Universidad de Málaga

Escuela Técnica Superior de Ingeniería de Telecomunicación

TESIS DOCTORAL

Analysis of QoS parameters in fading channels based on the effective bandwidth theory

Autora:

Beatriz Soret Álvarez

Directores:

Mª Carmen Aguayo Torres

José Tomás Entrambasaguas Muñoz

UNIVERSIDAD DE MÁLAGA
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA DE TELECOMUNICACIÓN

Reunido el tribunal examinador en el día de la fecha, constituido por:

Presidente: Dr. D. _____

Secretario: Dr. D. _____

Vocales: Dr. D. _____

Dr. D. _____

Dr. D. _____

para juzgar la Tesis Doctoral titulada *Analysis of QoS parameters in fading channels based on the effective bandwidth theory* realizada por Dª Beatriz Soret Álvarez y dirigida por los Prof. Dr. Dª Mª Carmen Aguayo Torres y Dr. D. José Tomás Entrambasaguas Muñoz, acordó por

_____ otorgar la calificación de

_____ y para que conste,

se extiende firmada por los componentes del tribunal la presente diligencia.

Málaga a _____ de _____ del _____

El Presidente: El Secretario:

Fdo.: _____ Fdo.: _____

El Vocal: El Vocal: El Vocal:

Fdo.: _____ Fdo.: _____ Fdo.: _____

To my parents

Table of Contents

Abstract

Providing Quality of Service (QoS) guarantees is an important challenge in the design of next generations of wireless networks. In particular, real-time services involving stringent delay constraints are expected to be increasingly popular among users of mobile equipments. This dissertation analyzes the joint influence of the channel fading and the data outsourcing process in the QoS metrics (with particular emphasis on the delay) of a wireless system. As basis for the analysis we propose the application of the effective bandwidth theory. This theory has been widely employed in wired networks and more recently adapted to wireless communication systems.

Within the effective bandwidth framework, we first do a revision on traffic modeling for wireless networks and study the effective bandwidth functions of several common traffic models. On the channel process side, expressions of the channel effective bandwidth function (also known as effective capacity) are obtained for uncorrelated and time-correlated flat Rayleigh channels, and under different adaptive rate policies. The procedure to obtain these functions is generic and could be applied to other channel models.

The effective bandwidth theory makes feasible the analysis of the distribution tail of the loss probability and the delay. For example, the percentile of the delay and the tradeoff between delay and source rate are obtained. The delay suffered by certain information flow depends not only on the transmission rate but also on the distribution and self-correlation of the information process. Even in wired systems (constant rate channels) different distributions of the information process having the same average rate will cause different delays. Indeed, the better conditions for the delay are obtained when the incoming user traffic is constant. For any other source process, the delay performance degrades.

Besides, we propose the definition of a new capacity in a wireless system, the Capacity with Probabilistic Delay Constraint, $C_{D^t,\,\varepsilon}$. This is the maximum information rate that can be transmitted over the channel under a target BER and a

probabilistic delay constraint given by the pair (D^t, ε), where D^t is the target delay and ε is the probability of exceeding D^t. In short, the new capacity reflects in a single metrics the tradeoff among the channel fading, the source parameters and the QoS requirements.

Next, an OFDM system is tackled with the same techniques applied in the flat channel, assuming now a frequency selective channel that exhibits not only time correlation but also frequency correlation among the subcarriers. The analysis of the delay shows how much the maximum allowable rate decreases with the time or frequency correlation of the channel.

In the last part of the thesis, multiplexing of users over multiple shared fading channel is addressed. A new element comes up in this case: the scheduling algorithm. We propose a multiuser formulation of the system model and redo the analysis, calculating the maximum rate that each user can transmit by fulfilling a target BER and its own delay constraint, and under a given scheduling discipline. The analysis is done first in a single channel link and later on generalized to multiple shared channels employing OFDMA as multiplexing mechanism. Now it is not only the delay constraint and the channel and source process that influence the source rate, but also the discipline that rules the system. Three representative multiplexing algorithms are analyzed: Round Robin, Best Channel and Proportional Fair. The tradeoff between fairness and disciplines is examined with numerical results.

In summary, this thesis shows that the delay exhibits high sensitivity to the burstiness of the traffic, to the time or frequency correlation of the channel and to the scheduling discipline. The results are validated by confronting to simulations. Moreover, the proposed procedure is generic and can be extended to other disciplines and traffic and channel models. Nevertheless, the effective bandwidth function of the source and the channel process cannot always be explicitly evaluated. For such cases, a semi-analytical strategy is also proposed.

Resumen

Garantizar calidad de servicio (Quality of Service, QoS) es un reto importante en el diseño de las próximas generaciones de redes inalámbricas. En concreto, se estima que los servicios de tiempo real, que implican fuertes restricciones de retardo, serán cada vez más populares entre los usuarios de equipos móviles. Esta tesis analiza la relación entre los desvanecimientos del canal, la fuente de información y los parámetros de calidad de servicio (con especial atención al retardo) en un sistema inalámbrico. Se propone la aplicación de la teoría del ancho de banda efectivo como base para el análisis. Se trata de una teoría ampliamente aplicada en redes cableadas y que recientemente ha sido adaptada para su uso en sistemas inalámbricos.

En el marco de la teoría del ancho de banda efectivo, en primer lugar se hace una revisión de modelos de tráfico y se estudia la correspondiente función de ancho de banda efectivo. En lo que se refiere al proceso del canal, se calculan las expresiones de la función ancho de banda efectivo (también conocida como capacidad efectiva) para canales Rayleigh incorrelados y con correlación temporal, y bajo distintas políticas de adaptación de la velocidad. El procedimiento para obtener la función ancho de banda efectivo del canal es genérico y podría ser aplicado a otros modelos de canal.

Una vez estudiadas las funciones ancho de banda efectivo de la fuente y del canal, es posible analizar la cola de la distribución del retardo. Así, se estudian parámetros como el percentil del retardo y la relación entre retardo y velocidad de fuente. El retardo que sufre cierta información no depende sólo de la velocidad de transmisión del canal, sino también de la distribución y la autocorrelación del proceso fuente. Incluso en sistemas cableados (con velocidades de transmisión constantes) sucede que dos procesos fuente con la misma velocidad media pero siguiendo una ditribución distinta darán lugar a distintos valores de retardo. En cualquier caso, las mejores condiciones de retardo se consiguen cuando el usuario genera tráfico a una velocidad constante. Para cualquier otro proceso fuente, las prestaciones en términos de retardo empeoran.

Además, se propone la definición de una nueva capacidad en un sistema inalámbrico, la capacidad con una restricción probabilística de retardo $C_{D^t,\varepsilon}$. Es la máxima velocidad de información que se puede transmitir por un canal bajo una restricción en la BER y una restricción en el retardo, dada esta última por la pareja (D^t, ε), donde D^t es el retardo objetivo y ε es la probabilidad de superar D^t. En definitiva, la nueva capacidad recoge en un único parámetro la relación entre los desvanecimientos del canal, los parámetros de la fuente y los requisitos de QoS.

A continuación, se analiza un sistema OFDM con las mismas técnicas empleadas en el canal plano, suponiendo ahora un canal selectivo en frecuencia que presenta no sólo correlación temporal sino también frecuencial entre las distintas portadoras. El análisis del retardo muestra cuánto decrece la máxima velocidad de fuente con la correlación (temporal o frecuencial) del canal.

En la última parte de esta tesis, se aborda la multiplexación de usuarios en un canal con desvanecimientos. Aparece un nuevo elemento en el sistema: el algoritmo de multiplexación. Se rehace el análisis con el nuevo modelo de sistema. Se calcula la máxima velocidad que puede transmitir cada usuario bajo una BER objetivo y una restricción de retardo (propia de cada usuario), y con una disciplina de multiplexación dada. El análisis se hace primero con un único canal (canal plano) y después se generaliza para multicanal, usando OFDMA como técnica de multiplexación. En este nuevo escenario, no sólo la restricción de retardo y el proceso canal y fuente influyen en la velocidad de fuente admisible, sino que también la disciplina con la que se selecciona al usuario que va a transmitir tiene un gran impacto en los resultados. Se han analizado tres algoritmos de multiplexación habituales: Round Robin, Best Channel y Proportional Fair. El balance entre justicia y disciplina se puede evaluar con los resultados aportados.

En resumen, en esta tesis se ha comprobado la alta sensibilidad del retardo a la variabilidad del tráfico, a la correlación temporal o frecuencial del canal y al algoritmo de multiplexación. Los resultados analíticos presentados se han validado comparándolos con simulaciones. Además, el procedimiento propuesto es genérico y podría extenderse a otras disciplinas, modelos de tráfico y modelos de canal. En

cualquier caso, las funciones de ancho de banda efectivo de una fuente y un canal no siempre se pueden evaluar analíticamente. Para esos casos, se aporta una estrategia semianalítica.

Acknowledgements

First and foremost, I would like to thank my supervisors, Dr. Mari Carmen Aguayo Torres and Dr. José Tomás Entrambasaguas Muñoz, for all they have taught to me. I feel very fortunate to have been working with them; their enthusiasm and dedication to this work were truly inspiring. Without their guidance, help and encouragement this thesis had not been possible.

