

Maximum Delay-Constrained Source Rate over a Wireless Channel

Beatriz Soret
Department of Ingeniería de
Comunicaciones
University of Málaga
Campus de Teatinos, E-29071
Málaga, Spain
bsoret@ic.uma.es

M. Carmen
Aguayo-Torres
Department of Ingeniería de
Comunicaciones
University of Málaga
Campus de Teatinos, E-29071
Málaga, Spain
aguayo@ic.uma.es

J. Tomás
Entrambasaguas
Department of Ingeniería de
Comunicaciones
University of Málaga
Campus de Teatinos, E-29071
Málaga, Spain
jtem@ic.uma.es

ABSTRACT

Providing Quality of Service (QoS) guarantees in the presence of delay-sensitive data streams demands the understanding of the delay's behavior. Moreover, the support of QoS over a wireless channel comes up against to the time-varying nature of the channel. In this work, we accommodate both issues and evaluate the maximum source rate such that certain delay bound D^t can be supported with a violation probability ϵ . We call this maximum source rate *Capacity with Probabilistic Delay Constraint* $C_{D^t, \epsilon}$. The effective bandwidth theory is a framework widely used to analyze wired networks and constitutes the basis of the work with the necessary adaptation for its use in a wireless system. The time-correlated nature of the wireless channel has been modeled with a Finite State Markov Chain (FSMC). As expected, the maximum $C_{D^t, \epsilon}$ increases for long allowed delays ($D^t \rightarrow \infty$) and diminishes when the delay constraint is more strict. The expected delay violation probability is compared to simulations in order to validate our results.

Categories and Subject Descriptors

G.3 [Probability and Statistics]: Queueing Theory-Large Deviations

General Terms

Theory, Performance

Keywords

Effective bandwidth, wireless systems, Quality of Service, delay

1. INTRODUCTION

The provisioning of Quality of Service (QoS) comes up against to the time-varying nature of the wireless channel. This time-variation or fading is caused by the existence of multiple paths for the signal travelling between the transmitter and the receiver.

The wireless channel has a complex mathematical characterization that turns out to be a handicap for the development of upper layer procedures (e.g. QoS provisioning) where it would be desirable to have information of the behavior of the underlying radio channel. For the incorporation of the channel effects in the design of high layer procedures, simplified models that capture the main properties of the wireless channel while avoiding the mathematical complexity have been developed. Finite State Markov Chain (FSMC) [1] [2] is one of these models and has been incorporated for the representation of flat fading. Thus, the FSMC models the wireless channel while avoiding the underlying mathematical complexity.

Broadband networks are expected to support various applications that can generate a mixture of heterogeneous traffic to the network. Quality of Service (QoS) requirements are different depending on the traffic class and may be mainly expressed in terms of throughput, Bit Error Rate (BER) and delay. Thus, the understanding of the delay's behavior becomes a main issue to be studied for providing QoS guarantees in a wireless systems, specifically when delay-sensitive applications are involved in the communication. In particular, it is feasible to provide probabilistic delay guarantees at the maximum constant source rate such that certain delay bound can be supported with probability $1-\epsilon$. In this paper, the problem of evaluating this maximum source rate, that we have called *Capacity with Probabilistic Delay Constraint*, is tackled based on the effective bandwidth theory.

The theory of large deviations, also known as the theory of "rare events", has provided a basis for the development of the effective bandwidth theory [3] [4], which has been widely used to analyze the statistical multiplexing in wired networks such as asynchronous transfer mode (ATM) networks. The effective bandwidth of a source express the bandwidth (i.e. the service rate) needed to satisfy some QoS requirements that can be expressed in terms of delay and loss constraints. In [5] Wu and Negi developed a link-layer channel model dual to the effective bandwidth model suitable for wireless communication systems. Two functions

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Valuetools '07, October 23-25, 2007, Nantes, France
Copyright 2007 ICST 978-963-9799-00-4.

are employed to defined their link-layer channel model: the probability of nonempty buffer, and the QoS exponent of a physical channel or, equivalently, its Effective Capacity. The main difference with the classical effective bandwidth application in ATM is that in wireless communications the service rate is no longer constant but variable, due to the time-varying nature of the physical channel.

Several works following [5] employed the link-layer model by Wu and Negi and showed that is capable of predicting the QoS metrics under various conditions. In [6] the effective capacity model is applied to the transmission of variable rate traffic and measurements are made for several source and channel realizations. In [7] information from an ON-OFF source is transmitted over a channel modelled as a FSMC and they evaluate the source effective bandwidth which can be transmitted with delay probabilistically bounded. In [8] authors also employ an FSMC to model a Multiple Input Multiple Output (MIMO) channel with adaptive modulation and coding and evaluate the effective capacity function parameterized by the QoS exponent. However, their approach differs of ours as they do not derive the maximum source rate for the achieving of certain QoS guarantees. In [9] they evaluate the optimum power control for an uncorrelated fading channel in order to maximize the effective capacity under certain QoS exponent.

We have already calculated the Capacity with Probabilistic Delay Constraint for the case of uncorrelated wireless channel, i.e., a channel with no correlation among samples [10]. Moreover, the extension to the general time-correlated was made based on a semianalytic procedure. Here, an analytical result is given in a time-correlated channel based on the employment of a FSMC for the modeling of the physical layer.

The remainder of the paper is organized as follows. Section 2 briefly describes the queueing system model including the effective bandwidth tool. Section 3 presents the main properties of the wireless channel and how it can be modelled with a FSMC. In Section 4 we make some previous calculations to evaluate the Capacity with Probabilistic Delay Constraint in Section 5. The capacity is also calculated for the case of uncorrelated channel in Section 6 and Section 7 compares both results. The expected delay violation probability is compared to simulations under several conditions in order to validate our results in Section 8. Finally, some concluding remarks are given in Section 9.

2. QUEUEING SYSTEM ANALYSIS

The analysis of the behavior of the delay over a wireless link is carried out through a queueing system as shown in Fig. 1. Attention must be paid to the fact that the server is representing the transmission rate in a wireless channel and so the service rate is a function of the instantaneous response of the channel which, as it will be detailed in Section 3, is an autocorrelated (and therefore time-variant) random process. Thus, the system under study is a G/G/1 queue with autocorrelated service demands.

Using a fluid model, the arrival process has an instantaneous rate $a(t)$ and the wireless channel can serve the queue with rate $c(t)$. Along this work, $a(t)$ is assumed to be constant. The model for $c(t)$, on the basis of the transmission with adaptive modulation and coding, is detailed in Section 3.2.

