The Effect of Feedback Quantization on the Throughput of a Multiuser Diversity Scheme

Florén, Fredrik; Edfors, Ove; Molin, Bengt-Arne

Published in:
Conference Record / IEEE Global Telecommunications Conference

DOI:
10.1109/GLOCOM.2003.1258287

2003

The Effect of Feedback Quantization on the Throughput of a Multiuser Diversity Scheme

Fredrik Florén[1,2], Ove Edfors[1], and Bengt-Arne Molin[3]

[1]Dept. of Electroscience, Lund University, Box 118, SE-221 00 Lund, Sweden

[2]TeliaSonera AB, Box 85, SE-201 20 Malmö, Sweden

[3]Axis Communications AB, Emdalav. 14, SE-223 69 Lund, Sweden

Email: fredrik.floren@es.lth.se

Abstract— **The impact of the quantization of SNR measurements on the throughput of a multiuser diversity scheme for constant-rate transmission is investigated under a block-Rayleigh fading assumption. In the downlink, each user measures its SNR, quantizes it, and feeds it back to the transmitter, which transmits a packet to the user with the highest quantized SNR. In the case of several users having the same quantized SNR, one of them is selected at random. It is concluded that using only a few quantization levels can yield a throughput that is only slightly less than the throughput obtained by using unquantized feedback.**

I. Introduction

Radio resources are scarce and expensive. These facts motivate research in techniques that better utilize the radio spectrum than existing systems do today. The goal is ultimately to increase the spectral efficiency and Quality of Service. One approach that has the potential of accomplishing this goal is Multiuser Diversity [1], where the fading of channels is exploited instead of combatted. This scheme can be employed in both the uplink and the downlink and transmissions should ideally be made to/by a user with a high Signal-to-Noise Ratio (SNR). Furthermore, multiuser diversity benefits from a large number of users in that many users in a system yields a high probability that one of them experiences a large SNR at any given moment. The drawback is that in the downlink the SNR at each user must be monitored and fed back to the transmitter and in the uplink the receiver must signal which user shall transmit. However, it can be argued that in practice this poses no major difficulties [1].

In this paper, we investigate the impact of the degree of quantization of the SNR measurements on the throughput of constant-rate transmission. We focus on the downlink of a simple multiuser diversity system where each user measures its instantaneous SNR, quantizes it, and feeds it back to the base station. At the base station the fed-back quantized SNRs are examined and a packet is transmitted to the user that experiences the highest quantized SNR. If several users lie within the highest quantization level, one of those users is selected at random.

It is assumed that the control signalling between the users and the base station occupies no time nor any other resources. No power control is employed and a block-Rayleigh fading AWGN channel is assumed. Additionally, frequency-flat fading and perfect channel side information is assumed at the receivers, i.e., the users.

As throughput measures both the ergodic capacity and the throughput of packets using uncoded binary transmission are used, although we focus on the former. For the block-fading channel the ergodic capacity can be used when no decoding delay constraints are imposed, and under this assumption the Shannon capacity in the ordinary sense is obtained [2]. Using the ergodic capacity gives an upper bound on the throughput, and using the throughput of uncoded packets gives an indication the behavior of real systems.

Related work can be found in [1], where the concept of Multiuser Diversity, as well as a technique for inducing fast fading, is described in greater detail. For the multiple access channel it was shown in [3] that the optimal transmission strategy in the uplink of multiuser system using power control was to only let the user with largest SNR transmit. Similar work is also presented in [4], where other schemes are investigated in order to improve the delay performance.

This paper is organized as follows. Section II presents the system model and assumptions. In Section III and Section IV the throughputs for the unquantized and quantized schemes are derived, respectively. In Section V numerical results are presented that illustrate the effect of quantization and, finally, the results are elaborated on in Section VI.

II. System Model

Consider the downlink of a single-cell wireless TDMA system with M mobile users. It is assumed that the channel is frequency-flat, i.e., single-ray, block-Rayleigh fading, where the fading level is constant over blocks of N symbols. Furthermore, the fading levels from block to block are independent and identically distributed for each user and the fading levels are independent between users. It is assumed that a block is equal in duration to a slot of the TDMA scheme. One frame is equal to M slots, and in a Round-Robin scheme each user would be assigned the same slot in every frame.

