Link Adaptation with Retransmissions for Partial Channel State Information

Stephan Pfletschinger and Monica Navarro
Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)
Parc Mediterrani de la Tecnologia
Av. del Canal Olímpic s/n
08860 Castelldefels (Barcelona), Spain

Abstract—We consider the performance of adaptive coding and modulation (ACM) over a fading channel for the case that the transmitter has only partial channel state information (CSI). We evaluate the degradation due to the partial CSI in terms of additionally required SNR by simulations and an analytical approximation. Then, we propose an ACM scheme with fast link-level retransmissions based on cyclic incremental redundancy and compare its performance to an equivalent scheme based on Chase combining. For both HARQ (hybrid automatic repeat request) schemes, we develop an analytical approximation for the throughput and the delay.

I. INTRODUCTION

One of the principal objectives of broadband wireless systems is to provide users with a fast, flexible and reliable access under varying channel and traffic conditions. This requires a highly flexible air interface which adapts its transmission parameters to the current radio environment. Therefore, fast link adaptation is now an integral part of upcoming system standards and proposals like WiMAX, 3GPP LTE (Long Term Evolution) or WINNER [1], [2]. A good overview about the key aspects of systems beyond 3G, focusing on the physical layer, is provided in [3].

Fast link adaptation requires accurate and timely knowledge of the channel state in order to select the appropriate modulation and coding scheme (MCS). However, since this information has to be provided over a feedback link after channel estimation at the receiver, there is an intrinsic delay. Additionally, since many data traffic types are bursty in nature, channel estimation has to rely on a relatively low number of samples and is therefore limited in accuracy. This situation is aggravated when bursty interference is present, which is the case for many packet-oriented systems with a low frequency reuse factor. These effects can be reduced to some extent by channel prediction [4], [5], but there remains some degree of uncertainty about the actual channel state.

In this paper, we first study the impact of incomplete channel knowledge on adaptive coding and modulation (ACM) and evaluate the required SNR margin to compensate the lack of exact CSI. Since the degradation due to incomplete CSI is quite severe, we evaluate the performance of two retransmission strategies based on packet repetition and on a generalized incremental redundancy scheme. For the former strategy, we develop an analytical approximation of the throughput and delay performance, which is confirmed by simulations.

II. SYSTEM MODEL

Throughout this paper, we consider some variants of the flat Rayleigh fading channel, in which the received signal is given by

\[ y = h \cdot x + w, \quad h \sim \mathcal{C}\mathcal{N}(0, 1), \quad w \sim \mathcal{C}\mathcal{N}(0, N_0) \]  

where \( x \) is the encoded and modulated transmit signal in equivalent baseband representation. We will consider the case where the fading coefficient \( h \) is constant during one codeword (block fading) and the case where \( h \) is drawn independently for each transmitted QAM symbol (fast i.i.d. fading). The latter case is often referred to as the perfectly interleaved channel since there is no correlation between subsequent fading coefficients.

The average SNR is given by \( \bar{\gamma} \triangleq \mathbb{E}[|h|^2 E_S/N_0] = E_S/N_0 \), and is assumed to be constant while the instantaneous SNR \( \gamma \triangleq |h|^2 \bar{\gamma} \) depends on the channel realization.

A. Transmitter Structure

The basic structure of the transmitter and the receiver, depicted in Fig. 1, is based on bit-interleaved coded modulation (BICM) [6] as it is found in nearly all current and upcoming wireless communication systems. The data bits \( u = (u_0, u_1, \ldots, u_{K-1}) \) are first encoded by a channel encoder with mother code rate \( R_m \). The code rate is adapted consequently by rate-compatible puncturing to the code rate \( R_2 \geq R_m \). The rate-compatible puncturing is implemented with an interleaver \( \pi_w \) which permutes the coded bits to a vector \( c = (c_0, c_1, \ldots, c_{K/R_m-1}) \) and a subsequent bit selection, which simply takes the first \( N \) bits of each permuted codeword: \( d = (c_0, c_1, \ldots, c_{N-1}) \). For the number of coded bits to be transmitted, it must hold \( K \leq N \leq K/R_m \). Hence, any code rate between \( R_m \) and 1 can be realized.

