COMBINING ADAPTIVE MODULATION AND CODING WITH TRUNCATED ARQ ENHANCES THROUGHPUT

Qingwen Liu, Shengli Zhou, and Georgios B. Giannakis

Dept. of ECE, Univ. of Minnesota, 200 Union Street SE, Minneapolis, MN 55455

ABSTRACT

We develop a cross-layer design which combines adaptive modulation and coding (AMC) at the physical layer with a truncated auto repeat request (ARQ) protocol at the data link layer, in order to maximize spectral efficiency under prescribed delay and performance constraints. We derive the achieved spectral efficiency in closed-form for transmissions over Nakagami-m block fading channels. Numerical results reveal that retransmissions at the data link layer alleviate stringent error-control requirements at the physical layer, and thereby enable considerable spectral efficiency gains. Diminishing returns appear on the spectral efficiency improvement as the number of retransmissions increases, which suggests that a small number of retransmissions strikes a desirable delay-throughput tradeoff.

1. INTRODUCTION

To enhance throughput in future wireless communication systems, adaptive modulation and coding (AMC) have been advocated at the physical layer, in order to match transmission rates to time-varying channel conditions based on channel state information (CSI) at the transmitter [1]. However, to achieve high reliability at the physical layer, one has to lower the transmission rate using either small size constellations, or, powerful but low-rate error-control codes.

An alternative way to mitigate channel fading is to rely on the ARQ protocol at the data link layer, that requests retransmissions only for those packets received in error. ARQ is quite effective in improving throughput relative to using only forward error coding (FEC) at the physical layer [5]. To minimize delays and buffer sizes, truncated ARQ protocols have been adopted to limit the maximum number of retransmissions [5]. However, only fixed modulation and coding have been considered in systems with truncated ARQ protocols [5].

Instead of considering AMC at the physical layer and ARQ at the data link layer separately, we pursue here a *cross-layer design*, that judiciously combines these two layers to maximize the system spectral efficiency, or throughput, under prescribed delay and performance constraints. With ARQ correcting occasional packet errors at the data link layer, the stringent error control requirement is alleviated for the AMC at the physical layer. Depending on the error-correcting capability of the truncated ARQ, that depends on the maximum allowable number of retransmissions, we design AMC transmissions that guarantee the required performance. We analyze the performance of this cross-layer design, and obtain the achieved average spectral efficiency in closed-form. Numerical results demonstrate that our joint AMC-ARQ design outperforms either usage of AMC alone at the physical layer, or, incorporation of ARQ with a fixed modulation and coding.

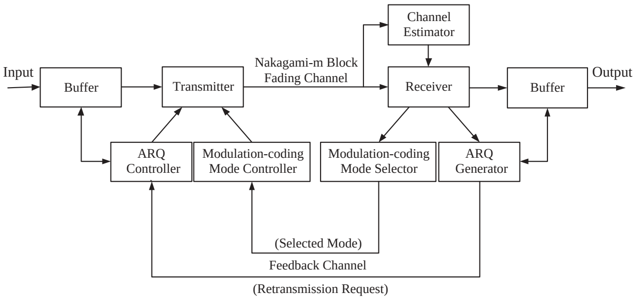

Fig. 1. The system and channel models

2. MODELING

Consider the single-transmit single-receive antennae system in Fig. 1. It consists of a joint adaptive modulation and coding module at the physical layer, and an ARQ module at the data link layer (more details available in [4]). The processing unit at the data link layer is a packet, which comprises multiple information bits. On the other hand, the processing unit at the physical layer is a frame, which is a collection of multiple transmitted symbols. The detailed packet and frame structures used in this paper will be clarified soon.

At the physical layer, we assume that multiple transmission modes are available, which comprising a specific modulation and FEC code pair (mode), as in HIPERLAN/2 and IEEE 802.11a standards [3]. Based on CSI acquired at the receiver, the AMC selector determines the mode, which is sent back to the transmitter through a feedback channel. The AMC controller then updates the transmission mode at the transmitter. Coherent demodulation and maximum-

---

This work was supported by the NSF grant no. 0105612, and by the ARL/CTA grant no. DAAD19-01-01-2-011.

