Queuing With Adaptive Modulation and Coding Over Wireless Links: Cross-Layer Analysis and Design

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

Abstract—Assuming there are always sufficient data waiting to be transmitted, adaptive modulation and coding (AMC) schemes at the physical layer have been traditionally designed separately from higher layers. However, this assumption is not always valid when queuing effects are taken into account at the data link layer. In this paper, we analyze the joint effects of finite-length queuing and AMC for transmissions over wireless links. We present a general analytical procedure, and derive the packet loss rate, the average throughput, and the average spectral efficiency (ASE) of AMC. Guided by our performance analysis, we introduce a cross-layer design, which optimizes the target packet error rate of AMC at the physical layer, to minimize thpacket loss rate and maximize the average throughput, when combined with a finite-length queue at the data link layer. Numerical results illustrate the dependence of system performance on various parameters, and quantify the performance gain due to cross-layer optimization. Our focus is on the single user case, but we also discuss briefly possible applications to multiuser scenarios.

Index Terms—Adaptive modulation and coding (AMC), cross-layer design, discrete time Markov chain, quality of service (QoS), queuing analysis, wireless networks.

I. INTRODUCTION

IN MULTIMEDIA wireline-wireless communication networks, the demand for high data rates and quality of service is growing at a rapid pace. The "bottleneck" in such networks is the wireless link, not only because wireless resources (bandwidth and power) are more scarce and expensive relative to their wireline counterparts, but also because the overall system performance degrades markedly due to multipath fading, Doppler, and time-dispersive effects introduced by the wireless propagation. In order to enhance the spectral efficiency while adhering to a target error performance over wireless channels, adaptive modulation and coding (AMC) has been widely used to match transmission parameters to time-varying channel conditions [4], [7], [9], [10], [16]. Due to its attractive rate and error performance characteristics, AMC has been adopted at the physical layer of several standards, e.g., 3GPP, 3GPP2, HIPERLAN/2, IEEE 802.11a, IEEE 802.15.3 and IEEE 802.16 [1]–[3], [8], [11].

However, existing AMC schemes rely on the salient assumption that data are continuously available at the transmitter: whatever modulation-coding modes (pairs) are chosen at the physical layer to match the wireless channel, there are sufficient data waiting to be transmitted in the queues (buffers) at the data link layer. However, in practical communication systems with randomly arriving data streams, the queues may be empty from time to time, even though the wireless channel can accommodate transmissions. For example, data streams from wide-area networks (WAN) arrive according to a Poisson or a Bernoulli distributed process [5], [20]. This leads to a dynamic behavior of the queue, the impact of which on AMC has not been investigated so far.

On the other hand, the service process of the queue feeding an AMC module, is no longer deterministic, which is not the case with nonadaptive modulations. Indeed, the service process is affected by the wireless medium, and depends on how the AMC module adapts its parameters to channel variations. Furthermore, the queue length (buffer size) is finite in practice. When the buffer is full and overflow occurs, excess packets have to be dropped. The effect of overflow also needs to be taken into account when analyzing the overall system performance.

The interaction of queuing at the data link layer with AMC at the physical layer provides interesting design problems. Consider for instance the packet loss rate of the end-to-end system, that is defined as the ratio of the number of incorrectly received packets at the destination over those transmitted from the source. The packet loss rate is affected by both the queuing overflow, and the packet reception error. Designing protocols with either overflow or reception error criteria only, may not lead to the "overall best" performance. It is thus important to optimize these two components jointly across layers.

In this paper, we analyze the joint effect of finite-length queuing and AMC. We first characterize the queuing service process dictated by AMC, and derive the queue state recursion. Then, we construct a finite state Markov chain (FSMC) with a state pair containing both the queue and the queue server states, and compute its stationary distribution. The latter enables us to obtain the packet loss rate, the average throughput, and the average spectral efficiency (ASE) of AMC. Based on this performance analysis, we develop a cross-layer design, which optimizes the target packet error rate of AMC at the physical layer, to minimize the system packet loss rate and maximize the average throughput, when combined with a finite-length queue at the data link layer. Numerical results illustrate the

Q. Liu and G. B. Giannakis are with the Department of Electrical Engineering, University of Minnesota, Minneapolis, MM 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).
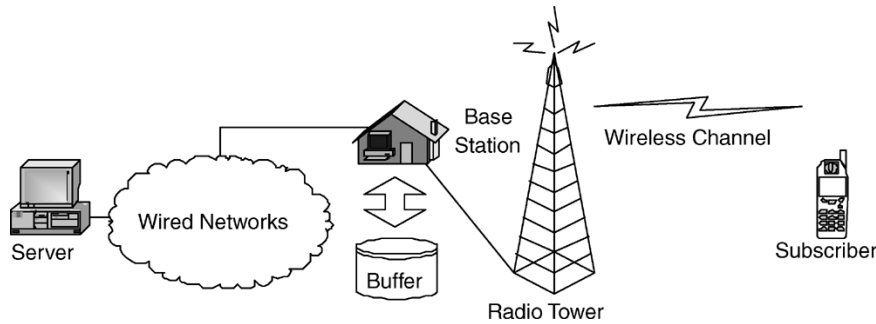
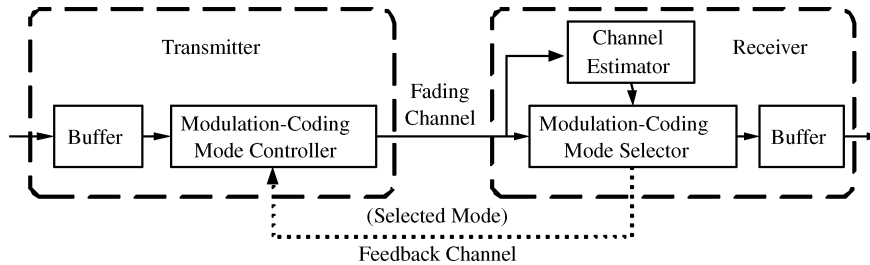Fig. 1.   End-to-end wired-wireless connection.

Fig. 2.   Wireless link with combined queuing and AMC.
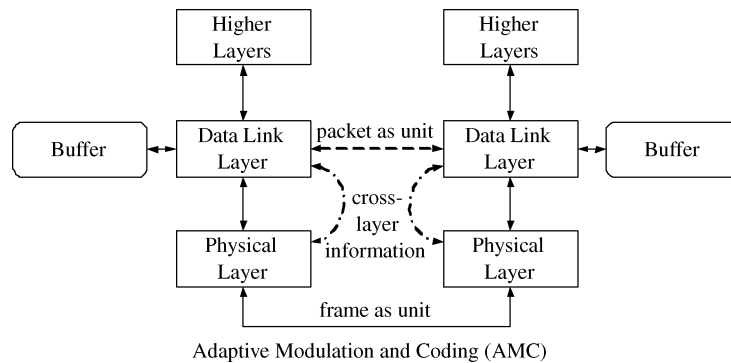

Fig. 3.   Cross-layer structure combining AMC with queuing.

dependence of system performance on various parameters, and quantify the performance gain due to cross-layer optimization.

As in [4], [7], [9], [10], and [16] for conventional AMC designs, we focus on the single user case, but we also touch upon possible applications of our analytical approach to multiuser scenarios.

The rest of this paper is organized as follows. We introduce the system model in Section II. We present an analytical procedure to investigate the performance of the combined queuing with AMC in Section III, and derive a cross-layer design paradigm in Section IV. We then discuss possible extensions in Section V, collect numerical results in Section VI, and finally draw concluding remarks in Section VII.

II. SYSTEM MODEL

A. System Description

Fig. 1 illustrates an end-to-end connection between a server (source) and a subscriber (destination), which includes a wireless link with a single-transmit and a single-receive antenna. We focus on the downlink here, although our results are applicable to the uplink as well. As depicted in Fig. 2, a finite-length

queue (buffer) is implemented at the transmitter, and operates in a first-in-first-out (FIFO) mode. The queue feeds the AMC controller at the transmitter. The AMC selector is implemented at the receiver. The layer structure of the system under consideration is shown in Fig. 3. 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 consists of multiple transmitted symbols. The packet and frame structures will be detailed soon.

We assume that multiple transmission modes are available, with each mode representing a pair of a specific modulation format, and a forward error correcting (FEC) code, as in the HIPERLAN/2 and the IEEE 802.11a standards. Based on channel state information (CSI) estimated at the receiver, the AMC selector determines the modulation-coding pair (mode), which is sent back to the transmitter through a feedback channel, for the AMC controller to update the transmission mode. Coherent demodulation and maximum-likelihood (ML) decoding are employed at the receiver. The decoded bit streams are mapped to packets, which are pushed upwards to layers above the physical layer.