Secondly, I am also grateful to all the colleagues that make the Department of Ingeniería de Comunicaciones of the University of Málaga a very friendly working environment. I thank all the bright people who provided me comments and feedback on some of the aspects analyzed in this thesis. Particularly, thanks to Jesús López, José Paris, Gerardo Gómez, Juan Sánchez, Maribel Delgado and Javi López.

The invitation of the University of Aalborg in 2007 has contributed substantially to this work. During this stay, I had the excellent opportunity to work with Dr. Ramjee Prasad and Dr. Neeli Prasad, who made me feel part of their team. I would like to express my special gratitude to Maurizio Motta, for his help with English and his continuous encouragement.

Last but not least, I am grateful to my family, my mother and my brother, for their unconditional support. They never let me forget that they are proud of me.

The work in this thesis was mainly performed within the context of the following projects:

- *'Adaptive multilayer resource allocation, with non ideal models, for wireless access networks'. Proyecto del Plan Nacional de I+D: TEC2007-67289*
- *'Modelado de la calidad de servicio extremo a extremo en redes heterogéneas con cooperación entre capas y acceso inalámbrico multicanal adaptativo'. Proyecto de Investigación de Excelencia de la Junta de Andalucía: TIC-03226*

Acronyms

ACF Auto-Correlation Function

AMC Adaptive Modulation and Coding

AR Auto-Regressive

ATM Asynchronous Transfer Mode

AWGN Additive White Gaussian Noise

BC Best Channel

BER Bit Error Rate

BPSK Binary Phase Shift Keying

CBR Constant Bit Rate

CDF Cumulative Distribution Function

CLT Central Limit Theorem

CSI Channel State Information

DQF Delay QoS Factor

EBF Effective Bandwidth Function

EBT Effective Bandwidth Theory

FCFS First Come First Served

FDMA Frequency-Division Multiplexing Access

FIR Finite Impulse Response

FSMC Finite State Markov Chain

GEC Gilbert-Elliot Channel

IDFT Inverse Discrete Fourier Transform

IDP Interrupted Deterministic Process

IFFT Inverse Fast Fourier Transform

IPP Interrupted Poisson Process

IRP Interrupted Renewal Process

LCR Level Crossing Rate

LDP Large Deviation Principle

LDP Large Deviation Principle

LDT Large Deviation Theory

L-MGF Log-Moment Generating Function

LOS Line Of Sight

LRD Long Range Dependence

MGF Moment Generating Function

MIMO Multiple Input Multiple Output

MMPP Markov Modulated Poisson Process

NLOS No Line Of Sight

NTX No Transmission

OFDM Orthogonal Frequency Division Multiplexing

OFDMA Orthogonal Frequency Division Multiple Access

pdf probability density function

PDP Power Delay Profile

PF Proportional Fair

PSD Power Spectral Density

QAM Qadrature Amplitud Modulation

QoS Quality of Service

QPSK Quaternary Phase Shift Keying

rms Root Mean Square

RR Round Robin

SoS Sum of Sinusoids

SNR Signal-to-Noise Ratio

SRD Short Range Dependence

TDMA Time-Division Multiplexing Access

VBR Variable Bit Rate

WSS Wide Sense Stationary

Notation

$a[n]$	Instantaneous source rate
$a_u[n]$	Instantaneous source rate of uth user
$A[n]$	Accumulated source rate
$A_u[n]$	Accumulated source rate of uth user
B_c	Coherence bandwidth
$c[n]$	Instantaneous channel rate
$c^f[n]$	Instantaneous channel rate at subcarrier f
$c_u[n]$	Instantaneous channel rate assigned to uth user
$c_u^f[n]$	Instantaneous channel rate assigned to uth user at subcarrier f
$C[n]$	Accumulated channel rate
$C_u[n]$	Accumulated channel rate assigned to uth user
$\mathrm{C}_{D^t,\varepsilon}$	Capacity with Probabilistic Delay Constraint
$D[n]$	Delay experienced in the queue
D^t	Target delay
$E_1(x) := \int_1^\infty \frac{1}{s} e^{-x} ds$	Exponential integral
f_D	Maximum Doppler frequency
$f_{\mathbf{z}}(z_n, z_{n+m})$	Bivariate pdf of the channel envelope z_n
F	Number of subcarriers in an OFDM system
${}_p\mathbf{F}_q(\boldsymbol{n}, \boldsymbol{d}, z)$	Hypergeometric function
$g \approx 0.5772$	Euler constant
$h[n]$	Complex channel gain at symbol nth
$\mathcal{K}_c(m)$	Autocovariance function of $c[n]$ for a time lag m
$\mathcal{K}_{\lvert h\rvert}(m)$	Autocovariance function of $\lvert h[n]\rvert$ for a time lag m

m_c	Mean of the instantaneous channel rate
$Q[n]$	Queue length at symbol n
$\mathrm{R}^{u}{}_{D^t,\varepsilon}$	Maximum achievable user rate for a delay constraint
$\mathcal{R}_{\lvert h\rvert}(m)$	Autocorrelation function of $\lvert h[n]\rvert$ for a time lag of m
$\mathcal{R}_{\lvert h\rvert^2}(m)$	Autocorrelation function of $\lvert h[n]\rvert^2$ for a time lag of m
T_S	Symbol period
$z_n = \lvert h_n\rvert$	Envelope of the complex channel gain at symbol nth
$\alpha(\upsilon)$	Effective bandwidth function of a stochastic process
$\alpha_A(\upsilon)$	Effective bandwidth function of the arrival process
$\alpha_C(\upsilon)$	Effective bandwidth function of the channel process
$\gamma[n]$	Instantaneous Signal to Noise Ratio
$\bar{\gamma}$	Average Signal to Noise Ratio
$\gamma_u[n]$	Instantaneous SNR of uth user
$\bar{\gamma}_u[n]$	Average SNR of uth user
$\Gamma(x)$	Gamma function
$\varepsilon = Pr\{D(\infty) > D^t\}$	Probability of exceeding the target delay
$\eta = Pr\{Q(\infty) > 0\}$	Probability of non-empty queue
θ	QoS exponent
$\Lambda(\upsilon)$	Asymptotic log-moment generating function of a stochastic process
σ_c^2	Variance of the instantaneous channel rate
$\psi(x) = \frac{d(\log(\Gamma(x)))}{dx}$	Digamma function

Chapter 1

Introduction

The last two decades have witnessed a tremendous growth in the wireless communication industry. First generation (analog voice) and second generation (digital voice and low-rate data) wireless networks have been vastly deployed. While first systems focused on voice services, future generation wireless networks are targeted to transmit multimedia applications, including real-time transport of image, voice and video. The traffic characteristics and real-time nature of these multimedia applications pose new challenges to the design, implementation and management of future high speed networks.

In a multimedia environment, different applications can have very diverse Quality of Service (QoS) requirements, making the provisioning of QoS a non trivial problem to be faced. In particular, real-time services involving stringent delay constraints are expected to be increasingly popular among the users of mobile systems. Furthermore, providing QoS is especially challenging in wireless networks, where low reliability of wireless links along with their time varying nature may result in severe QoS violations.

This thesis addresses the analysis of QoS parameters in a wireless system. Specifically, the joint influence of channel fading and data outsourcing processes in QoS

metrics such as delay, Bit Error Rate (BER) and throughput is analyzed, getting fundamental limits that can be of great use in the planning and design of future wireless networks.

This introductory chapter is organized as follows. The motivation for this research is presented in Section 1.1, while Section 1.2 states the addressed problem. Section 1.3 collects the contributions of the thesis. The structure of the dissertation is summarized in Section 1.4. Finally, the publications resulting from this thesis are listed in Section 1.5.

1.1 Motivation

Providing QoS guarantees to different applications is a key issue in the design of next generation of high-speed wireless networks. Usually, the wireless channel is much more hostile than the models employed in wired communications. The low reliability of the channel and its time varying nature make the provisioning of QoS especially challenging in this kind of networks.

The QoS metrics of interest are likely to vary from one application to another, but are predicted to include measures such as throughput, BER and delay. Unlike traditional data communication, where system performance is largely measured in terms of the average overall throughput and loss rate, real-time communications may require QoS guarantees expressed in terms of the delay, not only the mean delay but other statistical indicators such as the jitter or the P-percentile of the delay [3GPP 2009] [Tang 2007a].

Some applications such as medical applications demand reliable and timely delivery of information. In these cases it is critical to guarantee that no packet is lost or delayed during the packet transmission. We refer to this type of QoS guarantees as *deterministic QoS guarantees* [Wrege 1996] [Wu 2003b]. On the other hand, for a

majority of multimedia applications such stringent QoS requirements are not essential. This is because these applications can in general tolerate a small fraction of lost or delayed packets while maintaining reasonably good quality. We call this type of QoS guarantees *statistical QoS guarantees.* Examples of these *soft* real-time applications are video conferencing, radio/TV broadcast or interactive games. In wireless networks, the channel response varies randomly with time and so does the capacity. Therefore, an attempt to provide deterministic QoS (i.e., requiring zero QoS violation probability) will most likely result in extremely conservative guarantees.

User multiplexing for QoS guarantees is an active research topic in wireless systems [Chen 2006], under different names such as subcarrier and slot allocation, resource allocation or scheduling. Most current traffic sources are variable-rate, and their requirements of transmission resources fluctuate. On the other hand, in a spatially distributed multiuser environment channel quality varies asynchronously for different users. In new wireless systems, adaptive schemes are applied in the transmission so that some parameters such as constellation size and coding rate are modified dynamically, trying to adapt to the time-varying conditions of the channel. Exploiting both the source diversity and the variations in channel conditions can increase the system throughput.