The coded bits \( (c_0, c_1, \ldots, c_{N-1}) \) are then interleaved and mapped to QAM symbols, which are transmitted over the fading channel. At the receiver side, the corresponding inverse operations are performed: the soft demapper derives an APP and evaluate the required SNR margin to compensate the lack of exact CSI. Since the degradation due to incomplete CSI is quite severe, we evaluate the performance of two retransmission strategies based on packet repetition and on a generalized incremental redundancy scheme. For the former strategy, we develop an analytical approximation of the throughput and delay performance, which is confirmed by simulations.
results can be obtained by applying another strong channel code like e.g. an LDPC code. An asymptotic analysis based on random coding and a multi-user collision channel can be found in [9], while a theoretical analysis focussing on LDPC codes is given in [10].

B. Modulation and Coding Schemes

For the link adaptation, a set of eight MCS has been selected based on square QAM and duo-binary turbo codes, which yields state-of-the-art performance for the message length of $K = 288$ (see also [8], [11]). While with the puncturing scheme presented above, any code rate of the form $i = 0, 1, \ldots, (1/R_{m} - 1)K$ can be implemented, the number of selected MCS has to be kept small due to the associated signalling overhead. Table I lists the number of bits per QAM symbol $R_{1}$ and the code rate $R_{2}$ for each MCS, which are identified by the indices $c = 1, 2, \ldots, 8$.

The throughput of MCS $c$ is defined as the average number of information bits per channel use in correctly received codewords:

$$\eta_{c} = \frac{R_{2}}{R_{1}} \left( 1 - p_{w,\gamma_{c}} \right), \quad c = 1, \ldots, 8 \quad (2)$$

where $p_{w,\gamma}$ denotes the word error ratio (WER) of MCS $c$ at SNR $\gamma$. The throughput curves are depicted in Fig. 2 for the AWGN and the fast, perfectly interleaved, Rayleigh fading channel.

Analytical approximations of simulation results are often helpful for further studies. To this end, we approximate the WER for the AWGN channel by an exponential function:

$$\tilde{p}_{w,\gamma} = \begin{cases} 1 & \gamma \leq \gamma_{c} \alpha_{c} \left( \gamma_{c} - \gamma \right), \\ \exp \left( \alpha_{c} \left( \gamma_{c} - \gamma \right) \right) & \gamma > \gamma_{c} \end{cases} \quad (3)$$

which shows a very good agreement with the simulation results as can be observed in Fig. 3. The parameters $\alpha_{c}$ and $\gamma_{c}$ are derived for each MCS such that the relative quadratic error between the simulated WER $p_{w,\gamma}$ and the approximated WER $\tilde{p}_{w,\gamma}$

$$\sum_{i=1}^{n} \left( \frac{p_{w,\gamma_{i}} - \tilde{p}_{w,\gamma_{i}}}{p_{w,\gamma_{i}}} \right)^{2}$$

is minimized. The parameters $\gamma_{c}$ and $\alpha_{c}$ for the Rayleigh fading channel are obtained in the same way.

This approximation is motivated by two simple observations: while for low SNR, the word error probability is close to unity, for high SNR it is dominated by the term $A_{d} \exp(-d_{\min})$, where $A_{d}$ denotes the distance spectrum of a binary linear block code (note that a duo-binary turbo code is a binary linear block code) [12]. Although we do not know these parameters for the employed turbo code, we can use the curve-fitting approach (4) to align the approximation to the simulation results.

