ADAPTIVE MODULATION AND SCHEDULING FOR FADING CHANNELS

Nilo Casimiro Ericsson
Signals and Systems, Uppsala University,
PO Box 528, SE-751 20 Uppsala, Sweden
nce@signal.uu.se

ABSTRACT

In future packet based wireless communication systems, transmission in the downlink will often dominate the traffic load. High bit-rate applications like WWW-browsing, file transfer, and full motion video will impose strong requirements on the system capacity. An obstacle in this context is the time-variability of the channel: For mobile users, frequently occurring fading dips will cause unnecessary and capacity degrading retransmissions.

To achieve a high throughput also over fading channels, adaptive methods for adjustment of e.g. the modulation alphabet, and the coding complexity, can be used. The idea is to make efficient use of the bits: Whenever the channel conditions are adequate, transmission of redundant bits should be avoided.

In this paper we shall investigate the effect of adaptive modulation, along with time-slot scheduling, in a scenario involving several mobiles and one base station.

We found that the bit error rate can be kept at an adjustable low level, well suited for Forward Error Correction (FEC) codes. Moreover, the more mobiles that share a frequency channel, the more efficiently the scheduler uses the channel bandwidth.

1 INTRODUCTION

Fading channels confront us with the problem of lost packets and the need for frequent retransmissions. One strategy to combat time-variability is to use averaging: Spread-spectrum signalling can average out variations of the noise and interference level, while coding and interleaving can compensate for the temporary loss of signal strength due to fading dips provided that modem synchronisation can be maintained. Such strategies can combat bad signalling conditions, but are inefficient when the channel conditions are good. In this paper we explore a strategy, where the time-variations of the channel, due to short-term fading, are estimated and the signalling scheme is adapted accordingly. We can exploit temporarily good transmission conditions to obtain higher throughput, while reducing the demands on the channel when its condition is bad. Assuming a system making use of either Frequency Division Duplex (FDD) or Time Division Duplex (TDD), with separate (ideal) control channels, the current channel parameters can be estimated and predictions about their future evolutions can be stored for subsequent transmission in the control channel. The bit-rate can be tailored to the current channel conditions by e.g. adjusting the modulation complexity, while keeping the transmitted symbol energy at a constant level. The further into the future the terminal can perform accurate predictions of the channel parameters, the more flexible and efficient the selection of the modulation alphabet will be. Moreover, the traffic on the control channel can be efficiently planned to minimize the signalling overhead.

For a predicted value of the Signal to Noise and Interference Ratio (SNIR) of each channel, the modulation level is maximized under the constraint of a certain probability of symbol error, for example, $P_M \leq 10^{-5}$. If no modulation level attains the required error probability, then transmission is deferred until later when the SNIR is higher, thus avoiding retransmissions. The reason for using this strategy is that it will stabilize the error probability, thus keeping the error rate at a low and constant level, avoiding retransmissions. The averaging strategies mentioned above do not have this feature. On the contrary, they would yield a higher traffic load when conditions are bad, since the increasing error rate would increase the requests for retransmissions. In the case when many mobiles are connected to the same base station, sharing the same frequency, the strategy used here will instead be to allocate the channel to the mobile that can make the best use of it.

Similar approaches to adaptive modulation have been proposed by Ue, Sampei and Morinaga [1], Chua and Goldsmith [2], and Webb and Steele [3].

We assume that accurate long-term predictions of the channel SNIR are available, allowing us to *schedule* the transmissions for one or more users. A non-linear method for achieving accurate broad-band channel condition predictions is described by Ekman and Kubin in [4]. It is demonstrated to be capable of making accurate predictions farther than $10ms$ ahead in time, thus seeing beyond the next fading dip.

The proposed scheme will result in some overhead due to the transmission of scheduling decisions over separate control channels. It is crucial to the performance of this system that the control information is correctly transmitted. Torrance and Hanzo discuss this problem in [5].

A nice feature of our approach is that the adaptive modulation strategy is embedded in the scheduling operation. The decision on the modulation format is merely a first natural step in the scheduling process.