For the ith user the SNR is described by the stochastic variable (SV) Γ_i. The probability density function (PDF) and cumulative distribution function (CDF) of Γ_i are denoted $f_{\Gamma_i}(\gamma)$ and $F_{\Gamma_i}(\gamma)$, respectively. Due to the Rayleigh fading assumption Γ_i is exponentially distributed with mean $\bar{\gamma}_i$ [5].

That is,

$$f_{\Gamma_i}(\gamma) = \frac{1}{\bar{\gamma}_i} e^{-\gamma/\bar{\gamma}_i}, \quad \gamma \geq 0 \tag{1}$$

Moreover, it is assumed that the SNR estimation is perfect and that the feedback of the SNRs is errorless and occurs without any loss in throughput. It is also assumed that all users experience the same mean SNR, and, hence, $\bar{\gamma}_i = \bar{\gamma}$ and $\Gamma_i = \Gamma$ for $i = 1, \ldots, M$. In the sequel, the index i is thus dropped since the expressions are valid for all users.

III. Throughput Using Unquantized Feedback

In this section, the throughput when feeding back unquantized measurements of the users' SNRs is derived. For comparison with the capacities, this is also done in terms of packets per slot using uncoded binary coherent transmission, e.g., BPSK.

For the Round-Robin scheme, where each user is statically assigned a slot in each frame, the probability of successful transmission of a block of N bits is

$$P_{\text{block}} = \left(1 - Q\left(\sqrt{2\Gamma}\right)\right)^N, \tag{2}$$

which makes P_{block} a SV. In (2), $Q(\cdot)$ is the area under the tail of a PDF for a zero-mean, unit-variance Gaussian variable. Averaging P_{block} over the PDF of the SNR yields the mean probability of successful transmission of a block and by normalizing with M, the mean throughput in terms of packets per slot per user becomes

$$\begin{aligned}\eta_{\text{RR}} &= \frac{\mathrm{E}\left[P_{\text{block}}\right]}{M} \\ &= \frac{1}{M} \mathrm{E}\left[\left(1 - Q\left(\sqrt{2\Gamma}\right)\right)^N\right] \\ &= \frac{1}{M} \int_0^\infty \left(1 - Q\left(\sqrt{2\gamma}\right)\right)^N f_\Gamma(\gamma)\, d\gamma,\end{aligned} \tag{3}$$

where $\mathrm{E}[\cdot]$ is the expectational operator. This expression is valid for all users since the mean SNRs of all users are equal.

The Multiuser Diversity scheme considered in this paper is referred to as the Max-SNR scheme. In every slot, it compares the SNRs of all users and the user with the maximum SNR is transmitted to. Since all users experience the same mean SNR, the random variable $\Gamma_{\text{max}} = \max_i \Gamma_i$ has CDF

$$\begin{aligned}F_{\Gamma_{\text{max}}}(\gamma) &= \left(F_\Gamma(\gamma)\right)^M \\ &= \left(1 - e^{-\gamma/\bar{\gamma}}\right)^M\end{aligned} \tag{4}$$

and PDF

$$\begin{aligned}f_{\Gamma_{\text{max}}}(\gamma) &= M f_\Gamma(\gamma) \left(F_\Gamma(\gamma)\right)^{M-1} \\ &= M \frac{e^{-\gamma/\bar{\gamma}}}{\bar{\gamma}} \left(1 - e^{-\gamma/\bar{\gamma}}\right)^{M-1}.\end{aligned} \tag{5}$$

The mean probability of successful transmission of a block, i.e., the throughput, per slot using the Max-SNR scheme becomes

$$\begin{aligned}\eta_{\text{MS}} &= \frac{1}{M} \mathrm{E}\left[\left(1 - Q\left(\sqrt{2\Gamma_{\text{max}}}\right)\right)^N\right] \\ &= \int_0^\infty \left(1 - Q\left(\sqrt{2\gamma}\right)\right)^N \times \\ &\quad \frac{e^{-\gamma/\bar{\gamma}}}{\bar{\gamma}} \left(1 - e^{-\gamma/\bar{\gamma}}\right)^{M-1} d\gamma.\end{aligned} \tag{6}$$