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

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5
Modulation	BPSK	QPSK	8-QAM	16-QAM	32-QAM
R_n (bits/sym.)	1	2	3	4	5
a_n	67.7328	73.8279	58.7332	55.9137	50.0552
g_n	0.9819	0.4945	0.1641	0.0989	0.0381
γ_{pn}(dB)	6.3281	9.3945	13.9470	16.0938	20.1103

TABLE II
TRANSMISSION MODES IN TM2 WITH CONVOLUTIONALLY
CODED MODULATION

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

We consider the following group of transmission modes.

TM1: Uncoded (without FEC) M_n-ary rectangular or square QAM modes, where $M_n = 2^n$, with $n = 1, 2, \ldots, 5$ [26].

TM2: Convolutionally coded M_n-ary rectangular or square QAM modes, adopted from the HIPERLAN/2 or the IEEE 802.11a standards [8].

The transmission modes in TM1 and TM2 are listed in Tables I and II, respectively, in a rate ascending order. Although we will focus on TM1 and TM2, other transmission modes can be similarly constructed [1]–[3], [11].

At the *physical layer*, we deal with frame by frame transmissions, where each frame contains a fixed number of symbols (N_s). Given a fixed symbol rate as in [1], [8], the frame duration (T_f seconds) is constant, and represents the time-unit throughout this paper. Each frame at the physical layer may contain one or more packets from the data link layer. The packet and frame structures are depicted in Fig. 4. Each packet contains a fixed number of bits (N_b), which include packet header, payload, and cyclic redundancy check (CRC) bits. After modulation and coding with mode n of rate R_n (bits/symbol), each packet is mapped to a symbol-block containing N_b/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 the HIPERLAN/2 and the IEEE 802.11a standards [8]. If mode n is used, it follows that the number of symbols per frame is $N_s = N_c + N_p N_b/R_n$, which implies that N_p (the number of packets per frame) depends on the chosen mode. The queue at the *data link layer* has finite-length (capacity) of K packets.

We next list all our operating assumptions.

1) 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 model, which is suitable for slowly-varying wireless channels. As a consequence, AMC is adjusted on a frame-by-frame basis.
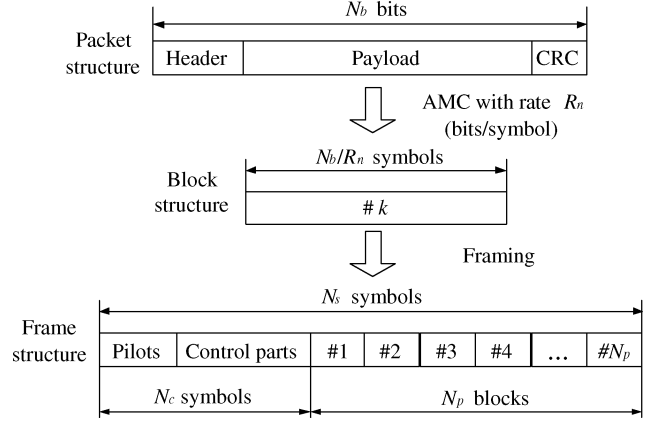

Fig. 4. Packet and frame structures.

2) 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 assumed also in [4], [9].

 The assumption that the feedback channel is error free and has no latency, could be (at least approximately) satisfied by using coded and fast feedback channels. On the other hand, the effects of imperfect CSI on AMC have been investigated in [4], [15], [29], and suggest further practical considerations that go beyond the scope of this paper.

3) If the queue is full, the additional arriving packets will be dropped, and will not be recovered by end-to-end (server-to-subscriber), or, link-layer retransmissions.

 This can be afforded by the user datagram protocol (UDP) for instance, where retransmission can be denied when delay or buffer-size constraints are violated [28].

4) Error detection based on CRC is perfect, provided that sufficiently reliable error detection CRC codes are used [14]. As in [14], the packet header and the CRC parity bits per packet are not included in the throughput calculation, because they introduce negligible redundancy relative to the number of payload bits.

 Although no retransmission is provided per A3, UDP employs CRC to verify the integrity of packets, and detect errors in the packet header or payload [28].

5) If a packet is not received correctly at the receiver after error detection, we drop it, and declare packet loss as in [28].

 Assumption A5 is reasonable and can be afforded by e.g., UDP-based video transmissions, because the underlying bit streams represent highly correlated contents.

For flat fading channels adhering to A1, the channel quality can be captured by a single parameter, namely the received signal-to-noise ratio (SNR) γ. Since the channel varies from frame to frame, we adopt the general Nakagami-m model to describe γ statistically [21]. 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, $\Gamma(m) := \int_0^\infty t^{m-1} \exp(-t)dt$ is the Gamma function, and m is the Nakagami fading parameter ($m \geq 1/2$). We choose the Nakagami-m channel model because it encompasses a large class of fading channels, e.g., it includes the Rayleigh channel as a special case when $m = 1$. An one-to-one mapping between the Ricean factor and the Nakagami fading parameter m allows also Ricean channels to be well approximated by Nakagami-m channels [21].

B. AMC

The objective of AMC is to maximize the data rate by adjusting transmission parameters to the available CSI, while maintaining a prescribed packet error rate P_0. Let N denote the total number of transmission modes available ($N = 5$ for both TM1 and TM2). As in [4], we assume constant power transmission, and partition the entire SNR range into $N + 1$ nonoverlapping consecutive intervals, with boundary points denoted by $\{\gamma_n\}_{n=0}^{N+1}$. Specifically

$$\text{mode } n \text{ is chosen when } \gamma \in [\gamma_n, \gamma_{n+1}). \tag{2}$$

To avoid deep channel fades, no data are sent when $\gamma_0 \leq \gamma < \gamma_1$, which corresponds to the mode $n = 0$ with rate $R_0 = 0$ (bits/symbol). The design objective for AMC is to determine the boundary points $\{\gamma_n\}_{n=0}^{N+1}$.

For simplicity, we approximate the packet error rate (PER) in the presence of additive white Gaussian noise (AWGN), as [12, eq. (5)]

$$\text{PER}_n(\gamma) \approx \begin{cases} 1, & \text{if } 0 < \gamma < \gamma_{pn} \\ a_n \exp(-g_n\gamma), & \text{if } \gamma \geq \gamma_{pn} \end{cases} \tag{3}$$

where n is the mode index, γ is the received SNR, and the mode-dependent parameters a_n, g_n, and γ_{pn} are obtained by fitting (3) to the exact PER.[1] With packet length $N_b = 1,080$, the fitting parameters for transmission modes TM1 and TM2 are provided in Tables I and II, respectively [12]. Based on (1) and (2), the mode n will be chosen with probability [4, eq. (34)]

$$\begin{aligned} \Pr(n) &= \int_{\gamma_n}^{\gamma_{n+1}} p_\gamma(\gamma)d\gamma \\ &= \frac{\Gamma\left(m, \frac{m\gamma_n}{\bar{\gamma}}\right) - \Gamma\left(m, \frac{m\gamma_{n+1}}{\bar{\gamma}}\right)}{\Gamma(m)} \end{aligned} \tag{4}$$

where $\Gamma(m, x) := \int_x^\infty t^{m-1} \exp(-t)dt$ is the complementary incomplete Gamma function. Let $\overline{\text{PER}}_n$ denote the average PER corresponding to mode n. In practice, we have $\gamma_n > \gamma_{pn}$ and thus obtain $\overline{\text{PER}}_n$ in closed-form as (cf. [4, eq.(37)])

$$\begin{aligned} \overline{\text{PER}}_n &= \frac{1}{\Pr(n)} \int_{\gamma_n}^{\gamma_{n+1}} a_n \exp(-g_n\gamma)p_\gamma(\gamma)d\gamma \\ &= \frac{1}{\Pr(n)} \frac{a_n}{\Gamma(m)} \left(\frac{m}{\bar{\gamma}}\right)^m \frac{\Gamma(m, b_n\gamma_n) - \Gamma(m, b_n\gamma_{n+1})}{(b_n)^m}, \\ &\quad n = 1, \ldots, N \end{aligned} \tag{5}$$

where $b_n := m/\bar{\gamma} + g_n$. The average PER of AMC can then be computed as the ratio of the average number of packets in error over the total average number of transmitted packets [4]

$$\overline{\text{PER}} = \frac{\sum_{n=1}^{N} R_n \Pr(n) \overline{\text{PER}}_n}{\sum_{n=1}^{N} R_n \Pr(n)}. \tag{6}$$

We want to find the thresholds $\{\gamma_n\}_{n=0}^{N+1}$, so that the prescribed P_0 is achieved for each mode: $\overline{\text{PER}}_n = P_0$, which naturally leads to $\overline{\text{PER}} = P_0$ based on (6). Given P_0, $\bar{\gamma}$ and m, the following threshold searching algorithm determines $\{\gamma_n\}_{n=0}^{N+1}$, and guarantees that $\overline{\text{PER}}_n$ is exactly P_0.