<table>
<thead>
<tr>
<th>$c$</th>
<th>$R_{1}$</th>
<th>$R_{2}$</th>
<th>$\gamma_{c}$</th>
<th>$\alpha_{c}$</th>
<th>$\gamma_{c}$</th>
<th>$\alpha_{c}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/3</td>
<td>1/3</td>
<td>-1.48</td>
<td>42.309</td>
<td>-5.44</td>
<td>25.38</td>
</tr>
<tr>
<td>2</td>
<td>1/3</td>
<td>1/3</td>
<td>-1.47</td>
<td>21.151</td>
<td>-0.43</td>
<td>12.69</td>
</tr>
<tr>
<td>3</td>
<td>1/2</td>
<td>1/2</td>
<td>0.75</td>
<td>13.009</td>
<td>2.55</td>
<td>5.54</td>
</tr>
<tr>
<td>4</td>
<td>1/3</td>
<td>1/3</td>
<td>3.20</td>
<td>5.806</td>
<td>4.70</td>
<td>3.34</td>
</tr>
<tr>
<td>5</td>
<td>1/2</td>
<td>1/2</td>
<td>5.96</td>
<td>3.226</td>
<td>7.95</td>
<td>1.33</td>
</tr>
<tr>
<td>6</td>
<td>2/3</td>
<td>2/3</td>
<td>8.50</td>
<td>1.650</td>
<td>11.03</td>
<td>0.43</td>
</tr>
<tr>
<td>7</td>
<td>2/3</td>
<td>2/3</td>
<td>13.35</td>
<td>0.441</td>
<td>15.83</td>
<td>0.125</td>
</tr>
<tr>
<td>8</td>
<td>6/7</td>
<td>6/7</td>
<td>16.85</td>
<td>0.144</td>
<td>18.41</td>
<td>0.014</td>
</tr>
</tbody>
</table>

Fig. 1. Structure of the transmitter and the receiver.

Fig. 2. Throughput of the eight MCS over an AWGN and an i.i.d. fast Rayleigh fading channel.

Fig. 3. Simulated WER (continuous lines) and analytical approximation (crosses).
III. LINK ADAPTATION WITH PARTIAL CSI

The transmitter obtains the knowledge about the SNR (which is equivalent to the channel power gain \(|h|^2\)) through a feedback channel. This incurs an inevitable delay and thus an error, which can be partially alleviated by channel prediction. We can account for the remaining error by the estimated channel coefficient

\[
\hat{h} = (1 - \beta) \cdot h + \sqrt{\beta(1 - \beta)} \cdot v, \quad v \sim \mathcal{CN}(0, 1)
\]

where \(\beta\) is the normalized mean square prediction error (NMSE) and \(v\) is an independent random variable (for details, please refer to [4]). The NMSE depends basically on the average SNR and the channel variability, mainly caused by user mobility. The predicted SNR is accordingly given by \(\hat{\gamma} = |\hat{h}|^2\).

According to this model, the instantaneous channel coefficient \(h\), conditioned on the estimation \(\hat{h}\) is \(\mathcal{CN}(h, \beta)\) distributed, which yields the conditional pdf of the instantaneous SNR

\[
p_{\gamma|\hat{\gamma}}(x|\hat{\gamma}) = \frac{1}{\hat{\gamma}\beta} \exp\left(-\frac{x + \frac{x}{\hat{\gamma}\beta}}{\frac{\beta}{\hat{\gamma}}}ight) I_0\left(\frac{2\sqrt{\gamma_x}}{\hat{\gamma}\beta}\right), \quad x \geq 0
\]

where \(I_0(\cdot)\) is the modified Bessel function of the first kind. It is interesting to see that the average SNR and the NMSE only appear as their product in this expression.

Since the transmitter does not have knowledge about the instantaneous SNR, it must base its selection of the MCS on the estimated SNR. We assume that it also has knowledge about the average SNR and the prediction error. The average word error probability can then be computed as

\[
P_{w}^{(c)}(\hat{\gamma}, \hat{\gamma}) = \int_{0}^{\infty} \hat{p}_w(x)p_{\gamma|\hat{\gamma}}(x) \, dx
\]

\[
= \frac{1}{\hat{\gamma}\beta} \int_{0}^{\infty} \hat{p}_w(x) e^{-x}\left(\frac{x^2}{\hat{\gamma}\beta^2}\right) I_0\left(\frac{2\sqrt{\gamma_x}}{\hat{\gamma}\beta}\right) \, dx
\]

\[+ e^{\alpha x} - \frac{x^2}{\hat{\gamma}\beta^2} \int_{0}^{\infty} e^{-x}\left(\frac{x^2}{\hat{\gamma}\beta^2}\right) I_0\left(\frac{2\sqrt{\gamma_x}}{\hat{\gamma}\beta}\right) \, dx
\]

Unfortunately, there is no closed solution for this integral, so we must resort to numerical integration. As can be seen in Fig. 4, the analytical approximation by (7) matches well with simulation results. We can also observe that even a relatively small prediction error has a considerable impact on the WER.