A scheduling scheme ideally should be able not only to handle the uncertainty of the channel but also to exploit it, i.e., opportunistically serve users with good channels. Using such an approach leads to a system capacity that increases with the number of users (due to multiuser diversity) [Knopp 1995]. The performance for a user will depend on the channel condition he experiences and hence different performance is expected when the same resource is assigned to different users.

Many questions regarding the performance of most used opportunistic algorithms are still open. For example, very few works consider the delay or study the treatment given to each user [Berry 2002] [Entrambasaguas 2007]. The main difficulty in obtaining analytical results comes from the fact that the classical queueing theory,

which usually employs fixed rate servers and uncorrelated processes, is no longer suitable. Moreover, the result is linked to the scheduling discipline and the analysis has to be done algorithm by algorithm. To the best of our knowledge, the papers with analytical results in the literature either use simple channel models [Ying 2006] [Neely 2009] or only provide bounds on the QoS metrics [Neely 2003] [Georgiadis 2006]. This observation motivated us to start this work. Thus, the aim of this thesis is to analyze the QoS metrics, in particular the delay, in multiuser wireless systems with variable-rate sources and shared fading channels.

1.2 Problem statement

In the first part of the thesis the single user system depicted in Figure 1.1 is addressed. In this model, data arrive from some higher layer application and are placed into a transmission buffer. Periodically the transmitter removes some of the data from the buffer to be transmitted over the wireless channel. It is assumed that the transmitter can allocate communication resources based on both the buffer occupancy and its perfect knowledge of the channel.

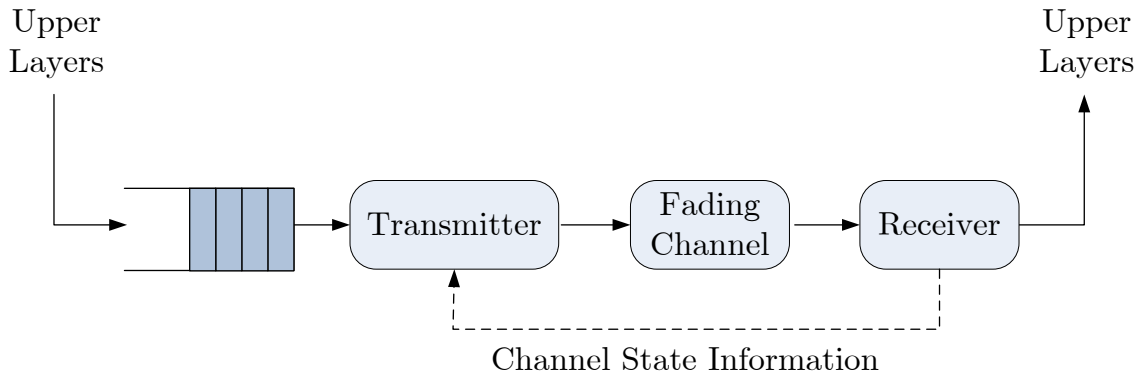


Figure 1.1: System model.

The equivalent discrete-time queueing model in Figure 1.2 is used to tackled the analysis of the information processing. The source process characterizes the

incoming user traffic and the server represents the information transmitted to the wireless channel. The channel is assumed to be a Rayleigh fading channel. Both processes (source and channel) are variable and possibly correlated, making the classical queueing theory not applicable. Following the work in [Wu 2003b] we propose the application of the effective bandwidth theory [Chang 1995a] to analyze the system.

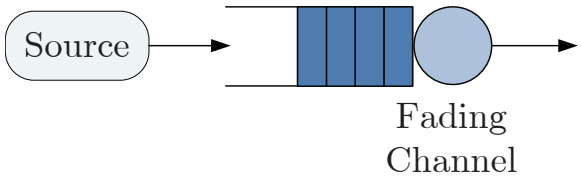


Figure 1.2: Queueing system.

One of the factors that affects the delay, but surely not the only one, is the variable channel transmission rate, which depends on the random response of the wireless channel. The bursty nature of the traffic source is also harmful to the delay suffered by an information flow. Even in wired systems (constant rate channels) different distributions of the information process having the same average rate will cause different delays [Entrambasaguas 2007]. By means of the effective bandwidth theory, we study the tradeoff among the following three elements: traffic source, channel process and QoS requirements (throughput, delay, BER).

In the second part of the dissertation, we extend the system making it multiuser (Figure 1.3), and redo the analysis.

The channel is shared among U users whose incoming traffics are characterized by U source processes, respectively. A new element comes up in this multiuser scenario: the scheduling discipline. The scheduler allocates the resources to users based on the buffer occupancy, the channel state information and the selected scheduling discipline. Together with the channel process and the traffic source, the scheduling scheme has deep influence in the QoS guarantees that each user can expect.

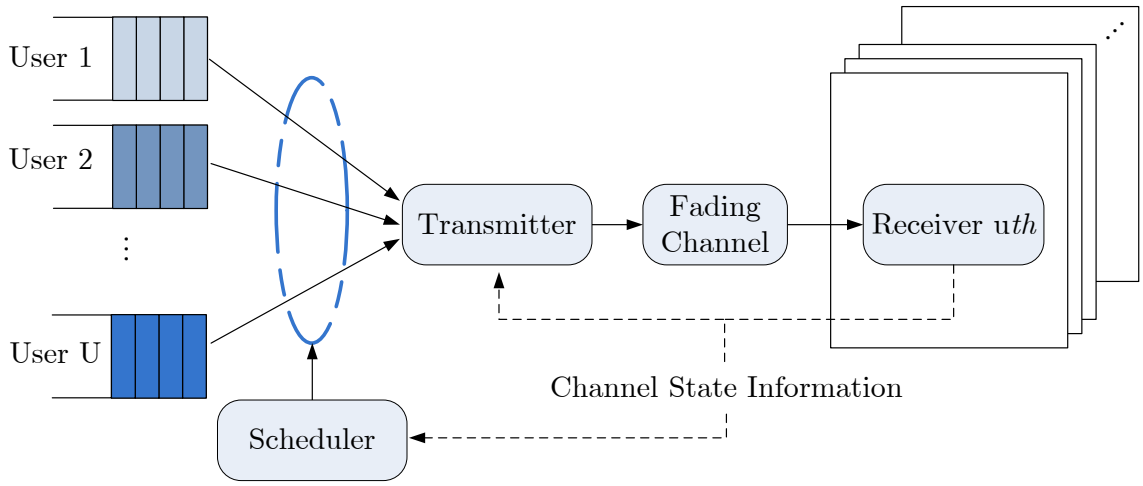


Figure 1.3: Multiuser system model.

The channel can be a single link or a set of physical channels. In this case, the transmission medium is divided into units (resources) to be allocated to users like, for example, in OFDMA (Orthogonal Frequency Division Multiple Access), where subcarriers and time slots are the resources.

1.3 Contributions of this dissertation

The contributions of this thesis are the following:

- We propose an analysis of the QoS in a wireless system based on the application of the effective bandwidth theory. The effective bandwidth theory has been widely employed in wired networks to analyze the statistical multiplexing. Recognizing that the classical queueing theory is not valid for queues using correlated physical-layer channel models, we follow the work in [Wu 2003b] and apply the effective bandwidth theory to a wireless systems, with the aim of getting fundamental limits in QoS parameters such as the delay and the loss probability.

- The effective bandwidth function of a frequency-flat Rayleigh channel (also known as *effective capacity* [Wu 2003b]) is analytically calculated under two different adaptive rate policies: continuous and discrete. This function is obtained first for an uncorrelated Rayleigh channel and later on generalized to time-correlated channels. To address the effective bandwidth function of the channel, the log moment generating function of the process is needed. By applying the Central Limit Theorem (CLT) the calculation comes down to the computation of the mean and the variance of the channel rate.

- With the effective bandwidth function of the channel and the source it is feasible to study the delay in a wireless system. First of all, the analysis is done for a flat Rayleigh channel. In particular, the percentile of the delay, the probability density function of the delay and the tradeoff among delay, source rate and channel conditions are investigated.

- A new definition of capacity in a wireless system is proposed. In wireless communications, the delay-limited capacity is associated to the maximum *fixed-rate* that can be deterministically guaranteed, which implicitly ensures delay. This capacity is zero for the case of Rayleigh fading channels. In such cases, we propose the definition of its probabilistic version, the *Capacity with Probabilistic Delay Constraint*, which indicates the maximum constant information rate that can be transmitted over the channel while accomplishing a target BER and a delay constraint. The delay constraint is given by the pair (D^t, ε), being D^t the target delay and ε the probability of exceeding D^t.

- The analysis performed over a flat channel is extended to a multichannel system over a frequency selective channel. The effective bandwidth of a frequency selective Rayleigh channel is obtained following the same procedure applied in the case of flat channels. Then, expressions of the percentile of the delay in this new scenario are given. Finally, the capacity with probabilistic delay constraint is obtained for an OFDM system over a frequency selective Rayleigh

channel.