2 SYSTEM DESCRIPTION

We outline and investigate a system that exploits the time-variations in the channel, instead of fighting their effects on the data being transmitted. The system presented here

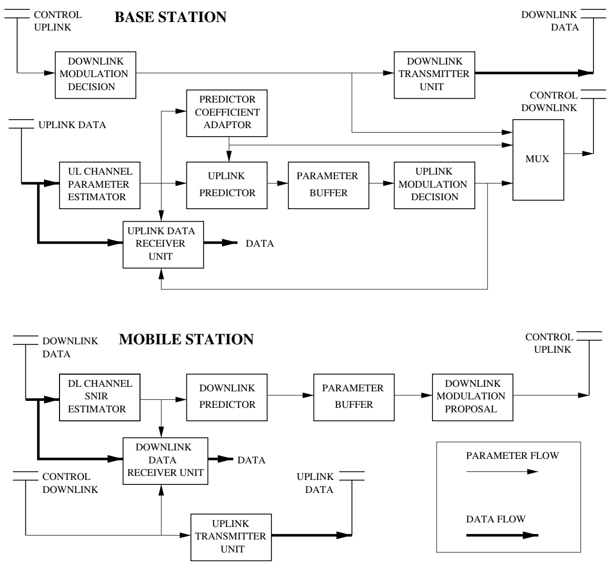

Figure 1: TDD downlink dominant system overview.

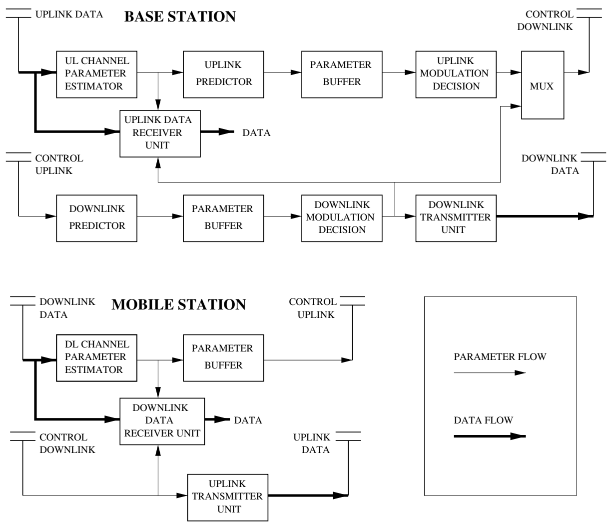

Figure 2: FDD symmetric system overview.

is intended to demonstrate the achievable performance gains when using an adaptive approach to the problem of digital transmission over time-varying channels.

We suggest two variants of the system: TDD for assymetric traffic, and FDD for symmetric traffic.

2.1 TDD system for asymmetric traffic

Time Division Duplex is a strategy for synchronizing uplink and downlink transmissions. The base station and the mobile share a common frequency, and the uplink and downlink transmissions are closely spaced in time. It can then be assumed that the channel conditions are correlated for the two transmissions. Under the assymetry assumption, the dominant part of the traffic will be carried in the downlink channel.

The current downlink SNIR is then estimated in the mobile and fed into its predictor. The predictions are performed regularly, so that the mobile is able to make a decision on which modulation level to prefer for each SNIR value. This decision only involves a table lookup, which is matching the SNIR to the required BER. The decisions of the mobile terminal are collected in a buffer before transmitting its contents to the base station via the control uplink channel. The final decision on the choice of the downlink modulation is made by the base station, which signals its decision to the mobile, which in turn prepares for reception accordingly.

Since we use a TDD system, the downlink channel can be estimated by analysis of the uplink transmission, but since we assume an asymmetric traffic, this would require frequent transmission of pilot signals in the uplink. We therefore propose that that the estimator and predictor of the downlink quality should be placed in the mobile in this type of system. Although the traffic in the uplink does not require much transmission time, its quality needs to be guaranteed. For this reason the uplink modulation is also adapted to the channel conditions. The total system is schematically described in Figure 1.

2.2 FDD system for symmetric traffic

A different scenario would be that of a high, symmetric, traffic load. This would require equal resources in both the uplink and the downlink. A natural way to accomplish such symmetric resource distribution is through Frequency Division Duplex. In FDD, the uplink and downlink channels are independent, acting on separate frequencies, so they can transmit simultaneously without affecting each other. In this case, we have two different channels. Their characteristics need to be estimated and predicted to make the adaptive modulation work efficiently.