Table 1. Transmission modes in TM with convolutionally coded modulation

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6
Modulation	BPSK	QPSK	QPSK	16-QAM	16-QAM	64-QAM
Coding rate R_c	1/2	1/2	3/4	9/16	3/4	3/4
Rate (bits/sym.)	0.50	1.00	1.50	2.25	3.00	4.50
a_n	274.7229	90.2514	67.6181	50.1222	53.3987	35.3508
g_n	7.9932	3.4998	1.6883	0.6644	0.3756	0.0900
γ_{pn}(dB)	-1.5331	1.0942	3.9722	7.7021	10.2488	15.9784

(The generator polynomial of the mother code is $g = [133, 171]$. Different coding rates are obtained with puncturing as in HIPERLAN/2.)

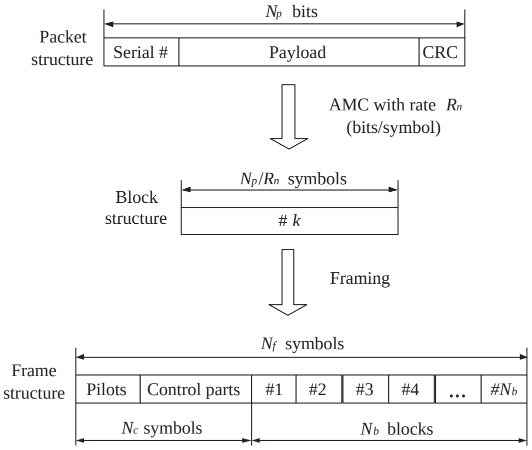

Fig. 2. The packet and frame structures

likelihood (ML) decoding are used at the receiver. The decoded bit streams are mapped to packets, which are pushed upwards to the data link layer.

At the data link layer, the selective repeat ARQ protocol is implemented. If an error is detected in a packet, a retransmission request is generated by the ARQ generator, and is communicated to the ARQ controller at the transmitter via a feedback channel; otherwise, no retransmission request is sent. The ARQ controller then arranges retransmission of the requested packet that is stored in the buffer.

We next detail the parameters of both physical and data link layers. At the physical layer, we consider here the following group of transmission modes:

TM: Convolutionally coded BPSK, QPSK, or QAM modulations, adopted from the HIPERLAN/2, or, IEEE 802.11a standards [3]. These transmission modes are listed in Table 1, in a rate ascending order.

Although we will focus on TM in this paper, other transmission modes can be similarly constructed [4].

At the physical layer, we deal with frame by frame transmissions, where each frame contains a fixed (N_f) number of symbols. Each frame at the physical layer may contain multiple packets from the data link layer. The packet and frame structures are depicted in Fig. 2. Each packet contains N_p bits, which include serial number, payload, and cyclic redundancy check (CRC) bits to facilitate error detection. After modulation and coding with mode n of rate R_n (bits/symbol), each packet is mapped to a symbol-block containing N_p/R_n symbols. Multiple such blocks, together with N_c pilot symbols and control parts, constitute one frame to be transmitted at the physical layer, as in HIPERLAN/2 and IEEE 802.11a standards [3]. If mode n is used, it follows that the number of symbols per frame is $N_f = N_c + N_b N_p / R_n$, which implies that N_b (the number of packets per frame) depends on the chosen mode.

We next list our operating assumptions:

A1: The channel is frequency flat, and remains invariant per frame, but is allowed to vary from frame to frame, as the block fading channel model [2]. Thus, AMC is adjusted on a frame-by-frame basis.

A2: Perfect CSI is available at the receiver and the corresponding mode selection is fed back to the transmitter without error and latency, as in [1].

A3: Error detection based on CRC is perfect, and the serial number and the CRC parity bits in each packet are not included in the throughput calculation, as in [6].