As the transmission works on a symbol by symbol ba-

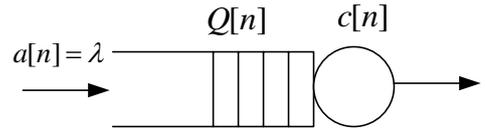


Figure 1: Queueing System.

sis, functions can be time-discretized with symbol period T , $f[n] = f(nT)$, so that $a[n]$ and $c[n]$ are the amount of bits per symbol generated by the source and transmitted by the server, respectively. Thus, $A[n]$ is the amount of bits generated by the source from 0 to time $n-1$, $A[n] = \sum_{m=0}^{n-1} a[m]$ and, analogously, $C[n]$ represents the accumulated transmission rate, $C[n] = \sum_{m=0}^{n-1} c[m]$.

The queue size is supposed to be infinite and $Q[n]$ denotes the amount of data in the buffer. The dynamic of a generic G/G/1 system is characterized by the discrete-time Lindley's equation:

$$Q[n] = \left(\sum_{m=0}^{n-1} (a[m] - c[m]) \right)^+ = (A[n] - C[n])^+ \quad (1)$$

with $(x)^+ \triangleq \max(0, x)$.

The asymptotic log-moment generating function of $Q[n]$ is defined as:

$$\Lambda(u) = \lim_{n \rightarrow \infty} \frac{1}{n} \log E \left[e^{uQ[n]} \right] \quad (2)$$

Since $a[n]$ and $c[n]$ are independent of each other, $\Lambda(u)$ may be decomposed in two terms, $\Lambda(u) = \Lambda_A(u) + \Lambda_C(-u)$, where $\Lambda_A(u)$ and $\Lambda_C(u)$ are the log-moment generating functions of $A[n]$ and $C[n]$, respectively.

If the arrival and service processes are stationary and the steady state queue length exists, (i.e. $\sup_n Pr \{Q[n] > B\} = Pr \{Q(\infty) > B\}$), then the workload process $Q[n]$ satisfies a Large Deviation Principle (LDP) and the following asymptotic behavior for the queue length exceeding a certain threshold B is satisfied [3]:

$$Pr\{Q(\infty) > B\} \asymp e^{-\theta B} \quad B \rightarrow \infty \quad (3)$$

where $f(x) \asymp g(x)$ means that $\lim_{x \rightarrow \infty} f(x)/g(x) = 1$ and θ is the solution to [11]:

$$\Lambda_A(u) + \Lambda_C(-u) |_{u=\theta} = 0 \quad (4)$$

Defining $\alpha(u) = \Lambda(u)/u$ (the so-called *effective bandwidth function*), the equation for the obtaining of θ (usually termed as QoS exponent) can be also expressed as:

$$\alpha(\theta) = \alpha_A(\theta) + \alpha_C(-\theta) = 0 \quad (5)$$

A more accurate approximation for small values of B includes the probability that the queue is not empty, denoted by η , and, therefore, the following less conservative approximation for the tail probability of the queue is satisfied:

$$Pr\{Q(\infty) > B\} \approx \eta \cdot e^{-\theta B} \quad (6)$$

with $\eta = Pr\{Q[n] > 0\}$

In this paper the arrival process has constant rate, $a[n] = \lambda$. In this particular case, the delay experienced by a bit leaving a First In First Out (FIFO) queue at time n will be $D[n] = Q[n]/\lambda$. Thus, the probability of exceeding a target delay D^t (referred throughout this paper as ε) can be written as follows:

$$\epsilon = Pr\{D(\infty) > D^t\} \approx \eta \cdot e^{-\theta\lambda D^t} \quad (7)$$

This approximation may be interpreted as the complementary of the tail of the probability distribution function of the delay, so that the distribution of the tail of the delay (PDF) is obtained as:

$$F(D^t) = Pr\{D(\infty) < D^t\} \approx 1 - \eta \cdot e^{-\theta\lambda D^t} \quad (8)$$

Notice that the fall follows an exponential rule as the parameter η approaches 1 when D^t increases. Moreover, the tail of the probability density function (pdf) is obtained by taking derivative of the PDF, resulting in the following expression:

$$f(D^t) = \frac{d(F(D^t))}{dD^t} \approx \theta \cdot \lambda \cdot \eta \cdot e^{-\theta\lambda D^t} \quad (9)$$

In (7), (8) and (9) the parameter to be obtained is θ . In order to evaluate this QoS exponent it is necessary to evaluate the log moment generating function for the source, $\Lambda_A(u)$, and the channel, $\Lambda_C(u)$.

Calculation of the $\Lambda_A(u)$ is straightforward. As the arrival process is assumed to be constant, evaluating the log moment generating function for $A[n]$ results in:

$$\Lambda_A(u) = \lim_{n \rightarrow \infty} \frac{1}{n} \log E \left[e^{uA[n]} \right] = \lambda \cdot u \quad (10)$$

Evaluation of $\Lambda_C(u)$ is more complex and is derived in Section 4. First, the main properties of the wireless channel are to be detailed.

3. WIRELESS CHANNEL MODELLING

In this section, we revise the main properties of the wireless channel and how it can be modelled with a Finite State Markov Chain.

3.1 Rayleigh Fading Wireless Channel

3.1.1 Multipath propagation

In wireless systems, signals travel through multiple paths between the transmitter and the receiver. Due to these multiple ways, the received signal is formed as the addition of different constructive and destructive components that the receiver perceives as variations of the amplitude, phase and angle of arrival of the signal. This phenomenon is known as multipath fading [13]. The received signal is therefore a set of attenuated, time-delayed, phase shifted replicas of the transmitted signal.

The fading is categorized into two groups: large-scale and small-scale fading. The large-scale fading refers to variations that occur over relatively large distances. The small-scale fading, characterizes the effects of small changes in the separation between a transmitter and a receiver. These changes can be caused by the mobility of the transmitter, the receiver or the intermediate objects in the path of the signal. Variation due to small-scale fading occurs over very short

distances, on the order of the signal wavelength. No deterministic model for the phenomenon exists and hence it is characterized statistically. When there is not predominant direct line of sight between the transmitter and the receiver (NLOS, Not Line Of Sight), the Rayleigh distribution approximates quite well the channel envelope and the fading is denoted Rayleigh fading. In the LOS case (Line of Sight, i.e., the direct line of sight dominates), the most suitable distribution to approximate the signal envelope is the Rician distribution.

