008, article ID 597072, 2008.

Jieying Chen received her B.E. degree in Electronic Engineering from Shanghai JiaoTong University in 2000 and the M.S. degree in Electrical Engineering from National University of Singapore in 2004. She is pursuing her Ph.d degree at EECS Department of Northwestern University. Her current research lies in wireless communication with emphases on resource allocation and limited feedback. She is the recipient of 2004 Walter P. Murphy Fellowship at Northwestern University.

Randall A. Berry (S'93-M'00) received the B.S. degree in Electrical Engineering from the University of Missouri-Rolla in 1993 and the M.S. and PhD degrees in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology in 1996 and 2000 respectively. He is currently an associate professor in the Department of Electrical Engineering and Computer Science at Northwestern University. In 1998 he was on the technical staff at MIT Lincoln Laboratory in the Advanced Networks Group. He is the recipient of a 2003 NSF CAREER award.

Michael L. Honig (S'80-M'81-SM'92-F'97) received a B.S. degree in electrical engineering from Stanford University in 1977, and M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1978 and 1981, respectively. He subsequently joined Bell Laboratories in Holmdel, NJ, where he worked on local area networks and voiceband data transmission. In 1983 he joined the Systems Principles Research Division at Bellcore, where he worked on Digital Subscriber Lines and wireless communications. Since the Fall of 1994, he has been with Northwestern University where he is a Professor in the Electrical Engineering and Computer Science Department. He has held visiting positions at Princeton University, the University of Sydney, and the University of California, Berkeley. He has also worked as a free-lance trombonist.

He has served as an editor for the IEEE Transactions on Information Theory and the IEEE Transactions on Communications. He has also served as a member of the digital signal processing technical committee for the IEEE Signal Processing Society, and as a member of the Board of Governors for the Information Theory Society. He is the recipient of a Humboldt research award for senior U.S. scientists, and the co-recipient of the 2002 IEEE Communications Society and Information Theory Society Joint Paper Award.