Diversity-Enhanced Equal Access
– Considerable Throughput Gains With 1-bit Feedback

Mathias Johansson
Signals and Systems
Uppsala University
Box 528, SE-75120
Uppsala, Sweden
E-mail: mj@signal.uu.se

Abstract — We investigate performance aspects of adaptive modulation and scheduling as the amount of channel feedback is reduced. We study throughput, fairness and the sensitivity to incorrect channel quantizations.

A main finding is that the throughput of a cellular downlink using strict multiuser diversity does not degrade significantly when the channel information is heavily quantized. On the other hand, unfairness increases and due to an inherent sensitivity to incorrectly chosen quantization levels there is a risk of occasional drastic performance drops.

Noting that fixed-access schemes do not have the bad properties of multiuser diversity, but achieve unsatisfactory throughput, we propose a scheme combining the good aspects of multiuser diversity with the desirable properties of fixed access schemes. The result is a low-complexity scheduler and quantization policy that achieve large throughput gains as compared to fixed access without compromising fairness.

I. INTRODUCTION
The use of scheduling and adaptive modulation based on the predicted channel quality is known to promise significant throughput gains [1] in the downlinks of cellular communication systems. The scheduling policy that maximizes system throughput is to transmit exclusively to the user that can receive at the highest rate at the particular time slot, provided that this user has at least as much data to send as his channel can support [2], [3]. Although throughput is then maximized, there is a risk that some users do not receive any service at all. In the extreme case, with one user having a channel that is constantly better than all other users’ channels, only the best user will receive any data. There is a rich literature of scheduling policies that try to also take some precautions so as to reduce the risk of unfairness, see e.g. [4], [5]. There is however no generally agreed upon definition of fairness, and the proposals generally cannot guarantee fairness other than with some probability, or asymptotically as time goes to infinity. Further, the inevitable price for increased fairness is throughput degradation.

In order to realize the performance gains promised by scheduling and adaptive modulation, channel information is required at the transmitter. Essentially, the base station must know which modulation level to use. As each user must feed back channel information for each scheduled time slot, the amount of feedback should be reduced to a minimum, in particular for multi-channel systems such as OFDM.

In this work, we study how the pure throughput-maximizing strategy is affected by 1-bit channel feedback as compared to a traditional fixed-access scheme where the users transmit in a fixed order regardless of channel quality. We find that the throughput is not as badly affected by reduced channel feedback as is the fixed-access strategy, but unfairness generally increases. Furthermore we note an inherent extreme sensitivity to the quantization. Small errors may lead to drastic throughput drops.

It was noted in [6] that the throughput degradation with pure multiuser diversity due to limited channel feedback is not as bad as that of traditional fixed access. That work did however not recognize the quite severe disadvantages of multiuser diversity with limited feedback that we find in this fuller analysis.

As a solution to the problematic aspects of multiuser diversity, we suggest a simple modification of fixed access which results in taking advantage of multiuser diversity while guaranteeing a maximum inter-access time of twice the number of users in the cell. The proposed scheme does not suffer from the throughput degradation associated with limited feedback and fixed access, and does not have the extreme sensitivity of pure multiuser diversity. Simulations show that the proposed scheme yields considerable throughput gains at low computational complexity.

II. QUANTIZATION FOR MAXIMUM EXPECTED THROUGHPUT
We consider a quantization scheme in which the mobile terminals predict their signal-to-noise ratio (SNR), then determine the corresponding attainable transmission rate, and send a quantized value of the rate to the base station. The transmission rate may be approximated by the general formula

\[ r_u = \log_2 \left( 1 + \frac{\text{SNR}_u}{\Gamma_u} \right), \]

where \( r_u \) is the transmission rate of the \( u \)-th user, \( \text{SNR}_u \) is the predicted SNR at the receiver of user \( u \), and \( \Gamma_u \) is a system-specific value which depends on the desired bit-error rate (BER) and the type of modulation and coding used. For instance, (1) is a good approximation of the attainable rate using Gray-coded M-QAM modulation [7] with

\[ \Gamma_u = -\ln(5\text{BER}_u) \frac{1.6}{\ln(5\text{BER}_u)} . \] (2)

Consider a simple quantization strategy in which a user sends a 1 to the base station if that user’s rate satisfies \( r_u \geq q \). Here, \( q \) is a rate that is determined by the base station and that in
this section is assumed to be equal for all users. If \( r_u < q \), the user sends a 0. The base station then transmits with rate \( q \) to any one of the users who signalled that their channel would support rate \( q \). In the remainder of this section we will examine the consequences of using a rate threshold \( q \) which maximizes the expected throughput. Such a scheme would require that the rate threshold \( q \) be updated at regular (but infrequent) intervals as the channel and the number of users vary.