A4: The fading channel coefficients corresponding to the original and the retransmitted packets are independent and identically distributed (i.i.d.) random variables.

The round trip delay in ARQ is the time elapsed from sending the retransmission request until receiving the retransmitted packet. It mainly depends on the coding and decoding delays, as well as the queuing delays at both the transmitter and the receiver. In general, round trip delays are larger than the channel coherence time. Hence, the channels experienced by the original packet and the possible subsequent retransmissions can be assumed i.i.d., which justifies well A4.

For flat fading channels adhering to A1, the channel quality can be captured by the received signal to noise ratio (SNR) γ. Because it models a large class of fading channels, we adopt the general Nakagami-m model to describe γ statistically with a probability density function [1]:

$$p_\gamma(\gamma) = \frac{m^m \gamma^{m-1}}{\bar{\gamma}^m \Gamma(m)} \exp\left(-\frac{m\gamma}{\bar{\gamma}}\right), \tag{1}$$

where $\bar{\gamma} := E\{\gamma\}$ is the average receive SNR, $\Gamma(m) :=$

$\int_0^\infty t^{m-1}e^{-t}dt$ is the Gamma function, and m is the Nakagami fading parameter ($m \geq 1/2$).

3. COMBINING AMC WITH TRUNCATED ARQ

Here, we develop our cross-layer design, which combines AMC at the physical layer with truncated ARQ at the data link layer. Because only finite delays and buffer sizes can be afforded in practice, the maximum number of ARQ retransmissions is bounded. Thus, we adopt the following delay constraint:

C1: The maximum number of retransmissions allowed per packet is $N_r^{\max}$.

Since only finite retransmissions are allowed, error-free delivery can not be guaranteed [5]. Therefore, if a packet is not received correctly after $N_r^{\max}$ retransmissions, we drop it, and declare packet loss. This is very reasonable and can be afforded in video/image transmissions because the underlying bit streams are highly correlated. To maintain an acceptable packet stream, we impose the following performance constraint:

C2: The probability of packet loss after $N_r^{\max}$ retransmissions is no larger than $\mathrm{P}_{\mathrm{loss}}$.

The delay constraint C1 dictates that truncated ARQ with up to $N_r^{\max}$ retransmissions should be performed at the data link layer. The special case with $N_r^{\max} = 0$ corresponds to no retransmission, and we term it AMC-only. Having specified ARQ at the data link layer, we next design the AMC at the physical layer.

3.1. Performance Requirement at the Physical Layer

We first determine how reliable performance is needed at the physical layer to meet C2, given that $N_r^{\max}$-truncated ARQ is implemented at the data link layer.

Suppose that AMC achieves an average packet error rate (PER) P_0, and a packet is dropped if it is not received correctly after $(N_r^{\max}+1)$ transmissions. Per A4, the packet loss probability at the data link layer can be computed as $\mathrm{P}_0^{N_r^{\max}+1}$. To satisfy C2, we need to impose: $\mathrm{P}_0^{N_r^{\max}+1} \leq \mathrm{P}_{\mathrm{loss}}$. Thus, we obtain:

$$\mathrm{P}_0 \leq \mathrm{P}_{\mathrm{loss}}^{\frac{1}{N_r^{\max}+1}} := \mathrm{P}_{\mathrm{target}}. \tag{2}$$

Therefore, if we use AMC to realize P_0 bounded as in (2), and implement a $N_r^{\max}$-truncated ARQ, then both C1 and C2 will be satisfied. Our remaining problem is to design AMC to maximize spectral efficiency while ensuring that P_0 satisfies (2).

3.2. AMC Design at the Physical Layer

Our objective here is to maximize the data rate, while maintaining the required performance at the physical layer. As in [1], we assume constant power transmission, and partition the total SNR range into $N+1$ non-overlapping consecutive intervals, with boundary points denoted as $\{\gamma_n\}_{n=0}^{N+1}$, where $N = 6$ is the total number of transmission modes for TM. Specifically,

$$\text{mode } n \text{ is chosen,} \quad \text{when } \gamma \in [\gamma_n, \gamma_{n+1}). \tag{3}$$

To avoid deep channel fades, no payload bits will be sent when $\gamma_0 \leq \gamma < \gamma_1$. What remains now is to determine the boundary points $\{\gamma_n\}_{n=0}^{N+1}$ to meet the required PER.