The throughput in terms of capacities for any user using the Round-Robin scheme becomes [2], [6]

$$\begin{aligned}C_{\text{RR}} &= \mathrm{E}\left[\frac{1}{2M} \log_2(1 + \Gamma)\right] \\ &= \frac{1}{M} \int_0^\infty \frac{1}{2} \log_2(1 + \gamma) f_\Gamma(\gamma) d\gamma \\ &= \frac{e^{\bar{\gamma}^{-1}}}{2M \ln 2} \mathrm{E}_1\left(\bar{\gamma}^{-1}\right),\end{aligned} \tag{7}$$

where $\mathrm{E}_1(x) = \int_x^\infty t^{-1} e^{-t} dt$ is the exponential integral function [7, p. 287].

The throughput for any user in terms of capacities for the Max-SNR scheme follows from (7) by averaging over the PDF of Γ_{max} in (5) instead of the PDF of Γ as in (7). Hence,

$$\begin{aligned}C_{\text{MS}} &= \mathrm{E}\left[\frac{1}{2M} \log_2(1 + \Gamma_{\text{max}})\right] \\ &= \int_0^\infty \frac{1}{2\bar{\gamma}} \log_2(1 + \gamma) e^{-\gamma/\bar{\gamma}} \left(1 - e^{-\gamma/\bar{\gamma}}\right)^{M-1} d\gamma.\end{aligned} \tag{8}$$

In Figure 1 the throughput for the Round-Robin and unquantized Max-SNR schemes are plotted for both throughput measures for $M = 10$ users. The gain available by using the Max-SNR scheme is the distance between the two curves, and we will investigate the how much of this gain can be achieved for a certain degree of quantization of the fed-back SNRs.

IV. Throughput Using Quantized Feedback

In this section, the throughput when feeding back quantized measurements of the users' SNRs is derived in terms of packets per slot and capacities. The SNR measured by the user is mapped into a quantization level, and the impact of the number of such levels when optimizing with respect to the thresholds separating those levels is investigated.

Let $Q_k,\ k = 0, \ldots, N-1$ be the quantization levels of the SNRs defined by

$$\begin{aligned}Q_k &= [q_k, q_{k+1}), \quad 0 \leq k \leq N-1 \\ q_0 &= 0 \\ q_N &= \infty,\end{aligned}$$

where the q_k are the values of the SNR that limit the quantization levels, i.e., the q_k are thresholds, as illustrated in Figure 2. For brevity, $\Gamma \in Q_k$ denotes $q_k < \Gamma \leq q_{k+1}$. It is noted that in order to describe which level a SNR belongs to $\log_2(N)$ bits are needed.

Since it is the quantized SNRs that are compared, the Max-SNR scheme now needs to be reformulated since several users might experience SNRs belonging to the same quantization level. Therefore, if several values that are equal are obtained

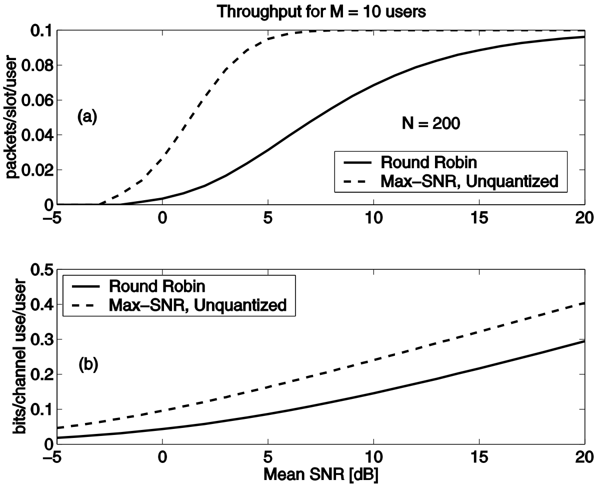

Fig. 1. Performance of Round Robin and Max-SNR schemes for (a) uncoded binary transmission and (b), in terms of capacity. The number of users is $M = 10$ for both plots and for plot (a) the number of bits per packet is $N = 200$.