The complexity and frequency of the predictor parameter updating is a topic for further research, but it would probably require considerable computations. For this reason, the adaptor is located at the base station.

For the case of FDD, see Figure 2, the predictor has also been moved to the base station, since there is a potential advantage in providing the base station with the downlink SNIR information: If the SNIR information is known to the base station, it can control the different connections more efficiently. The price to be paid is an increase in traffic on the control uplink channel caused by the transmission of estimated downlink SNIRs from the mobile channel estimator to the base station channel predictor.

Thus the base station decides on both the uplink and downlink modulation formats, and signals the result to the mobile over the control downlink channel.

3 ADAPTIVE MODULATION

In traditional communications systems, the radio transmission method is designed to cope with channel variations in a worst-case manner. For wireless systems this implies the use of a simple modulation scheme, and a complex error-correcting code. When the coding fails to compensate for temporary bad conditions, higher layers in the protocol will ensure that the information is correctly and completely transmitted, by requiring a retransmission of the erroneous data. We wish to avoid this by adapting our demands on the channel as it varies. By changing the modulation format as the channel SNIR (SIR) varies, we

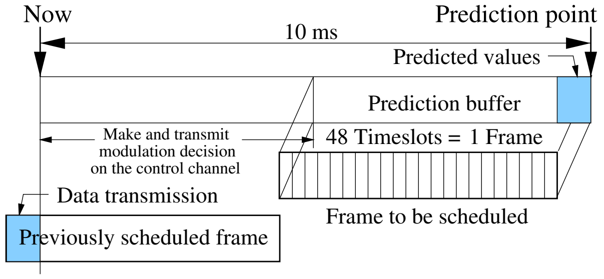

Figure 3: Based on measured channel SNIRs, predictions of future SNIR are made and stored in a buffer until a frame of $5ms$ is filled. The decision algorithm then has $5ms$ to make a sequence of decisions and transmit them to the other end of the link. Then, based on these decisions, the data transmission starts.

accomplish less retransmissions.

The modulation alphabet is chosen in advance, since we assume accurate predictions of the channel quality to be available. The decision is made on a frame-by-frame basis, each of which contains 48 time-slots, where each time-slot corresponds to one output value from the predictor. For each time-slot, a burst of 512 symbols is transmitted. Using a prediction horizon of $10ms$ we can collect the predictions during $5ms$, then use the remaining $5ms$ to make the decision and transmit it to the other side of the link, see Figure 3.

3.1 *Finding the decision thresholds*

For a given symbol error probability we can calculate the required SNIR for the different modulation formats used. Thus, we can decide the thresholds where we should change from one modulation format to another. A tight upper bound on the symbol error probability for M-QAM modulation is given by [6]:

$$P_M \leq 1 - \left[1 - 2Q\left(\sqrt{\frac{3E_{av}}{(M-1)N_0}}\right)\right]^2. \quad (1)$$

Here the average symbol energy E_{av}, the noise power, N_0, and the modulation format, M, are assumed to be known. The Gaussian cumulative distribution function, $Q(x)$, can be calculated according to:

$$Q(x) = \frac{1 - \mathrm{erf}(\frac{x}{\sqrt{2}})}{2}, \quad (2)$$

where $\mathrm{erf}(x)$ is the error function:

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \quad (3)$$

By solving (1) for $\frac{E_{av}}{N_0}$ and using (2) and (3), we get the SNIR required for a certain symbol error probability, P_M, and a given M:

$$\frac{E_{av}}{N_0} \geq \frac{2(M-1)}{3} \left[\mathrm{erf}^{-1}\left(\sqrt{1-P_M}\right)\right]^2. \quad (4)$$

Furthermore, since we measure and predict the average signal energy, whereas the constraints[1] have to be applied to the maximum signal energy, the value E_{av} has to be modified by a factor ξ_M, according to Table 1, yielding $E_{max} = \xi_M \times E_{av}$.

M	2	4	16	64
ξ_M	1	1	$\sqrt{9/5}$	$\sqrt{7/3}$

Table 1: Conversion factor from average energy to peak energy for M-QAM.

