Exploiting Multiuser Diversity Using Multiple Feedback Thresholds

Vegard Hassel*, Mohamed-Slim Alouini†, David Gesbert‡, Geir E. Øien*
*Norwegian University of Science and Technology, Dept. of Electronics and Telecommunications,
Trondheim, Norway, emails: {vegard, oien@iet.ntnu.no}
†University of Minnesota, Dept. of Electrical and Computer Eng., Minneapolis, USA, email: {alouini@ece.umn.edu}
‡Eurecom Institute, Sophia-Antipolis, France, email: {gesbert@eurecom.fr}

***Abstract*— This paper describes a novel scheduling algorithm that takes advantage of multiuser diversity to obtain the maximum system spectral efficiency and uses multiple feedback thresholds to reduce the feedback load to a minimum. In this scenario the relevant users are probed with a set of carrier-to-noise ratio (CNR) thresholds. The users are first probed with the highest threshold. If none of the users are above this threshold the threshold value is sequentially lowered until one or more users are found. A closed-form expression for the average, normalized feedback load (NFL) is found. It can be argued that it is wise to minimize this average NFL to minimize the guard time needed for the feedback process. Consequently, the optimal CNR thresholds which minimize the average NFL are found. We also develop closed-form expressions for the overall capacity using quantized feedback. Plots show that the number of transmitted symbols between feedback queries has great impact on the overall capacity and that one bit feedback is optimal in all cases. The scheduling outage probability has also been analyzed, and the results show that the scheduling outage probability increases dramatically when a scheduling deadline is exceeded.**

I. Introduction

The ever-increasing demand for new applications in wireless communication systems makes efficient transmission scheduling between users a priority. One approach is to take advantage of *multiuser diversity* (MUD) [1]. Ignoring capacity degradation caused by the feedback traffic, the scheduling algorithm that maximizes the average system spectral efficiency among all time division multiplexing (TDM) based algorithms, is the one where the user with the highest carrier-to-noise ratio (CNR) is served at all times [2]. We refer to this rate-optimal algorithm with full feedback as *Max CNR scheduling* (MCS).

To reduce the feedback load, the *Selective multiuser diversity* (SMUD) algorithm was introduced [3]. In that scenario, only the users that have a CNR above a threshold should send feedback to the scheduler. If the scheduler does not receive feedback, a random user is chosen.

The new algorithm proposed here employs multiple feedback thresholds. The thresholds are denoted as $\gamma_{th,L} > \gamma_{th,L-1} > \cdots > \gamma_{th,0}$. For convenience we choose the highest threshold, $\gamma_{th,L}$, to be infinity and the lowest threshold, $\gamma_{th,0}$, to be zero. To initiate the feedback process, the base station sends out a query containing the number of thresholds employed and a list of the relevant users. From the number of feedback thresholds (L) and the number of users (N) in the query, each user can look up the threshold values in predefined tables. After the query is sent, each threshold value is assigned the duration of a *mini-slot* (T_{MS}). In the first mini-slot the base station requests feedback from those users whose CNR is above $\gamma_{th,L-1}$. If there are none, the threshold is successively lowered to $\gamma_{th,L-2}$, $\gamma_{th,L-3}$ down to $\gamma_{th,0}$. Because $\gamma_{th,0}$ equals zero, the best user is always found, and the average feedback load is significantly reduced compared to the MCS algorithm.

We assume that the feedback is transmitted over a *contention channel*. If a user is above a threshold, this user should try to feed back channel state information with probability one in the mini-slot assigned for this threshold value. Consequently, the thresholds are sequentially lowered until a successful feedback transmission or a collision occur. If only one user is above a threshold, it will successfully feed back its channel state information and the guard period will be over. However, if more users try to transmit feedback at the same time a collision will occur. The problem with colliding feedback transmission can be solved by using the *exponential backoff* scheme [4]. Such algorithms give a fast and adaptive solution to the contention problem.