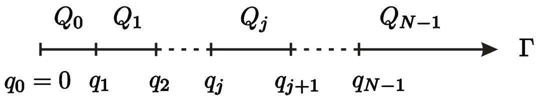

Fig. 2. Quantization regions and thresholds of the SNRs.

after taking the maximum of the quantized SNRs one of the users in the maximum level is chosen at random. In order to derive the throughput when the feedback information is quantized, we let $\mathcal{A}$ denote the event that an arbitrary reference user is transmitted to in a slot. That is, the event that no user's SNR falls in a higher quantization level than the reference user's SNR *and* that the reference user is chosen among all users in its quantization level. By deriving the PDF of the SNR of a user conditioned on that this user is transmitted to, the throughput for both throughput measures can be written as

$$\int_0^\infty g(\gamma) \Pr(\mathcal{A}|\Gamma) f_\Gamma(\gamma) d\gamma. \tag{9}$$

Here, $g(\gamma)$ is either $\frac{1}{2}\log_2(1+\gamma)$ or $\left(1 - Q\left(\sqrt{2\gamma}\right)\right)^N$. For the unquantized case we have

$$\Pr(\mathcal{A}|\Gamma) = \begin{cases} \frac{1}{M}, & \text{Round Robin} \\ \Pr(\Gamma > \max_{k \neq i} \Gamma_k), & \text{Max-SNR} \end{cases} \tag{10}$$

For the quantized case, the integral in (9) can be split into several integrals, one for each quantization level j, i.e.,

$$\begin{aligned} &\int_0^\infty g(\gamma) \Pr(\mathcal{A}|\Gamma) f_\Gamma(\gamma) d\gamma \\ &= \sum_{j=0}^{N-1} \int_{\gamma \in Q_j} g(\gamma) \Pr(\mathcal{A}|\Gamma \in Q_j) f_\Gamma(\gamma) d\gamma. \end{aligned} \tag{11}$$

This is done since the probability that the reference user is transmitted to in a slot given that the reference user's SNR lies in quantization level j, i.e., $\Pr(\mathcal{A}|\Gamma \in Q_j)$, is independent of the instantaneous SNR γ. $\Pr(\mathcal{A}|\Gamma \in Q_j)$ is given by

$$\begin{aligned} \Pr(\mathcal{A}|\Gamma \in Q_j) = \sum_{m=0}^{M-1} \frac{1}{m+1} \binom{M-1}{m} \left[\Pr(\Gamma \in Q_j)\right]^m \times \\ \left[\Pr\left(\Gamma \in \bigcup_{k<j} Q_k\right)\right]^{M-m-1}. \end{aligned} \tag{12}$$

This expression is obtained since if the SNRs of the reference user and m other users lie in quantization level j, and the SNRs of $M-m-1$ users lie in lower quantization levels, the reference user is transmitted to with probability $1/(m+1)$. The quantization levels that are lower than level j are the Q_k for which $k < j$. For every m, there are $\binom{M-1}{m}$ combinations giving rise to this event, and summing over m yields the total probability of $\mathcal{A}$. By using that

$$\left[\Pr(\Gamma \in Q_j)\right]^m = \left[F_\Gamma(q_{j+1}) - F_\Gamma(q_j)\right]^m \tag{13}$$

and

$$\Pr\left(\Gamma \in \bigcup_{k<j} Q_k\right) = \left[F_\Gamma(q_j)\right]^{M-m-1}, \tag{14}$$

the expression in (12) reduces to

$$\begin{aligned} P(\mathcal{A}|\Gamma \in Q_j) &= \sum_{m=0}^{M-1} \frac{1}{m+1} \binom{M-1}{m} \times \\ &\left[F_\Gamma(q_{j+1}) - F_\Gamma(q_j)\right]^m \left[F_\Gamma(q_j)\right]^{M-m-1} \\ &= \frac{\left[F_\Gamma(q_{j+1})\right]^M - \left[F_\Gamma(q_j)\right]^M}{M\left(F_\Gamma(q_{j+1}) - F_\Gamma(q_j)\right)}. \end{aligned} \tag{15}$$