Step 1)   Set $n = N$, and $\gamma_{N+1} = +\infty$.
Step 2)   For each n, search for the unique $\gamma_n \in [0, \gamma_{n+1}]$ that satisfies

$$\overline{\text{PER}}_n = P_0. \tag{7}$$

Step 3)   If $n > 1$, set $n = n - 1$, and go to Step 2; otherwise, go to Step 4.
Step 4)   Set $\gamma_0 = 0$.

The SNR region $[\gamma_n, \gamma_{n+1})$ corresponding to transmission mode n constitutes the channel state indexed by n. To describe the transition of these channel states, we rely on a FSMC model, which we develop next.

C. FSMC Channel Model

As in [17], we adopt an FSMC channel model to analyze the performance of our system. Assuming slow fading conditions so that transition happens only between adjacent states, the probability of transition exceeding two consecutive states is zero [17], [23], [27], i.e.,

$$P_{l,n} = 0, \quad |l - n| \geq 2. \tag{8}$$

The adjacent-state transition probability can be determined by [17, eqs. (5) and (6)]:

$$\begin{aligned} P_{n,n+1} &= \frac{N_{n+1}T_f}{\Pr(n)}, \quad \text{if } n = 0, \ldots, N-1 \\ P_{n,n-1} &= \frac{N_n T_f}{\Pr(n)}, \quad \text{if } n = 1, \ldots, N \end{aligned} \tag{9}$$

where N_n is the cross-rate of mode n (either upward or downward), which can be estimated as [25, eq. (17)]

$$N_n = \sqrt{2\pi \frac{m\gamma_n}{\bar{\gamma}}} \frac{f_d}{\Gamma(m)} \left(\frac{m\gamma_n}{\bar{\gamma}}\right)^{m-1} \exp\left(-\frac{m\gamma_n}{\bar{\gamma}}\right) \tag{10}$$

where f_d denotes the mobility-induced Doppler spread. The probability of staying at the same state n is [23]

$$P_{n,n} = \begin{cases} 1 - P_{n,n+1} - P_{n,n-1}, & \text{if } 0 < n < N \\ 1 - P_{0,1}, & \text{if } n = 0 \\ 1 - P_{N,N-1}, & \text{if } n = N. \end{cases} \tag{11}$$

---

[1]A similar approximation was adopted in [10] but for the bit error rate.

In summary, we model the channel as a FSMC with an $(N + 1) \times (N + 1)$ state transition matrix, which is banded as

$$\mathbf{P}_c = \begin{bmatrix} P_{0,0} & P_{0,1} & \cdots & & 0 \\ P_{1,0} & P_{1,1} & P_{1,2} & & \vdots \\ 0 & \ddots & \ddots & & 0 \\ \vdots & P_{N-1,N-2} & P_{N-1,N-1} & P_{N-1,N} \\ 0 & \cdots & & P_{N,N-1} & P_{N,N} \end{bmatrix}. \quad (12)$$

III. PERFORMANCE ANALYSIS

In this section, we will quantify the joint effects of finite-length queuing and AMC. The end-to-end performance in systems including both wired and wireless connections is usually dominated by the performance over the wireless link. As figures of merit, we will focus on the packet loss rate and the packet throughput, for transmissions over the wireless link.

Due to finite-length queuing, arriving packets will be dropped when the queue is full. Let P_d denote the packet dropping (overflow or blocking) probability upon queuing. A packet from the source is correctly received by the subscriber, only if it is not dropped from the queue (with probability $1 - P_d$), and it is correctly received through the wireless channel (with probability $1 - P_0$). Hence, the probability of a packet received correctly is $(1 - P_d)(1 - P_0)$, and the packet loss rate can be expressed as

$$\xi = 1 - (1 - P_d)(1 - P_0). \quad (13)$$

The average throughput can then be evaluated as:

$$\eta = (\lambda T_f)(1 - \xi) = (\lambda T_f)(1 - P_d)(1 - P_0) \quad (14)$$

where λT_f is the packet arrival rate upon the queue at the base station. Hence, to evaluate the system performance in terms of packet loss rate and average throughput, the key is to find P_d.

In order to obtain P_d, we pursue the following queuing analysis.

A. Queuing Analysis

We will subsequently model and analyze the queuing arrival process, the service process and the queue state recursion. We will then construct an enlarged FSMC containing both the queue and queue server states and derive the joint stationary distribution.

1) Arrival Process: Let t index the time units and A_t denote the number of packets arriving at time t. The process A_t is stationary with $E\{A_t\} = \lambda T_f$ and independent of the queue state as well as the channel state. For simplicity, we assume that A_t is Poisson distributed with parameter λT_f [5, pp. 164]

$$P(A_t = a) = \begin{cases} \frac{(\lambda T_f)^a \exp(-\lambda T_f)}{a!}, & \text{if } a \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

which leads to $E\{A_t\} = \lambda T_f$. Then, we have $A_t \in \mathcal{A} := \{0, 1, \ldots, \infty\}$.

2) Queue Service Process: Different from nonadaptive modulations, AMC dictates a dynamic, rather than deterministic, service process for the queue, with a variable number of packets transmitted per time unit. Let C_t (packets/time-unit) denote the number of packets transmitted using AMC at
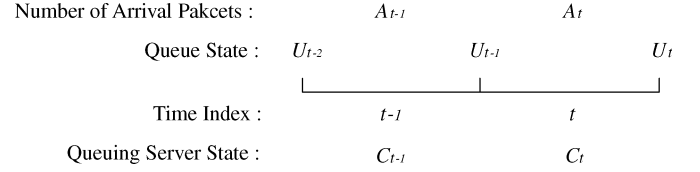

Fig. 5. Recursive queuing model.

time t. Corresponding to each transmission mode n, let c_n (packets/time-unit) denote the number of packets transmitted per time-unit. We then have

$$C_t \in \mathcal{C}, \qquad \mathcal{C} := \{c_0, c_1, \ldots, c_N\} \quad (16)$$

where c_n takes positive integer values (see also Fig. 4). Suppose that for the rate $R = 1$ transmission mode (e.g., Mode 1 in TM1, or, Mode 2 in TM2), a total of b packets can be accommodated per frame, i.e.,

$$c_n = bR_n.$$

Parameter b depends on the system resource allocated per user, which is up to the designer's choice.

As specified in (16), the AMC module yields a queue server with a total of $N + 1$ states $\{c_n\}_{n=0}^N$, with the service process C_t representing the evolution of server states. Since the AMC mode n is chosen when the channel enters the state n, we model the service process C_t as a FSMC with transition probability matrix given by (12).

3) Queue State Recursion: Having modeled the queue service process, we now focus on the queue itself. Let U_t denote the queue state (the number of packets in the queue) at the end of time-unit t, or, at the beginning of time-unit $t + 1$, as shown in Fig. 5. It is clear that $U_t \in \mathcal{U} := \{0, 1, \ldots, K\}$.

We assume that the transmitter first moves packets out of the queue at the beginning of time t, based on the server state C_t. Arriving packets are placed in the queue throughout the time t. After moving C_t packets out of the queue, the number of packets left in the queue is

$$L_t = \max\{0, U_{t-1} - C_t\}. \quad (17)$$

The number of free slots in the queue at the beginning of time t is thus

$$F_t = K - L_t = K - \max\{0, U_{t-1} - C_t\}. \quad (18)$$

Let us now focus on the packets arriving at time t. If $A_t \leq F_t$, all arrivals enter the queue and the queue state becomes $U_t = L_t + A_t$. On the other hand, if $A_t > F_t$, only F_t packets enter the queue and the remaining $A_t - F_t$ packets are dropped. The corresponding queue state becomes $U_t = K$. The recursion of the queue state can, therefore, be summarized as follows:

