Abstract—In this paper, we consider a deadline constrained traffic and analyze the joint effects of packet queuing and adaptive modulation (AM) for packet transmission in MIMO systems. We present a general analytical procedure, and derive the achievable delay-bound violation probability, the packet loss rate, and the average throughput of AM in MIMO wireless links employing orthogonal space-time block coding (STBC) over Nakagami fading channels. In the proposed analytical procedure, we use statistical delay-bound for deadline constrained traffic based on the effective bandwidth theory. Based on our performance analysis, we propose a cross-layer design, which selects the optimal target packet error rate of AM at the physical layer, to minimize the packet loss rate and maximize the average throughput, when combined with packet queuing at the data link layer. Numerical results demonstrate that the system performance depends on various parameters, and show the performance gain due to cross-layer design. We also discuss briefly possible applications of our proposed modeling to quality of service guarantee in multiuser scenarios.

I. INTRODUCTION

Cross-layer design for packet transmissions over fading channels has recently emerged as an appealing area of research [1]. The aim is to improve the spectral efficiency of wireless systems while guaranteeing prescribed quality of service (QoS) requirements such as delay and error rate constraints. In this paper, we analyze the joint effects of deadline constrained packet queuing at the data link layer and adaptive modulation at the physical layer for transmit diversity systems employing orthogonal space time block coding (STBC) over MIMO Nakagami fading channels.

Link adaptation at the transmitter, in particular adaptive modulation (AM), is a promising approach towards high throughput wireless communications. In order to optimize the wireless system performance, a link adaptation algorithm adjusts transmission parameters such as power and rate based on the time-varying channel conditions [2]. However, it was shown in [2] that rate adaptation is the key factor for increasing the system spectral efficiency. In this paper, we use constant-power discrete-rate adaptive modulation based on M-QAM signal constellations as a simple technique for link adaptation at the frame level (physical layer). This scheme adapts the modulation constellation size to the variations of the fading channel, by using higher or lower order of modulation.

Traditionally, AM schemes at the physical layer have been designed separately from higher layers. In fact, in designing these schemes it is assumed that there is always sufficient data available at the transmitter queue to be transmitted and the queuing effects are not considered. However, in practical wireless packet networks, data arrivals have a random and bursty nature, resulting in the dynamic behavior of the queue and queuing delay. For example, traffic sources from wide-area networks are modeled as Poisson (or Bernoulli) arrival processes [3]. Therefore, it is necessary to investigate the impact of dynamic behavior of the queue on AM. In [4], the authors studied the effect of finite-length queuing on AM. Specifically, it is assumed that the packets in the queue are not deadline constrained and only the effect of packet dropping due to buffer overflow is taken into account [4]. However, real-time services such as video and audio require a bounded delay, that is, if a packet is not successfully transmitted by a certain deadline; it is dropped from the transmitter buffer. For instance, video-conferencing, as a real-time application, is subject to a delay bound of 40-90 ms. In this work, we propose a solution by providing a statistical delay-bound guarantee. There are a significant amount of research results on statistical QoS guarantees in wired networks. Among them, the effective bandwidth concept is an efficient approach for characterizing the statistical behavior of traffic sources [5]. In this paper, we use this concept for modeling the statistical delay-bound violation probability over multi-rate MIMO wireless networks. We consider a deadline constrained traffic and analyze the joint effects of packet queuing and adaptive modulation. We derive the achievable delay-bound violation probability, the packet loss rate, and the average throughput of AM in MIMO wireless links employing orthogonal space-time block coding (STBC) over Nakagami fading channels.