The expected system throughput with a pure throughput-maximizing strategy becomes

\[
\langle x \rangle = q P(q \mid I),
\]

with \( P(q \mid I) \) denoting the probability that there is at least one user that can receive at a certain rate \( q \), conditional on any relevant information \( I \) at hand regarding the users’ channels.

Note that, assuming logical independence between different users’ channels,

\[
P(q \mid I) = 1 - \prod_{u=1}^{U} P(r_u < q \mid I),
\]

where \( U \) denotes the number of users, and \( r_u \) the rate with which user \( u \) can receive data at the particular time slot to be scheduled. Hence, the expected throughput is

\[
\langle x \rangle = q \left( 1 - \prod_{u=1}^{U} P(r_u < q \mid I) \right). \tag{5}
\]

We can draw some interesting conclusions about the behavior of a throughput-maximizing policy already from this expression (5). If all the users have identical independent rate probability distributions then the probability that there is at least one user who can receive with an arbitrary rate \( q \) becomes

\[
P(q \mid I) = 1 - P(r_u < q \mid I)^{U}. \tag{6}
\]

Under the assumption that the SNR pdf for each user is exponential (corresponding to the case of a Rayleigh fading channel) with known mean \( \text{SNR}_{u} \),

\[
P(\text{SNR}_u \mid I) = \frac{1}{\text{SNR}_u} \exp \left( - \frac{\text{SNR}_u}{\text{SNR}_{u}} \right), \tag{7}
\]

and that the relation between SNR and rate is given by (1), the rate pdf \( P(r_u \mid I) \) for each user is obtained by a variable transformation:

\[
P(r_u \mid I) = P(\text{SNR}_u \mid I) \frac{\text{dSNR}_u}{dr_u} = P(\text{SNR}_u \mid I) \Gamma_u 2^{\text{SNR}_u} \ln 2 \]

\[
= \Gamma_u 2^{\text{SNR}_u} \ln 2 \exp \left( - \frac{\text{SNR}_u}{\text{SNR}_u} \left( 2^{\text{SNR}_u} - 1 \right) \right). \tag{8}
\]

The probability that a user can receive at a rate in the interval \( q_1 < r_u < q_2 \) is then

\[
P(q_1 < r_u < q_2 \mid I) = \int_{q_1}^{q_2} P(r_u \mid I)dr_u
\]

\[
= \int_{q_1}^{q_2} \Gamma_u 2^{\text{SNR}_u} \ln 2 \exp \left( - \frac{\text{SNR}_u}{\text{SNR}_u} \left( 2^{\text{SNR}_u} - 1 \right) \right) \]  

\[
= \exp \left( - \frac{\text{SNR}_u}{\text{SNR}_u} \left( 2^{\text{SNR}_u} - 1 \right) \right) \exp \left( - \frac{\text{SNR}_u}{\text{SNR}_u} \left( 2^{\text{SNR}_u} - 1 \right) \right). \tag{9}
\]

With \( q_1 = 0 \) as in (6), (9) becomes

\[
P(q_1 < q \mid I) = 1 - \exp \left( - \frac{\Gamma_u \left( 2^{x} - 1 \right)}{\text{SNR}_u} \right). \tag{10}
\]

We can easily find the throughput-maximizing value of \( q \), by inserting (10) in (5) and find the integer \( q \) which maximizes (5). For \( U = 30 \) users, with mean individual SNR \( \text{SNR}_{u} = 13 \text{db} \) and Gray-coded M-QAM with a desired BER of \( 10^{-3} \), i.e. \( \Gamma_u = 2 \) determined by (2), we find the optimum to be \( q = 4 \), yielding an expected throughput of \( \langle x \rangle = 3.71 \) bits per symbol. With perfect channel information at the transmitter (i.e. without quantization) and adaptive modulation supporting any integer positive rate, the expected throughput becomes 4.09 bits per symbol. The performance drop by going from unlimited resolution to a 1-bit quantization is thus only 10%.