Small-scale fading can be further divided into two types: frequency selective and frequency non-selective fading. The latter is also known as flat fading because all the frequency components of the transmitted signal are affected by the channel in approximately the same way. In this case, the effect of the channel over the transmitted signal can be represented through the low-pass equivalent as in Fig. 2. $x(t)$ and $y(t)$ are complex signals that correspond to the low-pass equivalent of the modulated transmitted and received signal, respectively. $h(t)$ denotes the complex low-pass equivalent of the response of the channel.

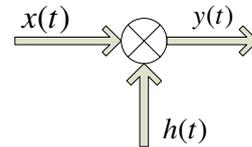


Figure 2: Low-pass equivalent of the flat fading channel.

In this paper we focus on channels with Rayleigh fading and non-frequency selective (flat fading).

3.1.2 Autocorrelation and Power Spectral Density

Another characteristic of the multipath channel is its time-varying nature, since either the transmitter or the receiver is in motion and therefore the location of reflectors in the transmission path, which give rise to multipath, will change over time. The coherence time of the channel T_C is defined as the time over which the channel response can be considered invariant.

Fading can also be studied in the frequency domain. Whenever there is relative motion between the receiver and the transmitter, the received signal suffers a frequency shift which is just the manifestation of the fading phenomenon in the frequency domain. The maximum frequency shift is characterized by the Doppler frequency, f_d , which is computed as v/λ , where v is the relative velocity between the transmitter and receiver and λ is the wavelength of the transmitted signal. The Doppler shift and the coherence time are inversely proportional to one another, that is:

$$T_C \approx \frac{1}{f_d} \quad (11)$$

The variability of the wireless channel over time is usually reflected through its autocorrelation function. This second-order statistic generally depends on the propagation geometry, the velocity of the mobile and the antenna characteristics. A uniform scattering environment is a common assumption that means that the channel consists of many scatterers densely packed with respect to angle. The impulse

response of the channel is a wide-sense stationary random process (WSS) and thus the continuous-time autocorrelation function of the received signal does not depend on the time t but just on the time difference τ :

$$R(\tau) = J_0(2\pi f_d \tau); \quad (12)$$

where $J_0(\cdot)$ is the zeroth order Bessel function of the first kind and f_d is the maximum Doppler frequency in Hertz. In the discrete-time domain, it is more convenient to express the Doppler frequency normalized by the sampling rate, i.e., multiplied by the symbol period T_S ($f_d \cdot T_S$).

Another representation of the autocorrelation is in the frequency domain. Thus, the power spectral density (PSD) is obtained by taking the Fourier transform of $R(\tau)$ relative to the time parameter τ :

$$S(f) = \begin{cases} \frac{1}{\pi f_d \sqrt{1 - (\frac{f}{f_d})^2}}, & |f| \leq |f_d| \\ 0 & elsewhere \end{cases} \quad (13)$$

From the observation of the autocorrelation function it can be concluded that a low value of the product $f_d \cdot T_S$ implies high correlation in the signal and, analogously, a high value of $f_d \cdot T_S$ means low correlation. In the limit ($f_d \cdot T_S \rightarrow \infty$) there is no correlation and the samples are independent of each others.

Before concluding this section, we introduce the concept of uncorrelated channel. The uncorrelated channel (also termed as block fading channel [15]) is used when the channel response is the same over constant-sized blocks of channel uses so that there is no correlation from one block to the next. The uncorrelated channel approximation suits well multicarrier systems, where different carriers (frequency diversity) play the role of time-separated blocks (time diversity). We refer to this case hereafter as block fading or uncorrelated or time-independent channel.

3.1.3 Signal to Noise Ratio

The instantaneous Signal to Noise Ratio at the receiver, $\gamma(t)$, is widely employed as a good indicator of the state of the channel. When the channel is "bad", the signal is severely degraded during its route between transmitter and receiver and the instantaneous SNR decreases. A high value of instantaneous SNR indicates that the channel is "good" and the signal is hardly affected by channel degradation.

With the previous definition of $h(t)$ as the complex response of the channel, the instantaneous SNR is proportional to the square of $|h(t)|$:

$$\gamma(t) = |h(t)|^2 \frac{E_s}{N_0} \quad (14)$$

where E_s is the average energy per symbol and N_0 is the noise power spectral density. If the noise is Additive White Gaussian Noise (AWGN), $\gamma(t)$ is exponentially distributed for Rayleigh channels, with probability density function:

$$p(\gamma) = \frac{1}{\bar{\gamma}} e^{-\frac{\gamma}{\bar{\gamma}}} \quad (15)$$

where $\bar{\gamma}$ is the average Signal to Noise Ratio.

3.2 Service rate in an Adaptive Modulation and Coding scheme

Adaptive Modulation and Coding schemes (AMC) is a technique widely employed in current wireless systems that modifies transmission parameters such as constellation size and coding rate dynamically, trying to adapt to the time-varying conditions of the channel. The basic premise is to estimate the channel at the receiver and feed this estimate back to the transmitter, so that the transmission scheme can be adapted relative to the channel characteristics.

A constellation of m_i symbols is employed within the fading region (Γ_i, Γ_{i+1}), $i = 0, 1, \dots, M$ (defining $\Gamma_0 = 0$). The SNR thresholds may be designed, for example, to fulfill an instantaneous Bit Error Rate (I-BER) condition [14]. In Fig. 4, the curves of the BER for several constellations are presented as a function of the received SNR. Thus, the curve BER^{BPSK} represents the instantaneous BER when using BPSK and analogously for the rest of the curves. A target BER, BER^t , is fixed as a possible QoS requirement (e.g. 10^{-2} in the figure). Then, the crossing of the different curves with the straight line set by BER^t determines the thresholds Γ_i that may be defined to keep the instantaneous BER below the desired BER^t whatever the time-varying SNR is.

