Cross-Layer Combining of Adaptive Modulation and Coding With Truncated ARQ Over Wireless Links

Qingwen Liu, *Student Member, IEEE*, Shengli Zhou, *Member, IEEE*, and Georgios B. Giannakis, *Fellow, IEEE*

***Abstract*—We developed a cross-layer design which combines adaptive modulation and coding at the physical layer with a truncated automatic repeat request protocol at the data link layer, in order to maximize spectral efficiency under prescribed delay and error 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 relieve stringent error control requirements at the physical layer, and thereby enable considerable spectral efficiency gain. This gain is comparable with that offered by diversity, provided that the maximum number of transmissions per packet equals the diversity order. Diminishing returns on spectral efficiency, that result when increasing the maximum number of retransmissions, suggest that a small number of retransmissions offers a desirable delay-throughput tradeoff, in practice.**

***Index Terms*—Adaptive modulation and coding (AMC), automatic repeat request (ARQ) protocol, cross-layer design, quality of service (QoS), wireless networks.**

I. Introduction

IN WIRELESS communication networks, the demand for high data rates and quality of service (QoS) is growing at a rapid pace. However, the performance of wireless links is severely degraded due to channel fading, which limits the overall system throughput considerably relative to wireline alternatives.

To enhance throughput in future wireless data communication systems, adaptive modulation and coding (AMC) have been studied extensively and advocated at the physical layer, in order to match transmission rates to time-varying channel conditions; see e.g., [1], [3], [6]–[9], [14], [16], [18], and references therein. However, to achieve high reliability at the physical layer, one has to reduce 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 automatic repeat request (ARQ) protocol at the data link layer, that requests retransmissions for those packets received in error. Since retransmissions are activated only when necessary, ARQ is quite effective in improving system throughput relative to using only forward error coding (FEC) at the physical layer [10]. To minimize delays and buffer sizes in practice, truncated ARQ protocols have been widely adopted to limit the maximum number of retransmissions [10]. However, only fixed modulation and coding at the physical layer have been considered in systems with truncated ARQ protocols [10].

Instead of considering AMC at the physical layer and ARQ at the data link layer separately, we pursue here a cross-layer design, that combines these two layers judiciously to maximize spectral efficiency, or throughput, under prescribed delay and error 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 then 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 only at the physical layer, or, incorporation of ARQ only with a fixed modulation and coding. Specifically, the spectral efficiency of the cross-layer design relative to AMC alone improves by about 0.25 b/transmitted symbol, using only one retransmission over Rayleigh-fading channels. This is a major gain for high symbol rate transmissions. Compared with ARQ applied to fixed modulation and coding schemes, a much larger throughput improvement is achieved, because our cross-layer design exploits the channel knowledge at the transmitter.

Since the improvement on spectral efficiency decreases as the maximum number of retransmissions increases, a small number of retransmissions yields a desirable delay-throughput tradeoff, because it achieves sufficient spectral efficiency gain with reduced delay and buffer-size requirements. The gain introduced by retransmissions is found comparable to that offered by diversity, provided that the maximum number of transmissions per packet equals the diversity order.

The rest of this paper is organized as follows. We introduce the system and channel models in Section II. We develop the cross-layer design in Section III, by combining AMC at the physical layer with ARQ at the data link layer. We analyze the achieved spectral efficiency in Section IV, present numerical results in Section V, and finally draw concluding remarks in Section VI.

Manuscript received December 5, 2002; revised March 16, 2003; accepted June 27, 2003. The editor coordinating the review of this paper and approving it for publication was H. Yanikomeroglu. This work was supported by the National Science Foundation under Grant 0105612, and by the Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-011. The material in this paper was presented in part at the IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC), pp. 110–114, Rome, Italy, June 15-18, 2003.

Q. Liu and G. B. Giannakis are with the Department of Electrical Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: qliu@ece.umn.edu, georgios@ece.umn.edu).

S. Zhou is with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269 USA (e-mail: shengli@engr.uconn.edu).

Digital Object Identifier 10.1109/TWC.2004.833474

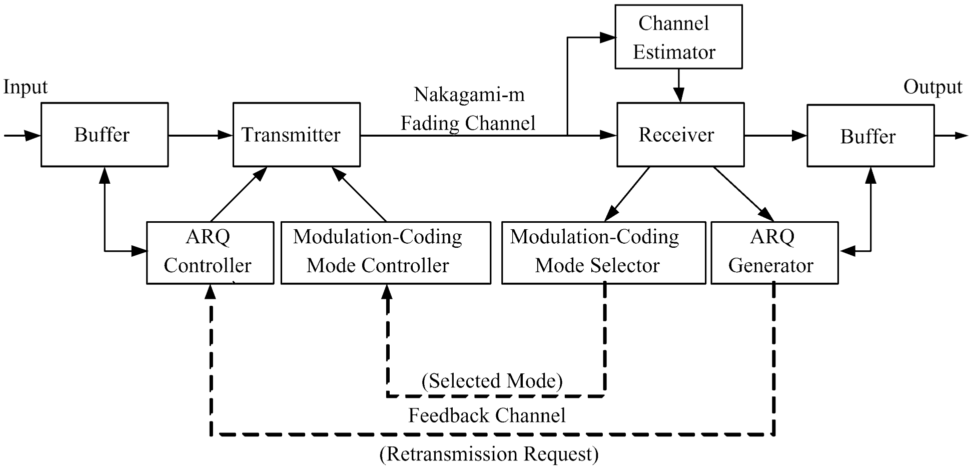

Fig. 1. System and channel models.

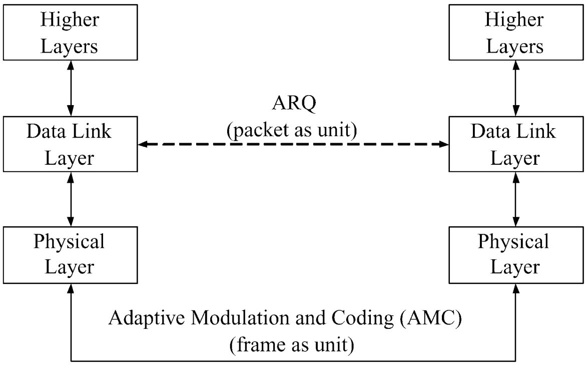

Fig. 2. Cross-layer structure combining AMC with ARQ.