To simplify the AMC design, we will rely on the following approximate PER expression [4]:

$$\mathrm{PER}_n(\gamma) \approx \begin{cases} 1, & \text{if } 0 < \gamma < \gamma_{pn}, \\ a_n \exp(-g_n\gamma), & \text{if } \gamma \geq \gamma_{pn}, \end{cases} \tag{4}$$

where n is the mode index. Parameters a_n, g_n, and γ_{pn} in (4) are mode-dependent, and are obtained by fitting (4) to the exact PER [4]. They are listed in Table 1 for the transmission modes in TM, with a packet length $N_p = 1,080$.

We set the region boundary (or the switching threshold) γ_n for the transmission mode n to be the minimum SNR required to achieve $\mathrm{P}_{\mathrm{target}}$. In general, the required PER in (2) satisfies $\mathrm{P}_{\mathrm{target}} < 1$. From (4), we obtain:

$$\begin{aligned} &\gamma_0 = 0, \qquad \gamma_{N+1} = +\infty, \\ &\gamma_n = \frac{1}{g_n} \ln\left(\frac{a_n}{\mathrm{P}_{\mathrm{target}}}\right), \quad n = 1, 2, \ldots, N. \end{aligned} \tag{5}$$

With γ_n specified by (5), one can verify that the AMC in (3) guarantees the PER in (2), and maximizes the spectral efficiency, for the given finite transmission modes.

Summarizing our results in the previous and this subsection, the AMC is designed following these steps:

Step 1: Given C1 and C2, determine the $\mathrm{P}_{\mathrm{target}}$ from (2);

Step 2: For this $\mathrm{P}_{\mathrm{target}}$, determine $\{\gamma_n\}_{n=0}^{N+1}$ via (5).

The proposed cross-layer design hence leads to the following operating stages in the overall system:

Stage 1: Update transmission modes for each frame by using AMC as in (3).

Stage 2: Retransmit error packets by $N_r^{\max}$-truncated ARQ.

4. PERFORMANCE ANALYSIS

In this section, we derive the average PER and the spectral efficiency of our cross-layer design, as well as those for truncated ARQ with a fixed mode; we term it ARQ-only.

4.1. Combined AMC with truncated ARQ

Because both the constant power and the transmission modes have finite granularity, the actual average PER at the physical layer will generally end up being lower than $\mathrm{P}_{\mathrm{target}}$.

According to the AMC rule in (3), the transmission mode and thus the instantaneous PER depend on the receive SNR

γ. Letting $\mathrm{PER}(\gamma)$ denote the instantaneous PER, we obtain the average PER as:

$$\overline{\mathrm{PER}} = \int_0^\infty \mathrm{PER}(\gamma) p_\gamma(\gamma) d\gamma. \tag{6}$$

Since $\mathrm{P}_{\mathrm{target}} < 1$, in general we have $\gamma_{pn} < \gamma_n$ for the γ_n chosen in (5). Based on (1) and (4), we simplify (6) as [4]:

$$\begin{aligned} \overline{\mathrm{PER}} &= \sum_{n=1}^{N} \int_{\gamma_n}^{\gamma_{n+1}} a_n \exp(-g_n\gamma) p_\gamma(\gamma) d\gamma \qquad (7) \\ &= \sum_{n=1}^{N} \left[\frac{a_n}{\Gamma(m)} \left(\frac{m}{\bar{\gamma}}\right)^m \frac{\Gamma(m, b_n\gamma_n) - \Gamma(m, b_n\gamma_{n+1})}{(b_n)^m} \right], \end{aligned}$$

where $b_n := m/\bar{\gamma} + g_n$, and $\Gamma(m,x) := \int_x^\infty t^{m-1}e^{-t}dt$ is the complementary incomplete Gamma function.