In short, the selection of region i means the utilization of certain transmission (or service) rate, which express the number of bits per second that are available for user transmission in that region. Thus, the service rate $c[n]$ is a function of the instantaneous SNR $\gamma[n]$ through the following expression:

$$r(\gamma) = \log_2(m_i), \Gamma_i \leq \gamma < \Gamma_{i+1}, i = 0..M \quad (16)$$

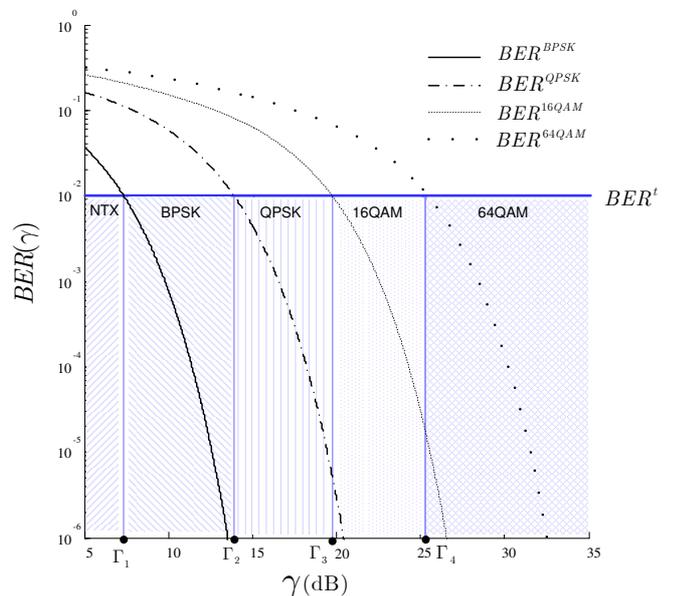


Figure 4: Adaptive Modulation.

With the employment of coding techniques, bit errors introduced by transmission can be either detected or corrected by the receiver. If γ is small, a stronger error correction may be used and consequently the number of bits for user data transmission given by (16) decreases. For high values of γ ,

Table 1: AMC parameters

	State 0	State 1	State 2	State 3	State 4	State 5
Modulation	BPSK	QPSK	QPSK	16-QAM	16-QAM	64-QAM
Coding Rate	1/2	1/2	3/4	9/16	3/4	3/4
Service Rate r_i (bits/symb)	0.50	1.00	1.50	2.25	3.00	4.50
Γ_i (dB)	-1.5331	1.0942	3.9722	7.7021	10.2488	15.9784

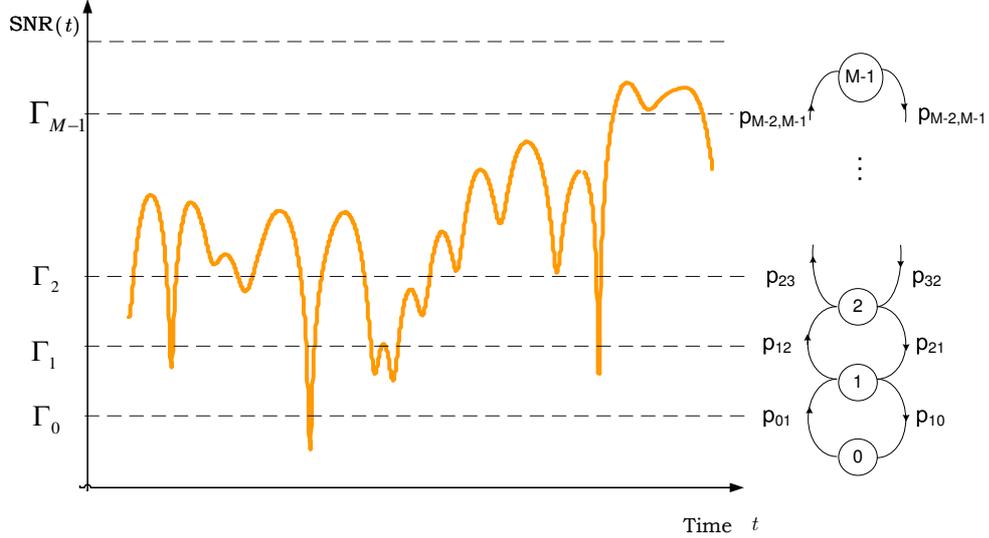


Figure 3: Representation of a Rayleigh channel with a FSMC.

a weaker error correction or no coding may be used.

For the numerical examples and simulations hereafter in this work, six AMC modes are defined with the parameters shown in Table 1 [12].

3.3 Channel Modeling with a FSMC

In a Finite State Markov Chain, the range of received SNR is divided into several consecutive regions. Region i is mapped into state i of the chain and is delimited by two thresholds, Γ_i and Γ_{i+1} (see Fig. 3). Thus, each state of the chain represents one fading region. The steady state probability for state i is just the probability that the exponentially distributed SNR is between the thresholds of the region:

$$\pi_i = \int_{\Gamma_i}^{\Gamma_{i+1}} p(\gamma) d\gamma = e^{-\frac{\Gamma_i}{\bar{\gamma}}} - e^{-\frac{\Gamma_{i+1}}{\bar{\gamma}}} \quad (17)$$

An AMC scheme over a fading channel can be directly mapped into a FSMC. Thus, each fading region is characterized by a constellation size and a coding rate.

The Level Crossing Rate (LCR) for the instantaneous signal to noise ratio process is the average number of times per unit interval that a fading signal crosses a given signal level Γ . For a random distribution of direction of motion providing a maximum Doppler frequency f_d , it can be shown that the level crossing rate of level Γ in the positive direction only (or in the negative direction only) is:

$$N(\Gamma) = \sqrt{\frac{2\pi \cdot \Gamma}{\bar{\gamma}}} f_d \cdot \exp\left(-\frac{\Gamma}{\bar{\gamma}}\right) \quad (18)$$

The transition probabilities from state i to state $i + 1$, $p_{i,i+1}$, can be approximated by the ratio of the level crossing rate at threshold Γ_{i+1} and the average number of packets per second staying in state i . Similarly, the transition probability $p_{i,i-1}$ is approximately the ratio of the LCR at threshold Γ_i and the average number of packets per second staying in state i :

$$p_{i,i+1} \approx \frac{N(\Gamma_{i+1}) \cdot T_S}{\pi_i} \quad (19)$$

$$p_{i,i-1} \approx \frac{N(\Gamma_i) \cdot T_S}{\pi_i} \quad (20)$$

where T_S is the period symbol.

It is assumed that transitions only happen to adjacent states, i.e. $p_{k,i} = 0$, $if |k - i| > 1$.