II. Modeling

Consider the single-transmit single-receive antenna system in Fig. 1. The layer structure of the system is shown in Fig. 2. It consists of a joint adaptive modulation and coding module at the physical layer, and an ARQ module at the data link layer. 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, with each mode consisting of a specific modulation and FEC code pair as in HIPERLAN/2, IEEE 802.11a, and 3GPP standards [1], [5]. Based on CSI acquired at the receiver, the AMC selector determines the modulation-coding pair (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-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 arranges retransmission of the requested packet that is stored in the buffer. The feedback for ARQ commands is different from that used for feeding back AMC information, although both ARQ and AMC related parameters are sent back via the same physical feedback channel.

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

TM1) Uncoded [without forward error correction (FEC)] M_n-ary rectangular or square quadrature amplitude modulation (QAM) modes, where $M_n = 2^n$, $n = 1, 2, \ldots, 7$ [20].

TM2) Convolutionally coded M_n-ary rectangular or square QAM modes, adopted from the HIPERLAN/2 or IEEE 802.11a standards [5].

The transmission modes in TM1 and TM2 are listed in Tables I and II, respectively, in a rate of ascending order. Although we will focus on TM1 and TM2, other transmission modes can be similarly constructed.

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. 3. 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 b/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 [5]. 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 modulation and coding pair.

We next list the operating assumptions adopted in this paper.

A1) The channel is frequency flat, and remains invariant per frame, but is allowed to vary from frame to frame. This corresponds to a block fading channel model, which is suitable for slowly-varying fading channels [4]. As a

TABLE I
TRANSMISSION MODES IN TM1 WITH UNCODED M_n-QAM MODULATION

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6	Mode 7
Modulation	BPSK	QPSK	8-QAM	16-QAM	32-QAM	64-QAM	128-QAM
Rate(bits/sym.)	1	2	3	4	5	6	7
a_n	67.7328	73.8279	58.7332	55.9137	50.0552	42.5594	40.2559
g_n	0.9819	0.4945	0.1641	0.0989	0.0381	0.0235	0.0094
γ_{pn}(dB)	6.3281	9.3945	13.9470	16.0938	20.1103	22.0340	25.9677

TABLE II
TRANSMISSION MODES IN TM2 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 $g = [133, 171]$. The coding rates are obtained from the puncturing pattern P2 in the HIPERLAN/2 standard.)

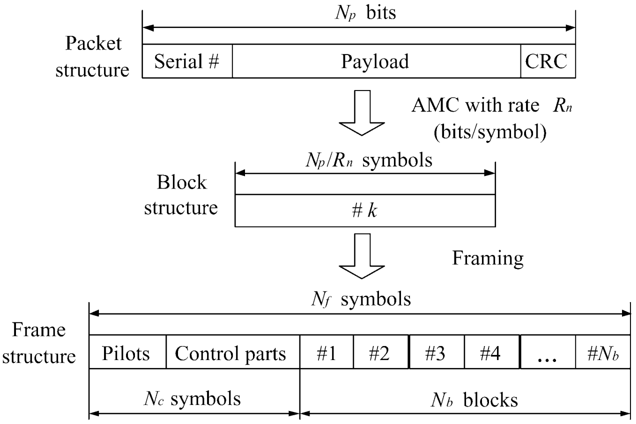

Fig. 3. Packet and frame structures.

consequence, AMC is adjusted on a frame-by-frame basis.

A2) Perfect channel state information (CSI) is available at the receiver using training-based channel estimation. The corresponding mode selection is fed back to the transmitter without error and latency, as in [3], [7]–[9].

The assumption that the feedback channel is error free and has no latency, could be at least approximately satisfied by using a fast feedback link with powerful error control for feedback information. Further considerations on system design with e.g., delayed, or, noisy CSI [6], [13], will be left for future investigation.

A3) Error detection based on CRC is perfect, provided that sufficiently reliable error detection CRC codes are used [12], [19]. As in [12], the serial number and the CRC parity bits in each packet are not included in the throughput calculation, because they introduce negligible redundancy relative to the number of payload bits.

For flat fading channels adhering to A1, the channel quality can be captured by a single parameter, namely the received SNR γ. Since the channel varies from frame to frame, we adopt the general Nakagami-m model to describe γ statistically [15]. The received SNR γ per frame is thus a random variable with a Gamma probability density function (pdf):

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

where $\bar{\gamma} := E\{\gamma\}$ is the average received SNR, $\Gamma(m) := \int_0^\infty t^{m-1} e^{-t} dt$ is the Gamma function, and m is the Nakagami fading parameter $(m \geq 1/2)$. We choose the Nakagami-m channel model because it applies to a large class of fading channels. It includes the Rayleigh channel as a special case when $m = 1$. An one-to-one mapping between the Ricean factor K and the Nakagami fading parameter m allows also Ricean channels to be well approximated by Nakagami-m channels [15].

III. COMBINING AMC WITH TRUNCATED ARQ

In this section, we develop our cross-layer design, which combines AMC at the physical layer with truncated ARQ at the data link layer.

Since only finite delays and buffer sizes can be afforded in practice, the maximum number of ARQ retransmissions has to be bounded. This number can be specified by dividing the maximum allowable system delay over the round trip delay required for each retransmission. Formally, 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 [10]. If a packet is not received correctly after $N_r^{\max}$ retransmissions, we will drop it, and declare packet loss. This is very reasonable and can be afforded in video/image transmissions for instance, because the underlying bit streams represent highly correlated image contents. On the other hand, the error packets can also be utilized if the receiver decides to do so. 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 P_{loss}.

As an example, in MPEG-4 audio/video transmissions, the quality of service requires less than 150–400 ms delays, and a bit-error rate (BER) range 10^{-4}–10^{-6} [2]. If the average round trip delay is 100 ms and the packet length is $N_p \approx 1000$, the maximum number of retransmissions $N_r^{\max}$ should be at most 4, and the packet loss probability P_{loss} should be less than 0.1–0.001 (assuming independent bit errors inside each packet). Hence, C1 and C2 can be derived from the required quality of service in the application at hand.

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. In other words, we address the following interesting question: with the aid of $N_r^{\max}$-truncated ARQ at the data link layer, how can we optimally design the AMC at the physical layer to maximize throughput, while guaranteeing the overall system performance dictated by C2?