Contributions. As described above, we have developed a generalization of the SMUD scheduling algorithm introduced in [3]. A closed-form expression for the normalized feedback load (NFL) is found. After a discussion of the importance of minimizing the NFL we find the CNR thresholds that give the minimum NFL for a fixed number of thresholds. From plots we show that with a sufficient number of CNR thresholds, the feedback load is minimized, i. e. feedback is received only from the user with the highest CNR. In addition, we have analyzed the overall capacity using quantized feedback. The results show that one-bit feedback is optimal for the scheduling algorithm employing multiple thresholds. It is also shown that the number of transmitted symbols between feedback queries has great impact on the overall capacity. Finally, we show that when there is a deadline for the polling process, the scheduling outage probability increases dramatically if the deadline is shorter than the time it takes to poll the users with all feedback thresholds.

Organization of the paper. The remainder of the paper is organized as follows. We outline the system model in Section II. In Section III it is argued why it can be reasonable to minimize the feedback load to obtain the shortest possible guard time. Following this discussion we derive an expression for the NFL and find the threshold values that give the minimum NFL for a fixed number of CNR thresholds. Next, in Section V we give an expression for the spectral efficiency when it is assumed that the time to collect feedback information and take a scheduling decision is negligible. To have a more realistic expression for the capacity, we develop an expression for the overall capacity using quantized feedback measured in [bits/channel use], in Section VI. Next, in Sections VII and VIII we analyze the expected number of polls before feedback transmission and the scheduling outage assuming a deadline for the polling process. Finally, in Section IX we list our conclusions.

II. System Model

We consider a TDM system with a single base station that schedules N users based on CNR measurements received from the users for each time-slot. The base station only notifies the relevant users, if the preferred user is different from the one in the previous time-slot. The channels of all users are i. i. d., slowly-varying, flat-fading channels with average CNR $\overline{\gamma}$. It is assumed that the duration of a time-slot is shorter than the coherence time of the channels.

III. Minimizing guard time versus minimizing feedback load

In previous publications regarding multiuser scheduling, it was usually assumed that the time it takes for the scheduler to conduct the polling process, take a scheduling decision, and distribute this decision is negligible. In practical systems this process will have to be conducted within a guard time at the beginning of the time-slots. Consequently, the theoretical spectral efficiencies should be reduced by a factor $\frac{T_{TS}-T_G}{T_{TS}}$, where T_G is the guard time interval, and T_{TS} is the total duration of a time-slot.

To have the highest possible spectral efficiency we have to assure that the user with the highest CNR is found and at the same time minimize the guard time. To analyze the guard time we have to take the nature of the feedback channel into consideration. Assuming a *contention channel*, the guard time increases both with the number of thresholds and with the number of users. If the number of threshold values increases, the average number of polls before a user is found will also increase, hence also the guard time. A large number of users on the other hand will lead to more collisions between feedback from different users and will hence also increase the guard time. If we assume that the additional guard time caused by collisions and feedback transmissions is much longer than the duration of a mini-slot, it will be reasonable to minimize the *feedback load* to have the highest possible throughput. The feedback load can be seen as the average number of users giving feedback. Consequently, by minimizing the feedback load the probability of having only the best user giving feedback is maximized. This will lead to a reduction of the guard time due to the minimization of the collision probability between users giving feedback. In the next section we will show that the probability of only having the best user giving feedback will rapidly converge to unity with a increase in the number of threshold values used. Hence, only a relatively small number of threshold values is needed to come close to zero collision probability.

IV. Analysis of the Feedback Load

To evaluate the performance of the scheduling algorithm with multiple feedback thresholds, we investigate the *normalized feedback load* (NFL), which expresses the average share of users that give feedback for each time-slot. It can be shown that the NFL can be expressed as:

$$\bar{F}=\frac{1}{N}\sum_{l=0}^{L-1}\sum_{n=1}^{N} n\binom{N}{n}(P_\gamma(\gamma_{th,l+1})-P_\gamma(\gamma_{th,l}))^n\cdot P_\gamma^{N-n}(\gamma_{th,l}), \tag{1}$$

where $P_\gamma(\gamma)$ is the cumulative distribution function (CDF) of the CNR for a single user. This expression was found by calculating the average number of users that give feedback for each threshold value, and summing up all these feedback load contributions. The expression is normalized by dividing by the number of users. By using the binomial expansion formula [5, Eq. (1.111)], (1) can be written as:

$$\bar{F}=\sum_{l=0}^{L-1}\left(P_\gamma(\gamma_{th,l+1})-P_\gamma(\gamma_{th,l})\right)\cdot P_\gamma^{N-1}(\gamma_{th,l+1}). \tag{2}$$