The remainder of this paper is organized as follows: Section II provides a background on the system model for MIMO channel, the source modeling and the effective bandwidth theory. Section III is dedicated to the performance analysis of the combined cross-layer design. Numerical results are provided in Section IV, while concluding remarks are presented in Section V.
II. System Model

A. Physical and Link Layer Parameters

We consider a wireless link with \( N_t \) transmit and \( N_r \) receive antennas. The packets generated by the traffic source are first queued at an infinite buffer and then divided into frames. Then, the frames are transmitted at the physical layer, where AM and MIMO space-time diversity are employed, respectively, to enhance the system performance. We adopt the packet and frame structures as in [4]. At the physical layer each frame contains a fixed number of symbols \( N_f \). Given a fixed symbol rate, the frame duration, \( T_f \) seconds, is constant, and sets the time-unit throughout this paper. Each frame contains a fixed number of symbols \( N_t \) at the physical layer and a variable number of packets \( N_p \) from the data link layer. Each packet contains a fixed number of bits \( N_b \), which include packet header, payload, and cyclic redundancy check (CRC) bits. Applying modulation and coding with rate \( R_c \) bits/symbol in mode \( n \), \( N_b \) bits of a packet are mapped into a block of \( N_b/R_c \) symbols. A frame comprises of multiple such symbol-blocks as well as \( N_f \) pilot symbols and control parts, as in HIPERLAN/2 or IEEE 802.11a standards [11]. In mode \( n \), the number of symbols per frame is \( N_f = N_t + N_m N_r / R_c \), which indicates that \( N_f \) depends on the chosen AMC mode.

We next list all our assumptions. (1) The channel is assumed to follow a block fading model, i.e., the gain remains invariant during a frame, but varies from frame to frame. This model is suitable for slowly-varying fading channels. As a consequence, AM is adjusted on a frame-by-frame basis. (2) Perfect channel state information (CSI) is available at the receiver. The feedback channel is assumed to be instantaneous and error free. (3) We are interested in real-time applications with strict delay bounds for arrived packets, that is, if a packet is not transmitted before a deadline \( D_{\text{max}} \), it is dropped. (4) We also assume strong CRC code, so that perfect packet error detection is available. If a packet is not received correctly at the receiver, it is dropped from the receiver buffer, and packet loss is declared.

B. MIMO Channel Model

We consider a MIMO wireless communication employing STBC with \( N_t \) transmit and \( N_r \) receive antennas subject to independent block Nakagami fading. Using STBC, transmit diversity over the wireless link is provided by mapping each \( R \leq T \) complex input symbols \( \{s_1, s_2, \ldots, s_R\} \), from a given signal set \( S \), into \( N_t \) orthogonal sequences of length \( T \), that are simultaneously transmitted by \( N_t \) transmit antennas. Since \( R \) input symbols from the signal set \( S \) are transmitted within \( T \) symbol durations (a frame period), the information code rate of the space-time block code is defined as \( R_c = R/T \). For a given channel realization, the MIMO system with diversity order \( K = N_t N_r \) can then be represented within a frame period by the channel matrix \( H = [h_{j,k}] \), where \( j,k \) is the path gain between \( k \)th transmit and \( j \)th receive antennas and is distributed according to Nakagami-\( m \) probability density function (PDF) with parameter \( m \) and mean \( \mathbb{E}[h^2_{j,k}] \). The channel input-output relationship is then given by \( Y = HX + V \), where the received signal \( Y \) is a \( N_r \times T \) matrix, \( X \) is the \( N_t \times T \) matrix of transmitted symbols, and the receiver noise \( V \) is \( N_r \times T \) matrix with elements modeled as i.i.d. complex circular Gaussian random variables, each with a \( CN(0, \sigma^2) \) distribution.

The channel quality, for flat fading channels, can be captured by a single parameter, namely the received signal-to-noise ratio (SNR) \( \gamma \). Let \( P_T \) be the average transmit power per symbol time over the \( N_t \) transmit antennas and define \( \gamma = mP_T / \sigma^2 \) as the average SNR per receive antenna. Using the STBC SISO equivalency [6], it is shown that the SNR at the output of the STBC decoder, \( \gamma^{\text{STBC}} \), is Gamma distributed with parameter \( mK \) and mean \( N_r \gamma / R_c \), according to the PDF, \( p_{\gamma^{\text{STBC}}} (\gamma) \), given by

\[
p_{\gamma^{\text{STBC}}} (\gamma) = \frac{\gamma^{mK-1}}{\Gamma(mK)} \left( \frac{mN_r R_c}{\gamma} \right)^{mK} e^{-\frac{mN_r R_c \gamma}{\gamma}}, \gamma \geq 0
\]

where \( \Gamma(m) = \int_0^\infty t^{m-1} e^{-t} dt \) is the Gamma function.