4. EVALUATION OF Λ_C

The wireless channel is represented in this paper with a FSMC. Moreover, AMC schemes are employed so that discrete transmission rates have to be assumed and the selection of region i (state i of the chain) means the utilization of a transmission rate r_i , which represents the number of bits per second that are available for user transmission in that region. Thus, the evaluation of the asymptotic log-moment generating function for the service process reduces to the well-known result for the asymptotic log-moment generat-

$$\mathbf{P} \cdot \mathbf{C}(u) = \begin{bmatrix} p_{00} \exp(r_0 u) & p_{01} \sqrt{\frac{\pi_0}{\pi_1}} \exp(u \frac{r_0+r_1}{2}) & 0 & 0 & \dots \\ p_{10} \sqrt{\frac{\pi_1}{\pi_0}} \exp(u \frac{r_0+r_1}{2}) & p_{11} \exp(r_1 u) & p_{12} \sqrt{\frac{\pi_1}{\pi_2}} \exp(u \frac{r_1+r_2}{2}) & 0 & \\ 0 & p_{21} \sqrt{\frac{\pi_2}{\pi_1}} \exp(u \frac{r_1+r_2}{2}) & p_{22} \exp(r_2 u) & p_{23} \sqrt{\frac{\pi_2}{\pi_3}} \exp(u \frac{r_2+r_3}{2}) & \\ 0 & 0 & p_{32} \sqrt{\frac{\pi_3}{\pi_2}} \exp(u \frac{r_2+r_3}{2}) & p_{33} \exp(r_3 u) & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (21)$$

ing function of a FSMC [16]:

$$\Lambda_C(u) = \log(\Omega(\mathbf{P} \cdot \mathbf{C}(u))) \quad (22)$$

where \mathbf{P} is the transition probabilities matrix, $\mathbf{C}(u)$ is a diagonal matrix whose diagonal elements are of the form $c_{ii} = \exp(r_i u)$, r_i are the bits per second transmitted when region i is selected and $\Omega(\cdot)$ is the spectral radius, defined as the maximum of the absolute values of the eigenvalues of the matrix.

Thus, the evaluation of $\Lambda_C(u)$ leads to the evaluation of the spectral radius of $\mathbf{P} \cdot \mathbf{C}(u)$. Recall that $\mathbf{P} \cdot \mathbf{C}(u)$ is a tridiagonal matrix with general form:

$$\mathbf{P} \cdot \mathbf{C}(u) = \begin{bmatrix} p_{00} e^{r_0 u} & p_{01} e^{r_1 u} & 0 & 0 & \dots \\ p_{10} e^{r_0 u} & p_{11} e^{r_1 u} & p_{12} e^{r_2 u} & 0 & \\ 0 & p_{21} e^{r_1 u} & p_{22} e^{r_2 u} & p_{23} e^{r_3 u} & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (23)$$

Symmetric matrices are normally preferred as they are much more tractable mathematically. $\mathbf{P} \cdot \mathbf{C}(u)$ can be easily symmetrized obtaining matrix (21).

The eigenvalues of a tridiagonal symmetric matrix S can be calculated recursively with the following recurrence:

$$\det S = s_{n,n} \cdot \det S_{1,\dots,n-1} - s_{n,n-1} \cdot s_{n-1,n} \cdot \det S_{1,\dots,n-2}$$

where $\det S_{1,\dots,k}$ denotes the k th principal minor, i.e., the submatrix formed by the k th first rows and columns of S .

Unfortunately, if the number of regions is greater than three, the resultant polynomial to be solved has not explicit solution and therefore the eigenvalues of the matrix cannot be derived in a close form. The usual number of regions employed in AMC schemes is between five and eight (six in this work as shown in Table 1). Thus, we conclude that the result for the log moment generating function of the service rate has to be given in terms of the spectral radius as in (22).

5. CAPACITY WITH PROBABILISTIC DELAY CONSTRAINT

In this section, the log-moment generating functions of the source and the channel are used to obtain the capacity with probabilistic delay constraint in a Rayleigh fading channel.

The application of the large deviation principle provides a probabilistic approximation of delay violation in the queue. We work out the value of λ in (7):

$$\lambda \approx - \frac{\log \frac{\varepsilon}{\eta}}{\theta \cdot D^t} \quad (24)$$

In a high load scenario, the probability that the buffer is empty approaches one, i.e. $\eta \rightarrow 1$.

The value of the QoS exponent θ is obtained by replacing (5) with (10) and (22):

$$\lambda \cdot \theta + \log(\Omega(\mathbf{P} \cdot \mathbf{C}(-\theta))) = 0 \quad (25)$$

When λ is evaluated with (24), with $\eta \approx 1$ and with the value of the QoS exponent that fulfills (25), the resulting source rate represents the maximum source rate that can be supported with a given delay QoS requirements, expressed in terms of target delay and probability of exceeding the target delay. We call this value Capacity with Probabilistic Delay Constraint and denote it C_{ε, D^t} :

$$C_{\varepsilon, D^t} \triangleq - \frac{\log \varepsilon}{\theta \cdot D^t} \quad (26)$$

where θ is the solution to:

$$- \frac{\log \varepsilon}{D^t} + \log(\Omega(\mathbf{P} \cdot \mathbf{C}(-\theta))) = 0 \quad (27)$$

Naturally, source rates below C_{ε, D^t} will carry out with the delay QoS requirements.

Some limitation on the values of C_{ε, D^t} are foreseen before evaluating (26). For low values of the target delay D^t or ε , the wireless channel tolerates lower traffic arrival rates in order to guarantee the delay constraints and $C_{D^t, \varepsilon}$ may decrease. On the other side, high D^t values or $\varepsilon \rightarrow 1$ means less strict QoS requirement and a higher value of $C_{D^t, \varepsilon}$ is possible. Nevertheless, it can never be above the ergodic capacity of the channel, m_c , which determines the maximum achievable rate of the fading channel and is the theoretical upper bound [15].

6. $C_{D^t, \varepsilon}$ FOR UNCORRELATED CHANNELS

Before evaluating numerically C_{ε, D^t} , the case of time-independent channel is studied in order to compare later with the correlated one. The calculation of Λ_C is much easier and a closed form for the Capacity with Probabilistic Delay Constraint can be derived. Notice that the time-independent channel represents the more optimistic case and it will provide the maximum source rate [5].