Inspired by [6], we take the gradient of (2) with respect to the threshold values and set it equal to zero, which gives the following expression for the optimal threshold values:

$$\gamma^*_{th,l}=P_\gamma^{-1}\left(S_l\cdot P_\gamma(\gamma^*_{th,l+1})\right),\quad l=1,2,3,\cdots,L-1, \tag{3}$$

where $P_\gamma^{-1}(\cdot)$ is the inverse CDF of the CNR for a single user, and the constants S_l are given by:

$$S_l=\begin{cases} N^{\frac{1}{1-N}}, & l=1\\ \left[N-(N-1)S_{l-1}\right]^{\frac{1}{1-N}}, & l=2,3,\cdots,L-1,\end{cases} \tag{4}$$

with $N\geq 2$. The set of equations in (3) has a recursive nature. One way to calculate all the threshold values is to start with calculating $\gamma_{th,L-1}$. This value can easily be found because $\gamma_{th,L}$ is defined to be infinity. Knowing $\gamma_{th,L-1}$, (3) can be used to calculate all threshold values down to $\gamma_{th,1}$.

It is also possible to express the threshold values as the sum of the average CNR and a constant (in dB). By writing (3) on the form:

$$P_\gamma(\gamma^*_{th,l})=S_l\cdot P_\gamma(\gamma^*_{th,l+1}),\quad l=1,2,3,\cdots,L-1, \tag{5}$$

and exploiting the fact that $P_\gamma(\gamma_{th,L})=1$, we can write (3) as:

$$\gamma^*_{th,l}=P_\gamma^{-1}\left(\prod_{i=l}^{L-1}S_i\right),\quad l=1,2,3,\cdots,L-1. \tag{6}$$

For the Rayleigh, Nakagami, and Rice distributions, the inverse CDF equals $\overline{\gamma}$ multiplied by a constant which is only dependent of the number of users [6]. Consequently, the

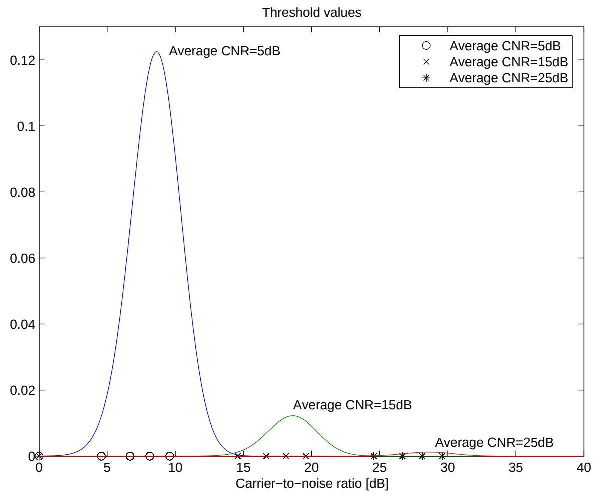

Fig. 1. Three sets of threshold values corresponding to Rayleigh channels with $\overline{\gamma}$=5 dB, $\overline{\gamma}$=15 dB, and $\overline{\gamma}$=25 dB, respectively. The PDF of the best user among 10 users is drawn for each set.