In this investigation we use 64-QAM as the maximum modulation level, thus transmitting six bits per symbol when the channel is as its best. When the channel de-

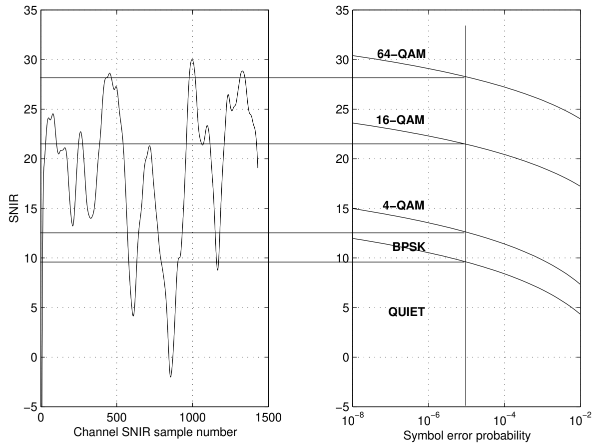

Figure 4: SNIR trend and modulation level related to the error probability.

grades, lower powers of two are used with BPSK being the lowest level. In Figure 4, the left hand diagram illustrates the SNIR-variation of a typical channel while the right hand part illustrates how the level of modulation can be selected for a pre-specified symbol error probability. As an example we note that for an SNIR $\geq 22dB$ we can transmit during 150 time-slots (time-slot 390 to time-slot 540) with a modulation level of 16-QAM at a symbol error probability of $P_M \leq 10^{-5}$.

Slow power control is assumed to compensate for long-term fading, thus holding the long-term average of the received power at a constant level.

4 SCHEDULING

One way to make use of the channel predictions is to simply choose the modulation format for a user during the nearest future, in order to satisfy the demand of a low and constant bit error rate. On the other hand, it can seem like a waste of time, and thus bandwidth, to choose a low modulation format (or not to transmit at all) when the channel condition for a specific link is poor: We could allocate that time-slot to another user, probably with better conditions. Moreover, a user may not need all its allocated bandwidth. Such inefficiencies can be avoided by the use of a scheduler. The scheduler works on a specific physical channel

[1] RF amplifiers have a constraint on the *maximum* input amplitude, for operation in the domain of *linear* amplification, which is required in multi-amplitude modulation

(such as a frequency band), where several links are maintained, and the users compete for transmission time.

In order to efficiently distribute the channel bandwidth between different concurrent users on a TDD/TDMA channel, we make use of the predictor in a way that is a natural extension from the original adaptive modulation approach. The predicted SNIR values are now not only used for the selection of modulation format, but also for time-slot distribution among the users so that the system throughput is maximized.

4.1 *Maximization of system throughput - Pass 1*

First, the scheduling procedure allocates each time-slot to the user that can transmit most efficiently in that particular slot. This approach actually maximizes the system throughput (for a given error probability), but it may be a very unfair way of allocating time-slots among different users. Even in this case, a user may be allocated more bandwidth than is required, whereas at the same time, another user may not be allowed to transmit at all. To compensate for this unfairness, a re-distribution of the time-slots takes place.

4.2 *Equalize user satisfaction - Pass 2*

In most cases there will be users that have received more time-slots than they need, and users that have received less. The re-distribution procedure starts by identifying the “richest” user, that is, the user with the largest over-allocation of bandwidth. The richest user offers its poorest[2] time-slots to the scheduler for distribution among the remaining users. The time-slots are given to the users that can use them best, only this time the set of competing users is limited to the ones that need more bandwidth.

This is only one example of many possible strategies for transmission scheduling. The method outlined above has a number of parameters that can be adjusted. One is the *order* in which the scheduler should prioritize between different features of each user and time-slot. The features implemented in the experiments described below are:

- Predicted SNIR
- Allowed modulation format
- User priority

Other methods to optimize the allocation decision can be considered, e.g. linear programming algorithms [7], and, generalizations of existing router-scheduling algorithms to take the varying channel quality into account. The drawback of linear programming methods is that the procedure is iterative. Thus no upper limit can be given for the number of operations required. Moreover, in the linear programming case, we would need to define a cost function that is to be minimized in order to optimize the scheduling decision.

5 EXPERIMENTS

To evaluate the proposed system solutions, a simulation series was conducted, assuming one base station transmitting to a number of mobile terminals. We assume that the channel conditions are predicted for 10 milliseconds ahead in time. Moreover, perfect synchronization and transmission at a constant maximum amplitude, regardless of the modulation alphabet, is assumed. The experiment is applicable to both TDD and FDD systems, provided that accurate predictions of the channel conditions exist.