In the uncorrelated channel the channel response is the same over constant-sized blocks of channel uses. Then, there is no correlation from one block to the next and the accumulated transmission rate $C[n]$ is simply the addition of n uncorrelated and identically distributed random variables:

$$C[n] = \sum_{i=0}^{n-1} c[i] \quad (28)$$

As $n \rightarrow \infty$ the law of the large numbers can be applied and the distribution of $C[n]$ approximates a gaussian, simplifying

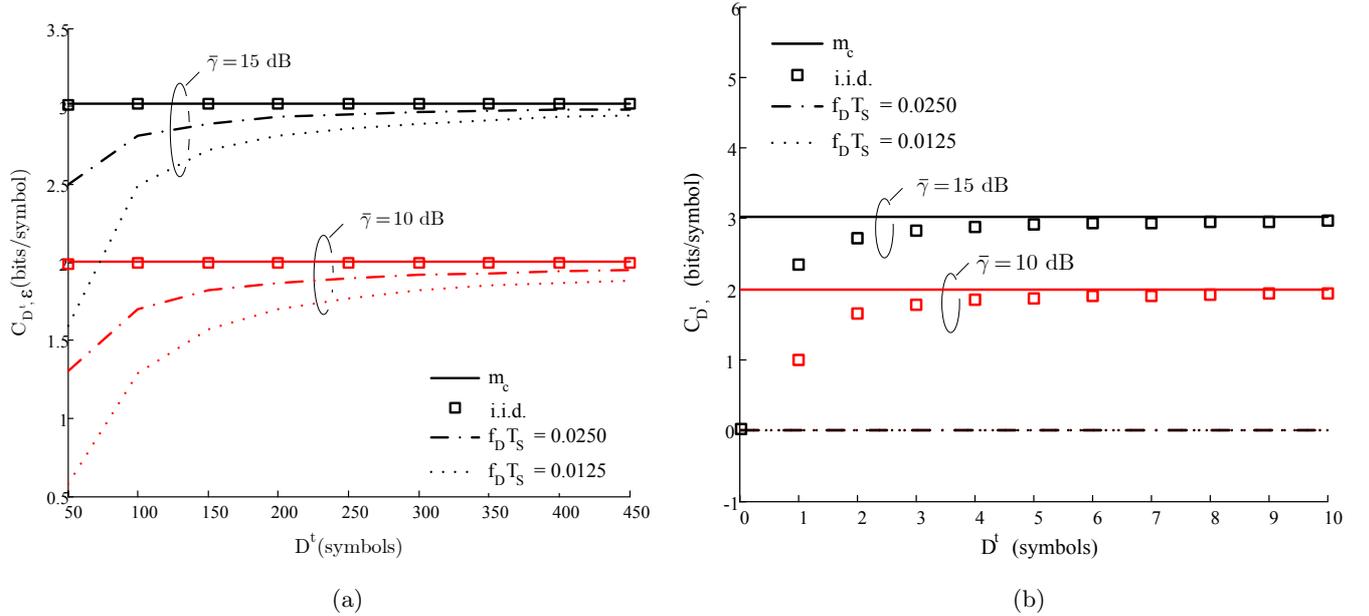


Figure 5: Comparison of the theoretical ergodic capacity with the delay-constrained capacity under several average Signal to Noise Ratios and delay parameters. (a) target delay from 50 to 450 symbols (b) target delay from 0 to 10 symbols

the mathematical processing. The log moment generating function for the resulting normal distribution of $C[n]$ is:

$$\Lambda_C(u) = \lim_{n \rightarrow \infty} \frac{1}{n} \log E \left[e^{u C[n]} \right] = m_c \cdot u + \frac{u^2}{2} \sigma_c^2 \quad (29)$$

where m_c and σ_c are the mean and the variance of the instantaneous channel capacity $c[n]$ and are easily computed for an AMC scheme [14] [17].

Once the asymptotic log moment generating function has been calculated, the value of the QoS exponent is derived by replacing (5) with (10) and (29):

$$\theta = \frac{2(m_c - \lambda)}{\sigma_c^2} \quad (30)$$

The probability that the buffer is empty is supposed again to approach 1, i.e. $\eta \approx 1$. Substituting (30) into (7) the maximum allowable source rate that fulfils the delay QoS requirements in an uncorrelated channel is:

$$C_{D^t, \epsilon} \triangleq \frac{m_c}{2} + \frac{1}{2} \sqrt{m_c^2 - 2\sigma_c^2 \frac{(-\log \epsilon)}{D^t}} \quad (31)$$

The same two foreseen limits predicted in (26) ought to fit in (31). With the close form of $C_{D^t, \epsilon}$ derived above, these limits can actually be checked from the observation of the equation. Thus, for high D^t values or $\epsilon \rightarrow 1$, $C_{D^t, \epsilon}$ approaches m_c , the ergodic capacity of the channel. On the other side, as the target delay D^t or ϵ become lower, the wireless channel tolerates lower traffic arrival rates for ensuring these more strict QoS guarantees. Again, $C_{D^t, \epsilon}$ keeps always below the ergodic capacity, the upper limit that cannot be exceeded.

7. EVALUATION OF $C_{D^t, \epsilon}$

Fig. 5 shows the evaluation of $C_{D^t, \epsilon}$ with equation (26) for the time-correlated channel and with equation (31) for the uncorrelated channel. The capacity is plot as a function of D^t and for different values of $\bar{\gamma}$. The violation probability ϵ is 0.1 and two different values of $\bar{\gamma}$, 10dB and 15dB, are presented. The product $f_d \cdot T_S$ takes two values, $25 \cdot 10^{-3}$ and $12.5 \cdot 10^{-3}$. The time-independent case *i.i.d.* is shown with squares, with the values of m_c and σ_c obtained as in [10].

It can be checked in Fig. 5 (a) that for small values of D^t , $C_{D^t, \epsilon}$ decreases and for high values of for D^t , it approaches m_c which is the limit. It is worth noting that for the same $\bar{\gamma}$ value the time-correlated capacity is always below the time-uncorrelated case, which is an upper bound very close to the limit m_c for the values of D^t drawn in the figure. Moreover, the higher the correlation between samples (i.e. the lower $f_d \cdot T_S$) the lower the capacity fulfilling the same delay QoS requirements. Obviously, for a given parameters of correlation, the maximum source rate for fulfilling certain QoS parameters (D^t, ϵ) is higher for a higher value of SNR.

In Fig. 5 (b) $C_{D^t, \epsilon}$ is again presented but now for smaller values of D^t . Thus, it can be checked that with these strict QoS requirements, the maximum source rate for the time-independent channel separates from the limiting m_c . Moreover, the time-correlated channel cannot fulfil the QoS requirements for the shown values of $f_d \cdot T_S$, leading to a capacity of zero.