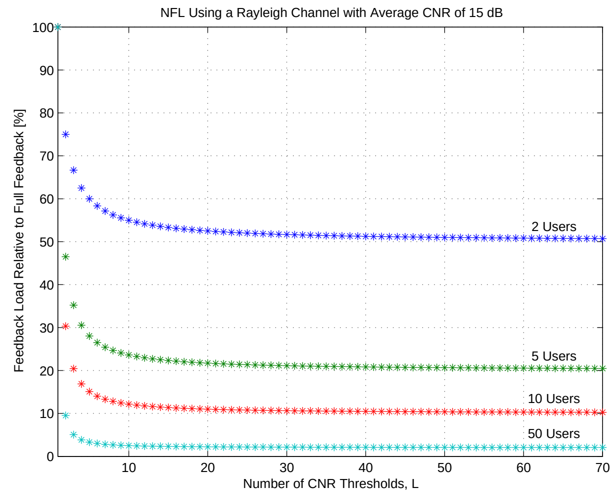

Fig. 2. Minimum normalized feedback load as a function of L for a Rayleigh fading channel with $\overline{\gamma} = 15$ dB.

threshold values in dB will be a sum of $\overline{\gamma}$ and a constant. The calculation of the threshold values in the terminals is therefore a simple operation.

Fig. 1 shows three sets of threshold values for 10 users having Rayleigh channels with $\overline{\gamma}$ = 5 dB, $\overline{\gamma}$ = 15 dB, and $\overline{\gamma}$ = 25 dB, respectively. Each set of threshold values contains 6 CNR values ($L = 5$). The highest threshold value is defined to be infinity and is hence not shown in the figure. The plot shows that the threshold values in dB are a sum of $\overline{\gamma}$ and a constant. The probability density functions (PDF) of the best user among 10 users is also shown [7, (5.85)] for each of the three $\overline{\gamma}$-values. These plots show that the probability of finding the best user below $\gamma_{th,1}$ is quite small. Consequently, the probability of full feedback is low.

A plot of the minimum NFL as a function of L is shown in Fig. 2 for a Rayleigh channel with $\overline{\gamma}$ = 15 dB. We see that the NFL converges to $1/N$ as L grows large. This is logical since the more thresholds there are in the system, the more likely is it that at most one user will have a CNR value between two adjacent thresholds.

V. Spectral Efficiency

The scheduling algorithm studied here will always obtain the same spectral efficiency as the MCS algorithm. Consequently, the expression for the maximum average system spectral efficiency (MASSE) in the case of Rayleigh channels can be written as [8, Eq. (44)]:

$$\frac{< C >_{ora}}{W} = \frac{N}{\ln 2} \sum_{n=0}^{N-1} \binom{N-1}{n} (-1)^n \frac{e^{\frac{(1+n)}{\overline{\gamma}}}}{1+n} E_1 \left(\frac{1+n}{\overline{\gamma}} \right), \tag{7}$$

where W [Hz] is the bandwidth and $E_1(x) = \int_1^\infty e^{-xt} dt$ is the first order exponential integral function. For the expression in (7) it is assumed that the guard time duration is zero, i.e. that feedback will not degrade the total system spectral efficiency. In the next section it will be shown that this assumption is not valid for practical systems.

VI. Capacity Degradation due to Feedback

To analyze the spectral efficiency degradation caused by the feedback traffic we can consider a practical scenario where the feedback is quantized [9], [10]. Instead of considering the MASSE given in [bits/s/Hz] we choose to analyze the capacity measured in [bits/channel use]. The scheduling algorithms analyzed here gives a better quantized feedback performance compared to the algorithms that only employ one threshold. For the algorithm employing multiple thresholds the scheduler knows that the CNR of each user that gives feedback lies between two adjacent thresholds $\gamma_{th,l}$ and $\gamma_{th,l+1}$. Consequently, the quantization regions only need to lie between these two feedback thresholds and the CNR can be fed back with a higher precision.