Compare this to the case of using a traditional fixed-access scheme, in which users transmit in the same order regardless of channel quality. Then multiuser diversity is completely lost, and, under the same assumptions as just described, the expected throughput with perfect channel knowledge becomes

\[
\langle x \rangle = \sum_{u=1}^{U} P(r_u < q \mid I)dr_u \approx 2.35 \text{ bits per symbol}.
\]

With a 1-bit quantization, the optimally adjusted \( q \) for maximum expected throughput is determined from (5) with \( U = 1 \). The result is \( q = 2 \), yielding an expected throughput of \( \langle x \rangle = 1.22 \). Evidently, with fixed access the expected throughput is approximately halved by a 1-bit quantization as compared with perfect channel knowledge. Hence, with regard to optimum throughput, it is clear that multiuser diversity-driven systems do not suffer at all as badly from reduced feedback as does the traditional fixed-access scheme.

Let us now discuss the sensitivity to erroneously set rate thresholds \( q \). Consider again a system employing pure multiuser diversity; at each time slot the user with currently highest rate is served. With a large number of users, the probability distribution for the rate that will be used may become extremely sharp\(^3\); up until a certain level there will be almost probability 1 that someone can receive at that rate, but then it suddenly drops down to zero. This drop will be extremely steep. For instance, consider the same scenario as in the preceding paragraph. Then the expected throughput with \( q = 4 \) is 3.71 bits per symbol. Increasing the threshold to \( q = 5 \) however yields an expected throughput of only 0.81 bits per symbol, a most dramatic performance decrease! The probability for being able to transmit at a particular rate is almost certainty; just adding one bit to that rate leads to a probability for transmission of only 16%. The expected throughput decreases by a factor of 4.56 if the selected threshold changes by a factor of only 1/4. The throughput degrades to below what can be expected from using fixed access!

In practice, the base station has very little information regarding individual channels and is therefore in the unenviable position of realizing the risk for potential performance breakdown (to a level well below that of ordinary fixed access) but having no information as to ensure its avoidance.

Furthermore, since a correctly chosen threshold \( q \) will rely heavily on the upper tails of the individual rate distributions, there is a large risk that the throughput-maximizing \( q \) will be set so high that only a very small number of users will ever be able to receive at that rate. Consider for example a case in which the mean SNRs of different users range from, say, 6–30db

\(^1\)We defer a discussion of individually adjusted rate thresholds to the next section, after we have acquired some more insight into the problem.

\(^2\)From (1) we have that \( \text{SNR}_u = \Gamma_u (2^{Q_u} - 1) \) and consequently \( \frac{\text{dSNR}_u}{dr_u} = \Gamma_u 2^{Q_u} \ln 2 \).

\(^3\)In particular, this happens when all users have the same mean SNR, e.g. due to slow power control.
III. Diversity-Enhanced Equal Access

The examples in the previous section, although not exhaustive, show that for systems employing strict multiuser diversity, if the quantization levels are optimally adjusted and the number of users is not too small, then the throughput decrease is negligible for a 1-bit quantization as compared to having perfect channel knowledge. On the other hand, in order to take full advantage of multiuser diversity the quantization levels need to be adapted as the number of users change and as the channel statistics vary. This brings out an inherent property of using pure multiuser diversity; a strong performance sensitivity to the rate threshold \( q \). A small change of \( q \) can result in a drastic performance loss, to the extent that traditional fixed access would actually achieve higher throughput. With limited individual channel knowledge at the base station, the risk of sometimes setting \( q \) too high is probably unavoidable. Further, we argued that unfairness is typically amplified as the feedback is reduced.