For notational brevity, let us define $p := \overline{\mathrm{PER}}$. Since $N_r^{\max}$-truncated ARQ is implemented, the average number of transmissions per packet can be found per A4 as [4]:

$$\overline{N}(p, N_r^{\max}) = 1 + p + p^2 + \cdots + p^{N_r^{\max}} = \frac{1 - p^{N_r^{\max}+1}}{1-p}. \tag{8}$$

With the average PER in (7), the actual packet loss probability at the data link layer with $N_r^{\max}$-truncated ARQ is:

$$\mathrm{P}_{\text{actual loss}} = p^{N_r^{\max}+1} \le \mathrm{P}_{\text{target}}^{N_r^{\max}+1} := \mathrm{P}_{\text{loss}}, \tag{9}$$

which guarantees C2.

With C1 and C2 satisfied, we are now ready to evaluate the achieved system spectral efficiency. From (3), the transmission mode n will be chosen with probability [1, 4]:

$$\mathrm{Pr}(n) = \int_{\gamma_n}^{\gamma_{n+1}} p_\gamma(\gamma) d\gamma = \frac{\Gamma(m, \frac{m\gamma_n}{\bar{\gamma}}) - \Gamma(m, \frac{m\gamma_{n+1}}{\bar{\gamma}})}{\Gamma(m)}. \tag{10}$$

When mode n is used, each transmitted symbol will carry $R_n = R_c \log_2(M_n)$ information bits. Therefore, the average spectral efficiency achieved at the physical layer without packet retransmission is (analogous to [1] where only physical layer AMC design is considered):

$$\overline{\mathrm{S}}_{\mathrm{e,physical}} = \sum_{n=1}^{N} R_n \mathrm{Pr}(n). \tag{11}$$

With the truncated ARQ, each packet, and thus each information bit, is equivalently transmitted $\overline{N}(p, N_r^{\max})$ times. The overall average spectral efficiency is thus:

$$\overline{\mathrm{S}}_{\mathrm{e}}(N_r^{\max}) = \frac{\overline{\mathrm{S}}_{\mathrm{e,physical}}}{\overline{N}(p, N_r^{\max})} = \frac{\sum_{n=1}^{N} R_n \mathrm{Pr}(n)}{\overline{N}(p, N_r^{\max})}. \tag{12}$$

When $N_r^{\max} = 0$, we have $\overline{N}(p, N_r^{\max} = 0) = 1$ from (8). Hence, the average spectral efficiency for the special case of AMC-only is:

$$\overline{\mathrm{S}}_e(N_r^{\max} = 0) = \sum_{n=1}^{N} R_n \mathrm{Pr}(n). \tag{13}$$

The form in (13) is in agreement with [1], where the AMC design is considered only at the physical layer.

4.2. Truncated ARQ-only

In the truncated ARQ-only, CSI is not exploited at the transmitter, for each transmission mode. We have to evaluate the spectral efficiency for each mode separately.

Suppose that the transmission mode n is adopted. The average PER at the physical layer can be computed based on (1) and (4) as [4]:

$$\begin{aligned} \overline{\mathrm{PER}}_n &= \int_0^\infty \mathrm{PER}_n(\gamma)\, p_\gamma(\gamma) d\gamma \qquad (14) \\ &= \int_0^{\gamma_{pn}} p_\gamma(\gamma) d\gamma + \int_{\gamma_{pn}}^\infty a_n \exp(-g_n\gamma)\, p_\gamma(\gamma) d\gamma \\ &= 1 - \frac{\Gamma(m, m\gamma_{pn}/\bar{\gamma})}{\Gamma(m)} + \frac{a_n}{\Gamma(m)} \left(\frac{m}{\bar{\gamma}}\right)^m \frac{\Gamma(m, b_n\gamma_{pn})}{(b_n)^m}. \end{aligned}$$