Modifying [9, Eq. (17)] it can be shown that the overall capacity can be given as:

$$\begin{aligned} C_{\text{QUANT}} &= \sum_{j=0}^{L \cdot J-1} \frac{P_\gamma^N(q_{j+1}) - P_\gamma^N(q_j)}{2(P_\gamma(q_{j+1}) - P_\gamma(q_j))} \\ &\times \int_{q_j}^{q_{j+1}} \log_2(1+\gamma) p_\gamma(\gamma)\, d\gamma \\ &- \frac{N \cdot \bar{F} \cdot \log_2(J)}{N_S}, \end{aligned} \tag{8}$$

where N_S is the number of symbols transmitted between feedback queries (symbols per time-slot), J is the number of quantization regions employed within each feedback region and q_j and q_{j+1} are the lower and upper quantization region thresholds, respectively. The last term in (8) arises because we

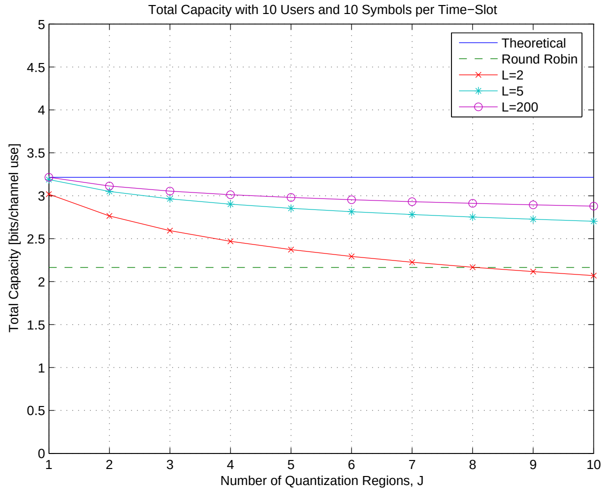

Fig. 3. Total capacity with quantized feedback for 10 users having Rayleigh fading channels with $\overline{\gamma} = 15$ dB. 10 symbols are transmitted per time-slot.

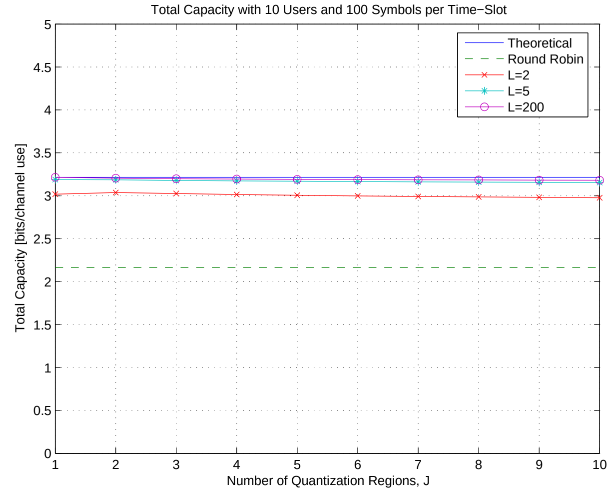

Fig. 4. Total capacity with quantized feedback for 10 users having Rayleigh fading channels with $\overline{\gamma} = 15$ dB. 100 symbols are transmitted per time-slot.

have to deduct the number of bits needed to be fed back to the scheduler before a time-slot is assigned to a user.

Figs. 3 and 4 show (8) plotted as a function of the number of quantization regions. Within each of the feedback regions we assume uniform quantization with J quantization regions. For the theoretical capacity it is assumed that perfect channel estimates are available and that there is no degradation due to feedback traffic. The theoretical capacity is found the same way as for (7). The Round Robin capacity shows the capacity when MUD is not exploited and no feedback is needed to perform scheduling. We see that the degradation due to feedback traffic is largest when few symbols are transmitted between every feedback query. The number of bits fed back, i.e. the number of quantization regions, will also influence the total capacity. The more quantization regions, the more feedback information will be fed back. Because the number of quantization regions has to be a multiple of 2, it is seen from the graphs that one bit feedback is optimal for all cases.

For the expression in (8), the increase in guard time due to the feedback polling process is neglected. Only the degradation due to the transmitted feedback is taken into account. In real life systems we often have to assume that there is a deadline for the polling process. In the next two sections we will investigate the average number of polls before a user is found, and the *scheduling outage probability* caused by a polling deadline.