Interestingly, the implications of reduced feedback for fixed access, where each user gets served in the same order regardless of channel quality, are logically opposite to the implications for multiuser diversity. Fairness is of course not affected at all, and is perfect in the sense that each user gets equal access to the channel with a constant inter-access time. We saw that throughput roughly decreased to half of the average achievable rate. Finally, since the optimum quantization level no longer depends either on the number of users or the details of the tail of the individual rate distributions, the sensitivity to erroneously chosen rate thresholds \( q \) is small.

Should we combine the good aspects – high throughput and little degradation due to rate quantization, fairness, and small sensitivity to the rate thresholds – of the two alternatives? First of all, note that we would in principle wish to use individual rate thresholds, so that sending a 1 to the base station would mean using rate \( q_1 \) for user 1 and for user 2 it would mean using rate \( q_2 \). Determining the optimal individual thresholds would however require accurate knowledge of every user’s individual channel distribution, which seems antithetical to the desideratum that feedback should be minimized. If we could decentralize the individual threshold determination, so that each mobile terminal adjusts its own threshold, and periodically, but infrequently, updates the base station of its threshold, then it is highly plausible that the sensitivity could be reduced and possibly also the unfairness. Such a procedure would however not result in maximum possible expected throughput, since the optimum levels still require taking into account all other users’ channel distributions. Even if we could somehow determine “good enough” individual rate thresholds, we cannot guarantee fairness with the pure multiuser-diversity strategy.

For these reasons, we propose a modified scheduling strategy, which combined with individual rate thresholds will guarantee fairness, high throughput, and low threshold sensitivity.

Scheduling Policy

Assume that there are \( U \) users in the cell and that the base station has channel knowledge (quantized or unquantized) only of one time-slot ahead. The proposed scheduling policy then works as follows:

1. In the first time slot, let the user \( u^* \) with maximum rate transmit.
2. In the second time slot, out of all users except for \( u^* \) let the user with maximum rate transmit.
3. In each of the following time slots, select the user with maximum rate out of the remaining users that have not yet accessed the channel.
4. After \( U \) time slots, all users have accessed the channel, and the procedure is restarted with all users again participating in the competition.

This simple strategy has the attractive feature that it has the same fairness properties of fixed access, guaranteeing a maximum inter-access time of \( 2U - 1 \) time slots, while still utilizing multiuser diversity.

At the first time slot, the proposed policy employs a pure multiuser-diversity strategy for \( U \) users; in the second slot it does so again but only among \( U - 1 \) users, and so on. Thus, over a period of \( U \) time slots the policy can be interpreted as taking full advantage of multiuser diversity among a number of users that is decreasing by one for every time slot. We would then expect that in terms of throughput the policy would on average achieve full multiuser diversity gain for a system of approximately \( U/2 \) users. This is the price that is paid by guaranteeing equal access. It can however be observed that the multiuser diversity gain increases more slowly the larger \( U \) becomes [1], [3]. Thus, with many users in the system, the proposed policy will not be far from the maximum throughput strategy.

Quantization Policy

Each mobile terminal will select its own threshold \( q_u \) based on previous measurements of its channel statistics. Every \( M \)th time slot (with \( M \) so large as to make its contribution to the overall amount of feedback negligible), an updated \( q_u \) is reported to the base station. At the base station, a table is kept containing the rate thresholds \( q_u \) for all users. When a user \( u \) sends a 1 to the base station, it means that it can receive at rate \( q_u \) in the next time slot.

Consider the determination of the rate threshold \( q_u \) for a particular user \( u \). A simple approach would be to maximize

\[
q_u P(r_u > q_u | I) ,
\]  

but note that this expression does not take into account the fact that a user on average competes for access over more than one time slot. In effect, the expression does not take full advantage of the multiuser diversity that is utilized by the proposed scheduling policy. If the user would know the number of slots, \( n_u \) \((1 \leq n_u \leq U)\), that this user has the highest rate of all users, then he should use the \( q_u \) that maximizes his expected throughput

\[
\langle x_u \rangle = q_u (1 - P(r_u < q_u | I)^{n_u}) ,
\]  

Throughout this paper we assume that a time slot consists of only one symbol, but the results are valid for any slot size.
where \(1 - P(r_u < q_u | I)\) is the probability that the rate \(r_u\) is larger than \(q_u\) at least one out of \(n_u\) time slots.5