$$U_t = \min\{K, \max\{0, U_{t-1} - C_t\} + A_t\}. \quad (19)$$

Notice that the queue state U_t depends on (U_{t-1}, C_t, A_t) from (19). Since A_t is independent of U_{t-1} and C_t, we can isolate A_t from the state pair (U_{t-1}, C_t), where U_{t-1} and C_t are closely related. In order to analyze the system behavior, we have to construct an augmented FSMC with a state pair (U_{t-1}, C_t) containing both the queue and the server states; a similar approach is taken in [20]. Once the joint stationary distribution of (U_{t-1}, C_t) is found, the stationary distribution of the queue U_{t-1} can be also found as a marginal distribution. However, this
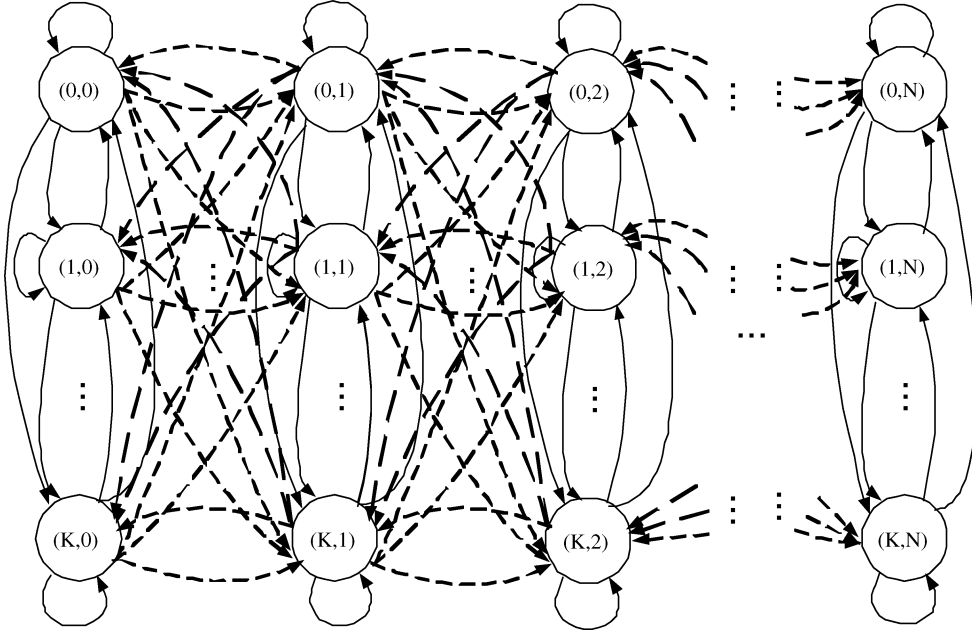
Fig. 6.   Markov chain diagram describing the transition of the pair containing both queue and channel states.

step is not necessary since our desired quantities will be directly evaluated based on the joint distribution.

4) Stationary Distribution: Let (U_{t-1}, C_t) denote the pair of queue and server states, and let $P_{(u,c),(v,d)}$ denote the transition probability from $(U_{t-1} = u, C_t = c)$ to $(U_t = v, C_{t+1} = d)$, where $(u,c) \in \mathcal{U} \times \mathcal{C}$, and $(v,d) \in \mathcal{U} \times \mathcal{C}$. We illustrate the Markov chain diagram in Fig. 6 and organize the state transition probability matrix in a block form

$$\mathbf{P} = \begin{bmatrix} \mathbf{A}_{0,0} & \mathbf{A}_{0,1} & \cdots & \mathbf{A}_{0,K} \\ \mathbf{A}_{1,0} & \mathbf{A}_{1,1} & \cdots & \mathbf{A}_{1,K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{K,0} & \mathbf{A}_{K,1} & \cdots & \mathbf{A}_{K,K} \end{bmatrix} \qquad (20)$$

where the submatrix $\mathbf{A}_{u,v}$ is defined as

$$\mathbf{A}_{u,v} = \begin{bmatrix} P_{(u,c_0),(v,c_0)} & \cdots & P_{(u,c_0),(v,c_N)} \\ \vdots & \ddots & \vdots \\ P_{(u,c_N),(v,c_0)} & \cdots & P_{(u,c_N),(v,c_N)} \end{bmatrix}. \qquad (21)$$

We next simplify $P_{(u,c),(v,d)}$ as

$$\begin{aligned} &P_{(u,c),(v,d)} \\ &:= P(U_t = v, C_{t+1} = d \mid U_{t-1} = u, C_t = c) \\ &= P(C_{t+1} = d \mid C_t = c) P(U_t = v \mid U_{t-1} = u, C_t = c) \\ &= P_{c,d} P(U_t = v \mid U_{t-1} = u, C_t = c) \end{aligned} \qquad (22)$$

where the last equality follows from the fact that C_{t+1} only depends on C_t. As explained in Section III-A2, $P(C_{t+1} = d \mid C_t = c)$ can be found from the entries of $\mathbf{P}_c$ in (12). Based on (19), one can easily verify that

$$\begin{aligned} &P(U_t = v \mid U_{t-1} = u, C_t = c) \\ &= \begin{cases} P(A_t = v - \max\{0, u - c\}), & \text{if } 0 \le v < K \\ 1 - \sum_{0 \le v < K} P(U_t = v \mid U_{t-1} = u, C_t = c), & \text{if } v = K. \end{cases} \end{aligned}$$
$$(23)$$

Combining (21)–(23), we obtain $\mathbf{P}$ in (20).

In Theorem 1 of the Appendix , we prove that the stationary distribution of the FSMC (U_{t-1}, C_t) exists and is unique. Let this stationary distribution be

$$P(U = u, C = c) := \lim_{t \to \infty} P(U_{t-1} = u, C_t = c). \qquad (24)$$

For notational convenience, let also $\pi_{(u,c)} := P(U = u, C = c)$, and define the row vector

$$\boldsymbol{\pi} = \begin{bmatrix} \pi_{(0,c_0)}, \ldots, \pi_{(0,c_N)}, \ldots, \pi_{(K,c_0)}, \ldots, \pi_{(K,c_N)} \end{bmatrix}. \qquad (25)$$

The stationary distribution of (U_{t-1}, C_t) can then be computed from the equality [5], [6]

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}, \qquad \sum_{u \in \mathcal{U}, c \in \mathcal{C}} \pi_{(u,c)} = 1 \qquad (26)$$

which implies that $\boldsymbol{\pi}$ is the left eigenvector of $\mathbf{P}$ corresponding to the eigen-value 1.

Based on the stationary distribution of $\boldsymbol{\pi}$, one can evaluate the desired performance measures. But first we need to show how to derive the packet dropping probability P_d.

B. Packet Dropping Probability and System Performance

Let D_t denote the number of packets dropped at time t. Using (18), we can express D_t as

$$\begin{aligned} D_t &= \max\{0, A_t - F_t\} \\ &= \max\{0, A_t - K + \max\{0, U_{t-1} - C_t\}\} \end{aligned} \qquad (27)$$

which depends on A_t, U_{t-1}, and C_t. We wish to find the stationary behavior of D_t, as $t \to \infty$. Let us define $(U, C) := \lim_{t \to \infty}(U_{t-1}, C_t)$, where the existence of a stationary distribution for the state pair (U_{t-1}, C_t) was established in Section III-A4. Since A_t is stationary, the limiting distribution of A_t exists as $t \to \infty$. Letting $A := \lim_{t \to \infty} A_t$, we have

$$P(A = a) = P(A_t = a) \qquad (28)$$

and $E\{A\} = E\{A_t\} = \lambda T_f$. From (27), the stationary distribution of D_t exists and is given by

$$
\begin{aligned}
D &:= \lim_{t \to \infty} D_t \\
&= \lim_{t \to \infty} \max\{0, A_t - K + \max\{0, U_{t-1} - C_t\}\} \\
&= \max\{0, A - K + \max\{0, U - C\}\}.
\end{aligned}
\tag{29}
$$

Using (24) and (29), the ensemble-average number of packets dropped per time-unit can be calculated as (30), shown at the bottom of the page, where the second equality follows because the arrival process is independent of the queue and queue server states. Based on (28) and (30), we can compute P_d as

$$
P_d := \lim_{T \to \infty} \frac{\sum_{t=1}^{T} D_t}{\sum_{t=1}^{T} A_t} = \frac{E\{D\}}{E\{A\}} = \frac{E\{D\}}{\lambda T_f}
\tag{31}
$$

where to establish the second equality we prove in the Appendix (Theorem 2) that the time-average dropping probability equals its ensemble-average counterpart.

With P_d, P_0 and λT_f available, the performance measures in terms of packet loss rate ξ and average throughput η are now ready in (13) and (14), respectively.

Remark 1: The packet loss rate ξ in (13) is obtained by coupling the FSMC channel model with queuing analysis, with each channel state having an average PER P_0. Notice that each individual packet may experience an instantaneous PER that is quite different from P_0. Taking this fact into account, one can alternatively evaluate the packet loss rate as the ratio of the time-average number of lost packets over the time-average number of arriving packets, where the lost packets include both dropped packets and erroneously received packets. The average number of dropped packets is found in (30), while the average number of erroneously received packets can be found following a similar derivation. This alternative approach leads to the same $\xi = P_d + (1 - P_d)P_0 = 1 - (1 - P_d)(1 - P_0)$, thus verifying (13). We omit the details for the lack of space.