8. SIMULATION COMPARISON

A queueing system has been simulated and compared with the analytical result. For certain values of D^t and ϵ , the Capacity with Probabilistic Delay Constraint is evaluated with

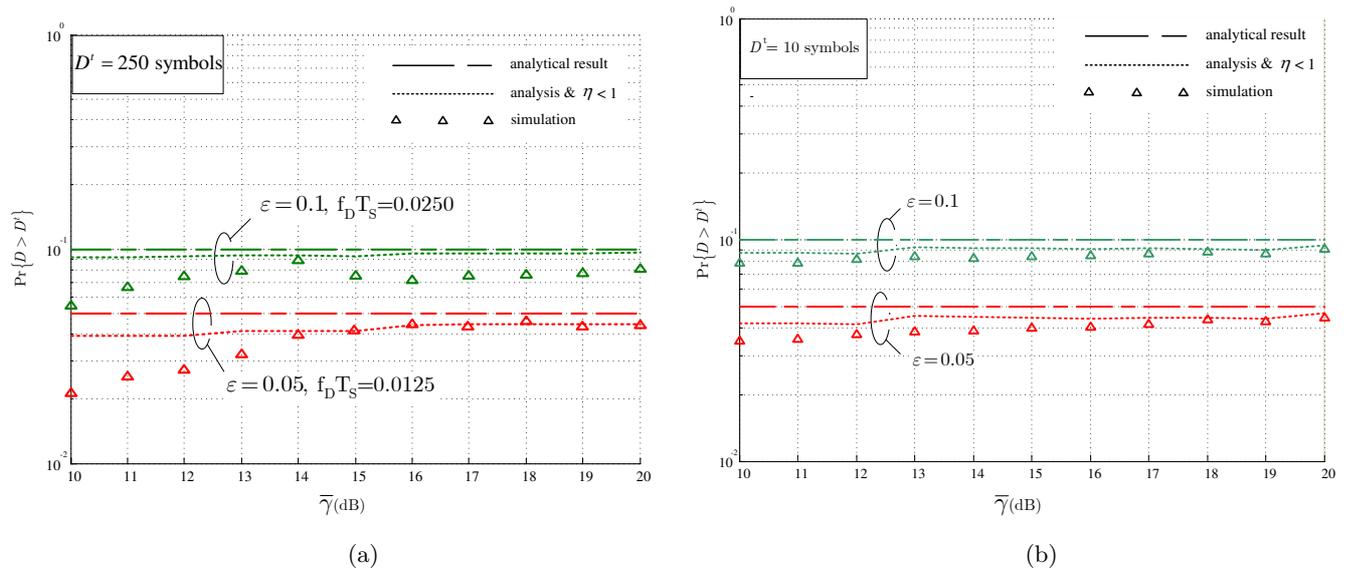


Figure 6: Simulated probability of delay violation versus analytical bound. (a) FSMC channel. (b) Uncorrelated channel.

(26). The arrival process generates source data at the resulting limiting rate, $C_{D^t, \varepsilon}$, and measures of the tail probability of exceeding the delay target are made. Notice that the expected value of this tail probability, $Pr\{D(\infty) > D^t\}$, is ε .

The generation of the service rate $c[n]$ leads to the generation of the process $\gamma[n]$, for which we have used an autoregressive model of order 10 for fading channels with specified parameters mean SNR $\bar{\gamma}$, maximum Doppler frequency f_d and sampling period T_S [18]. In Fig. 6 (a), $Pr\{D(\infty) > D^t\}$ is represented for target delay of 250 symbols and violation probabilities $\varepsilon=0.1$ and $\varepsilon=0.05$. The values of the product $f_d \cdot T_S$ are $25 \cdot 10^{-3}$ and $12.5 \cdot 10^{-3}$. The analytical result modified by the probability of non-empty queue is also shown. It can be observed that the measured violation probability is always lower than ε as the capacity was obtained for $\eta = 1$. It was also checked that as D^t or $\bar{\gamma}$ increases, the term η approaches one as expected.

Fig. 6 (b) presents the results of the simulation of the uncorrelated channel. The service rate $c[n]$ has been now generated from an uncorrelated Rayleigh wireless channel of certain value of $\bar{\gamma}$. $Pr\{D(\infty) > D^t\}$ is represented for target delay of 10 symbols and violation probabilities $\varepsilon=0.1$ and $\varepsilon=0.05$. Once more, the curve including the term η approximates more precisely to simulated results. It can be observed that the approximation is more accurate for the uncorrelated channel but nevertheless the QoS requirements are fulfilled in both cases, which is shown in the fact that the measured violation probability remains always below the target probability ε .

Finally, the tail of the probability density function of the delay predicted by the large deviation approximation has been studied. Fig. 7 compares (9) with the measured values in the simulated scenario, which are represented in form of histogram (a) for the channel modelled with a FSMC and (b) for the uncorrelated channel. $\bar{\gamma}$ is 15dB and the violation probability is $\varepsilon=0.1$. In the FSMC channel, $f_d \cdot T_S$ is set to

$12.5 \cdot 10^{-3}$. The figure shows that the pdf tail approximates precisely the measured histograms for large values of D^t for both cases. In fact, Fig. 7 (b) shows that the approximation is even exact for small values of D^t when the channel is uncorrelated. In the case of a correlated channel, Fig. 7 (a), the approximation is very accurate for delays over 160 symbols for the parameters specified in the simulation.

9. CONCLUSIONS

In this work, the Capacity with Probabilistic Delay Constraint for a Rayleigh fading channel modelled with a FSMC is derived. This analytical solution is based on the effective bandwidth theory and provides the maximum source rate that can be supported for a given delay bound D^t and a probability ε of exceeding the delay bound. Simulation results demonstrate the accuracy of the approximation and have been also compared with the simplest case of uncorrelated channel, where there is no correlation among samples and which represents the more optimistic limit for the source rate, that is, the maximum capacity with probabilistic delay constraints. It has been also observed the precision of the approximation given by the large deviation theory for the tail of the probability density function of the delay.

10. ACKNOWLEDGMENTS

This work has been partially supported by the Spanish Government and the European Union under project TEC2007-67289/TCM.

11. REFERENCES

- [1] H. Shen, N. Moayeri "Finite-State Markov Channel-A Useful Model for Radio Communication Channels", *IEEE Trans. on Vehicular Technology*, vol. 44, no. 1, Feb. 1995.