- A multiuser system under several scheduling algorithms is addressed. For that purpose, the single user system model is generalized, so that now each user has his own queue and the shared channel has to empty the queues according to the set scheduling discipline. The channel can be a single link or a multiple shared channel. Three representative allocation algorithms are analyzed: Round Robin [Hanssen 2004], Best Channel [Knopp 1995] and Proportional Fair [Shakkottai 2001]. The analytical results of the maximum achievable rate of each user under a delay constraint show the tradeoff between performance and fairness for the three disciplines. The employed procedure is generic and can be extended to the study of any other discipline.

1.4 Outline of the dissertation

A diagram with the organization of this dissertation is shown in Figure 1.4. In Chapter 2 we do a literature review on the effective bandwidth theory, which provides the basis for the analysis of the system, and present the system model employed along this thesis.

In Chapters 3 and 4 we investigate the effective bandwidth functions of several traffic sources and of a flat Rayleigh channel.

The calculation of the effective bandwidth functions makes possible the analysis of the delay in flat Rayleigh channels, as presented in Chapter 5.

In Chapter 6 the previous results are generalized to frequency selective Rayleigh channels.

In Chapter 7 we extend the system model to a multiuser system under different scheduling disciplines and redo the analysis. In particular, Round Robin, Best

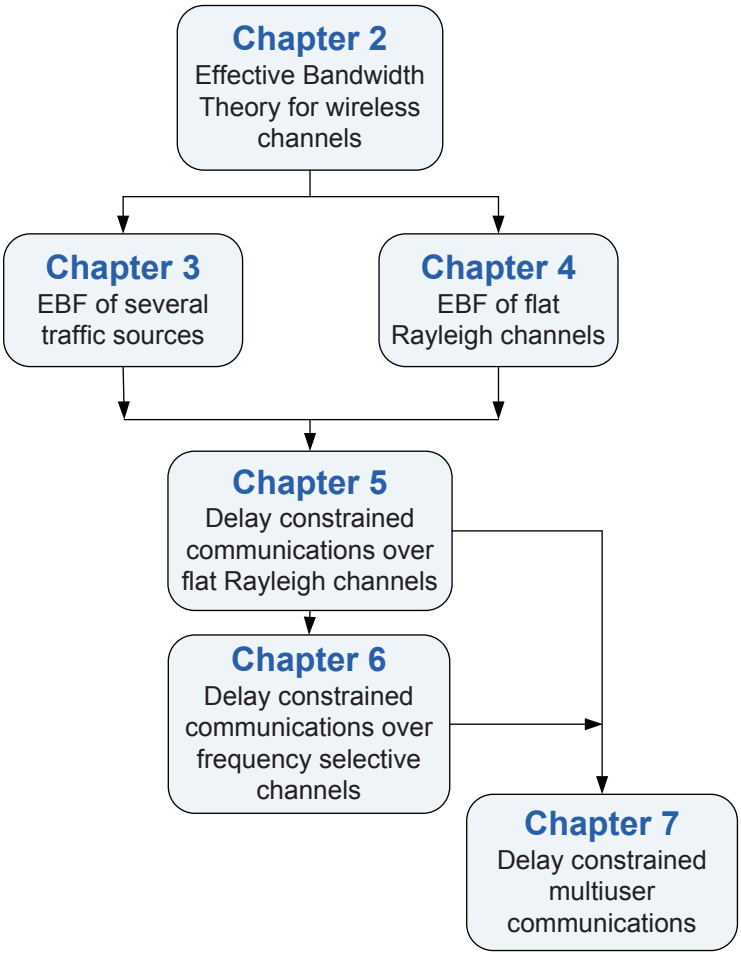


Figure 1.4: Organization of the thesis.

Channel and Proportional Fair schemes are investigated. The analysis is done first for a single link and then generalized to a multiple shared fading channel.

In Chapter 8 we summarize the dissertation and point out future research directions.

1.5 Author's publication list

The research conducted during this thesis resulted in the following publications:

- Journal Papers:
  - B. Soret, M. C. Aguayo-Torres, J. T. Entrambasaguas, 'Capacity with explicit delay guarantees for generic sources over correlated Rayleigh channel', *IEEE Transaction on Wireless Communications*, vol. 9, no. 6, pp. 1901–1911, June 2010
  - B. Soret, M. C. Aguayo-Torres, J. T. Entrambasaguas, 'Analysis of the tradeoff between delay and source rate in multiuser wireless systems', *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, Article ID 726750, 13 pages, doi:10.1155/2010/726750, 2010
- Conference Papers:
  - B. Soret, M. C. Aguayo-Torres, J. T. Entrambasaguas, 'Maximum delay-constrained source rate over a wireless channel', *Proc. Valuetools'07*, Nantes (France), October 2007
  - B. Soret, M. C. Aguayo-Torres, J. T. Entrambasaguas, 'Capacity with probabilistic delay constraint for Rayleigh channels', *Proc. IEEE Globecom'07*, Washington (USA), November 2007
  - B. Soret, M. C. Aguayo-Torres, J. T. Entrambasaguas, 'On the effective bandwidth function of 802.16 traffic models', *Proc. I International Interdisciplinary Technical Conference of Young Scientists Intertech'08*, Poznan (Poland), April 2008
  - J. G. Ruiz, B. Soret, M. C. Aguayo-Torres, J. T. Entrambasaguas, 'On Finite state Markov chains for Rayleigh channel modeling', *Proc. Wireless Vitae'09* (Best Paper Award), Aalborg (Denmark), May 2009
  - B. Soret, M. C. Aguayo-Torres, J. T. Entrambasaguas, 'Capacity with probabilistic delay constraint for voice traffic in a Rayleigh channel', *Proc. IEEE ICC'09*, Dresden (Germany), June 2009

- B. Soret, M. C. Aguayo-Torres, J. T. Entrambasaguas, ' Multiuser capacity for heterogeneous QoS constraints in uncorrelated Rayleigh channels', *Proc. Information Theory Workshop ITW'09*, Taormina (Italy), October 2009
- B. Soret, M. C. Aguayo-Torres, J. T. Entrambasaguas, 'Analysis of delay constrained communications over OFDM systems', *Proc. IEEE Globecom'09*, Honolulu (USA), November 2009

Chapter 2

Effective Bandwidth Theory for wireless channels

The theory of large deviations, known as the theory of *rare events*, has served as a basis for the development of the *effective bandwidth theory* (EBT) [Chang 1995a], which is applied in this thesis to address the analysis of QoS parameters in a wireless system. The EBT has been widely employed for analyzing the statistical multiplexing in wired networks such as Asynchronous Transfer Mode (ATM) networks. More recently, its adaptation to the wireless world has been developed by means of a link-layer channel model dual to the effective bandwidth model and suitable for wireless communication systems [Wu 2003b].

The fundamentals of the large deviations theory and the effective bandwidth theory can be found in Appendix A.

In this chapter, we present the application of the EBT to the analysis of the wireless system addressed in this thesis. The chapter is organized as follows. Section 2.1 reviews the existing literature on EBT. Section 2.2 describes the proposed single-user single-channel system model. The effective bandwidth analysis of this system model is detailed in Section 2.3. Generalizations of the model to multi-user multi-channel systems are presented in Section 2.4. Finally, some concluding

remarks are done in Section 2.5.

2.1 Bibliography review

The effective bandwidth theory was initiated by Chang [Chang 1994] [Chang 1995a]. The system model consists of a queueing system with the server modeling the channel and the queue storing user's data. Defined for ATM networks, it expresses the bandwidth (i.e. the constant channel rate) that a source needs to satisfy certain QoS requirements. In [Chang 1995b], Chang and Zajic extend the results to queues with time varying capacity in the server.

In [Kelly 1996] a unifying definition of the effective bandwidth is introduced. The definition summarizes the statistical characteristics of traffic over different time and space scales, and builds on earlier work on effective bandwidths and asymptotic models. Illustrative examples which demonstrate the unifying property of the definition include periodic sources, fractional Brownian input, policed and shaped sources, and deterministic multiplexing.

Several works evaluate the effective bandwidth function of various traffic sources [Kessidis 1996] [Chou 1996]. In [Courcoubetis 1994] the authors propose a simple measure of the effective bandwidth function of stationary traffic sources including the mean rate, the index of dispersion and the size of the buffer.

In [Choudhury 1996] the theory is applied to the connection admission control. The idea is to assign each source an effective bandwidth requirement, and then consider any subset of source feasible if the sum of the required effective bandwidths is less than the total available bandwidth.

The effective bandwidth function has been of great help in the study of the impact of the burstiness of real broadband traffic on the performance of the network and on its resource sharing capabilities. In [Courcoubetis 1999] the theory is applied under

various mixes of real traffic, with the goal of clarifying the effects of the time scales of traffic burstiness and of the traffic control mechanisms on the link performance.

Unlike its wired counterpart, a wireless connection suffers fading. In [Wu 2003b], Wu and Negi developed a link-layer channel model dual to the effective bandwidth model which is suitable for wireless communication systems. Additionally, they termed the effective bandwidth function of the channel process as *effective capacity* and proposed an algorithm for estimating it from the channel dynamics, which was applied by some of their followers [Quimi 2005].

The same authors in [Wu 2003b] made an attempt to evaluate the effective bandwidth function of the channel process but the result is given only for very low Signal to Noise Ratio (SNR).