C. ASE

Besides packet loss rate and average throughput, we are also interested in the ASE of AMC when combined with queuing. When $U_{t-1} < C_t$, no sufficient packets are available for transmission. Hence, the ASE of AMC in the presence of queuing should be less than that of traditional AMC schemes assuming always-available data [4], [9]. Specifically, given P_0, f_d, $\bar{\gamma}$, m, b, K, λ and T_f, we can compute the ASE of AMC as (32), shown at the bottom of the page, where $S(U = u, C = c)$ is the spectral efficiency given $(U = u, C = c)$. In conventional AMC schemes, the spectral efficiency for each transmission mode is the number of transmitted bits per symbol [4], [9]. Taking into account queuing effects, we express $S(U = c, C = c)$ as

$$
S(U = u, C = c) = \begin{cases} R_n, & \text{if } c = c_n, u \geq c \\ \frac{u}{c} R_n, & \text{if } c = c_n, u < c. \end{cases}
\tag{33}
$$

The ASE calculation for the AMC in (32) not only depends on the channel state, but also on the queue state, which is different from that of traditional AMC designs. We underscore that the average ASE in (32) depends on various system parameters: P_0, f_d, $\bar{\gamma}$, m, b, K, λ, T_f.

In summary, given a target packet error rate P_0, Doppler spread f_d, average SNR $\bar{\gamma}$, Nakagami parameter m, system resource parameter b, queue length K, packet arrival rate λ and frame length T_f, our performance analysis follows these steps.

1) Determine the boundary points of AMC $\{\gamma_n\}_{n=0}^{N+1}$ by the threshold searching algorithm.
2) Build the transition matrix $\mathbf{P}_c$ in **(12)** for the channel, and the queue server states.
3) Derive the recursion of the finite-length queue with AMC as in **(19)**
4) Construct the FSMC transition matrix $\mathbf{P}$ of the state pair (U_{t-1}, C_t) in **(20)**, and compute its stationary distribution $P(U = u, C = c)$ as in **(24)**.
5) Calculate the ensemble-average number of packets dropped per time-unit $E\{D\}$ as in **(30)**, and the dropping probability P_d as in **(31)**.
6) Compute the packet loss rate ξ from **(13)**, the average throughput η from **(14)**, and the ASE of AMC $\overline{S}$ from **(32)**.

Remark 2: The parameters P_0, f_d, $\bar{\gamma}$, m, b, K, λ, T_f, can be divided in two categories: i) the channel condition parameters, which include f_d, $\bar{\gamma}$ and m; and ii) the system design parameters, which include P_0, b, K, λ and T_f. The parameters in the first category are decided by the propagation environment, while the parameters in the second category are controllable through system design.

$$
\begin{aligned}
E\{D\} &= \sum_{a \in \mathcal{A}, u \in \mathcal{U}, c \in \mathcal{C}} D P(A = a, U = u, C = c) \\
&= \sum_{a \in \mathcal{A}, u \in \mathcal{U}, c \in \mathcal{C}} \left[\max\{0, a - K + \max\{0, u - c\}\} \times P(A = a) \times P(U = u, C = c) \right]
\end{aligned}
\tag{30}
$$

$$
\begin{aligned}
\overline{S} &= \lim_{t \to \infty} \sum_{u \in \mathcal{U}, c \in \mathcal{C}, c \neq 0} \left[S(U_{t-1} = u, C_t = c) \times P(U_{t-1} = u, C_t = c) \right] \\
&= \sum_{u \in \mathcal{U}, c \in \mathcal{C}, c \neq 0} S(U = u, C = c) P(U = u, C = c)
\end{aligned}
\tag{32}
$$

IV. CROSS-LAYER DESIGN EXAMPLE

Given certain parameters f_d, $\bar{\gamma}$ and m, it is desirable to optimize system performance (here ξ and η), through optimizing P_0, b, K, λ and T_f, subject to practical constraints. Our performance analysis in Section III provides a solid framework for cross-layer optimization.

As a simple example, we can fix (b, K, λ, T_f) and optimize P_0 in AMC to minimize ξ, which also maximizes η as shown in (14). Notice that P_0 affects ξ in (13) in two different ways: It directly controls the packet error rate over wireless links at the physical layer and also indirectly affects the packet dropping probability P_d at the data link layer. Hence, a simple approach is to numerically find the optimal P_0 (denoted by P_0^{opt}) among all possible choices, as we detail next.

Step 1) Compute $\xi(P_0)$ from (13) for each $P_0 \in \mathcal{P}$, where $\mathcal{P}$ is the set of possible target PER values, e.g., $\mathcal{P} = \{P_0 : 0 < P_0 < 1\}$.
Step 2) Determine the optimal P_0 as

$$P_0^{\text{opt}} = \arg\min_{P_0 \in \mathcal{P}} \xi(P_0). \tag{34}$$

These steps are repeated, each time the parameters are updated. Notice that $\xi(P_0)$ in Step 1 is affected by parameters at the data link layer, namely K and λ, through P_d. Thus, the optimal P_0^{opt} at the physical layer in Step 2 takes into account the data link layer characteristics. This simple example offers indeed a *cross-layer design*. As we will confirm by numerical results, such designs improve the system performance relative to the case where P_0 is set "blindly" at the physical layer without taking into account upper layers.

The original goal of introducing a layered structure is to facilitate the implementation and management of complicated communication systems. Therefore, in addition to performance gain relative to separate-layer designs, practical cross-layer designs should also minimize the system complexity as well as the amount of *cross-layer information* exchange across the layer interfaces. Our design fulfills this objective because of the following.

1) P_0^{opt} only needs to be updated based on slow-varying parameters, i.e., f_d, $\bar{\gamma}$, m, b, K, λ and T_f;
2) Since $\xi(P_0)$ is computed analytically, P_0^{opt} can be easily obtained by numerical search via (34), which could be simply realized by look-up tables in a practical implementation;
3) As thresholds $\{\gamma_n\}_{n=0}^{N+1}$ corresponding to P_0^{opt} are determined at the physical layer, the required cross-layer information is only λ and K coming from the data link layer, which can be afforded by practical systems; and
4) The proposed cross-layer design is compatible with existing separate-layer designs.

In a nutshell, our cross-layer design has low-complexity, requires minimal cross-layer information exchange and is backward compatible.

V. FURTHER ISSUES

A. Multiuser Scenario

So far, we only considered the single user case. Here we present general guidelines on how to apply our single-user re-

sults in a multiuser setting. We consider a time division multiplexing/time division multiple access (TDM/TDMA) system, where users are multiplexed to transmit in different time slots. We consider the following two cases:

Case 1) *(fixed scheduling)*: each user in this case is allocated a fixed number of time slots after being admitted, as in GSM systems.
Case 2) *(dynamic scheduling)*: the number of time slots per user may vary during transmission, depending on system variables such as channel states, queuing status and the number of time slots occupied by other users. Examples of such systems include the IEEE 802.15.3 WirelessPAN standard [11], the IEEE 802.16 WirelessMAN standard [3] and the 3G cellular systems with TDM mode [1], where the scheduling or admission control algorithms are not specified allowing for various implementation options [24].

Our performance results and cross-layer design for the single user case are directly applicable to the multiuser scenario in Case 1. Depending on how many slots are allocated to each user, one can figure out how many packets (b in Section III-A2) can be transmitted per frame with an $R = 1$ transmission mode. For Case 2, scheduling algorithms and performance analysis involving all users should be developed, which are interesting future research topics. Our results in this paper can be applied to develop novel scheduling algorithms, as we outline next.

The users admitted will be classified in two categories depending on the quality-of-service (QoS): QoS-guaranteed and best-effort [19]. For a QoS-guaranteed user i, we first associate d_i time slots as in the fixed-scheduling case. Based on our queuing analysis in Section III, we can derive the packet loss rate ξ and the average packet delay (see our preliminary results in [13] for delay analysis); thus, determining the relationship between QoS (loss and delay) and the number of time slots d_i. In order to guarantee the prescribed QoS for user i, we allocate the minimal value of d_i, denoted by d_i^*, time slots. The admission control should guarantee $\sum_{i \in I} d_i^* \leq N_{\text{total}}$ at any time, where I is the set of QoS-guaranteed user indices, and N_{total} is the total number of time slots per frame. This admission control policy guarantees QoS for all QoS-guaranteed users. For the best-effort users in the system, there are no associated time slots. To maximize system efficiency, we propose the following scheduling policies.