The throughput in terms of packets per slot using the Max-SNR scheme with quantized feedback can now be written as

$$\begin{aligned} \eta_{\text{QMS}} = \sum_{j=0}^{N-1} \int_{Q_j} \left(1 - Q\left(\sqrt{2\gamma}\right)\right)^N \times \\ \frac{\left[F_\Gamma(q_{j+1})\right]^M - \left[F_\Gamma(q_j)\right]^M}{M\left(F_\Gamma(q_{j+1}) - F_\Gamma(q_j)\right)} f_\Gamma(\gamma) d\gamma, \end{aligned} \tag{16}$$

and the throughput in terms of capacity using the Max-SNR scheme with quantized feedback becomes

$$\begin{aligned} C_{\text{QMS}} = \sum_{j=0}^{N-1} \int_{Q_j} \frac{1}{2}\log_2(1+\gamma) \times \\ \frac{\left[F_\Gamma(q_{j+1})\right]^M - \left[F_\Gamma(q_j)\right]^M}{M\left(F_\Gamma(q_{j+1}) - F_\Gamma(q_j)\right)} f_\Gamma(\gamma) d\gamma. \end{aligned} \tag{17}$$

V. Numerical Results

In this section the impact of the number of thresholds is illustrated for both single and multiple-thresholds cases.

We first examine the simplest case where only one threshold is used, i.e., there are two quantization levels. Figure 3 shows the throughput in terms of capacity versus the mean SNR when no quantization of the feedback information takes place and when a two-level quantization is used for 2, 10, and 20 users.

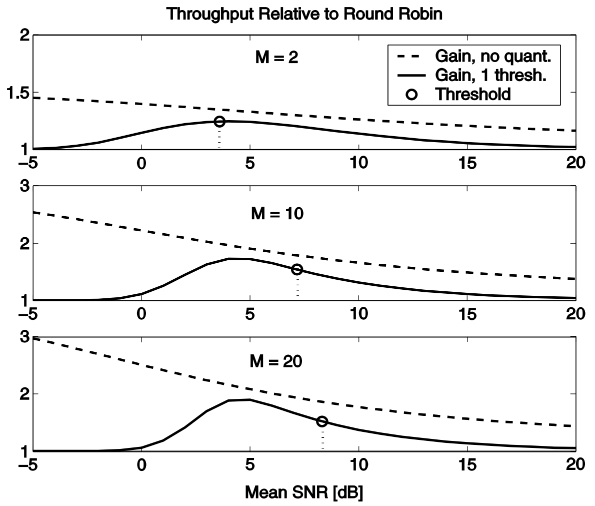

Fig. 3. Gain in throughput in terms of capacity relative to the Round-Robin scheme for 2, 10, and 20 users for the single-threshold case.

For illustrative purposes, the throughput for these two cases have been normalized with that of the Round-Robin scheme, which is given by (7). The threshold has been chosen to maximize the throughput at a mean SNR of 5 dB, and the optimum quantization threshold is marked with a circle on the curve for the throughput using quantized feedback. The optimum thresholds were obtained by numerically optimizing (17) with respect to q_1 when $N = 2$, $q_0 = 0$, and $q_2 = \infty$.

In Figure 3, it is firstly noted that the gain compared to the Round-Robin scheme for the unquantized case increases with the number of users. This is intuitively clear since more users give a higher probability that some user experiences a high SNR. Therefore, the PDF of the SNR conditioned on that a transmission occurs has a mean that increases with the number of users. Secondly, it is seen that using one threshold achieves throughputs close to the unquantized limit, although this occurs only around the mean SNR for which the threshold is optimized. Specifically, one threshold achieves approximately 93% of the throughput using unquantized feedback for $M = 2$ users and approximately 91% for $M = 10$ and 20 users. Interestingly, in the case of two users the throughput is relatively closer to the unquantized throughput compared to that of the case of 10 and 20 users. This is due to that the throughput using unquantized feedback increases with the number of users. As the difference in throughput between the Round-Robin scheme and the Max-SNR scheme using unquantized feedback increases, the number of thresholds required to attain a certain fraction of the maximum gain also increases.