Several works following [Wu 2003b] employed this link-layer model and showed that is capable of predicting the QoS metrics under various conditions. In [Quimi 2005] the effective capacity model is applied to the transmission of variable rate traffic and measurements are made for several source and channel realizations. In [Liu 2007] the tradeoff among spectral bandwidth, power and code rate is studied with the wireless channel modeled through a Markov chain. Plenty of authors have opted to use a Finite State Markov Chain (FSMC) to model the wireless channel [Hassan 2004] [Park 2006] [Tang 2006]. In [Tang 2007b] the optimum power control for an uncorrelated fading channel is evaluated in order to maximize the effective capacity under certain QoS. In [Park 2006] the authors propose a cumulative distribution-based scheduling algorithm and analyze its performance using the effective capacity model. In [Femenias 2009] a two-dimensional Markov model is considered to address a cross-layer design in a wireless network combining adaptive modulation and coding (AMC) with an automatic repeat request (ARQ) protocol.

The effective capacity model has also been applied to systems exploiting the diversity gain with multiple antennas or relay techniques. In [Zhang 2006] the authors

employ a FSMC to model a Multiple Input Multiple Output (MIMO) channel with adaptive modulation and coding and evaluate the effective capacity function. In [Ren 2009] a resource allocation scheme subject to a QoS requirement in a multi-relay cooperative wireless network is proposed.

To sum up, the effective bandwidth theory has been an active topic in the last twenty years, for wired networks during the 90's and for wireless networks with the *effective capacity model* in the last years.

2.2 System model

The analysis of the single-user single-channel system model presented in Section 1.2 is tackled by means of the discrete-time queueing model in Figure 2.1. Here the source process characterizes the incoming user traffic and the server represents the information transmitted to the wireless channel. The bits generated by the source are first put into a buffer to accommodate the mismatch between the time-variant source and the time-variant channel rate. Physical time is divided into units referred to as symbol periods and represented by the transmission discrete time unit, n. The channel response of each user is assumed to be constant over the symbol. The channel is allocated on a symbol-by-symbol basis: every new symbol, the transmitter removes some of the data from the buffer and transmits it over the wireless channel.

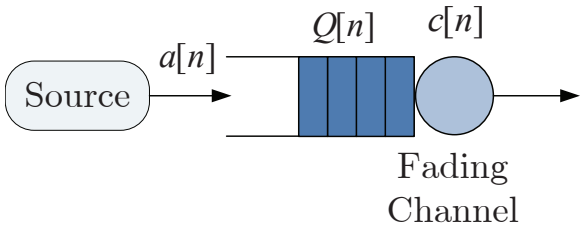


Figure 2.1: Queueing system.

There are two broad categories of traffic models. One option is to consider each packet individually. The random process used as a traffic model generates values representing the arrival times of individual packets (or equivalently the inter-arrival times between successive packets). The other alternative is a fluid model, where the source is described by the rate at which work arrives at the buffer.

On the other hand, the instantaneous gain of the wireless channel is, in general, a time-variant and autocorrelated random process. It is assumed that the transmitter employs adaptive techniques with constant transmitted power [Chung 2001], so that some transmission parameters such as the constellation size are modified dynamically, seeking to adapt to the time-varying conditions of the physical channel. Roughly speaking the service rate is also time-variant, and thus the system under study is a queue with autocorrelated arrival and service rates. The connection of the service rate with the channel process is detailed in Chapter 4.

In the fluid model in Figure 2.1, the incoming user traffic has an instantaneous rate $a[n]$ and the wireless channel can transmit at an instantaneous rate of $c[n]$. The processes $a[n]$ and $c[n]$ are not necessarily white and represent the amount of bits per symbol generated by the source and transmitted by the server, respectively.

The accumulated source rate $A[n]$ is the amount of bits generated by the source from 0 to instant $n-1$:

$$A[n] = \sum_{m=0}^{n-1} a[m] \tag{2.2.1}$$

Similarly, the accumulated channel process is:

$$C[n] = \sum_{m=0}^{n-1} c[m] \tag{2.2.2}$$

The queue size is assumed to be infinite and $Q[n]$ denotes the length of the queue at time n. The dynamics of the system is characterized by the discrete-time

equation:

$$Q[n] = (Q[n-1] + a[n] - c[n])^+ \tag{2.2.3}$$

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

We refer to the delay of the bits leaving the queue system at time n as $D[n]$. It is the time from the moment a bit arrives to the buffer until its service is completed. Notice that no encoding of the data is considered. Hence the delay experienced by data is due to the time spent in the buffer. If encoding/decoding were included, a second component of the delay would correspond to the time from when data is encoded until it is decoded. Information theory treatments typically consider only the delay due to the encoding and decoding processes. Buffer delay is usually considered a network layer problem and separated from physical layer considerations. From now on, the term *delay* refers to the buffer delay.

On the other hand, loss ratio is the ratio between the total number of bits lost and the total number of bits arriving. The underlying assumption in the definition of loss ratio is that the buffer is finite, and bits arriving when the buffer is full are lost. Furthermore, to define overflow probability we usually consider an infinite buffer queue, and define overflow probability as the probability (or the proportion of time) that the number of bits in the buffer exceeds a certain threshold.

2.3 Effective bandwidth analysis

The single-user single-channel scenario is analyzed by means of the effective bandwidth theory. The asymptotic log-moment generating function of the process $Q[n]$ is defined as [Chang 1995a]:

$$\Lambda(\upsilon) = \lim_{n\to\infty} \frac{1}{n} \log \mathrm{E}\left[e^{\upsilon Q[n]}\right] \ \forall \upsilon \geq 0 \tag{2.3.1}$$

Since $a[n]$ and $c[n]$ are independent from each other, $\Lambda(\upsilon)$ may be decomposed

into two terms:

$$\Lambda(\upsilon) = \Lambda_A(\upsilon) + \Lambda_C(-\upsilon) \tag{2.3.2}$$

where $\Lambda_A(\upsilon)$ and $\Lambda_C(\upsilon)$ are the log-moment generating functions of the accumulated source process $A[n]$ and the accumulated channel process $C[n]$, respectively.

Besides, the *effective bandwidth function* (EBF) of the process $a[n]$ is defined as:

$$\alpha_A(\upsilon) = \Lambda_A(\upsilon)/\upsilon = \lim_{n\to\infty} \frac{1}{n\upsilon} \log \mathrm{E}\left[e^{\upsilon A[n]}\right] \ \forall \upsilon \geq 0 \tag{2.3.3}$$

and, similarly, the EBF of the channel process $c[n]$ is:

$$\alpha_C(\upsilon) = \Lambda_C(\upsilon)/\upsilon = \lim_{n\to\infty} \frac{1}{n\upsilon} \log \mathrm{E}\left[e^{\upsilon C[n]}\right] \ \forall \upsilon \geq 0 \tag{2.3.4}$$

If the source and the channel processes are stationary and the steady state queue length exists (i.e. given a threshold B, $\sup_n Pr\{Q[n] > B\} = Pr\{Q(\infty) > B\}$), then the workload process $Q[n]$ satisfies a Large Deviation Principle (LDP) and the following asymptotic behavior for the queue length exceeding B is fulfilled [Chang 1995a]:

$$Pr\{Q(\infty) > B\} \asymp e^{-\theta B} \qquad B \to \infty \tag{2.3.5}$$

where $f(x) \asymp g(x)$ means that $\lim_{x\to\infty} f(x)/g(x) = 1$ and θ, the QoS exponent, is the solution to [Chang 1995b]:

$$\Lambda(\upsilon) \mid_{\upsilon=\theta} = 0 \Rightarrow \Lambda_A(\upsilon) + \Lambda_C(-\upsilon) \mid_{\upsilon=\theta} = 0 \tag{2.3.6}$$

With the definition of EBFs in (2.3.3) and (2.3.4), the equation to obtain θ can also be expressed as:

$$\alpha_A(\upsilon) - \alpha_C(-\upsilon) \mid_{\upsilon=\theta} = 0 \tag{2.3.7}$$

In related literature [Wu 2003b], $\alpha_C(-\upsilon)$ is referred to as the *effective capacity function* but, since it is the effective bandwidth function of the channel process, we

prefer to keep the term effective bandwidth function for all the processes involved in the analysis.

A more accurate approximation for small values of B includes the probability that the queue is not empty, denoted by $\eta = Pr\{Q[n] > 0\}$. Then, the following less conservative approximation for the tail probability of the queue is satisfied:

$$Pr\{Q(\infty) > B\} \approx \eta \cdot e^{-\theta B} \tag{2.3.8}$$

As in the queue length process, the steady state solution for the delay process exists, which means that given a delay D^t, $\sup_n Pr\{D[n] > D^t\} = Pr\{D(\infty) > D^t\}$. Moreover, the probability of exceeding D^t, denoted throughout this thesis as target delay, can be written as follows [Wu 2003a]:

$$\varepsilon = Pr\{D(\infty) > D^t\} \approx \eta \cdot e^{-\theta \cdot \alpha_A(\theta) D^t} = \eta \cdot e^{-\theta \cdot \alpha_C(-\theta) D^t} \tag{2.3.9}$$

where ε is the probability of exceeding the target delay D^t and provides a measure of the percentile of the delay directly as $1 - \varepsilon$. The expression in (2.3.9) is the starting point of our analysis of probabilistic QoS guarantees in a wireless system. It is linked to the EBF of the arrival and the service processes through the QoS exponent θ. Next, we dig deeper into the meaning of $\alpha_A(\upsilon)$ and $\alpha_C(\upsilon)$.