A. 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.

Notice that a packet is dropped if it is received incorrectly after a maximum number of $(N_r^{\max} + 1)$ transmissions; i.e., after $N_r^{\max}$ retransmissions. Let us suppose that the instantaneous packet error rate (PER) is guaranteed to be no greater than P_0 for each chosen AMC mode at the physical layer. Then, the packet loss probability at the data link layer is no greater than $P_0^{N_r^{\max}+1}$. To satisfy C2, we need to impose

$$P_0^{N_r^{\max}+1} \le P_{\text{loss}}. \tag{2}$$

From (2), we obtain

$$P_0 \le P_{\text{loss}}^{(1/N_r^{\max}+1)} := P_{\text{target}}. \tag{3}$$

Therefore, if we design AMC to satisfy a PER upper-bound as in (3) at the physical layer, and implement a $N_r^{\max}$-truncated ARQ at the data link layer, both delay and performance requirements C1 and C2 will be satisfied. Our remaining problem is to design AMC to maximize spectral efficiency while maintaining (3).

B. AMC Design at the Physical Layer

Our objective here is to maximize the data rate, while maintaining the required performance, through AMC at the physical layer. As we already mentioned, the transmission modes TM1 or TM2 are arranged so that the rate is increasing as the mode index n increases. Let N denote the total number of transmission modes available ($N = 7, 6$ for TM1 and TM2, respectively). As in [3], we assume constant power transmission, and partition the total SNR range into $N + 1$ nonoverlapping consecutive intervals, with boundary points denoted as $\{\gamma_n\}_{n=0}^{N+1}$. Specifically,

$$\text{mode } n \text{ is chosen}, \quad \text{when } \gamma \in [\gamma_n, \gamma_{n+1}). \tag{4}$$

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

The boundary points are specified in [3] for a given target BER. Finding the target BER through the required PER, and then specifying the boundaries as in [3], is certainly a possibility. However, as we verify in Appendix, BER is not easily determined from PER, and *vice versa*, especially for coded transmissions. Furthermore, since our system uses packets as processing units, we will henceforth specify the boundary points to meet the required PER.

Exact closed-form PERs for the uncoded modulations in TM1 are provided in Appendix, while exact closed-form PERs for the coded modulations in TM2 are not available. To simplify the AMC design, we will rely on the following approximate PER expression:

$$\text{PER}_n(\gamma) \approx \begin{cases} 1, & \text{if } 0 < \gamma < \gamma_{pn}, \\ a_n \exp(-g_n \gamma), & \text{if } \gamma \ge \gamma_{pn} \end{cases} \tag{5}$$

where n is the mode index and γ is the received SNR. Parameters a_n, g_n, and γ_{pn} in (5) are mode-dependent, and are obtained by fitting (5) to the exact PER, as explained in the Appendix, where the accuracy of this PER approximation is also verified. With a packet length $N_p = 1080$, the fitting parameters for transmission modes in TM1 and TM2 are provided in Tables I and II, respectively. Using the approximate yet simple expression (5) facilitates the mode selection. This approach has also been adopted by [3], [7], and [9], although these works used BER as their figure of merit.

We set the region boundary (or the switching threshold) γ_n for the transmission mode n to be the minimum SNR required to achieve P_{target}. In general, the required PER in (3) satisfies $P_{\text{target}} < 1$. Inverting the PER expression in (5), we obtain

$$\begin{aligned} \gamma_0 &= 0, \\ \gamma_n &= \frac{1}{g_n} \ln\left(\frac{a_n}{P_{\text{target}}}\right), \quad n = 1, 2, \ldots, N, \\ \gamma_{N+1} &= +\infty. \end{aligned} \tag{6}$$

With the γ_n specified by (6), one can verify that the AMC in (4) guarantees(3). Maintaining the target performance, the proposed AMC with (4) and (6) then maximizes the spectral efficiency, with the given finite transmission modes.

Summarizing our results in Section III-A and this subsection, we design AMC at the physical layer following these steps:

Step 1) Given C1 and C2, determine P_{target} from (3).

Step 2) For the P_{target} found, determine $\{\gamma_n\}_{n=0}^{N+1}$ via (6).

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

Step 1) Update modes per frame by using AMC as in (4).

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

IV. Performance Analysis

In this section, we derive the average PER and the spectral efficiency of our cross-layer design in Section III. We also derive the performance for truncated ARQ with a fixed modulation and coding pair; we term this scheme ARQ-only.

For analytical convenience, we further assume:

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 of 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, which may be greater than 100–200 ms in some practical systems [11]. In general, these round trip delays exceed the channel coherence time. Hence, the original packet and possible subsequent retransmissions experience channels that can be assumed i.i.d., which justifies well A4. For applications where A4 is violated, even though C1 and C2 are still guaranteed, evaluating the performance in closed-form becomes involved.

A. Combined AMC With Truncated ARQ

Since the instantaneous PER is upper-bounded by P_{target} in our AMC design, the average PER at the physical layer will be lower than P_{target}. We first evaluate this average PER at the physical layer.

According to the AMC rule in (4), the transmission mode, and thus the instantaneous PER, depend on the received SNR γ. Since $P_{\text{target}} < 1$ in general, we have $\gamma_{pn} < \gamma_n$ for the γ_n chosen in (6). Each mode n will be chosen with probability (cf. [3, eq. (34)])

$$\begin{aligned}\Pr(n) &= \int_{\gamma_n}^{\gamma_{n+1}} p_\gamma(\gamma)d\gamma \\ &= \frac{\Gamma\left(m, \frac{m\gamma_n}{\bar{\gamma}}\right) - \Gamma\left(m, \frac{m\gamma_{n+1}}{\bar{\gamma}}\right)}{\Gamma(m)} \end{aligned} \tag{7}$$

where $\Gamma(m,x) := \int_x^\infty t^{m-1}e^{-t}dt$ is the complementary incomplete Gamma function. Let $\overline{\text{PER}}_n$ denote the average packet error rate for mode n (the ratio of the number of incorrectly received packets over those transmitted using mode n). From (1), (5), and (6), we can derive $\overline{\text{PER}}_n$ as