1) For the QoS-guaranteed user $i \in I$, d_i^* time slots are reserved at any time.
2) The number of *actually scheduled* time slots $d_{t,i}$ for the QoS-guaranteed user i at time t, depends on the data availability in the queue, i.e.,

$$\bar{d}_{t,i} = \begin{cases} d_i^*, & \text{if } U_{t-1,i} \geq C_{t,i} \\ \left\lceil d_i^* \frac{U_{t-1,i}}{C_{t,i}} \right\rceil, & \text{if } U_{t-1,i} < C_{t,i} \end{cases} \tag{35}$$

where $\lceil x \rceil$ denotes the smallest integer not less than x; $U_{t-1,i}$ is the queue state of user i at the beginning of time t; and $C_{t,i}$ is the queuing server (channel) state of user i at time t.
3) The total number of unused time slots $N_{\text{total}} - \sum_{i \in I} \bar{d}_{t,i}$, will then be shared by the best-effort users, to improve the system efficiency.

These policies guarantee QoS for all QoS-guaranteed users, while at the same time, they improve the overall system efficiency. For example, based on (35), if the queue is empty $U_{t-1,i} = 0$, then user i will not use any time slot since $\bar{d}_{t,i} = 0$. Those saved time slots will be shared by best-effort users. The admission and scheduling policies for best-effort users shall depend on the stationary behavior of $\{\bar{d}_{t,i}\}_{i \in I}$. The complete scheduler design is currently under investigation and results will be reported elsewhere. In short, our preliminary discussion suggests that our single-user results can be applied to multiuser scenarios.

B. Shadowing Effects

We adopted the Nakagami-m channel model in Section II, because it captures well the small-scale channel fading. Large-scale channel variations, such as shadowing effects, may also be included in our analysis, if e.g., we use the Nakagami-lognormal model [22]. The level crossing rate in [22, eq. (8)] for Nakagami-lognormal channels will then be used in Section II-C to establish the channel state transition matrix.

VI. NUMERICAL RESULTS

In this section, we present numerical results based on our analytical expressions developed in Section III. However, we also verified the accuracy of queuing analysis by measuring the stationary distribution through simulations. We consider both TM1 and TM2 and set the packet length $N_b = 1080$, with the PER approximation parameters of (3) listed in Tables I and II. We assume that the frame length is $T_f = 2$ ms and $b = 2$ for all test cases.

Test Case 1 (The Dropping Probability P_d Versus the Target Packet Error Rate P_0): We fix the set of reference parameters for TM1 as follows: average SNR $\bar{\gamma} = 20$ dB ; Doppler frequency $f_d = 10$ Hz, i.e., $f_d T_f = 0.02$; Nakagami fading parameter $m = 1$; queue length $K = 50$ packets, and Poisson arrival rate $\lambda T_f = 2$ packets/time-unit. The corresponding reference parameters for TM2 are the same as in TM1 except that $\bar{\gamma} = 15$ dB. We plot the dependence of the dropping probability P_d on the target packet error rate P_0 for both TM1 and TM2 in Fig. 7, where P_0 varies from 10^{-4} to 10^{-1}.

Fig. 7 reveals that P_d decreases as P_0 increases, which justifies our intuition that relieving the error performance requirement at the physical layer increases the queuing service rate and, thus, decreases P_d at the data link layer.

Test Case 2 (The Packet Loss Rate ξ Versus the Target Packet Error Rate P_0): We plot ξ as a function of P_0 in Fig. 8 for TM1 and in Fig. 9 for TM2. For each curve, we modify only one parameter from the reference parameters. We consider the following parameters:

1) Poisson arrival rate $\lambda T_f = 1.6$ (packets/time-unit);
2) Nakagami parameter $m = 1.1$;
3) queue length $K = 100$ packets;
4) Doppler frequency $f_d = 5$ Hz, i.e., $f_d T_f = 0.01$.

Comparing the curves in Figs. 8 and 9, we infer the following.

1) Reducing the arrival data rate via decreasing λ leads to small P_d and, thus, lowers ξ;
2) Improving channel quality via increasing m reduces packet reception errors; so, ξ decreases;
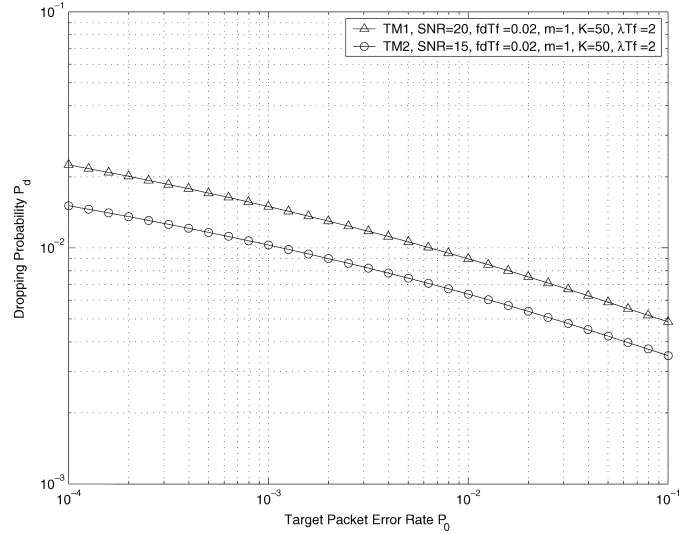

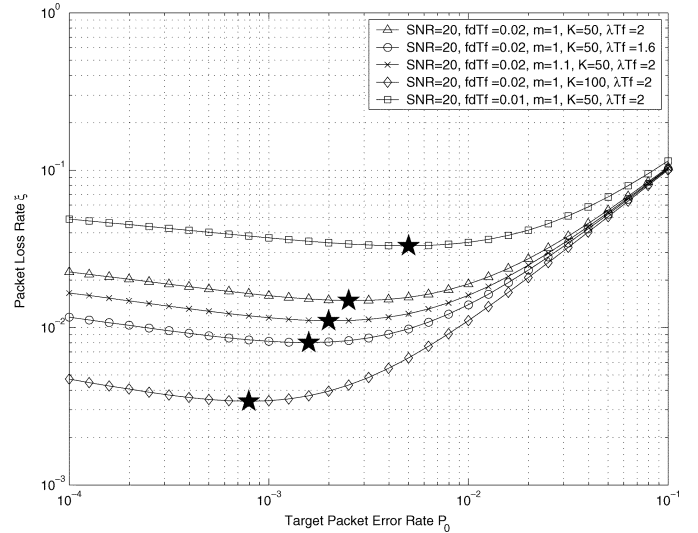Fig. 7.   Dropping probability versus target packet error rate.

Fig. 8.   Packet loss rate for TM1 (stars denote the minimum packet loss rate).

3) Increasing queue length K decreases P_d, which decreases ξ; and
4) A lower Doppler frequency f_d leads to longer fading-duration, which increases ξ.

On each curve in Figs. 8 and 9, the minimum value of ξ is depicted by star; the corresponding P_0 is the solution of our cross layer design in (34).

Test Case 3 (The Packet Loss Rate ξ With Available Average SNR $\bar{\gamma}$): Figs. 10 and 11 depict the packet loss rate ξ with the reference parameters at different $\bar{\gamma}$ for TM1 and TM2, respectively. The minimum values of ξ are again indicated by stars. Comparing Fig. 10 with Fig. 11, we deduce that TM2 achieves about 5–6 dB SNR gain relative to TM1 in terms of the system performance ξ, thanks to the coding advantage.

Test Cases 1, 2, and 3 illustrate the dependence of the system performance P_d, ξ (as well as η via (14)) on different parameters P_0, f_d, m, K, λ and T_f. This dependence is quantified by the analytical procedure of Section III.