VII. Number of Feedback Polls before Transmission

To investigate the practical number of threshold values that can be included in the polling process, we choose to investigate the number of thresholds, M, that has to be employed before getting feedback from at least one user. M can be modeled as a discrete random variable (RV), and its probability mass function (PMF) can be expressed as:

$$\Pr[M = k] = P_{\gamma}^{N}(\gamma_{th,L+1-k}) - P_{\gamma}^{N}(\gamma_{th,L-k}), \quad k = 1, 2, \cdots L. \quad (9)$$

Inserting (3) into (9) yields:

$$\Pr[M = k] = (1 - S_{L-k}^{N}) \cdot P_{\gamma}^{N}(\gamma_{th,L+1-k}), \quad k = 1, 2, \cdots L, \quad (10)$$

where S_0 is defined to be zero. In Fig. 5 the PMF of M is displayed for 10 users and 10 threshold values. It is seen that the probability of finding the best user is largest for the first poll and decreases almost linearly with the number of polls. The threshold values are found so that the feedback load is minimized. Consequently, the probability of having L number of polls, i.e. having full feedback load, is very low.

The average number of polls before transmission can be expressed as:

$$\begin{aligned} \mu_M &= \sum_{l=1}^{L} l \left[P_{\gamma}^{N}(\gamma_{th,L+1-l}) - P_{\gamma}^{N}(\gamma_{th,L-l}) \right] \\ &= \sum_{l=1}^{L} l \, (1 - S_{L-l}^{N}) \cdot P_{\gamma}^{N}(\gamma_{th,L+1-l}). \end{aligned} \quad (11)$$

The CDF of the RV in (10) can be written as:

$$P_M(k) = \sum_{l=1}^{k} (1 - S_{L-l}^{N}) \cdot P_{\gamma}^{N}(\gamma_{th,L+1-l}), \quad k = 1, 2, \cdots L. \quad (12)$$

The expression in (12) will be used in the next section to evaluate the scheduling outage probability when there is a deadline for the probing process.

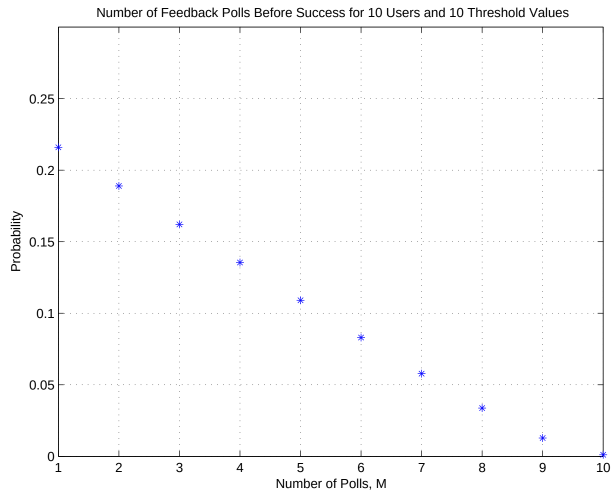

Fig. 5. Probability mass function for the number of feedback polls before a user is found. Each of the 10 users have Rayleigh fading channels with $\overline{\gamma} = 15$ dB and 10 threshold values are used.

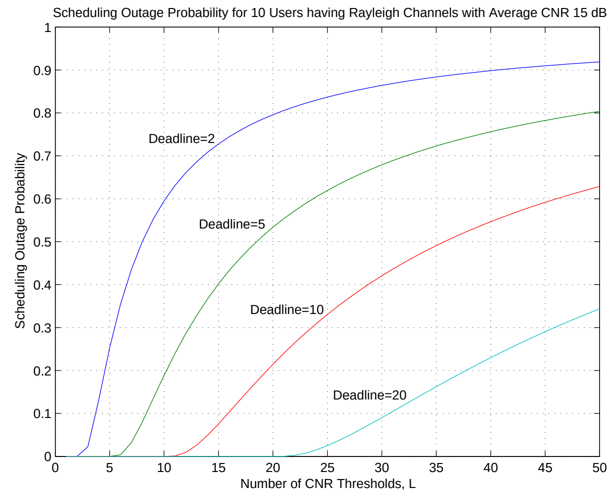

Fig. 6. Scheduling outage probability for 10 users having Rayleigh fading channels with $\overline{\gamma} = 15$ dB.