$$\begin{aligned}\overline{\text{PER}}_n &= \frac{1}{\Pr(n)} \int_{\gamma_n}^{\gamma_{n+1}} \text{PER}_n(\gamma) p_\gamma(\gamma)d\gamma \\ &= \frac{1}{\Pr(n)} \int_{\gamma_n}^{\gamma_{n+1}} a_n \exp(-g_n\gamma) p_\gamma(\gamma)d\gamma \\ &= \frac{1}{\Pr(n)} \frac{a_n}{\Gamma(m)} \left(\frac{m}{\bar{\gamma}}\right)^m \\ &\quad \times \frac{\Gamma(m, b_n\gamma_n) - \Gamma(m, b_n\gamma_{n+1})}{(b_n)^m} \end{aligned} \tag{8}$$

where the third equality follows from [3, eq. (37)], and

$$b_n := \frac{m}{\bar{\gamma}} + g_n.$$

The average PER of AMC can then be computed as the ratio of the average number of incorrectly received packets over the total average number of transmitted packets (cf. [3, eq. (35)])

$$\overline{\text{PER}} = \frac{\sum_{n=1}^{N} R_n \Pr(n) \overline{\text{PER}}_n}{\sum_{n=1}^{N} R_n \Pr(n)}. \tag{9}$$

Since truncated ARQ is implemented at the data link layer, the packets in error during the original reception may be retransmitted, up to a maximum of $N_r^{\max}$ times. For notational brevity, let us define $p := \overline{\text{PER}}$. Per A4, the average number of transmissions per packet can be found as (cf. [19, p. 397])

$$\begin{aligned}\overline{N}(p, N_r^{\max}) &= 1 + p + p^2 + \cdots + p^{N_r^{\max}} \\ &= \frac{1 - p^{N_r^{\max}+1}}{1-p}. \end{aligned} \tag{10}$$

When $N_r^{\max} = 0$, we have $\overline{N}(p, N_r^{\max} = 0) = 1$, which corresponds to the special case of AMC-only. With the average PER in (9), the actual packet loss probability at the data link layer with $N_r^{\max}$-truncated ARQ is

$$P_{\text{actual loss}} = p^{N_r^{\max}+1} \le P_{\text{target}}^{N_r^{\max}+1} = P_{\text{loss}} \tag{11}$$

which verifies C2.

With C1 and C2 satisfied, we are now ready to evaluate the achieved system spectral efficiency. When mode n is used, each transmitted symbol will carry $R_n = R_c \log_2(M_n)$ information bits for the mode adhering to a M_n-QAM constellation, and a rate R_c FEC code. For uncoded transmission modes in TM1, we set $R_c = 1$. As in [3], [7], and [8], we assume a Nyquist pulse shaping filter with bandwidth $B = 1/T_s$, where T_s is the symbol rate. Therefore, the average spectral efficiency (bit rate per unit bandwidth) achieved at the physical layer without considering possible packet retransmission is (similar to [3] where only physical layer AMC design is considered)

$$\overline{S}_{e,\text{physical}} = \sum_{n=1}^{N} R_n \Pr(n). \tag{12}$$

When truncated ARQ is implemented, each packet, and thus each information bit, is equivalently transmitted $\overline{N}(p, N_r^{\max})$ times. Hence, the overall average spectral efficiency, as a function of $N_r^{\max}$, is obtained as

$$\begin{aligned}\overline{S}_e(N_r^{\max}) &= \frac{\overline{S}_{e,\text{physical}}}{\overline{N}(p, N_r^{\max})} \\ &= \frac{1}{\overline{N}(p, N_r^{\max})} \sum_{n=1}^{N} R_n \Pr(n). \end{aligned} \tag{13}$$

Setting $N_r^{\max} = 0$ in (13), we obtain the average spectral efficiency for AMC-only as

$$\overline{S}_e(N_r^{\max} = 0) = \sum_{n=1}^{N} R_n \Pr(n). \tag{14}$$

The form in (14) is in agreement with [3], where the AMC design is considered only at the physical layer.

B. Truncated ARQ-Only

In a system using truncated ARQ-only, CSI is not exploited at the transmitter. The transmission mode is now fixed, and one has to evaluate the average 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 (5) as

$$\begin{aligned}\overline{\mathrm{PER}}(n) &= \int_0^\infty \mathrm{PER}_n(\gamma) p_\gamma(\gamma) d\gamma \\ &= \int_0^{\gamma_{pn}} p_\gamma(\gamma) d\gamma + \int_{\gamma_{pn}}^\infty a_n \exp(-g_n \gamma) p_\gamma(\gamma) d\gamma \\ &= 1 - \frac{\Gamma(m, \frac{m\gamma_{pn}}{\bar{\gamma}})}{\Gamma(m)} \\ &\quad + \frac{a_n}{\Gamma(m)} \left(\frac{m}{\bar{\gamma}}\right)^m \frac{\Gamma(m, b_n \gamma_{pn})}{(b_n)^m}. \end{aligned} \tag{15}$$

For notational brevity, let $q_n := \overline{\mathrm{PER}}(n)$. In accordance to A4, the average number of transmissions per packet is,

$$\begin{aligned}\overline{N}(q_n, N_r^{\max}) &= 1 + q_n + q_n^2 + \cdots + q_n^{N_r^{\max}} \\ &= \frac{1 - q_n^{N_r^{\max}+1}}{1 - q_n}. \end{aligned} \tag{16}$$

Mimicking the derivation of (13), the average spectral efficiency of mode n with $N_r^{\max}$-truncated ARQ is found to be

$$\overline{S}_{e,n}(N_r^{\max}) = \frac{R_n}{\overline{N}(q_n, N_r^{\max})}. \tag{17}$$

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

$$P_{n,\mathrm{ARQ}} = q_n^{N_r^{\max}+1}. \tag{18}$$

Notice that without adapting to the instantaneous SNR, the actual packet loss probability, $P_{n,\mathrm{ARQ}}$ for mode n, is not guaranteed to be less than P_{loss}, for the entire range of the average SNR $\bar{\gamma}$. For each average SNR $\bar{\gamma}$, we evaluate (15) and (18), 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 $\bar{\gamma} \geq \bar{\gamma}_{n,\mathrm{th}}$, 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 $\bar{\gamma} \geq \bar{\gamma}_{n,\mathrm{th}}$, that C2 is guaranteed. In summary, the average spectral efficiency for the truncated ARQ-only under C1 and C2 is determined as

$$\overline{S}_{e,n}(N_r^{\max}) = \begin{cases} 0, & \bar{\gamma} < \bar{\gamma}_{n,\mathrm{th}} \\ \frac{R_n}{\overline{N}(q_n, N_r^{\max})}, & \bar{\gamma} \geq \bar{\gamma}_{n,\mathrm{th}}. \end{cases} \tag{19}$$