Fig. 7 in Test Case 1 illustrates how P_d is affected when adjusting P_0. However, the overall packet loss rate ξ depends on
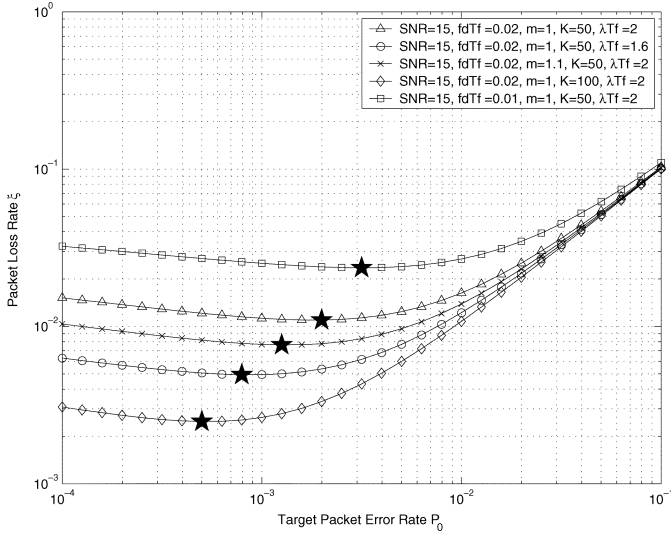
Fig. 9.   Packet loss rate for TM2 (stars denote the minimum packet loss rate).
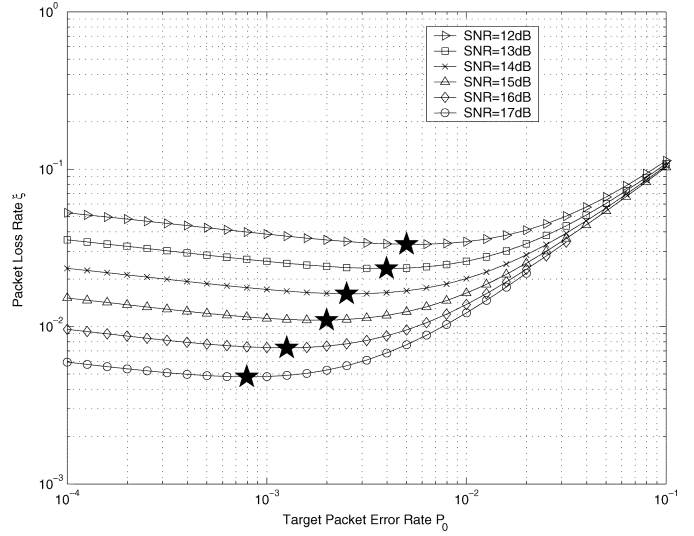

Fig. 11.   Packet loss rate with different SNR's for TM2 (stars denote the minimum packet loss rate).
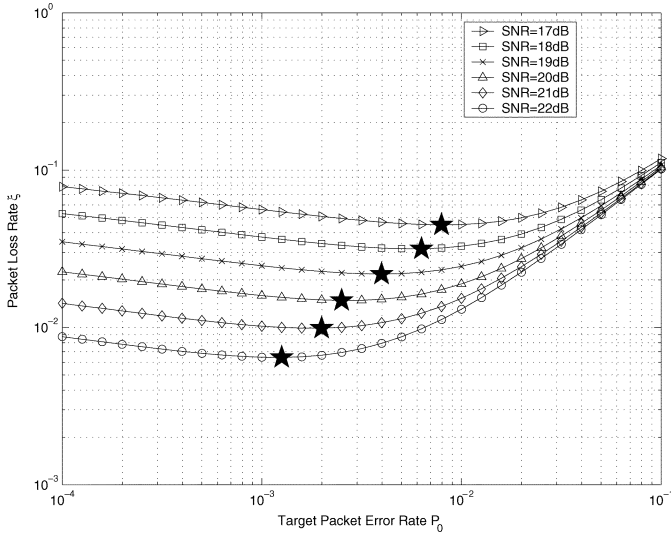

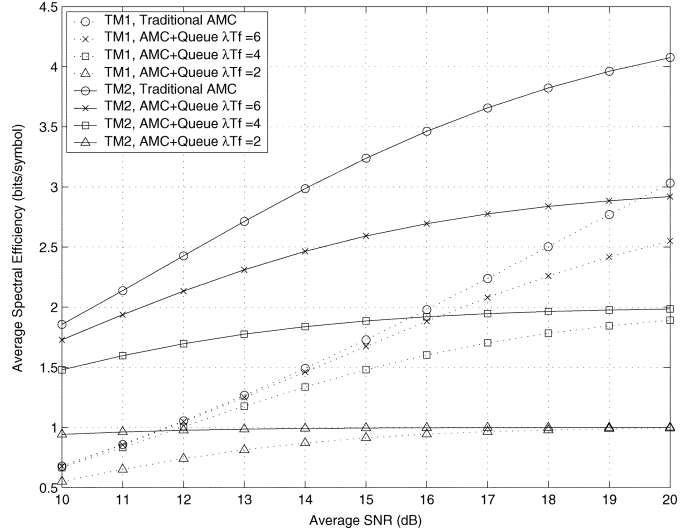Fig. 10.   Packet loss rate with different SNR's for TM1 (stars denote the minimum packet loss rate).

Fig. 12.   ASE of the AMC module.

both P_0 and P_d [cf. (13)]; their joint effects on ξ are shown by Figs. 8–11, in Test Cases 2 and 3. From the shapes of those plots, one can infer that P_0 dominates ξ when P_0 has large values, while P_d dominates ξ when P_0 has small values. This observation agrees with the intuition behind (13).

The stars on each curve in Figs. 8–11, show the optimal performance corresponding to P_0^{opt} in our cross-layer design of Section IV. The performance gain relative to the case where P_0 is set "blindly" is quantified. For example, the minimal ξ with $K = 100$ in Fig. 9 is less than 2×10^{-3}, which provides considerable performance gain compared with $\xi \approx 10^{-1}$ when P_0 approaches 10^{-1}.

Test Case 4 (The ASE of AMC): With $P_0 = 10^{-3}$, $f_d = 10\,\mathrm{Hz}$, and $m = 1$, the ASE of traditional AMC [4], [9] is depicted in Fig. 12 for both TM1 and TM2, with respect to different average SNR values $\bar{\gamma}$. The ASE of AMC in the presence of queuing with the reference parameters and $\lambda T_f = 2, 4, 6$ packets/time-unit are also plotted in Fig. 12 for both TM1 and TM2.

Fig. 12 confirms that the ASE of AMC in the presence of queuing is less than that of traditional AMC, which is based on the assumption that data are continuously available for transmission. The reason for the spectral efficiency loss is precisely the dynamic queuing behavior, which causes the queue to be empty from time to time. Increasing the arrival rate λT_f improves the average ASE; however, it may lead to larger packet dropping probability and larger packet loss rate as demonstrated in Test Case 2.

VII. CONCLUDING REMARKS AND FUTURE WORK

In this paper, we developed a general analytical procedure to investigate the performance for transmissions over the wireless link, where finite-length queuing at the data link layer is coupled with AMC at the physical layer. We first characterized the queuing service process provided by AMC and derived a recursion for the queue state. Then we constructed an FSMC with a state pair containing both the queue and the queue server states, and computed its stationary distribution. The latter enabled us to

obtain the packet loss rate, the average throughput and the ASE of AMC. Guided by our performance analysis, we developed a cross-layer design, which optimizes the target packet error rate P_0 in AMC at the physical layer, to minimize the packet loss rate and maximize the average throughput, when combined with a finite-length queue at the data link layer. Our cross-layer design enjoys low-complexity, requires minimal cross-layer information exchange and is compatible with existing separate-layer designs.

Our focus in this paper has been on the single user case. Application of our results to the development of efficient scheduling algorithms in multiuser scenarios is currently under investigation. Other possible research directions may include buffer scheduling for queue-length K, flow/congestion control for arrival rate λ and framing/packing for frame-length T_f. Furthermore, it is of interest to pursue cross-layer analysis and design considering, e.g., 1) frequency-selective wireless channels; 2) the automatic repeat request (ARQ) protocol adopted at the data link layer; and 3) the transmission control protocol (TCP) that is installed at the transport layer. [13].

APPENDIX

In order to prove existence of the stationary distribution of the enlarged FSMC with the transition probability matrix $\mathbf{P}$ in (20), we need the following lemmas.

Lemma 1: The Markov chain of (u, c), where $(u, c) \in \mathcal{U} \times \mathcal{C}$, is irreducible.

Proof: We need to show that there exists a multitransition path with nonzero transition probability for each transition from (u, c) to (v, d), which is denoted by $(u, c) \longrightarrow (v, d)$, for any $(u, c) \in \mathcal{U} \times \mathcal{C}$ and $(v, d) \in \mathcal{U} \times \mathcal{C}$ [18, pp. 169]. We verify the following cases.

1) $c \longrightarrow d$ for any $c \in \mathcal{C}$, $d \in \mathcal{C}$, because the individual FSMC corresponding to the server state c is irreducible by (12).
2) $(u, c) \longrightarrow (u, u)$. Based on (19), it follows that $A_t = u - \max\{0, u - C_t\} \geq 0$, which has nonzero probability $P(A_t = u - \max\{0, u - C_t\}) > 0$ for any $u \in \mathcal{U}$. Because $c \longrightarrow u$ from (i), each transition of the path from (u, c) to (u, u) has nonzero probability from (22).
3) $(u, u) \longrightarrow (v, u)$. Letting $A_t = v - \max\{0, u - u\} = v \geq 0$, we infer that $P(A_t = v) > 0$. Since $u \longrightarrow u$ with $P_{u,u} > 0$ in (11), we have $P_{(u,u),(v,u)} > 0$ from (22).
4) $(v, u) \longrightarrow (v, d)$ is the same as 2).