For each prediction of the SNIR at the receiver, the modulation alphabet is selected for 512 consecutive symbols (one time-slot) for each user. This implies that the channel estimator and the predictor work at a rate of $\frac{bw}{512}$, where bw is the channel bandwidth. The data bit stream is then modulated and transmitted with a constant maximum amplitude over the noisy channel. White Gaussian Noise (AWGN) with varying variance is added to simulate good and bad channel conditions. At the receiver, the signal is demodulated and the obtained bit stream is compared to the original one. The number of errors is counted, as well as the number of transmitted bits.

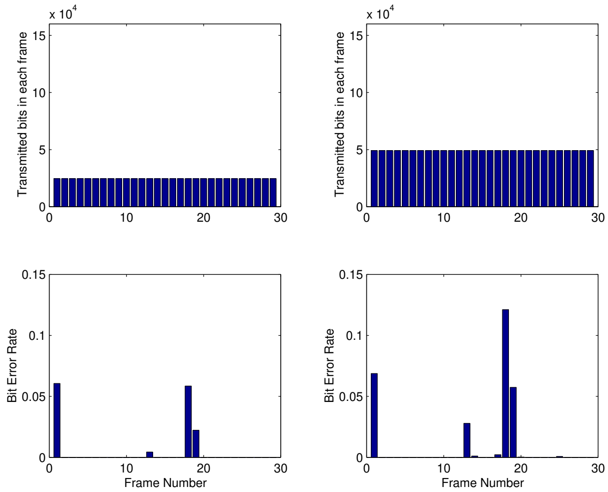

Figure 5: Using a constant modulation results in error bursts, with a high error rate, when the channel experiences deep fades.

In Figures 5-7 the outcome of six simulations are plotted. The top graphs show the bit throughput represented by a vertical bar for each frame. The maximum achievable throughput for 64-QAM is just below 15×10^4 bits per frame. The bottom diagrams show the resulting bit error rate (BER). Each frame consists of 48 time slots. Each time slot corresponds to one output sample from the channel predictor. The scheduling is optimizing the transmission within each frame.

As expected, the adaptive modulation approach results in a relatively constant (adjustable) error rate, see Figure 6. On the other hand, the use of non-adaptive modulation results in high peaks in the error rate when the receiver encounters a fading dip, see Figure 5.

6 CONCLUSIONS AND FUTURE WORK

The adaptive modulation approach provides a relatively constant error rate (Figure 6), which in turn provides an excellent basis for Forward Error Correction (FEC) codes, such as convolutional codes or block codes. Note that

[2] In the sense of poor transmission rate or poor transmission quality

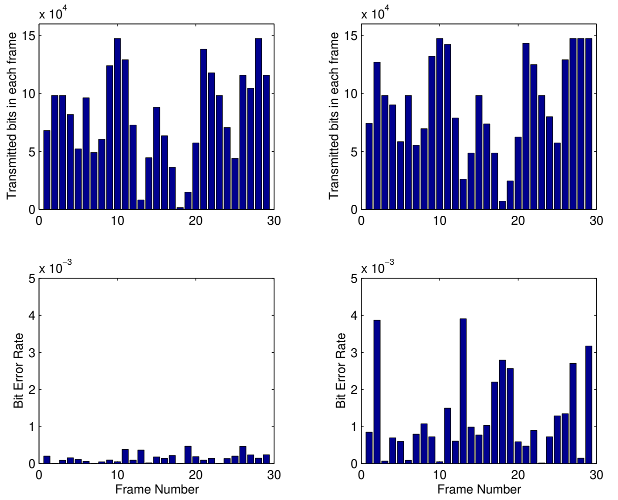

Figure 6: Transmission performance for two transmissions using adaptive modulation with different error-probability thresholds for the choice of modulation alphabet. P_M is set to 10^{-3} at the leftmost simulation, and 10^{-2} at the right.