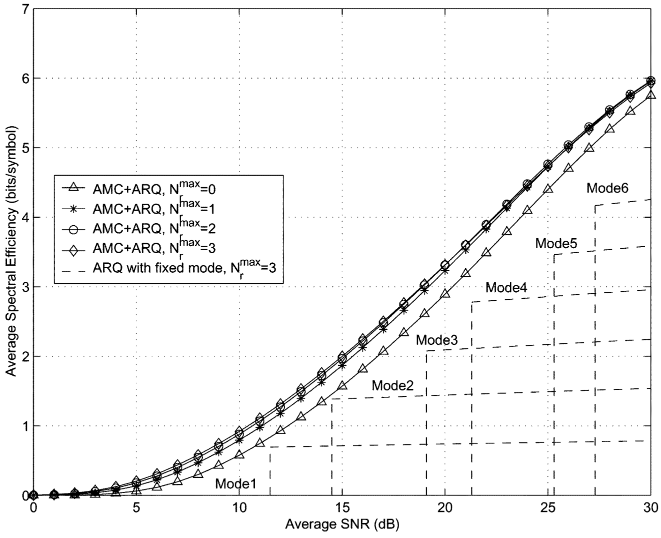

Fig. 4. Average spectral efficiency for TM1 versus average SNR.

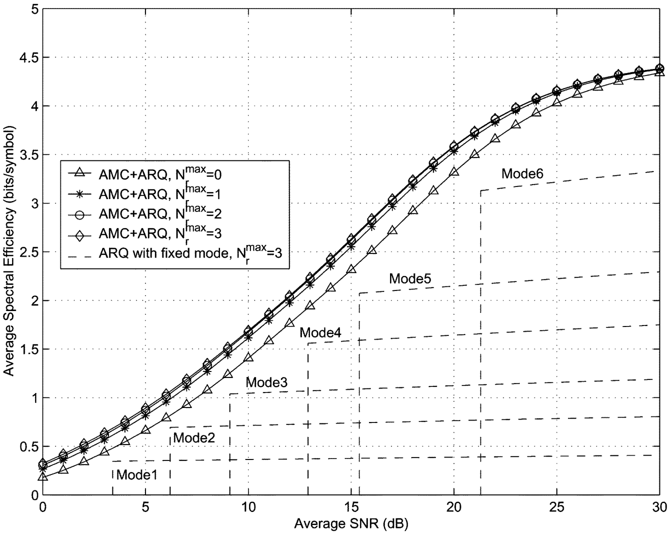

Fig. 5. Average spectral efficiency for TM2 versus average SNR.

Our closed-form expressions for the spectral efficiencies in (13), (14), (19), and for the average PER in (9), will facilitate the numerical performance testing that we pursue in Section V.

V. Numerical Results

In this section, we present numerical results, where both TM1 and TM2 are considered. We set the packet length $N_p = 1080$, with the PER approximation parameters of (5) listed in Tables I and II. Specific numerical values will be affected if one chooses a different N_p. However, similar observations are expected, because of the unifying development in Sections III and IV.

Test Case 1 (Dependence on the maximum number of retransmissions $N_r^{\max}$). Let the performance constraint in C2 be $P_{\mathrm{loss}} = 0.01$. 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 average spectral efficiencies in Figs. 4 and 5, for modes in TM1 and TM2, respectively.

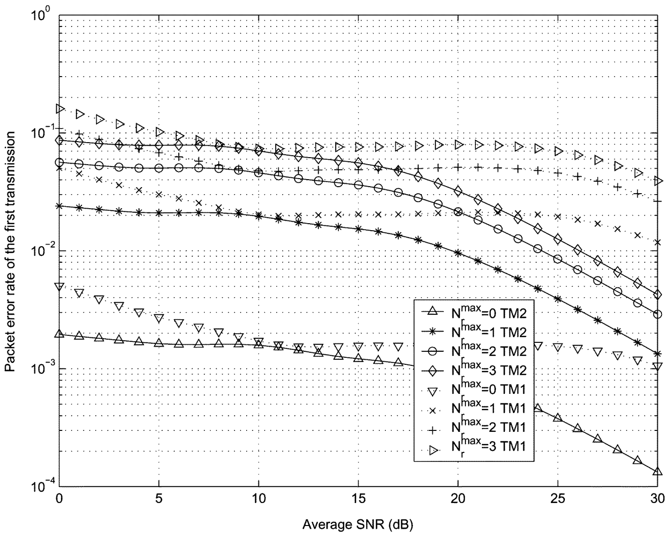

Fig. 6. Packet error rate at the physical layer versus average SNR.

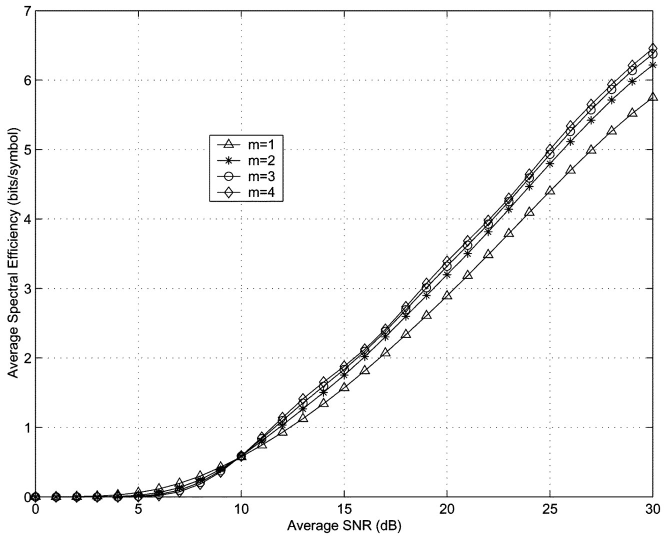

Fig. 7. Average spectral efficiency for TM1 versus average SNR.