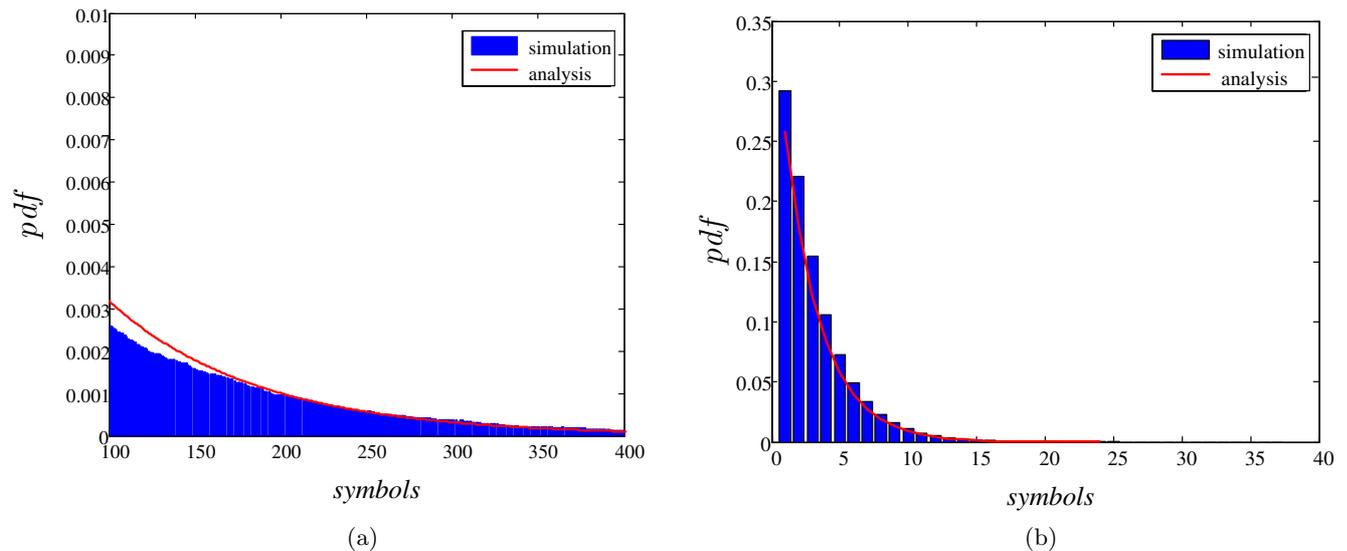


Figure 7: Comparison of the tail of the probability density function and the measured histogram. (a) FSMC channel. (b) Uncorrelated channel.

- [2] Q. Zhang, S. A. Kassam “Finite-State Markov Model for Rayleigh Fading Channels”, *IEEE Trans. on Communications*, vol. 47, no. 11, Nov. 1999.
- [3] C. S. Chang, J. A. Thomas, “Effective Bandwidth in High-Speed Digital Networks”, *IEEE J. Sel. Areas Commun.*, vol. 13, no. 6, Aug. 1995.
- [4] F. P. Kelly, “Notes on Effective Bandwidth”, in F. P. Kelly, S. Zachary, and I. Zeidins (editors) *Stochastic Networks: Theory and Applications*, Royal Statistical Society Lecture Notes Series, 4, pp. 141-168, Oxford University Press, 1996
- [5] D. Wu, R. Negi, “Effective capacity: A wireless link model for support of Quality of Service”, *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 630–643, July 2003
- [6] A. M. A. Quimi, A. Saiadian, A. Mirzaee, “Explicit statistical QoS guarantees for video stream over wireless fading channel”, *Proceedings of IEEE International Conference on Wireless, Networking and Mobile Computing*, vol. 2, pp. 1245-1248, September 2005
- [7] M. Hassan, M. M. Krunz, I. Matta, “Markov-based channel characterization for tractable performance analysis in wireless packet networks”, *IEEE Transactions on Wireless Communications*, vol. 3, no. 3, pp. 821–831, May 2004
- [8] X. Zhang, J. Tang, H. H. Chen, S. Ci, M. Guizani, “Cross-layer-based modeling for quality of service guarantees in mobile wireless networks”, *IEEE Communications Magazine*, pp. 100-106, January 2006
- [9] J. Tang, X. Zhang, “Quality-of-service driven power and rate control in mobile wireless networks”, *Proceedings of IEEE International Conference on Communications*, June 2006
- [10] B. Soret, M. C. Aguayo-Torres, J. T. Entrambasaguas, “Capacity with Probabilistic Delay Constraint for Rayleigh channel”, to be published in *Proc. IEEE GLOBECOM’07*.
- [11] C. S. Chang, T. Zajic, “Effective Bandwidths of Departure Processes from Queues with Time Varying Capacities”, *Proc. IEEE INFOCOM’95.*, pp. 1001-1009, Boston, USA, Apr. 1995.
- [12] Q. Liu, S. Zhou, G. B. Giannakis, “Cross-Layer combining of Adaptive Modulation and Coding with truncated ARQ over wireless links”, *IEEE Transactions on Wireless Communications*, vol. 3, no. 5, pp. 1746–1755, September 2004
- [13] Theodore S. Rappaport, “Wireless Communications: Principles and Practice”, *Prentice-Hall*, New Delhi (India), 2002
- [14] S. T. Chung, A. J. Goldsmith, “Degrees of freedom in adaptive modulation: A unified view”, *IEEE Transactions on Communications*, vol. 49, no. 9, pp. 1561–1571, September 2001
- [15] E. Biglieri, J. Proakis, S. Shamai, “Fading channels: Information theoretic and communication aspects”, *IEEE Transactions on Information Theory*, vol. 44, pp. 2619–2692, October 1998
- [16] G. Kessidis, J. Walrand, C. S. Chang, “Effective Bandwidth for multiclass Markov fluids and other ATM sources”, *IEEE/ACM Transactions on Networking*, vol. 1, no. 4, pp. 424–428, August 1993
- [17] B. Soret, M. C. Aguayo-Torres, J. T. Entrambasaguas, “Evaluation of log moment generation functions for Rayleigh channels”, *Tech. Rep.*, Dept. Ing. Commun., Univ. Malaga, Malaga, Spain, April 2006
- [18] K. E. Baddour, N. C. Beaulieu, “Autoregressive modeling for fading channel simulation”, *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1650–1662, July 2005.