Based on 1)-4), we can find a multitransition path with nonzero transition probability for each transition from (u, c) to (v, d), i.e., $(u, c) \longrightarrow (v, d)$. ∎

Lemma 2: The Markov chain of (u, c), $(u, c) \in \mathcal{U} \times \mathcal{C}$, is homogeneous and positive recurrent.

Proof: Since the transition probability of the FSMC in (12) is independent of t, the Markov chain of (u, c) is homogeneous [6, pp. 54]. Since the Markov chain has finite state space $\mathcal{U} \times \mathcal{C}$, it is positive recurrent, because Lemma 1 and Theorem 3.3 in [6, pp. 105] assert that the irreducible homogeneous Markov Chain with finite state space is positive recurrent. ∎

Based on Lemmas 1 and 2, we have that:

Theorem 1: The stationary distribution $\boldsymbol{\pi}$ of the Markov chain of (u, c), where $(u, c) \in \mathcal{U} \times \mathcal{C}$, exists; $\boldsymbol{\pi}$ is unique, and $\boldsymbol{\pi} > 0$.

Proof: Following Theorem 3.1 in [6, pp. 104], the proposition is valid if and only if the Markov chain is irreducible, homogeneous, and positive recurrent, which has been verified in Lemmas 1 and 2. ∎

To verify the (31), we also establish:

Theorem 2: The time-average dropping probability equals the ensemble-average dropping probability

$$P_d := \lim_{T \to \infty} \frac{\sum_{t=1}^{T} D_t}{\sum_{t=1}^{T} A_t} = \frac{E\{D\}}{E\{A\}} = \frac{E\{D\}}{\lambda T_f}. \quad (36)$$

Proof: From (29), we know that D is a function of $(A = a, U = u, C = c)$, $a \in \mathcal{A}$, $u \in \mathcal{U}$, $c \in \mathcal{C}$, and satisfies

$$D(u, c) = \sum_{a \in \mathcal{A}} D P(A = a) \leq \sum_{a \in \mathcal{A}} a P(A = a) = \lambda T_f \quad (37)$$

because the number of dropping packets is less than a per time-unit. From (24), we have

$$\sum_{u \in \mathcal{U}, c \in \mathcal{C}} |D(u, c)| \pi_{(u,c)} \leq \lambda T_f \sum_{u \in \mathcal{U}, c \in \mathcal{C}} \pi_{(u,c)} < \infty. \quad (38)$$

Based on (38) and [6, Theorem 4.1], we obtain

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} D_t = E\{D\} \quad (39)$$

for any initial distribution of (u, c). On the other hand, because A_t is a stationary process, we have

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} A_t = E\{A\} = \lambda T_f. \quad (40)$$
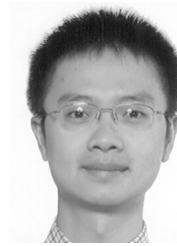
Based on (39) and (40), we obtain (36). ∎

REFERENCES

[1] *Physical Layer Aspects of UTRA High Speed Downlink Packet Access (Release 4)*, 3GPP TR 25.848 V4.0.0, 2001.
[2] *Physical Layer Standard for CDMA2000 Spread Spectrum Systems*, 3GPP2 C.S0002–0 Ver. 1.0, 1999.
[3] *IEEE Standard 802.16 Working Group, IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, 2002.
[4] M. -S. Alouini and A. J. Goldsmith, "Adaptive modulation over Nakagami fading channels," *J. Wireless Commun.*, vol. 13, no. 1–2, pp. 119–143, May 2000.
[5] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1992.
[6] P. Brémaud, *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. New York: Springer-Verlag, 1999.

[7] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: A unified view," *IEEE Trans. Commun.*, vol. 49, no. 9, pp. 1561–1571, Sep. 2001.

[8] 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, no. 5, pp. 172–180, May 2002.

[9] A. J. Goldsmith and S. -G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218–1230, Oct. 1997.

[10] K. J. Hole, H. Holm, and G. E. Oien, "Adaptive multidimensional coded modulation over flat fading channels," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 7, pp. 1153–1158, Jul. 2000.

[11] J. Karaoguz, "High-rate wireless personal area networks," *IEEE Commun. Mag.*, vol. 39, no. 12, pp. 96–102, Dec. 2001.

[12] 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.*, vol. 2, no. 5, pp. 1746–1775, Sep. 2004.

[13] ——, "TCP performance in wireless access with adaptive modulation and coding," in *Proc. Int. Conf. Commun.*, Paris, France, Jun. 20–24, 2004.

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

[15] 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 Commun. (ISWC-2002)*, Victoria, BC, Canada, Sep. 23–24, 2002.

[16] M. B. Pursley and J. M. Shea, "Adaptive nonuniform phase-shift-key modulation for multimedia traffic in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 8, pp. 1394–1407, Aug. 2000.

[17] J. Razavilar, K. J. R. Liu, and S. I. Marcus, "Jointly optimized bit-rate/delay control policy for wireless packet networks with fading channels," *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 484–494, Mar. 2002.

[18] S. M. Ross, *Introduction to Probability Models*, 7nd ed. San Diego, CA: Academic, 2000.

[19] S. Shakkottai, T. S. Rappaport, and P. C. Karlsson, "Cross-layer design for wireless networks," *IEEE Commun. Mag.*, vol. 41, no. 10, pp. 74–80, Oct. 2003.

[20] H. K. Shiu, Y. H. Chang, T. C. Hou, and C. S. Wu, "Performance analysis of TCP over wireless link with dedicated buffers and link level error control," in *Proc. Int. Conf. Commun.*, vol. 10, Amsterdam, Netherlands, Jun.–Jul. 2001, pp. 3211–3216.

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

[22] T. T. Tjhung and C. C. Chai, "Fade statistics in Nakagami-lognormal channels," *IEEE Trans. Commun.*, vol. 47, no. 12, pp. 1769–1772, Dec. 1999.

[23] H. S. Wang and N. Moayeri, "Finite-state Markov channel – A useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163–171, Feb. 1995.

[24] K. Wongthavarawat and A. Ganz, "IEEE 802.16 based last mile broadband wireless military networks with quality of service support," in *Proc. MILCOM Conf.*, Boston, MA, Oct. 2003, pp. 779–784.

[25] M. D. Yacoub, J. E. Vargas Bautista, and L. G. de R. Guedes, "On higher order statistics of the Nakagami-m distribution," *IEEE Trans. Veh. Technol.*, vol. 48, no. 3, pp. 790–794, May 1999.

[26] D. Yoon and K. Cho, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1074–1080, Jul. 2002.

[27] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no. 11, pp. 1688–1692, Nov. 1999.

[28] H. Zheng and J. Boyce, "An improved UDP protocol for video transmission over internet-to-wireless networks," *IEEE Trans. Multimedia*, vol. 3, no. 3, pp. 356–365, Sep. 2001.

[29] S. Zhou and G. B. Giannakis, "How accurate channel prediction needs to be for transmit-beamforming with adaptive modulation in Rayleigh MIMO channels?," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1285–1294, Jul. 2004.

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

His research interests lie in 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 (USTC), Hefei, China, in 1995 and 1998, respectively. He received the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis, in 2002.

He joined the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, in 2003. His research interests lie in 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.

Dr. Zhou currently serves as an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS .

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, in 1981, and the M.Sc. degree in electrical engineering, the M.Sc. degree in mathematics, and the Ph.D. degree in electrical engineering, all from the University of Southern California (USC), Los Angeles, in 1983 and 1986, respectively.

After lecturing for one year at USC, he joined the University of Virginia 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, 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 interests include transmitter and receiver diversity techniques for single- and multiuser fading communication channels, complex-field and space-time coding, multicarrier, ultrawide-band wireless communication systems, cross-layer designs, and distributed sensor networks.

Dr. Giannakis is the (co-)recipient of six paper awards from the IEEE Signal Processing (SP) and Communications Societies (1992, 1998, 2000, 2001, 2003, 2004). He also received the SP Society's Technical Achievement Award in 2000. He served as Editor-in-Chief for the IEEE SIGNAL PROCESSING LETTERS, as Associate Editor for the IEEE 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 Election Committee. He has also served as a member of the 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.