The effective bandwidth curves

The stochastic behavior of an arrival process or a service process is modeled asymptotically by its effective bandwidth. On the one hand, the effective bandwidth of the source expresses the minimum constant service rate required by a given arrival process in order to guarantee certain QoS requirements. A high value of the parameter υ of the function indicates a more severe QoS requirement (i.e. smaller D^t or smaller ϵ); hence a higher value of the service rate will be needed to guarantee the QoS requirement. Likewise, a small value of υ symbolizes loose QoS requirements.

Thus, the effective bandwidth curve of a traffic source $\alpha_A(\upsilon)$ (Figure 2.2) increases with υ, starting always at the source mean rate, for $\upsilon = 0$, and tending towards the peak rate of the source as $\upsilon \to \infty$.

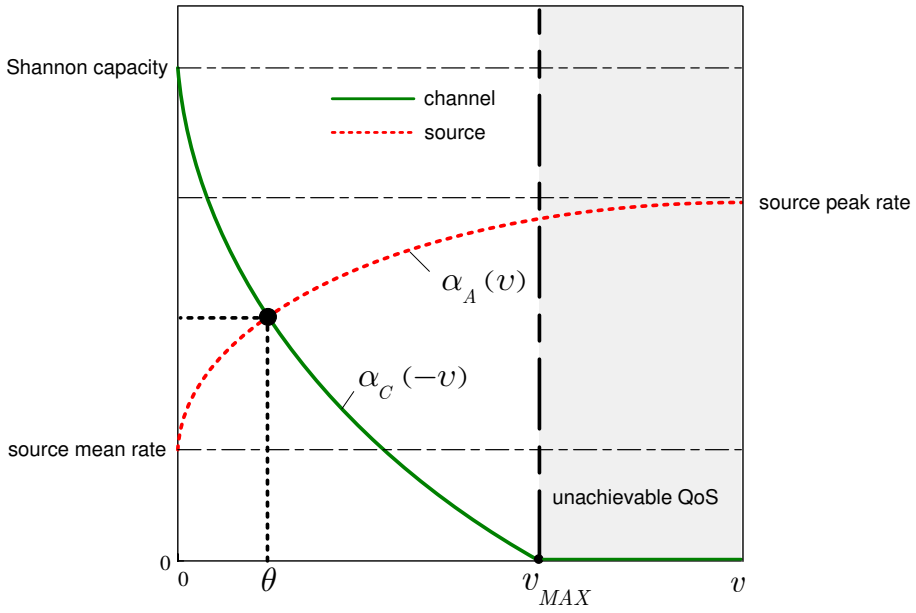


Figure 2.2: Intersection of the effective bandwidth functions.

On the other hand, the effective bandwidth of the channel $\alpha_C(-\upsilon)$ indicates the maximum arrival rate that a given service process can support by fulfilling the QoS requirements. The curve in this case starts at Shannon's capacity when $\upsilon = 0$ (no delay constraints are imposed) and decreases asymptotically with υ, since as the QoS requirement becomes more stringent, the maximum constant source rate the channel can support decreases. In the case of Rayleigh channels, where delay-limited capacity is zero, the curve reaches zero at a certain point denoted as υ_{MAX} in Figure 2.2. Higher values of υ imply QoS requirements that are not achievable by that channel, whatever the traffic source is.

When both concepts are joined and the whole scenario is studied, a working point of the system can be defined and will correspond to the intersection of the two

curves, as shown in Figure 2.2. This point is exactly the QoS exponent θ defined in equation (2.3.7). It symbolizes the point in which both the traffic process and the service process will be able to accomplish the QoS requirements. Probabilistic delay guarantees can then be provided for a given channel and source thanks to the analytical evaluation of θ, which requires computing the effective bandwidth functions $\alpha_A(\upsilon)$ and $\alpha_C(\upsilon)$. In the case of the source, many source examples and its corresponding EBF can be found in the literature. The difficulty lies in the channel side, where the expectation in (2.3.4) has to be obtained for the channel process.

Several representative models for the source and the computation of its corresponding EBF $\alpha_A(\upsilon)$ are presented in Chapter 3. In Chapter 4 we carry out the evaluation of the EBF for flat Rayleigh channels. Both results are combined in Chapter 5 to analyze the delay in a wireless system. The procedure is extended to frequency-selective channels and multiuser systems in Chapters 6 and 7, respectively.

2.4 Generalization

In this Section, the generalization of the system model to multiuser and multichannel scenarios is presented. The multi-channel case will be studied in Chapter 6, and the multi-user system in Chapter 7. In both cases, the effective bandwidth theory is applied in a similar way as the one presented in this Chapter.

Single-user multi-channel

The term multi-channel refers here to the organization of the resources into a set of physical channels. Each physical channel or subchannel or subcarrier can be allocated to a specific user during a session (dedicated physical channel) or time-shared by several users (user multiplexing). In this first generalization of the system only one user is considered.

The queueing system is shown in Figure 2.3. It is an extension of that applied for the single-channel. The source process characterizes the incoming user traffic and F servers in parallel model F subchannels.

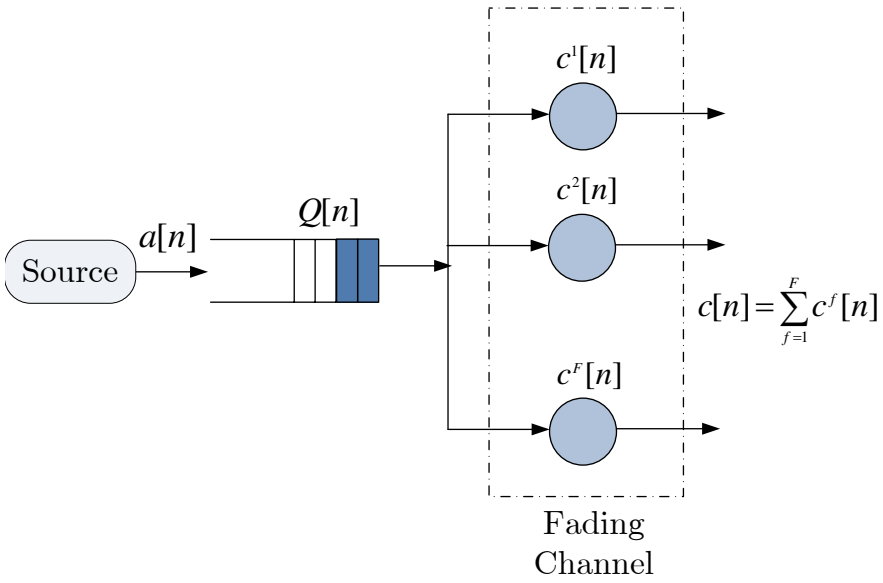


Figure 2.3: Multi-channel system model.

With the same discrete-time fluid model applied in flat channels, $a[n]$ is the instantaneous source rate. On the other hand, subchannel f can transmit at an instantaneous rate $c^f[n]$, so that the total instantaneous rate of the wireless channel is:

$$c[n] = \sum_{f=1}^{F} c^f[n] \tag{2.4.1}$$

Multi-user single-channel

The multi-user single-channel case is depicted in Figure 2.4. The channel is shared among U users, whose incoming traffics are characterized by U source processes. Each user has its own queue where the data are stored before being transmitted.

The scheduler allocates the channel to users: every new symbol, a user is selected for transmission.

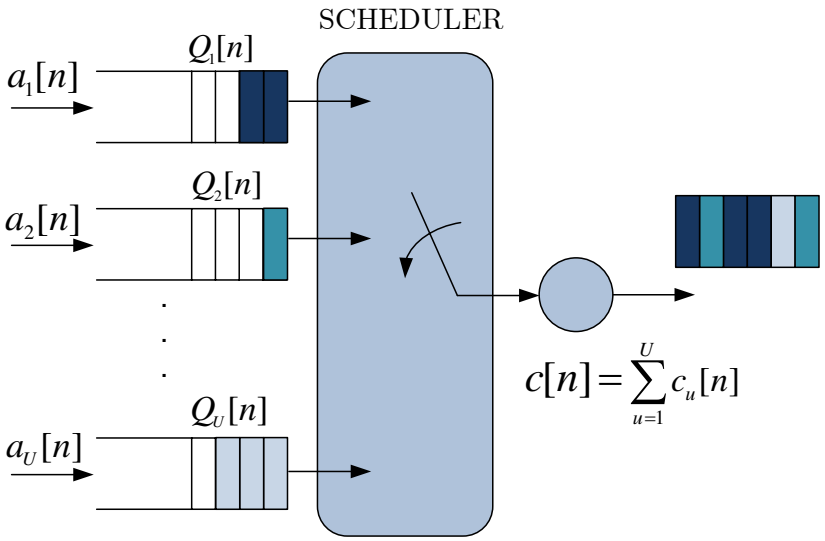


Figure 2.4: Multi-user system model

Each incoming user traffic has an instantaneous rate $a_u[n]$. On his side, the wireless channel can transmit at an instantaneous rate $c[n]$. Each user has a potential rate $r_u[n]$, which represents the channel rate that he may use if the channel is assigned to him, and which depends on his channel conditions. Moreover, the instantaneous channel rate of user uth, $c_u[n]$, is given by:

$$c_u[n] = \begin{cases} r_u[n] \text{ if channel is assigned to user u} \\ 0 \text{ in other case} \end{cases} \tag{2.4.2}$$

Since the channel is shared among U users, $c[n]$ can be expressed:

$$c[n] = \sum_{u=1}^{U} c_u[n] \tag{2.4.3}$$

Notice that in the sum above only one of the terms is non-zero, corresponding to the user allocated to the channel.