That the optimum threshold increases with the number of users is intuitive since, for a given threshold, there will be an increasing number of users above the threshold, in a mean sense, as the number of users increase, giving a higher probability that a user with a low SNR is chosen when this user is above the threshold. However, when the threshold is increased, the probability that a user is chosen when he is above the threshold increases as well as the expected SNR given that particular user is chosen. On the other hand, the threshold cannot be positioned too high since then the probability that none of the users' SNRs is mapped to the highest quantization level becomes too large. This will result in that the lowest quantization level contains users with both high and low SNRs and, in effect, users will be selected randomly regardless of their SNRs. In this case, the throughput will therefore approach that of the Round-Robin scheme.

Unfortunately, no closed form expression that explicitly shows the dependence of the optimum threshold on the number of users could be found except for the simplest case of two users and one threshold. For this case, straightforward differentiation of (17) gives, after some algebra, the optimum threshold, q^*, for a two-user, single-threshold case

$$q^* = e^{\mathrm{E}_1(\bar{\gamma}^{-1})e^{\bar{\gamma}^{-1}}} - 1 \tag{18}$$

For a mean SNR of 5 dB, (18) evaluates to $q^* = 3.6$ dB, which agrees with the result of the numerical optimization in the upper-most plot in Figure 3.

Figure 1a indicates that it would be interesting to investigate the number of thresholds required in order to achieve a minimum throughput for a certain mean SNR range, say 0-10 dB, as compared to optimizing the throughput for only one mean SNR. This can also be interpreted as finding a quantizer that performs well over a range of mean SNRs. Therefore, the minimum fraction of the throughput using unquantized feedback over a mean SNR range is maximized for different degrees of quantization. That is,

$$\max_{\mathbf{q}_N} \min_{a \leq \bar{\gamma} \leq b} \frac{C_{\text{QMS}}(\bar{\gamma}, \mathbf{q}_N)}{C_{\text{MS}}(\bar{\gamma})} \tag{19}$$

and

$$\max_{\mathbf{q}_N} \min_{a \leq \bar{\gamma} \leq b} \frac{\eta_{\text{QMS}}(\bar{\gamma}, \mathbf{q}_N)}{\eta_{\text{MS}}(\bar{\gamma})} \tag{20}$$

where $\mathbf{q}_N = [q_1, q_2, \ldots, q_{N-1}]$ denotes the vector of variable quantization thresholds, are sought. In (19) and (20), a and b are the upper and lower limits, respectively, of the mean SNR range of interest, and the dependence on the mean SNR and the quantization thresholds have been indicated explicitly. Note that in the numerical optimizations, the expressions above were evaluated for a mean SNR of 0 to 10 dB with 1 dB increments.

Figures 4 and 5 show the results of the optimization for uncoded transmission and in terms of capacities, respectively. In the former figure it is seen that with one threshold a throughput greater than 94% of the unquantized gain is achieved for the mean SNR range for $M = 2$ users, and for $M = 10$ and 20 users that figure is 80%. For two thresholds a throughput of more 90% of that of the unquantized case is obtained for 2, 10, and 20 users. As observed earlier in connection with Figure 3, the number of thresholds required in order to achieve a certain fraction of the maximum throughput increases with the number of users.

In Figure 5 it is seen that by using one threshold the gain when using quantized feedback is at least around 70% of the

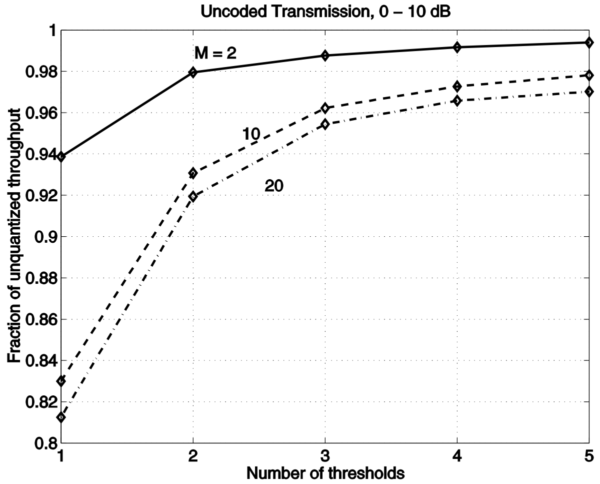

Fig. 4. Fraction of throughput using unquantized feedback for uncoded transmission ($N = 200$) for 1 to 5 thresholds and for 2, 10, and 20 users.