C. Adaptive Modulation

The objective of AM is to improve the system spectral efficiency by adjusting transmission parameters to the available CSI, while satisfying a target packet error rate \( P_e \) in the physical layer. In our system model, AM is performed on a frame-by-frame basis by dividing the range of the received SNR \( \gamma^{\text{STBC}} \) into \( N+1 \) non-overlapping consecutive intervals, denoted by \( [\gamma_n, \gamma_{n+1}] \), \( n = 0, 1, \ldots, N \), where \( \gamma_0 = 0 \), \( \gamma_{N+1} = \infty \), and \( N \) is the number of AM transmission modes. Whenever the CSI fed back to the transmitter falls within the interval \( [\gamma_n, \gamma_{n+1}] \), the mode \( n \) is chosen, data is transmitted with rate \( R_c \) bits/symbol. No data is sent when \( \gamma \leq \gamma_n < \gamma_{n+1} \), which corresponds to deep channel fades or the outage mode with rate \( R_c = 0 \) bits/symbol. In section IV, we consider a full rate STBC \( R_c = 1 \), therefore, the overall code rate is determined by AM module. In this paper, the rates of the available transmission modes are \( 1, 2, 3, 4, 5 \) bits/symbol, and their corresponding constellations are BPSK, QPSK, 8-QAM, 16-QAM and 32-QAM, respectively. The design objective for AM is to determine the mode switching levels. The exact packet error rate (PER) for \( M_n \)-QAM (\( M_n = 2^n \)) transmissions over additive white Gaussian noise (AWGN) channels is well approximated by [4]

\[
\text{PER}_n(\gamma) = \begin{cases} 
1, & 0 \leq \gamma < \Gamma_n^0 \\
\frac{a_n \exp(-g_n \gamma)}{\Gamma_n}, & \gamma \geq \Gamma_n
\end{cases}
\]

where \( n \) is the mode index, \( \gamma \) is the received SNR, and parameters, \( \{a_n, g_n, \Gamma_n\} \) are mode and packet-size dependent constants. These parameters can be obtained by least square fitting the PER expression of (2) to the exact PER [4]. In this work, we set packet length \( N_b = 1080 \) bits. Table I shows the set of fitting parameters for the selected transmission modes. Using (1), mode \( n \) is chosen with probability \( P_n \), given by
TABLE I
Transmission Modes for AM and their Corresponding Fitting Parameters

<table>
<thead>
<tr>
<th>Modulation</th>
<th>$R_{\text{bits/sym.}}$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPSK</td>
<td>107.97</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>QPSK</td>
<td>109.08</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>8-QAM</td>
<td>93.46</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>16-QAM</td>
<td>85.03</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>32-QAM</td>
<td>50.06</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
</tr>
</tbody>
</table>

The probability that the buffer is called effective bandwidth packets per frame. For this source, from [5] we have

$$P_e = \int_{\gamma_0}^{\gamma_{\text{max}}} p_{\gamma_{\text{trunc}}} (\gamma) d\gamma = \frac{1}{\Gamma(mK)} \left[ \Gamma(mK, mN, R_c) \frac{Y_{n+1}}{\gamma} - \Gamma(mK, mN, R_c) \frac{Y_{n}}{\gamma} \right]$$

where $\Gamma(m, x) = \int_0^x t^{m-1} e^{-t} dt$ is the complementary incomplete Gamma function. Having set the target PER for the physical layer to $P_0$, the mode switching levels $\{\gamma_n\}$ are set to the minimum SNR required to achieve $P_0$ over a non-fading AWGN channel. Inverting the PER expression in (2), leads to the following mode switching levels

$$\gamma_0 = 0,$$
$$\gamma_n = 1 / g_n \ln (a_n / P_0), \quad n = 1, 2, \ldots, N$$
$$\gamma_{N+1} = +\infty$$