Fig. 5 depicts the additional SNR to maintain the same target WER of \(P_w = 10^{-2}\) as a function of the prediction error \(\beta\) for the MCS \(c = 4\). In mobile communication systems with high user mobility and high carrier frequencies like 3G systems and beyond, the NMSE for many users might be around 0.1 or even higher, which severely degrades the system performance. Since these uncertainties about the actual channel state are unavoidable, few possibilities remain to increase the reliability of the transmission without sacrificing rate or spending much more transmit power.

A way out of this problem consists in the introduction of a mechanism for fast link-level retransmissions, as will be outlined in the following section.

IV. LINK ADAPTATION WITH HARQ

In a wireless communication system, it is not possible to foresee sudden channel variations and therefore it may happen frequently that a transmitted codeword cannot be decoded correctly at the receiver. In this context, fast link-level retransmissions can complement the ACM strategy. We consider HARQ from a link level perspective and as a complement to forward error correction. Therefore, the receiver applies soft combining of the retransmitted information.

For the following, we assume that only the long-term average SNR \(\hat{\gamma}\) is known at the receiver. Since the channel coefficient is Rayleigh distributed, the instantaneous SNR is exponential distributed with pdf

\[
p_\gamma(x) = \frac{1}{\hat{\gamma}} \exp\left(\frac{-x}{\hat{\gamma}}\right), \quad \text{for } x \geq 0
\]

A. Repetition Coding and Cyclic Incremental Redundancy

We assume that the MCS is selected by the ACM strategy and is not changed for the retransmissions. Depending on the selected MCS, the number of coded bits for the initial
transmission and all retransmissions is given by $N_c = K/R_2^{(c)}$. The message length $K$ is assumed fixed according to the frame structure of higher layers. The simplest retransmission method consists in the simple repetition of these $N_c$ bits, namely the bits $(c_0, c_1, \ldots, c_{N_c-1})$ in Fig. 1. At the receiver side, the L-values corresponding to the same coded bits are added and passed to the decoder (this can be implemented before or after de-interleaving). This simple strategy is also known as Chase combining or, from a coding perspective, as repetition coding.

Another method for selecting the retransmitted bits is cyclic incremental redundancy (CIR), which has been proposed for the WINNER system concept [8]. This strategy generalizes the classical incremental redundancy scheme which only retransmits the remaining parity bits by retransmitting the coded bits in a cyclic fashion: for the initial transmission, the bits $c_0, c_1, \ldots, c_{N_c-1}$ are transmitted. The first retransmission continues with the next $N_c$ coded bits. If less than $N_c$ parity bits are left, then the missing bits are taken from the start of the codeword. The index range for the $n$-th retransmission ($n = 0$ indicates the initial transmission) can be expressed concisely by a modulo operation:

$$\text{mod}(nN_c, N_m), \ldots, \text{mod}((n+1)N_c - 1, N_m)$$

where $N_m = K/R_m$; hence for the chosen mother code rate we have $N_m = 3K$). Note that for cyclic incremental redundancy, it is irrelevant whether the channel encoder is systematic or not since no distinction is made between systematic and parity bits. Another feature of this scheme is that there is no restriction on the size of the retransmission unit (RTU). Due to the cyclic repetition of the coded bits, the retransmission unit can be made as large as desired, it is also possible to vary the RTU size with the retransmission number according to some predefined strategy.

Thus, depending on the code rate of the selected MCS, at the first retransmission the ratio of repeated and remaining coded bits varies. E.g. for $R_2^{(c)} = 1/2$, at the first retransmission, only bits which have not been sent initially are selected, while at the second retransmission the same bits as in the initial transmission are sent. If the initial transmission uses the mother code rate $R_m$, then CIR becomes the same as repetition coding.

As a consequence, CIR can be seen as a seamless combination of classical incremental redundancy, which retransmits only the remaining coded bits until none is left, and simple packet repetition. The advantage of CIR is that while it exploits all the benefits of IR, it is as simple to implement as Chase combining and it is not restricted to a fixed size of the retransmission unit.