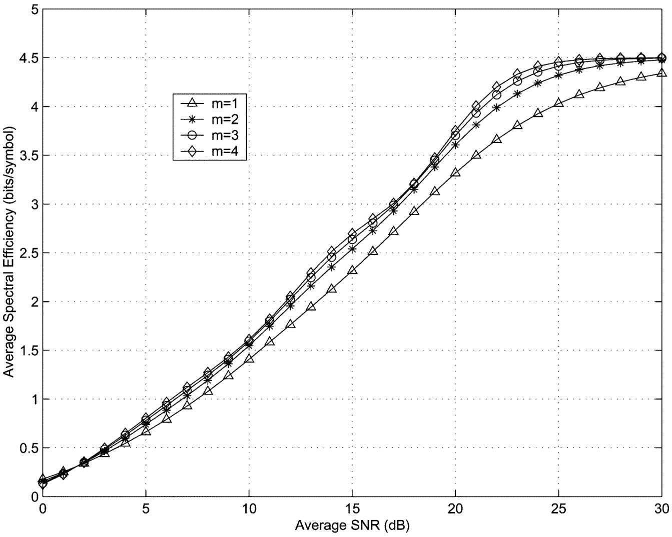

Fig. 8. Average spectral efficiency for TM2 versus average SNR.

From Figs. 4 and 5, we observe that the spectral efficiency gain when combining AMC with truncated ARQ exceeds that of the AMC-only scheme by about 0.25 b/transmitted symbol, using only one retransmission ($N_r^{\max} = 1$) for both TM1 and TM2. This implies a significant rate enhancement. For example, in the HIPERLAN/2 standard, where the symbol rate is 12 Msymbols/s [5], combining AMC with truncated ARQ leads to an approximate 3 Mb/s increase in transmission rate with only one retransmission. Due to the FEC advantage, TM2 has better error performance than TM1 for the same data rate, which results in higher spectral efficiency than TM1 for low and moderate average SNR. However, at high average SNR, TM1 achieves higher spectral efficiency than TM2, because its corresponding modes support higher data rates. The highest rate mode has $R_n = 7$ b/symbol in TM1, which is much greater than $R_n = 4.50$ b/symbol in TM2. This means that adopting high-rate modes benefits spectral efficiency at high average SNR. To improve spectral efficiency over the entire SNR range, a practical system could also optimally combine uncoded and coded transmission modes from TM1 and TM2.

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 gain. They incur smaller delay and buffer-size penalties, and thus lead to improved delay-throughput tradeoffs. For instance, $N_r^{\max} = 1$ yields the highest gain, while $N_r^{\max} = 2$ is highly recommended for practical systems.

The average packet error rate at the physical layer is depicted in Fig. 6. It explains why increasing $N_r^{\max}$ improves spectral efficiency. 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 (Dependence on the Nakagami fading parameter m). We set $P_{\text{loss}} = 0.01$, $N_r^{\max} = 0$ (no retransmission), and vary m from 1 to 4. The average spectral efficiencies are plotted in Figs. 7 and 8 for TM1 and TM2, respectively.

As established in [17], a Nakagami-m fading channel provides a diversity order of m. This is intuitively true, because the Nakagami-m fading channel is equivalent, after maximum ratio combining (MRC), to a set of m independent Rayleigh fading channels. As the channel diversity order increases, the spectral efficiency also improves with AMC at the physical layer [3]. From Figs. 7 and 8, we confirm that the spectral efficiency gain increases with the diversity order m. However, the increment of spectral efficiency decreases as m grows large, which is similar to what we have observed in Test Case 1 by varying $N_r^{\max}$.

Comparing Fig. 4 with Fig. 7 for TM1, and Fig. 5 with Fig. 8 for TM2, one recognizes that increasing the number of retransmissions $N_r^{\max}$ and the diversity order m by the same amount, renders the spectral efficiency gain by $N_r^{\max}$ comparable to that offered by increasing m. Hence, increasing $N_r^{\max}$ has similar impact on spectral efficiency as increasing the diversity order m.

Although $N_r^{\max}$ and m are parameters of different layers (link and physical, respectively), this comparison provides an alternative means of corroborating the advantages of combining AMC with ARQ relative to AMC-only.

Test Case 3 (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 TM1 and TM2 are depicted in Figs. 4 and 5, respectively. 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.

VI. Conclusion and Future Work

In this paper, we developed a cross-layer design, which combines adaptive modulation and coding at the physical layer with truncated ARQ at the data link layer, in order to maximize system throughput under prescribed delay and performance constraints. We derived a closed-form expression of the average spectral efficiency for packets transmitted over Nakagami-m block fading channels. Numerical results demonstrated the rate improvement of our cross-layer design over AMC alone, as well as ARQ with fixed modes. Retransmissions alleviate stringent error-control requirements on modulation and coding, and bring considerable spectral efficiency gain. This gain is comparable to that offered by diversity. 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 in practice.

Although they benchmark theoretical performance, the assumptions that perfect CSI is available at the receiver, and that the feedback channel has zero delay and is error free, may not always hold true. One possible extension of this work is to design and analyze our cross-layer design with imperfect CSI at the transmitter. A single-user link with single-transmit and single-receive antennas was considered in this paper. Generalizations to multiuser links with multitransmit multireceive antennas could also be investigated. Finally, the impact of our design on other parameters at the physical and higher layers is also worth pursuing.

Appendix
PER Approximation

If each bit inside the packet has the same BER and bit-errors are uncorrelated, the PER can be related to the BER through

$$\mathrm{PER} = 1 - (1 - \mathrm{BER})^{N_p} \tag{20}$$

for a packet containing N_p bits. However, the information bits incur different error probabilities for large-size QAM constellations [20]. Furthermore, bit-errors are usually correlated when coded transmissions are decoded in the maximum-likelihood (ML) sense. For these reasons, the PER evaluation through BER in (20) may not be always accurate. We next evaluate the exact PER for our transmission modes in TM1 and TM2, compare it with that in (20), and with the exponential approximation in (5).

A. Uncoded Modulations

We first derive an exact closed-form PER expression for the uncoded modulation modes in TM1. The exact PER for M_n-QAM transmissions over AWGN channels is found based on the approach of calculating the exact BER in [20].