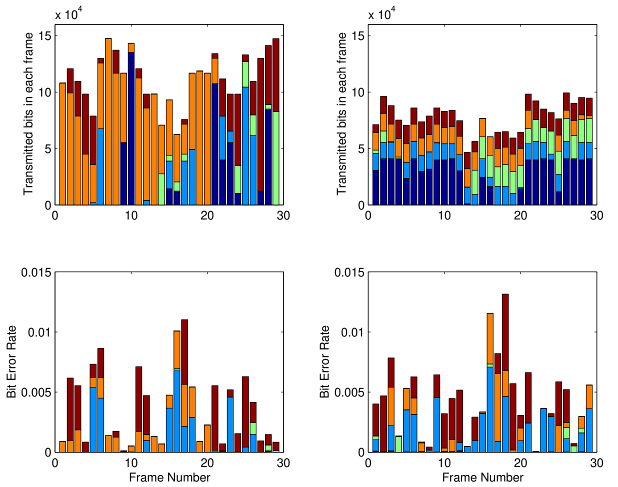

Figure 7: Transmission performance for five users with adaptive modulation and competitive scheduling. The different users are depicted in different nuances of grey. On the left we see the scheduling result after *Pass1* (see Section 4.1), and on the right, the resulting throughput and BER after *Pass2* (see Section 4.2) with the same data.

FEC has not been included in this work, so the achievable BER is considerably lower. For the non-adaptive case, clearly, the error rate peaks when the receiver enters a fading dip (Figure 5). This behavior can probably not be compensated for by FECs, unless long interleaving is used to average out the errors over time. The remaining errors after the decoding process will result in retransmissions, invoked by higher layers in the communication system. Obviously, such transmissions will increase the traffic over the channel.

By introducing adaptive modulation we gain:

1. The error rate is kept at a constant level, thus feeding the FEC algorithms with manageable data.
2. Radio transmission is postponed when channel conditions are bad, thus reducing the interference affecting other terminals.

By adding the multiple access scheduler, keeping multiple links on a single frequency, we gain two more things:

1. System throughput can be maximized for a given frequency band (Figure 7, left). The more users we add, the more efficiently we use the frequency band.
2. User throughput becomes the central issue, rather than the allocation of some fixed number of timeslots in a varying environment (Figure 7, right).

The following topics will be investigated in the future:

- Scheduling involving more than one frequency band at a time. High-priority users choose between several independent channels on different frequencies.
- More realistic traffic (IP) and channel (Multipath) models will be incorporated in order to find better scheduling criteria.
- The performance gains as a function of prediction error levels and prediction horizon will be quantified.
- A deeper analysis of the required signalling overhead, the predictor initialization procedure, and the required hardware, will be carried out and included in the evaluations of the proposed systems.

7 ACKNOWLEDGEMENTS

This work was supported by the Foundation for Strategic Research, within the PCC-program, in Sweden.

The simulations were carried out using the IT++ package developed at the Communications Group, Department of Signals and Systems, Chalmers University of Technology, Sweden.

REFERENCES

[1] Toyoki Ue, Seiichi Sampei, and Norihiko Morinaga, “Adaptive Modulation Packet Radio Communication System using NP-CSMA/TDD Scheme”, *IEEE Vehicular Technology Conference Proceedings*, pp. 416–420, May 1996.

[2] Soon-Ghee Chua and Andrea Goldsmith, “Variable-Rate Variable-Power MQAM for Fading Channels”, *IEEE Vehicular Technology Conference Proceedings*, pp. 815–819, May 1996.

[3] W.T. Webb and R. Steele, “Variable Rate QAM for Mobile Radio”, *IEEE Transactions on Communications*, Vol. 43, No. 7, July 1995.

[4] Torbjörn Ekman and Gernot Kubin, “Nonlinear prediction of mobile radio channels: Measurements and MARS model designs”, *IEEE International Conference on Acoustics Speech and Signal Processing*, Phoenix Arizona, March 1999.

[5] J.M. Torrance and L. Hanzo “Adaptive Modulation in a Slow Rayleigh Fading Channel”, *PIMRC’96*, Oct. 1996, Taipei, Taiwan.

[6] John G. Proakis, *Digital Communications*, 3rd ed. McGraw-Hill Book Co., Singapore 1995

[7] Frederick S. Hillier and Gerald J. Lieberman *Introduction to Operations Research*, 5th ed. McGraw-Hill Book Co., Singapore 1990