In the following, we consider two fading models: fast fading and block fading, which both have the same fading statistics but different coherent times.

**B. Fast Fading**

We first consider the case of fast fading, where for each transmitted QAM symbol, a new channel coefficient $h$ is drawn. Hence, the instantaneous SNR values are i.i.d. exponentially distributed, and one transmission involves $Q = K/R_1^{(c)}R_2^{(c)}$ QAM symbols and the same number of channel coefficients. The average SNR after $n$ retransmissions is thus

$$\xi = \frac{1}{(n+1)Q} \sum_{i=1}^{(n+1)Q} \gamma_i$$

which is $\Gamma\left((n+1)Q, \frac{\bar{\gamma}}{n+1}\right)$ distributed with pdf

$$p_\xi(x) = \frac{x^{\nu-1} \exp\left(-\frac{x\bar{\gamma}}{\nu}\right)}{\left(\frac{x}{\nu}\right)^\nu \cdot \Gamma(\nu)}$$

(10)

For repetition coding, the SNR seen by the decoder is the sum of the SNR values of each transmission and thus the WER curve according to (3) is shifted after each transmission. With these assumptions, we can calculate the probability for a word error at the $n$-th retransmission as

$$P^{(c)}(n) = \int_0^\infty P_\xi(x) p_W^{(c)}((n+1)x) \, dx = P\left(\nu, \frac{\bar{\gamma} n^{\nu-1}}{\gamma}\right) + \exp\left(\frac{\alpha_1^{(c)} \gamma_1^{(c)}}{1 + \frac{\alpha_2^{(c)} \gamma_2^{(c)}}{\nu, \gamma}}\right) \cdot \left[1 - P\left(\nu, \frac{\gamma_1^{(c)} \bar{\gamma}}{\gamma} + \alpha_1^{(c)} \gamma_1^{(c)}\right)\right]$$

(11)

where $P(a, x) \triangleq \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} \, dt$ is the incomplete lower Gamma function.

The probability that $n$ retransmissions are required for successful reception of a codeword is then

$$P_t(n) = \left(1 - P^{(c)}(n)^{n-1} \sum_{i=0}^{n-1} P^{(c)}(i)\right)$$

(12)

Finally, the average delay (i.e. the average number of retransmissions) and the throughput are given by

$$\tau^{(c)}(\bar{\gamma}) = \sum_{n=0}^\infty n \cdot P_t(n)$$

(13)

$$\eta^{(c)}(\bar{\gamma}) = \frac{R^{(c)}_1 R^{(c)}_2}{1 + \tau^{(c)}(\bar{\gamma})}$$

(14)

The throughput and the average delay have been evaluated by Monte-Carlo simulations for the eight MCS detailed in Table I. A duo-binary turbo code as outlined in Section II has been applied and it has been assumed that the receiver detects all codeword errors. This is a good approximation for a system where the probability for a detection error is much lower than the desired residual error probability, which is the case for the usually employed cyclic redundancy check (CRC) codes. In the simulation, up to 100 retransmissions were performed until the codeword was detected without error.

The outcome of the simulations for the fast i.i.d. fading channel and both retransmission methods is depicted on the left handside of Fig. 6. The analytical approximation for repetition coding based on (11) and (12) is also depicted.

From the diagrams, we can observe that the analytical approximation for repetition coding is reasonably accurate for
the first and second transmission, but for more retransmissions, or lower SNR, it loses accuracy. This behavior can be attributed to two effects:

- After some retransmissions, the fading statistics of the channel seen by the receiver is not Rayleigh anymore. Instead of being exponentially distributed, the received SNR is Gamma distributed and hence less variable. This explains why the simulations results show better performance than the analytical approximation when retransmissions occur.

- While the approximation (3) is accurate for high and low error rates, it is less exact for moderate WER, i.e. \( 0.01 \lesssim P_w < 1 \). However, this range plays an important role for the approximation of the retransmission probabilities.

We can also observe from the simulation results that CIR outperforms repetition coding significantly, especially for MCS \( c = 8 \). For \( c \in \{1, 2, 4\} \) the code rate is the mother code rate and hence CIR degenerates to repetition coding and the two methods are identical. For MCS with higher code rates, the performance difference is considerable. The reason for this is the coding gain of the punctured code with respect to simple repetition coding. Especially for higher code rates, CIR exploits the full potential of the coding scheme.