VIII. Scheduling Outage Probability having a Scheduling Deadline

When no feedback is received the scheduler will have to select a random user and we say that we have a scheduling outage. We want to investigate the scheduling outage probability when there is a deadline for the polling process. Having a deadline of D polls, the scheduling outage probability is given by:

$$\begin{aligned} P_{\text{OUT}} &= 1 - P_M(D) \\ &= 1 - \sum_{l=1}^{D} (1 - S_{L-l}^N) \cdot P_\gamma^N(\gamma_{th,L+1-l}) \\ &= \sum_{l=1}^{L-D} (1 - S_{l-1}^N) \cdot P_\gamma^N(\gamma_{th,l}), \end{aligned} \tag{13}$$

where $P_M(\cdot)$ is given by (12) and $D \leq L - 1$.

Fig. 6 shows (13) plotted as a function of the number of feedback thresholds. The graph illustrates that the scheduling outage probability increases dramatically if the deadline is shorter than the time it takes to poll the users with all feedback thresholds.

IX. Conclusions

We have described a novel scheduling algorithm that takes advantage of multiuser diversity to obtain the maximum average system spectral efficiency and uses multiple feedback thresholds to reduce the feedback load to a minimum. By minimizing the NFL for this algorithm we find the CNR thresholds when we have a fixed number of thresholds. From plots we show that with a sufficient number of CNR thresholds, the feedback load is minimized, i. e. feedback is received only from the user with the highest CNR. For quantized feedback it is shown that one bit feedback is optimal. It is also shown that the number of symbols transmitted between feedback queries will have a great impact on the overall capacity. Finally, we show that when there is a deadline for the polling process, the scheduling outage probability increases dramatically if the deadline is shorter than the time it takes to poll the users with all feedback thresholds.

References

[1] P. Viswanath, D. N. C. Tse, and R. Laroia, “Opportunistic beamforming using dumb antennas,” *IEEE Trans. Inform. Theory*, vol. 48, pp. 1277–1294, June 2002.

[2] R. Knopp and P. A. Humblet, “Information capacity and power control in single cell multiuser communications,” in *IEEE Int. Conf. on Communications (ICC’95)*, (Seattle, WA), pp. 331–335, June 1995.

[3] D. Gesbert and M.-S. Alouini, “How much feedback is multi-user diversity really worth?,” in *IEEE Int. Conf. on Communications (ICC’04)*, (Paris, France), pp. 234–238, June 2004.

[4] D. G. Jeong and W. S. Jeon, “Performance of an exponential backoff scheme for slotted-ALOHA protocol in local wireless environment,” *IEEE Trans. on Veh. Technol.*, vol. 44, pp. 470–479, Aug. 1995.

[5] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. San Diego, CA: Academic Press, 6th ed., 2000.

[6] H. Holm, G. Øien, M.-S. Alouini, D. Gesbert, and K. J. Hole, “Optimal design of adaptive coded modulation schemes for maximum average spectral efficiency,” in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2003)*, (Rome, Italy), pp. 403–407, June 2003.

[7] G. L. Stüber, *Principles of Mobile Communications*. Norwell, MA: Kluwer, 1996.

[8] M.-S. Alouini and A. J. Goldsmith, “Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques,” *IEEE Trans. on Veh. Technol.*, vol. 48, pp. 1165–1181, July 1999.

[9] F. Florén, O. Edfors, and B.-A. Molin, “The effect of feedback quantization on the throughput of a multiuser diversity scheme,” in *Proc. IEEE GLOBECOM*, vol. 1, (San Francisco, CA), pp. 497–501, Dec. 2003.

[10] M. Johansson, “Benefits of multiuser diversity with limited feedback,” in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2003)*, (Rome, Italy), pp. 155–159, June 2003.