D. Source Model and Effective Bandwidth

We model the traffic source as a stochastic arrival stream $\{X_t\}$, where $X_t$ indicates the number of packets arrived at frame $t$. Input packets, arrived from higher layers of stack, are queued at an infinite buffer, and served with rate $C$ in a first-in-first-out manner. In this paper, we are interested in real-time applications with strict bound on packet delay. Therefore, it is of interest to study the packet delay statistics and their dependence on the statistical characteristics of the arrival process. Let $Q(\infty)$ and $D$ denote the steady-state queue length and the maximum acceptable bound on queue length, respectively. The key result from effective bandwidth and large deviation theory [5] is that, in the asymptotic regime for large $B$, the tail probability of the form $P\{Q(\infty) \geq B\}$ decays exponentially with $B$, i.e.

$$P\{Q(\infty) \geq B\} \approx e^{-\delta(C)B} \quad \text{as} \quad B \to \infty$$

where the tail probability exponent $\delta(C)$ is determined as $\delta(C) = \max \{ s \geq 0 : \alpha(s) \leq C \}$, and the increasing function $\alpha(s)$, is called effective bandwidth of traffic source and $s$ denotes the space variable [5]. However, it is found that for smaller values of $B$, the following approximation is more accurate [8]:

$$P\{Q(\infty) \geq B\} \approx \rho(C) e^{-\delta(C)B}$$

where $\rho(C) = P\{Q(\infty) \geq 0\}$ is the probability that the buffer is non-empty. If the quantity of interest is the delay $D(t)$ experienced by a source packet departing at time $t$, then the probability of $D(t)$ exceeding a delay bound $D_{\text{max}}$ satisfies:

$$\sup_t P\{D(t) \geq D_{\text{max}}\} = P\{D(\infty) \geq D_{\text{max}}\} = \rho(C) e^{-\delta(C)C D_{\text{max}}}$$

where $D(\infty)$ denotes the steady state of $D(t)$. Thus, the main result is that, for a source that has a packet delay bound of $D_{\text{max}}$, and can tolerate a delay-bound violation probability of at most $\varepsilon$, if we model this source with the pair $\{\rho(C), \delta(C)\}$, then the effective bandwidth theory shows that the constant channel capacity should be at least $C$, where $C$ is the solution to $\varepsilon = \gamma(C) e^{-\delta(C)C D_{\text{max}}}$. In this paper, we consider a compound Poisson source that has fixed-length Poisson distributed packet arrivals. A compound Poisson is defined as $X_t = \sum_{n=1}^N Y_n$, where $Y_1, Y_2, \ldots$ are i.i.d. random variables with distribution $F$, and $N(t)$ is an independent Poisson process of rate $\nu$ packets per frame. For this source, from [5] we have

$$\alpha(s) = \frac{\nu}{s} \int (e^{-sx} - 1) dF(x)$$

In the scenario of our interest, the packet length is assumed constant and of size $1/\mu$, i.e., $Y_n = 1/\mu$, $\forall n$; or $F(x)$ will be a step function at $x=1/\mu$ point and thus the effective bandwidth for this source is $\alpha(s) = \nu (e^{s/\mu} - 1) / s$.

III. System Performance Analysis

Real-time services such as video and audio require a bounded delay. For instance, video-conferencing, which is a real-time variable-bit-rate application, is subject to a delay bound of 40-90 ms and an acceptable loss rate of $10^{-3}$ [10]. However, in practice a hard delay-bound guarantee is infeasible to be achieved over wireless links due to the effect of time-varying fading channel. An alternative solution can be considered by providing the statistical QoS guarantees, where the delay-bound is guaranteed with a small violation probability. In the context of wired networks, many research results on statistical QoS guarantees have been proposed. Among them, the effective bandwidth concept is an efficient approach that has received extensive research attention in the literature [5]. In this work, we generalize this concept for modeling the statistical delay-bound violation probability over multi-rate MIMO wireless systems. We consider the delay bound violation probability and packet loss rate as two QoS metrics for deadline constrained packet services. Due to delay constraint, packets will be dropped if they are not transmitted before the expiration of their deadlines. Let $P_d$ denote the probability of packet loss rate due to delay bound violation in data link layer and $P_0$ denote the PER in physical layer due to time-varying fading and noise effects. A packet from the source is correctly received by the receiver, only if it is not dropped from the queue, with probability $1 - P_q$, and it is correctly received through the wireless channel, with
probability $1-P_q$. As a result, the probability of a packet received correctly is $(1-P_q)(1-P_\nu)$, and the packet loss rate can be obtained as