For notational brevity, let $q_n := \overline{\mathrm{PER}}_n$. Under C1, the average number of transmissions per packet is thus:

$$\overline{N}(q_n, N_r^{\max}) = 1 + q_n + q_n^2 + \cdots + q_n^{N_r^{\max}} = \frac{1 - q_n^{N_r^{\max}+1}}{1 - q_n}. \tag{15}$$

Mimicking the derivation of (12), the average spectral efficiency of mode n with $N_r^{\max}$-truncated ARQ is:

$$\overline{\mathrm{S}}_{\mathrm{e,n}}(N_r^{\max}) = \frac{R_n}{\overline{N}(q_n, N_r^{\max})}. \tag{16}$$

The packet loss probability after a maximum number of $N_r^{\max}$ retransmissions is:

$$\mathrm{P}_{n,\mathrm{ARQ}} = q_n^{N_r^{\max}+1}. \tag{17}$$

Notice that without adapting to the instantaneous SNR, the actual packet loss probability $\mathrm{P}_{n,\mathrm{ARQ}}$ for mode n is not guaranteed to be less than $\mathrm{P}_{\mathrm{loss}}$, for the entire range of the average SNR $\bar{\gamma}$. For each average SNR $\bar{\gamma}$, we evaluate (14) and (17), and test if C2 is satisfied. Numerically, we have identified that there exists a threshold $\bar{\gamma}_{n,\mathrm{th}}$ for mode n, for which C2 is guaranteed when $\bar{\gamma} \ge \bar{\gamma}_{n,\mathrm{th}}$, while C2 is not satisfied when $\bar{\gamma} < \bar{\gamma}_{n,\mathrm{th}}$. We are only interested in finding the average spectral efficiency for the SNR range $\bar{\gamma} \ge \bar{\gamma}_{n,\mathrm{th}}$, that C2 is guaranteed. In summary, the average spectral efficiency for the truncated ARQ-only under C1 and C2 is determined as:

$$\overline{\mathrm{S}}_{\mathrm{e},n}(N_r^{\max}) = \begin{cases} 0, & \bar{\gamma} < \bar{\gamma}_{n,\mathrm{th}} \\ \frac{R_n}{\overline{N}(q_n, N_r^{\max})}, & \bar{\gamma} \ge \bar{\gamma}_{n,\mathrm{th}} \end{cases}. \tag{18}$$

Our closed-form expressions for the spectral efficiencies in (12), (13), (18), and for the average PER in (7), will facilitate the numerical performance testing in the next section.