C. Block Fading

Another fading model is block fading: the channel coefficient \( h \) is constant during one transmission and is drawn independently for the next transmission (or retransmission). This model corresponds to sudden changes in the channel and presents a greater challenge for the link adaptation scheme since the FEC cannot average over the fading states. Therefore, retransmissions play an even more important role than in the fast fading scenario.
For this fading model, the SNR after \( n \) retransmissions is 
\[
\xi = \sum_{i=0}^{n} \gamma_i, \text{ which is } \Gamma(n+1, \bar{\gamma}) \text{ distributed.}
\]
The probability of word error at the \( n \)-th retransmission is hence
\[
P_w^{(c)}(n) = E[p_w^{(c)}(\xi)] = \int_0^\infty p_\xi(x)\tilde{p}_w^{(c)}(a_\xi) \, dx
\]
\[
= P \left( n + 1, \frac{\gamma_a}{\gamma} \right) + \frac{\exp(\alpha_a(\gamma_a)\gamma)}{(1 + \alpha_a(\gamma_a))^{n+1}} \left[ 1 - P \left( n + 1, (\alpha_a(\gamma_a) + \frac{1}{\gamma})\gamma_a \right) \right].
\]
(15)

From this expression, the delay and the throughput can be derived as in the previous case by (12)–(14). In contrast to the fast fading case, which led to (11), for block fading, the channel seen by the decoder is constant and therefore we have to use the approximation for the AWGN channel, given by the parameters \( \alpha_a, \gamma_a \).

The simulation results for throughput and delay in Fig. 6 show that despite of the high uncertainty of the channel state, a surprisingly high throughput can be achieved. Again, CIR performs significantly better than repetition coding if the code rate of the selected MCS is much higher than the mother code rate. Similar results have been found earlier by Frenger [13] and Cheng [14], although the outcomes have to be compared with care since the employed incremental redundancy strategies are not identical.

D. Interpretation of the Simulation Results

From the obtained simulation results for both types of fading channels and the two HARQ strategies, we can make various interesting observations:

- The cyclic incremental redundancy scheme clearly outperforms repetition coding by taking advantage of the coding gain of the FEC scheme. This performance gain comes at hardly any additional implementation complexity.
- Incremental redundancy has no benefit if the initial code rate equals the mother code rate. Thus, additional gains can be obtained if the mother code rate is designed to be lower than the code rate of any MCS.
- The results allow to fix SNR thresholds for the MCS selection, taking both throughput and delay into account.
  - For instance, one can maximize the throughput while constraining the average delay. It is interesting to see that even when maximizing the throughput without any delay constraint, on the average very few retransmissions will be required.
- The throughput and the delay for repetition coding can be approximated analytically. This facilitates optimization for any criterion involving throughput and delay.
- The achieved throughput with retransmissions comes relatively close to the throughput of an ergodic fading channel, at the expense of few retransmissions. This is an encouraging result, considering the fact that the HARQ scheme relies only on the long-term average SNR \( \bar{\gamma} \) and no predicted SNR value has been used.

V. Conclusions

We have evaluated the impact of partial channel knowledge on adaptive coding and modulation schemes and evaluated the required additional SNR in order to maintain the same error rate. Since this SNR margin is considerable even for moderate channel prediction errors, we propose the incorporation of a fast link-level retransmission scheme in conjunction with link adaptation. To this end, we evaluated the performance of a cyclic incremental redundancy scheme, which exploits the full coding gain of the FEC scheme and offers full flexibility with respect to the size of the RTU sizes. For two fading models, it has been found that cyclic incremental redundancy performs significantly better than Chase combining while the complexity is comparable. For the latter scheme, we derived an analytical approximation for the throughput and the delay.

ACKNOWLEDGMENT

This work has been performed in the framework of the IST project IST-4-027756 WINNER II, which is partly funded by the European Union. The authors would like to acknowledge the contributions of their colleagues. This work has also been partially funded by the Catalan Government under grant SGR2005-00690 and it was supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM++ (contract n. 216715).

REFERENCES