Multi-user multi-channel

Finally, the last step is to consider a multi-user wireless system over a fading multi-channel. Figure 2.5 illustrates the queueing system model.

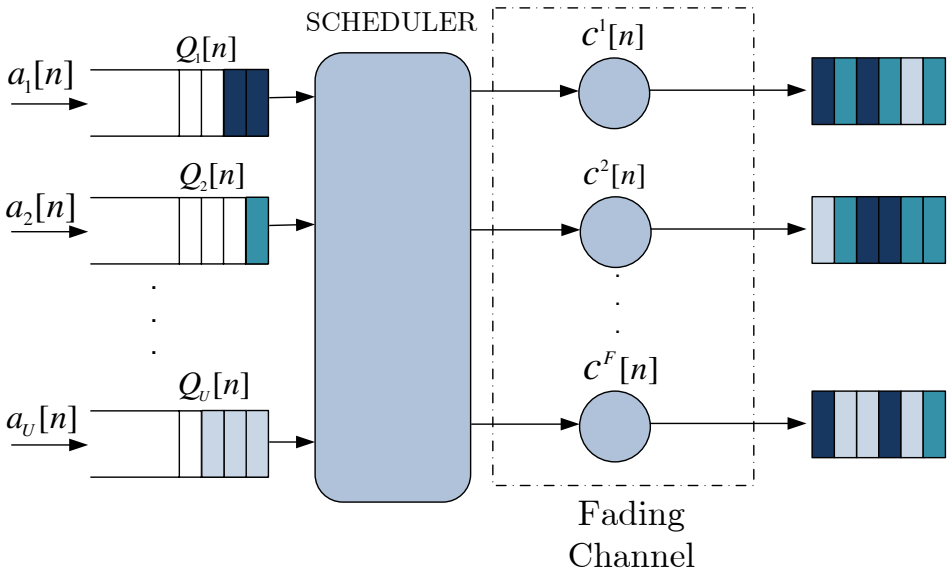


Figure 2.5: Multi-user multi-channel system model

F servers model the F subchannels. The scheduler assigns resources (subchannels) to users at every symbol. Each user has a potential rate $r_u^f[n]$ at each subcarrier, which represents the channel rate that he may use at subchannel f if it is assigned to him. The instantaneous transmission rate seen by user uth is:

$$\begin{aligned} c_u[n] &= \sum_{f=1}^{F} c_u^f[n], \\ c_u^f[n] &= \begin{cases} r_u^f[n] \text{ if subcarrier } f \text{ is assigned to user u} \\ 0 \text{ in other case} \end{cases} \end{aligned} \tag{2.4.4}$$

Moreover, the total transmission rate at subcarrier f is:

$$c^f[n] = \sum_{u=1}^{U} c_u^f[n] \tag{2.4.5}$$

2.5 Summary

The study of QoS provisioning over wireless links demands a simple and effective wireless model that can capture the nature of the key elements involved in the system. In this thesis, the wireless transmitter is modeled by means of a queueing system. The Effective Bandwidth Theory is applied to do the analysis. In a single-user single-channel system, the arrival process is modeling the instantaneous traffic generated by the user. and the server is representing the instantaneous transmission rate of the channel, capturing physical layer procedures such as adaptive modulation and coding.

We have provided key insights about the meaning of the EBF of the two processes involved in the system: source and channel. We have introduced the difficulty related to the analysis of the system, which basically lies in the computation of the effective bandwidth function of the channel process. Finally, we have presented the generalization to multi-channel and multi-user systems, to be addressed in the last chapters of the thesis.

Chapter 3

Effective Bandwidth Function of several traffic sources

The effective bandwidth theory provides the basis for analyzing several QoS parameters in a wireless system. Thereby, we first need to investigate the effective bandwidth functions of the two processes involved: the arrival and the channel process. Thoughout this chapter, the study of the arrival process is addressed.

In wireless systems, many efforts have been done in the area of applying the effective bandwidth theory to different upper layer procedures. Nevertheless, little work is devoted to include realistic traffic sources in the scenario. In this chapter we present a review on traffic modeling including the evaluation of the EBF of the models.

Section 3.1 introduces the problem of modeling traffic in a wireless network. In Section 3.2 a revision on the most usual traffic models is done, and their corresponding effective bandwidth function is discussed. As example of mix of traffic, the proposal by the IEEE 802.16 Broadband Wireless Access Working Group [802.16 2001b] is investigated. Section 3.3 presents two methods for the estimation of the effective bandwidth of a source, which have been used to validate the analytical solutions of this chapter. Finally, some concluding remarks are discussed in Section 3.4.

The evaluation of the effective bandwidth function of the 802.16 traffic models was published in [Soret 2008].

3.1 Traffic modeling

Traffic modeling and characterization have been intense areas of study in recent years, and together have an enormous impact on provisioning of QoS guarantees [Adas 1997]. Without the knowledge of traffic characteristics, it is impossible to provide traffic specification or schedule packets intelligently to satisfy the QoS requirements of applications.

Two categories of traffic modeling can be distinguished depending on the approach: approximating models and bounding models. Approximating models attempt to characterize the behavior of the source traffic using some mathematical model. Within them, stochastic processes have been mostly used to model traffic within a network. A simple example is the two-state on-off source typically applied to model voice traffic. In addition to stochastic processes, time series models have also been used, for example, autoregression models for video traffic and Internet traffic.

Unlike approximating models, bounding models attempt to upper bound the amount of traffic arriving from a source during a time interval. The models can be either deterministic (with the theory started by Cruz [Cruz 1991] and continued with the main contribution of LeBoudec [LeBoudec 2001]) or stochastic [Kurose 1992].

This thesis will focus on the approximating models. Specifically, stochastic processes and autoregressive models are studied. Particular emphasis is given on the so-called real-time services, for example video or audio services, wherein real-time data streaming is provided.

The following traffic models are presented next, together with the discussion of

their corresponding EBF's:

- First of all, the simplest model is addressed: Constant Bit Rate (CBR) traffic.
- Secondly, a review of the classical Markov models is presented.
- Streaming traffic is often modeled through a simple autoregressive model. It belongs to the most generic regression models, which define explicitly the next random variable in the sequence by previous ones within a specified time window and a moving average of a white noise.
- Lastly, as example of mix of traffic in a current wireless system, we study the proposal by the IEEE 802.16 Broadband Wireless Access Working Group [802.16 2001b]. Each subscriber is characterized by his set of services, that includes HTTP/TCP, FTP, voice and streaming activity. All the models are based on the superposition of up to four Interrupted Poisson Processes (IPP).

3.2 EBF of several traffic models

3.2.1 Constant source

The easiest case that may come to one's mind is a constant rate source, i.e., source traffic that arrives to the buffer at a constant rate:

$$a[n] = \lambda \tag{3.2.1}$$

where λ and $a[n]$ are expressed in bits per symbol. In this case the term Constant Bit Rate (CBR) is employed. The effective bandwidth of this source is constant:

$$\alpha_A(\upsilon) = \lambda \tag{3.2.2}$$

Figure 3.1 shows three CBR sources with rate 2, 5 and 10 bits/symbol. Figure 3.1 (a) shows a realization of the process. $a[n]$ is plot as a function of the symbol n. In this case the traffic is constant and so $a[n]$ is. In Figure 3.1 (b), the EBF of the source, which is also constant, is plot.

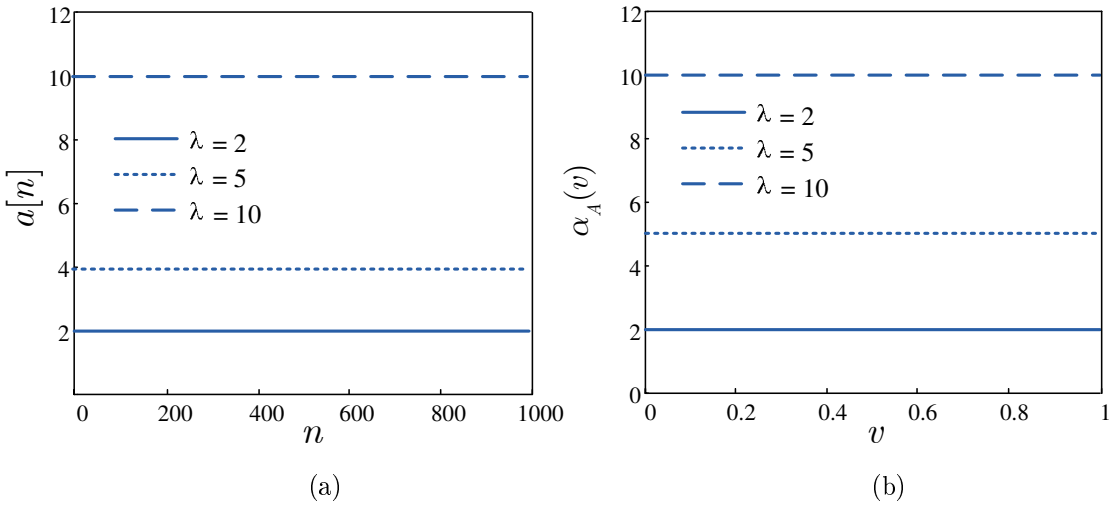


Figure 3.1: Constant traffic. (a) Realization of the process (b) Effective bandwidth function

Despite its simplicity, this model is widely employed in this thesis. It is well known that if sources are more bursty, queueing performance is degraded [Ryu 1996], and the service rate has to be increased in order to meet a fixed level of QoS. In particular, delay suffered by an information flow depends not only on the transmission rate but also on the distribution and self-correlation of the information rate. Therefore, the results for a constant source will represent the upper bound in the analysis of the QoS, since the best conditions are always obtained with a CBR source with regard to a Variable Bit Rate (VBR) source.