$$P_{loss} = 1 - (1 - P_q)(1 - P_\nu) \quad (9)$$

We can also obtain the average throughput of the system as $\eta = v T_f (1-P_{loss})$. Hence, to evaluate the system performance in terms of packet loss rate and average throughput, we need to find $P_q$. To do this, we will use the effective bandwidth concept. Intuitively, we expect that $P_q$ is inversely proportional to the channel bit rate. In fact, the higher rate the queue is served, the less time the packets spend waiting in the queue. On the other hand, from AM design, $P_q$ is increasing on the channel bit rate. Therefore, we expect and will show that there is a tradeoff between these two types of losses in the system. Our primary objective is to find the optimal $P_0$ that minimizes total packet loss rate. Also, we will show how one can adjust $P_0$ and allocate bandwidth to a wireless link for satisfying QoS requirement of real-time services.

### A. Queue Service Process

When AM is employed over wireless links, the service process of the queue has a random nature, rather than deterministic, with a variable number of packets transmitted per frame. Let $C_t$ packets/frame indicate the number of packets transmitted using AM at time $t$. Let $c_n$ packets/frame denote the number of packets transmitted per frame, corresponding to each transmission mode $n$. Therefore, we have $C_t \in \{c_0, c_1, \ldots, c_N\}$. Let $b$ denote the number of packets that can be accommodated per frame corresponding to mode one of AM scheme with $R=1$. Under such assumption, we obtain $c_n = b R_n$. Parameter $b$ depends on the resource allocated per connection, which can be determined by a resource allocation strategy. As specified in section II-C, the AM scheme yields a queue service process with a total of $N+1$ states $\{c_n\}_{n=0}^N$, each with probability $P(C_t = c_n) = P_n$, as specified by (3).

### B. Achievable Delay-Bound Violation Probability

For a stationary arrival traffic process, the required service bandwidth, denoted by $C$, that can upper-bound the delay-bound violation probability $\nu$, is determined by equation

$$\nu = \gamma(C) e^{-\delta(C) C D_{max}}$$

where $\gamma(C)$ and $D_{max}$ are defined in section II-D. Due to the time-varying nature of fading channels, service process is a random variable and consequently $\nu$ is also a random variable. When the channel provides a service rate larger than $C$, the user’s statistical delay bound is guaranteed. However, such service rate can be provided with a specific probability. The average achievable delay-bound violation probability denoted by $P_{q}$ can be derived as

$$P_{q} = \sum_{n=0}^{N} P(C_t = c_n) \gamma(c_n) e^{-\delta(c_n) c_n D_{max}} \quad (10)$$

In practice, $\gamma(c_n) = P[Q(\infty) \geq 0 | C_t = c_n]$ can be estimated by simply averaging a number of samples of source traffic based on sign bits indicating whether a packet is being transmitted [9]. It is necessary to note that the term $n=0$ in summation (10), corresponds to the outage state of wireless channel and is independent from traffic parameters and $D_{max}$. Using MIMO diversity mode for transmission over wireless channel a very small outage probability is guaranteed and consequently, acceptable $P_q$ can be guaranteed.

### IV. PERFORMANCE EVALUATION

In this section, we provide numerical results based on our analytical expressions developed in previous sections. Source traffic is modeled as a compound Poisson process that has fixed-length Poisson packet arrivals, for which the effective bandwidth is derived in section II-D. We set the packet length $N_b=1080$ bits and the available number of M-QAM constellations $N=5$, with the PER approximation parameters of (2) listed in Tables I. We select the set of reference parameters as follows: frame duration $T_f=2$ ms, Nakagami fading parameter $m=1$, Poisson arrival rate $v T_f=2$ packets/frame, and delay-bound $D_{max}=4T_f=8$ ms. For MIMO scheme, we consider a full rate STBC ($R_c=1$) used with $N_r=2$, since it represents the transmit diversity scheme adopted by current 3GPP specifications. In our performance analysis, we first study the tradeoff between average delay-bound violation probability $P_q$ and target packet error rate $P_0$ and its effect on total packet loss rate $P_{loss}$. We then discuss briefly possible applications of the proposed analysis to QoS guarantee in multiuser scenarios.