In practice, however, \(n_u\) is not only unknown, it further depends on the other users’ channels and their choices of thresholds. But if we admit the proofs of the Bayesian paradigm [8], [9] that probabilities can represent an honest description of a state of knowledge concerning some incompletely known quantity, and that the only set of rules that is internally consistent and satisfy the properties of an idealized uncertain reasoner are the usual product rule and sum rule of probability theory, then we could assign a probability for \(n_u\) which represents our uncertainty concerning its actual value. The expected throughput is then obtained by multiplying (12) by \(P(n_u | I)\) and then integrating out \(n_u\) as a nuisance parameter. As no value of \(n_u\) within the range 1 . . . \(U\) is more likely than any other the principle of indifference applies, and we assign a uniform probability distribution to \(n_u\):

\[
P(n_u | I) = \frac{1}{U - 1}.
\]

The expected throughput with unknown \(n_u\) thus becomes

\[
\langle x_u \rangle = \frac{q_u}{U - 1} \int_1^U \left(1 - P(r_u < q_u | I)\right) dn_u
\]

\[
= q_u \left(1 - \frac{P(r_u < q_u | I) | r_u - q_u | I}{\ln(P(r_u < q_u | I) | U - 1)}\right).
\]

The expression is not fully determined until we have definite expressions for \(P(r_u < q_u | I)\). If we allow only integer rates, then

\[
P(r_u < q_u | I) = \sum_{r_u < q_u} P(r_u | I),
\]

where \(P(r_u | I)\) is the probability distribution for the individual rates \(r_u\).

As an example of a simple but useful model for \(P(r_u | I)\) consider the following. Assume that the channel statistics are stationary over an interval that is at least \(N + M\) time slots, where \(N\) is the length (in time slots) of a measurement window upon which we base the determination of \(P(r_u | I)\), and \(M\) is the period between updates of \(q_u\). If the mobile terminal measures the receiver SNR in each time slot over a period of the \(N\) most recent time slots, and calculates the corresponding attainable rate \(r_u\) based on the rate-SNR relation that is applicable for the particular system, e.g. (1), then the probability \(P(r_u | I)\) is (Chapter 18 [8])

\[
P(r_u | I) = \frac{n_{r_u} + 1}{N + K},
\]

where \(n_{r_u}\) is the number of measured time slots in which the rate \(r_u\) (but not the rate \(r_u + 1\)) was attainable, and \(K\) is the number of possible modulation levels in the system. (If for instance \(r_u\) can take on any integer value between 0 and 8 bits per symbol, then \(K = 9\).)

In summary, the rate threshold \(q_u\) is determined from maximizing (14) using (15) and (16). The maximization is easily carried out by a one-dimensional numerical search over \(K\) integers. For instance, try \(q_u = 1\), then increase \(q_u\) by 1 until the expected throughput (14) drops. It should be emphasized that this procedure requires only a small number of arithmetic operations.

Note that this set-up is not equivalent to a uniform user distribution over the cell area, but was chosen for simplicity. The results are however representative also for other user distributions, as briefly mentioned in the end of the section.

A fuller study using more elaborate time-varying models of channels would be of great interest for specific network architectures, but the results would be harder to interpret and generalize.

Fig. 1: The optimized rate thresholds for 16 users having exponentially distributed SNR with mean SNR ranging from 30 - 60db. The users are ordered by decreasing mean SNR. The dark color refers to the optimum fixed-access thresholds, while the light color refers to the optimum thresholds using (14).

IV. Simulations

In this section we aim to verify that the scheme proposed in Section III does indeed overcome the problems of fixed access and those associated with the pure multiuser-diversity policy.

A set of simulations was carried out, in which 16 users were spread out uniformly over the cell radius6, and where each individual user’s SNR was exponentially distributed with a fixed mean proportional to \(d^{-2}\) where \(d\) is the distance to the base station. The proportionality constant was chosen so that the mean SNR of the 16 users ranged from 30db down to 6db. The rate-SNR relation (1) was used with \(\Gamma_u = 2\).