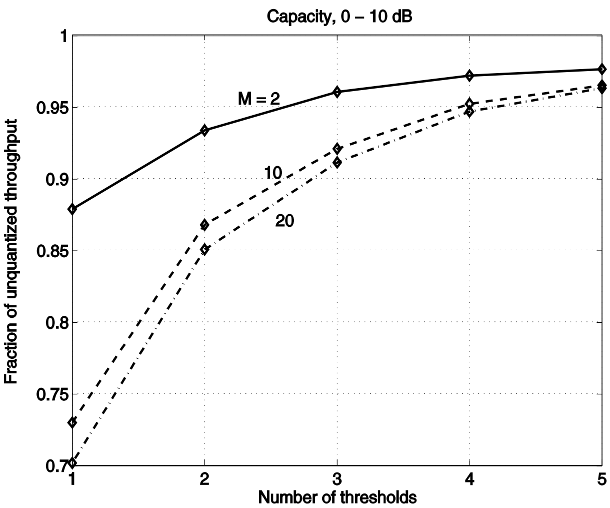

Fig. 5. Fraction of throughput using unquantized feedback in terms of capacities for 1 to 5 thresholds and for 2, 10, and 20 users.

maximum gain for 10 and 20 users and 87% for the case of two users. These values can be compared to those of Figure 3, which were higher and obtained by optimizing for a single mean SNR. By increasing the number of thresholds to two, a significant improvement is made and for 10 and 20 users and approximately 85% of the maximum gain is achieved. If a throughput of at least 90% of the maximum is desired, three thresholds is sufficient for 2, 10, and 20 users. Also when using the capacity as a throughput measure, it is observed that increasing the number of users requires that the number of thresholds be increased in order to achieve a certain fraction of the maximum gain. As mentioned above, this is due to the fact that the gain increases with the number of users.

It is also noted that, for both performance measures, the difference in throughput between 2 and 10 users is much larger than between 10 and 20 users, indicating that the difference will be even smaller for more than 20 users. The reason is that the gain when using unquantized feedback increases more slowly as the number of users grow.

By calculating the throughput in terms of packets per slot using uncoded binary transmission, a behavior more similar to that of a practical system is obtained although this performance measure is rather pessimistic. This is in contrast to using capacities, i.e., expressing the throughput in bits per channel use, which is a measure that gives a theoretical upper bound on the throughput. However, for both performance measures it is seen that using quite few thresholds gives a throughput greater than 90% of that of the unquantized case.

VI. Conclusions

We have in this paper investigated the impact of quantization of feedback information on the throughput of a Multiuser Diversity scheme for constant rate transmission. In the downlink, this scheme compares the users' SNRs, which are fed back over an error-free return channel, and transmits to the user with the highest SNR. The throughput was investigated in terms of ergodic capacities and throughput of uncoded packets, and the channel was assumed to be block-Rayleigh fading. The effect of the degree of quantization of the fed-back SNRs was also investigated, and for a case representing a practical scenario it was illustrated how the performance of a quantizer that yielded high gain over a range of mean SNRs was found by numerical optimization. It was concluded that only a few quantization levels were needed in order to achieve a large fraction of the diversity available in the multiuser dimension. Moreover, the number of thresholds required in order to achieve a certain fraction of the throughput for the unquantized case increased with the number of users, although the difference was small for a high number of users.

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] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.

[3] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE ICC '95*, vol. 1, pp. 331–335, 1995.

[4] I. Bettesh and S. Shamai, "A low delay algorithm for the multiple access channel with Rayleigh fading," in *Proc. IEEE PIMRC '98*, vol. 3, pp. 1367–1372, 1998.

[5] J. G. Proakis, *Digital Communications*. McGraw-Hill, fourth ed., 2001.

[6] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.

[7] L. Råde and B. Westergren, *Mathematics Handook for Science and Engineering*. Studentlitteratur, fourth ed., 2001.