An arbitrary $M_n = I \times J$ rectangular QAM can be viewed as two independent pulse amplitude modulation (PAM), i.e., I-ary and J-ary PAM's, through two quadrature branches. Gray bit-to-symbol mapping is assumed and all symbols are taken to be equally likely. The error probability of the kth bit in I-ary PAM, where $k \in \{1, 2, \ldots, \log_2 I\}$, is obtained from [20, eq. (20)]

$$\begin{aligned} P_I(k) = {} & \frac{1}{I} \\ & \times \sum_{i=0}^{(1-2^{-k})I-1} \Bigg\{ (-1)^{\lfloor (i \cdot 2^{k-1}/I) \rfloor} \\ & \qquad \times \left[2^{k-1} - \left\lfloor \frac{i \cdot 2^{k-1}}{I} + \frac{1}{2} \right\rfloor \right] \\ & \qquad \times \operatorname{erfc}\left((2i+1) \sqrt{\frac{3 \log_2(I \cdot J) \cdot \gamma_b}{I^2 + J^2 - 2}} \right) \Bigg\} \end{aligned} \tag{21}$$

where $\lfloor x \rfloor$ denotes integer floor of x, and $\gamma_b := E_b/N_0$ is the bit signal-to-noise ratio. Similarly, the bit error probability of the lth bit in J-ary PAM, where $l \in \{1, 2, \ldots, \log_2 J\}$, is obtained from [20, eq. (21)] as

$$\begin{aligned} P_J(l) = {} & \frac{1}{J} \\ & \times \sum_{j=0}^{(1-2^{-l})J-1} \Bigg\{ (-1)^{\lfloor (j \cdot 2^{l-1}/J) \rfloor} \\ & \qquad \times \left[2^{l-1} - \left\lfloor \frac{j \cdot 2^{l-1}}{J} + \frac{1}{2} \right\rfloor \right] \\ & \qquad \times \operatorname{erfc}\left((2j+1) \sqrt{\frac{3 \log_2(I \cdot J) \cdot \gamma_b}{I^2 + J^2 - 2}} \right) \Bigg\}. \end{aligned} \tag{22}$$

All information bits in I-ary and J-ary PAMs are equally likely, and independent from each other. Since each packet of N_p bits is mapped to $N_p / \log_2(I \cdot J)$ symbols, as depicted in Fig. 3, the exact PER with rectangular QAM symbols can be obtained as

$$\begin{aligned} \mathrm{PER}_{\mathrm{QAM}} = {} & 1 - \prod_{k=1}^{\log_2 I} [1 - P_I(k)]^{(N_p/\log_2(I \cdot J))} \\ & \times \prod_{l=1}^{\log_2 J} [1 - P_J(l)]^{(N_p/\log_2(I \cdot J))}. \end{aligned} \tag{23}$$

Note that when $M_n = 2$ for binary phase shift keying (BPSK) and $M_n = 4$ for quarternary phase shift keying (QPSK), all bits

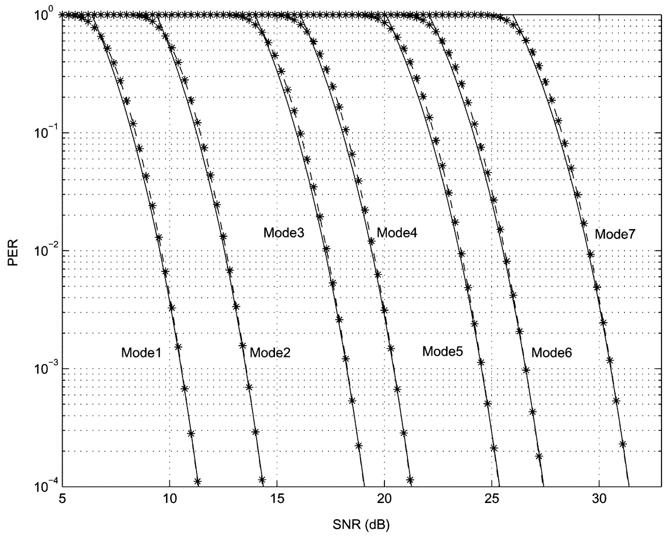

Fig. 9. Packet error rate of the transmission modes in TM1 (stars denote exact PER, solid lines are fitting curves to the exact PER, and dashed lines depict PER based on BER).

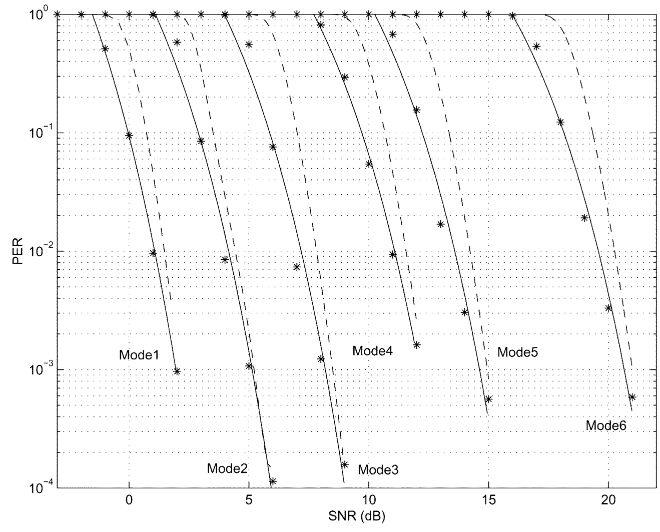

Fig. 10. Packet error rate of the transmission modes in TM2 (stars denote exact PER, solid lines are fitting curves to the exact PER, and dashed lines depict PER based on BER).

have the same BER, and thus (23) coincides with (20) in these two cases.

By least-squares fitting the approximate PER in (5) to the exact PER in (23), we can obtain the (a_n, g_n, γ_{pn}) parameters of each mode n. Equation (23) shows that the PER is dependent on the packet length N_p. With $N_p = 1080$, we compare in Fig. 9 the exact PER in (23), the exponential approximation in (5), and the PER in (20) based on the exact average BER in [20]. The fitting parameters a_n, g_n, and γ_{pn} for transmission modes in TM1 are listed in Table I. It is clear from Fig. 9, that these three PER expressions are close to each other, which justifies the usage of (20) for uncoded transmissions even with large QAM constellations. However, this is not the case for coded transmissions.