In order to simplify the comparison, the distributions were assumed to be stationary and the system assumed to have been started in an infinite past, so as to ensure that the probabilities \(P(r_u | I)\) were set correctly for all users. The simulation was run for 1600 time slots and the reported results are averages from 100 simulation runs. In order to make a fair comparison, the throughput was set to zero in time slots when none of the remaining users could transmit at their rate threshold. In reality, one would obviously choose to transmit to another user who has already received service in such cases7.

Figure 1 shows the rate thresholds obtained from maximizing (14) and the thresholds obtained from maximizing (11), i.e. the optimum quantization for a fixed access scheme that does not utilize multiuser diversity. The general tendency in using (14) is, as expected, to set the levels somewhat higher since a

5A useful analogy is to consider the probability of obtaining at least one 5, say, or higher when throwing a regular die. As the number of trials increase, the probability increases correspondingly.

6Note that this set-up is not equivalent to a uniform user distribution over the cell area, but was chosen for simplicity. The results are however representative also for other user distributions, as briefly mentioned in the end of the section.

7With such a mechanism, the proposed scheme would have an even bigger performance advantage than the present simulations suggest.
user typically competes for more than one time slot, thereby increasing his chances for obtaining a higher rate at least once in the $U$ slots.

In Figure 2, the total individual throughput obtained from using the proposed scheduling and quantization policy is plotted and compared to the throughput obtained by using the same scheduling policy but with the rate thresholds obtained from (11). It can be seen that almost every user obtains increased throughput by choosing the more aggressive quantization strategy. The total throughput summed over all users increases by approximately 27% by using the higher rate thresholds.

In order to see how the use of individual thresholds affect the performance, we also tested using a common quantization level optimized for the median user. With individually optimized thresholds using (14), the throughput increase was approximately 80%.

The multiuser-diversity gain was quantified by comparing the obtained throughput to a fixed-access schedule with a common rate threshold. The throughput increase was now 168%. In comparison to a fixed-access scheme with individually and for fixed-access optimally adjusted rate thresholds, the throughput increase was 90%.

Under somewhat different channel assumptions, with $U = 16$ users having identical but independent rate distributions (8) with mean SNR 15db and $\Gamma_u = 2$, the performance gain of using the proposed scheduling and quantization policy was about 25% as compared to using the same scheduling policy but with the rate thresholds obtained from (11). The optimum rate thresholds were found to be $q_u = 4$ for all users. In this scenario, it is possible to determine how much throughput is lost by using the proposed scheme in comparison to using a pure multiuser-diversity strategy. A numerical search found the optimum common rate threshold for pure multiuser-diversity to be $q = 5$. In order to carry out a fair comparison between the two approaches we let our proposed policy be augmented by a mechanism for avoiding transmitting zero bits in the time slots when none of the remaining users can reach their rate threshold. In such time slots, the policy instead transmits to an arbitrarily chosen user with non-zero rate. The throughput increase from using the pure multiuser-diversity strategy with the optimum quantization is then just below 25%, as expected.

In the previous section we conjectured that the proposed scheduling and quantization policy would be roughly equivalent to a pure multiuser-diversity strategy with $U/2$ users. With 8 users, the optimum $q$ for pure multiuser diversity in the current simulation scenario is $q = 4$, which is also the individual optimum for the proposed policy for 16 users, and as predicted, there is no throughput difference.

V. Conclusions

We have shown that by using individual rate thresholds and enhancing the traditional Round Robin fixed schedule to take advantage of multiuser diversity, we obtain a multiuser diversity-gain corresponding to $U/2$ users, while every user receives guaranteed access within at most $2U - 1$ slots from the previously accessed slot. The inherent sensitivity of a pure multiuser-diversity strategy to the choice of rate threshold was avoided by decentralizing the determination of rate thresholds. Using the proposed scheduling and 1-bit quantization policy was seen to yield considerable throughput gains over fixed access, in simulations yielding a factor of 2–3 throughput increase depending on how the fixed-access rate quantization was carried out.

Finally, the proposed scheme is relatively easy to implement as it is decentralized, and only relies on counting the number of times that different rate levels have been attainable during a recent time period. Letting $K$ be the number of possible modulation levels, the optimum level can be found by comparing on average $K/2$ values of (14).

References


Note that in this case, since all users have identical independent rate distributions, nothing would be gained by having individual rate thresholds. This applies to both strategies.