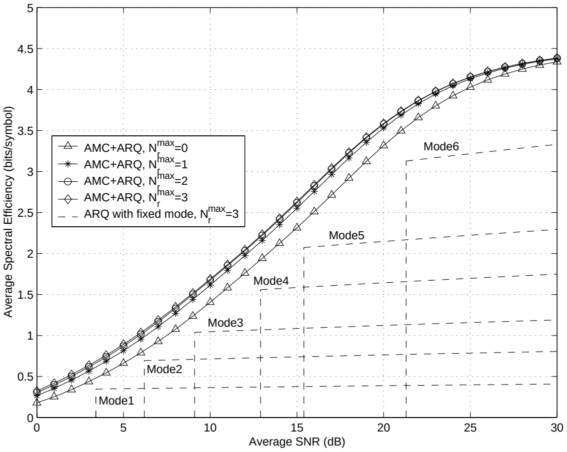

Fig. 3. The achieved average spectral efficiency

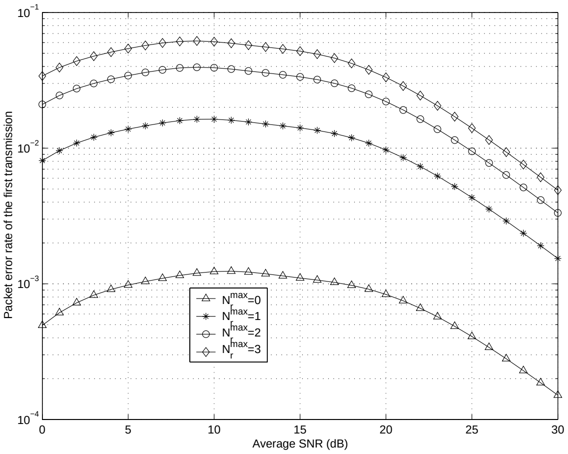

Fig. 4. The packet error rate at the physical layer

5. NUMERICAL RESULTS

Test Case 1 (Dependence on the maximum number of retransmissions $N_r^{\max}$): We adopt $N_p = 1,080$, and let the performance constraint in C2 be $\mathrm{P_{loss}} = 0.01$. We set the Nakagami fading parameter $m = 1$, which corresponds to a Rayleigh fading channel. With $N_r^{\max}$ varying from 0 to 3, we depict the corresponding spectral efficiencies in Fig. 3.

From Fig. 3, we observe that the average spectral efficiency with combined AMC-ARQ exceeds that of the AMC-only by 0.25 bits per transmitted symbol, using only one retransmission ($N_r^{\max} = 1$). This implies a significant rate enhancement. For example, in the HIPERLAN/2 standard, where the symbol rate is 12M symbols/s [3], combining AMC with truncated ARQ leads to an approximate 3Mbits/s increase in transmission rate with only one retransmission.

We observe that the spectral efficiency improves with increasing $N_r^{\max}$. However, the increment degrades quickly, and "diminishing returns" appear. This implies that the maximum number of retransmissions, need not be arbitrarily large. Small number of retransmissions can achieve sufficient spectral efficiency gains. They incur smaller delay and buffer-size penalties, and thus lead to improved delay-throughput tradeoffs.

The average packet error rate at the physical layer is depicted in Fig. 4. It explains why increasing $N_r^{\max}$ allows for spectral efficiency improvement. As $N_r^{\max}$ increases, the error-correcting capability of the truncated ARQ increases, which relieves the physical layer from stringent error correction requirements. With a lower performance requirement, transmission rates can be increased at the physical layer, which in turn leads to the overall spectral efficiency improvement. This gain is introduced by relaxing the system delay requirement in C1; thus, a tradeoff between delay and throughput emerges.

Test Case 2 (Comparison with truncated ARQ-only): With $N_r^{\max} = 3$ and $m = 1$, the spectral efficiencies of the truncated ARQ-only with different modes from TM are depicted in Fig. 3. Allowing for more retransmissions, the spectral efficiency with $N_r^{\max} = 3$ should be greater than those corresponding to $N_r^{\max} < 3$, whose plots are omitted due to lack of space. It is clear that combining AMC with truncated ARQ offers much higher spectral efficiency than the truncated ARQ with fixed transmission mode, thanks to the exploitation of the channel knowledge at the transmitter.

6. REFERENCES

[1] M. -S. Alouini and A. J. Goldsmith, "Adaptive modulation over Nakagami fading channels," *Kluwer Journal on Wireless Commun.*, vol. 13, no. 1–2, pp. 119–143, May 2000.

[2] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1273–1289, May 2001.

[3] A. Doufexi, S. Armour, M. Butler, A. Nix, D. Bull, J. McGeehan, and P. Karlsson, "A comparison of the HIPERLAN/2 and IEEE 802.11a wireless LAN standards," *IEEE Commun. Magazine*, vol. 40, no. 5, pp. 172–180, May 2002.

[4] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links," *IEEE Trans. Wireless Commun.*, submitted Dec. 2002.

[5] E. Malkamaki and H. Leib, "Performance of truncated type-II hybrid ARQ schemes with noisy feedback over block fading channels," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1477–1487, Sep. 2000.

[6] H. Minn, M. Zeng, and V. K. Bhargava, "On ARQ scheme with adaptive error control," *IEEE Trans. Veh. Tech.*, vol. 50, no. 6, pp. 1426–1436, Nov. 2001.