#### A. Delay-Bound and Packet Loss Rate Performance

![Fig. 1. Delay-bound violation probability versus target packet error rate](image-url)
Fig. 2 depicts $P_{\text{loss}}$ versus $P_0$ with the reference parameters and $b=2$ at different average SNR per receive antenna. On each curve in Fig. 2, the minimum value of $P_{\text{loss}}$ is indicated by an arrow; the corresponding $P_0$ is the solution of the following optimization problem

$$P_0^{\text{opt}} = \arg \min_{P_0} P_{\text{loss}}(P_0) \quad (11)$$

From (4), we observe that $P_0$ determines the mode switching levels of AMC scheme, and consequently affects the probability of mode selection in (3). As a result, $P_0$ indirectly affects $P_q$ in (10). On the other hand, $P_0$ directly controls the packet error rate over wireless links at the physical layer and consequently affects $P_{\text{loss}}$ in (9) in two different ways. Hence, a simple approach is to numerically find the optimal $P_0^{\text{opt}}$, denoted by $P_0^{\text{opt}}$, among all possible choices, such that $P_{\text{loss}}$ in (9) is minimized. Note that the optimal $P_0^{\text{opt}}$ in (11) is obtained from a packet loss rate QoS measure that is defined in the data link layer, and therefore, it offers a cross-layer design indeed. In fact, such designs improve the system performance in comparison to the case, where $P_0$ is set only based on the physical layer characteristics without taking into account the upper layers.

In Figs. 3 and 4, we plot $P_{\text{loss}}$ and $P_q$ versus $P_0$ for three different values of bandwidth coefficient $b$. In these figures, we set the average SNR per receive antenna to 15dB. From Fig. 4, we see that increasing the value of $b$, results in increased service rate for the queue, which in turn leads to the reduction of $P_q$ and thus, $P_{\text{loss}}$.

B. Resource Allocation for QoS Guarantees in Multi-user Scenario

One application of our analytical model is guaranteeing QoS for real-time services in multi-user environments such as time division multiplexing/time division multiple access (TDM/TDMA) system, where users are multiplexed to transmit in different time slots. In such systems, simultaneously guaranteeing QoS and utilizing resources efficiently (for instance, allocating minimal required bandwidth $b_k$ to user $k$) is an interesting research topic. Here, we introduce a simple solution to this problem. Suppose that the prescribed QoS metrics for a real-time user are $P_{\text{loss}} \leq 0.01$; and $P_q \leq 0.01$, corresponding to $D_{\text{max}}=4T_f=8$ ms (delay through wireless link), which are typical QoS requirements for real-time video transmissions [10]. If we use a STBC(2×2) system with the reference parameters and average SNR per receive antenna equal to 15dB, from Fig. 3 and 4, we see that with $b=3$ (minimal required bandwidth) and $P_0$ in the range $[2\times10^{-4}, 6\times10^{-3}]$, the prescribed QoS metrics of the user can be guaranteed (see the blue curves in Figs. 3 and 4). Interestingly,
if we choose $P_0 = 2 \times 10^{-3}$, both QoS metrics are guaranteed and the average throughput of the user is maximized, since $P_{\text{loss}}$ is minimized.

V. CONCLUSIONS

In this paper, we developed a general analytical procedure to investigate the performance of packet transmissions over the MIMO wireless links, where a deadline constrained traffic is considered and packet queuing at the data link layer is coupled with AM at the physical layer. In the proposed modeling, we use the effective bandwidth concept to describe statistical delay-bound for real-time services. Based on our performance analysis, we developed a cross-layer design, which selects the optimal target packet error rate in AM at the physical layer, to minimize the packet loss rate and maximize the average throughput. The proposed cross-layer design has low-complexity and requires minimal cross-layer information exchange.

While infinite buffer length is considered in our performance analysis, the buffer size is of special interest since buffer overflow will introduce a third source of packet loss in the system.

ACKNOWLEDGEMENT

This work is supported by Iran Telecommunications Research Center (ITRC).

REFERENCES