B. Convolutionally Coded Modulations

We will focus on transmission modes with convolutionally coded modulations from TM2. Exact closed-form expressions for PER and BER are not available. We, hence, obtain the exact PER and BER through Monte Carlo simulations. With $N_p = 1080$, the simulated PER, the fitting curves to the simulated PER, and the PER in (20) derived from the simulated BER, are compared in Fig. 10. We deduce that the PER in (20) is no longer as accurate, primarily due to bursty errors that emerge with ML sequence detection. The fitting parameters for TM2 are listed in Table II.

Evident from Figs. 9 and 10, the PERs in (5) approximate well the exact PERs for both transmission modes TM1 and TM2. This approximate PER expression simplifies the AMC design, and facilitates the performance analyzes of III and IV, respectively.

References

[1] "Physical layer aspects of UTRA high speed downlink packet access (release 4)," 3GPP TR 25.848 V4.0.0, 2001.

[2] "Error resilience in real-time packet multimedia payloads," 3GPP TSG-S4 Codec Working Group, 1999.

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

[4] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1273–1289, May 2001.

[5] 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. Mag.*, vol. 40, pp. 172–180, May 2002.

[6] D. L. Goeckel, "Adaptive coding for time-varying channels using outdated fading estimates," *IEEE Trans. Commun.*, vol. 47, pp. 844–855, June 1999.

[7] A. J. Goldsmith and S.-G. Chua, "Adaptive coded modulation for fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 595–602, May 1998.

[8] ——, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, pp. 1218–1230, Oct. 1997.

[9] K. J. Hole, H. Holm, and G. E. Oien, "Adaptive multidimensional coded modulation over flat fading channels," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1153–1158, July 2000.

[10] 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, pp. 1477–1487, Sept. 2000.

[11] E. Malkamaki, D. Mathew, and S. Hamalainen, "Performance of hybrid ARQ techniques for WCDMA high data rates," in *Proc. Vehicular Technology Conf.*, vol. 4, Rhodes Island, Greece, Spring 2001, pp. 2720–2724.

[12] H. Minn, M. Zeng, and V. K. Bhargava, "On ARQ scheme with adaptive error control," *IEEE Trans. Veh. Technol.*, vol. 50, pp. 1426–1436, Nov. 2001.

[13] G. E. Oien, H. Holm, and K. J. Hole, "Adaptive coded modulation with imperfect channel state information: System design and performance analysis aspects," in *Proc. IEEE Int. Symp. Advances in Wireless Communications*, Victoria, BC, Canada, Sept. 23–24, 2002, pp. 19–20.

[14] M. B. Pursley and J. M. Shea, "Adaptive nonuniform phase-shift-key modulation for multimedia traffic in wireless networks," *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1394–1407, Aug. 2000.

[15] G. L. Stüber, *Principles of Mobile Communication*, 2nd ed. Norwell, MA: Kluwer, 2001.

[16] T. Ue, S. Sampei, N. Morinaga, and K. Hamaguchi, "Symbol rate and modulation level-controlled adaptive modulation/TDMA/TDD system for high-bit-rate wireless data transmission," *IEEE Trans. Veh. Tech.*, vol. 47, pp. 1134–1147, Nov. 1998.

[17] Z. Wang and G. B. Giannakis, "A simple and general parameterization quantifying performance in fading channels," *IEEE Trans. Commun.*, vol. 51, pp. 1389–1398, Aug. 2003.

[18] W. T. Webb and R. Steele, "Variable rate QAM for mobile radio," *IEEE Trans. Commun.*, vol. 43, pp. 2223–2230, July 1995.

[19] S. B. Wicker, *Error Control Systems for Digital Communication and Storage*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[20] D. Yoon and K. Cho, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Trans. Commun.*, vol. 50, pp. 1074–1080, July 2002.

Qingwen Liu (S'04) received the B.S. degree in electrical engineering and information science from the University of Science and Technology of China, Hefei, in 2001. He is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, the University of Minnesota, Minneapolis.

His research interests include the areas of communications, signal processing, and networking, with emphasis on cross-layer analysis and design, quality of service support for multimedia applications over wired-wireless networks, and resource allocation.

Shengli Zhou (M'03) received the B.S. and M.Sc. degrees in electrical engineering and information science from the University of Science and Technology of China, Hefei, in 1995 and 1998, respectively, and the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis, in 2002.

He has been an Assistant Professor with the Department of Electrical and Computer Engineering, the University of Connecticut, Storrs, since 2003. His research interests lie the areas of communications and signal processing, including channel estimation and equalization, multiuser and multicarrier communications, space–time coding, adaptive modulation, and cross-layer designs.

Georgios B. Giannakis (S'84–M'86–SM'91–F'97) received the Diploma in electrical engineering from the National Technical University of Athens, Greece, 1981, and the MSc. degree in electrical engineering, the MSc. degree in mathematics, and the Ph.D. degree in electrical engineering, from the University of Southern California (USC), Los Angeles, in 1983, 1986, and 1986, respectively.

After lecturing for one year at USC, he joined the University of Virginia, Charlottesville, in 1987, where he became a Professor of Electrical Engineering in 1997. Since 1999, he has been a Professor with the Department of Electrical and Computer Engineering at the University of Minnesota, Minneapolis, where he now holds an ADC Chair in Wireless Telecommunications. His general interests span the areas of communications and signal processing, estimation and detection theory, time-series analysis, and system identification—subjects on which he has published more than 200 journal papers, 350 conference papers, and two edited books. His current research focuses on transmitter and receiver diversity techniques for single- and multiuser fading communication channels, complex-field and space–time coding, multicarrier, ultrawideband wireless communication systems, cross-layer designs, and distributed sensor networks.

Dr. Giannakis is the (co-) recipient of six best paper awards from the IEEE Signal Processing (SP) Society (1992, 1998, 2000, 2001, 2003, 2004). He also received the Society's Technical Achievement Award in 2000. He has served as Editor in Chief for the IEEE SIGNAL PROCESSING LETTERS, as Associate Editor for theIEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS, as Secretary of the SP Conference Board, as Member of the SP Publications Board, as Member and Vice-Chair of the Statistical Signal and Array Processing Technical Committee, as Chair of the SP for Communications Technical Committee, and as a Member of the IEEE Fellows Electrion Committee. He has also served as a member of the IEEE-SP Society's Board of Governors, the Editorial Board for the PROCEEDINGS OF THE IEEE, and the steering committee of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.