3.2.2 Markov sources

In many situations, the activities of a source can be modeled by a finite number of states. In general, increasing the number of states results in a more accurate model at the expense of increased computational complexity.

A Markov process with a discrete state space is referred to as Markov chain [Kleinrock 1975]. A set of random variables $\{X_n\}$ forms a Markov chain if, given the current state x_n, the probability that the next state is x_{n+1} depends uniquely upon x_n and not upon any previous values. Thus we have a random sequence in which the dependency extends backwards one unit in time. This is known as Markov property and can be expressed:

$$\begin{aligned} &Pr[X(t_{n+1}) = x_{n+1} | X(t_n) = x_n, X(t_{n-1}) = x_{n-1}, ..., X(t_1) = x_1] \\ &= Pr[X(t_{n+1}) = x_{n+1} | X(t_n) = x_n] \end{aligned} \tag{3.2.3}$$

If state transitions occur at integer values $\{0, 1, ..., n, ...\}$ then the Markov chain is discrete time. Otherwise, the Markov chain will be continuous time. Markov property implies that the way in which the entire past history affects the future of the process is completely summarized in the present, and does not depend on previous states nor on the time already spent in the current state. This imposes a heavy constraint on the distribution of time that the process may remain in a given state. In fact, if we have a continuous time chain, the state time must be exponentially distributed, since the exponential distribution is the only *memoryless* continuous distribution. On the other hand, in the discrete time Markov chain the process may remain in the given state follows a geometric distribution, the memoryless counterpart of the exponential distribution in the discrete time domain.

In an Interrupted Poisson Process (IPP) there are two states (Figure 3.2). Data are generated during ON state according to a given distribution and with average rate h bits per symbol. During OFF state, there is no traffic. μ is the average

number of transitions from the ON state to the OFF state per unit of time and, similarly, λ is the average number of transitions from the OFF state to the ON state per unit of time. The transitions among ON and OFF state are exponentially distributed whereas the distribution of the interarrival time during the active state (ON) gives rise to different types of IPP processes.

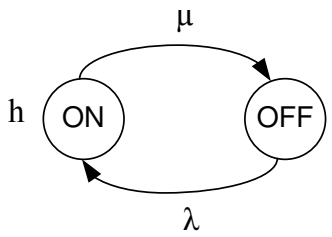


Figure 3.2: IPP process.

The instantaneous source rate of an IPP process is:

$$a[n] = \begin{cases} H & \text{if the process is in ON state} \\ 0 & \text{if the process is in OFF state} \end{cases} \tag{3.2.4}$$

$$H \sim Poiss(h) \tag{3.2.5}$$

The effective bandwidth of an IPP is [Kessidis 1996]:

$$\alpha_A(\upsilon) = \frac{1}{\upsilon} \log \left(\frac{\lambda + \mu\phi(\upsilon) + \sqrt{(\lambda + \mu\phi(\upsilon))^2 - 4(\lambda + \mu - 1)\phi(\upsilon)}}{2} \right) \tag{3.2.6}$$

where $\phi(\upsilon)$ is the moment generating function of the interarrival process during ON state (deterministic, exponential...).

There is a special kind of IPP process in which the rate during ON state is deterministic (Figure 3.3). Therefore, during sojourn time in the ON state the process generates data with fixed rate h and the time spent in ON and OFF states is exponentially distributed with average rate μ and λ, respectively. This model is

Figure 3.3: ON-OFF process.

the classical ON-OFF process, which has been widely used in the literature to model voice traffic. It is also called IDP (Interrupted Deterministic Process).

In this case, the instantaneous source rate $a[n]$ can be written:

$$a[n] = \begin{cases} h & \text{if the process is in ON state} \\ 0 & \text{if the process is in OFF state} \end{cases} \tag{3.2.7}$$

The mean arrival rate of an ON-OFF source is $m_A = h\frac{\lambda}{\lambda+\mu}$ and the EBF yields [Kelly 1996]:

$$\alpha_A(\upsilon) = \frac{1}{2\upsilon}\left[h\cdot\upsilon - \mu - \lambda + \sqrt{(h\cdot\upsilon - \mu + \lambda)^2 + 4\cdot\lambda\cdot\mu}\right] \tag{3.2.8}$$

Figure 3.4 shows the ON-OFF process. In Figure 3.4 (a) a realization of the instantaneous source rate $a[n]$ for the ON-OFF process is plot. Figure 3.4 (b) illustrates the effective bandwidth curve for different values of the parameters λ and μ. The EBF exhibits the expected behaviour of an effective bandwidth curve, starting at the mean rate ($m_A = 1$) of the source when no QoS is required (for $\upsilon = 0$) and asymptotically approaching the peak rate ($h = 2$) when $\upsilon \to \infty$.

It can be observed that shorter ON-OFF periods (higher values of the transition rates λ and μ) implies lower effective bandwidths. On the other hand, less variable sources, with longer ON-OFF periods, demand higher channel rates in order to accomplish certain QoS restriction.

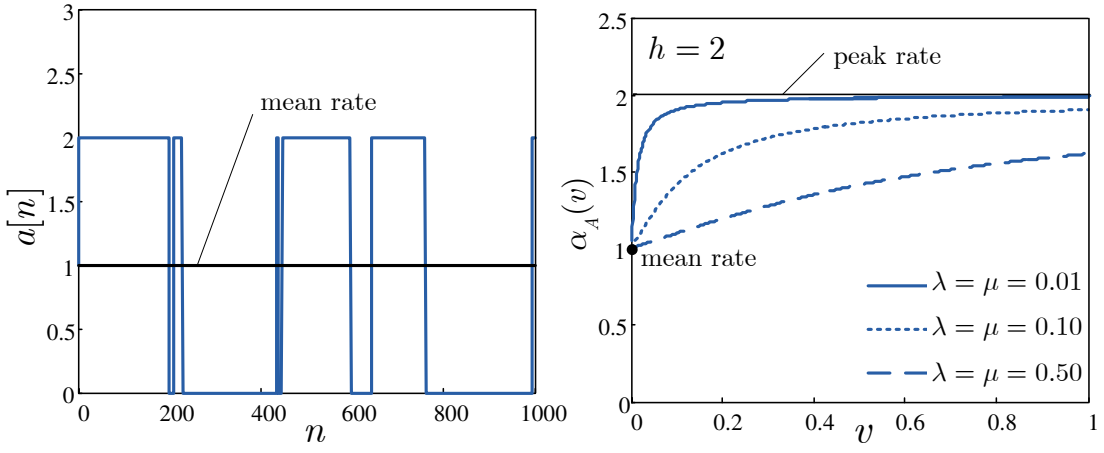


Figure 3.4: ON-OFF traffic. (a) Realization of the process (b) Effective bandwidth function

3.2.3 Autoregressive traffic

Autoregressive models belong to the most generic regression models, which define explicitly the next random variable in the sequence by previous ones within a specified time window and a moving average of a white noise.

An Autoregressive model of order p AR(p) has the following recurrence:

$$X_n = a_1 X_{n-1} + a_2 X_{n-2} + ... + a_p X_{n-p} + \epsilon_n \tag{3.2.9}$$

where ϵ_n is white noise and a_n are real numbers.

The literature on the modeling of streaming flows is extensive. Although most proposals are highly complex and take into account the long range dependence effect present in this kind of traffic, simple AR models have been proved to be appropriate (and sufficient) when the goal is simply the study of the queueing performance (see [Heyman 1996] [Ryu 1996]). Thus, AR models have been used to model the output bit rate of VBR encoders, where successive video frames do not vary much visually. For example, a video source is modeled through an autoregressive model of order

1 in [Maglaris 1998]. The video source is approximated by a continuous fluid flow model that assumes that the output bit rate within a symbol period is constant and changes from symbol to symbol according to the following AR(1) recurrence:

$$a[n] = \rho_A \cdot a[n-1] + q \cdot w[n] \tag{3.2.10}$$

where:

- $a[n]$ is the bit rate at time n
- $w = \mathcal{N}(m_w, 1)$ is Gaussian white noise of mean m_w and variance 1
- ρ_A and q are constants, with $|\rho_A| < 1$

$w[n]$ is chosen such that the probability of $a[n]$ being negative is very small. Nevertheless, there is always a probability to obtain a negative value of bit rate. In such cases, $a[n]$ is set to zero.

The effective bandwidth of this process is [Courcoubetis 1994]:

$$\alpha_A(\upsilon) = r \cdot m_w + \frac{r^2}{2} \cdot \upsilon \tag{3.2.11}$$

where:

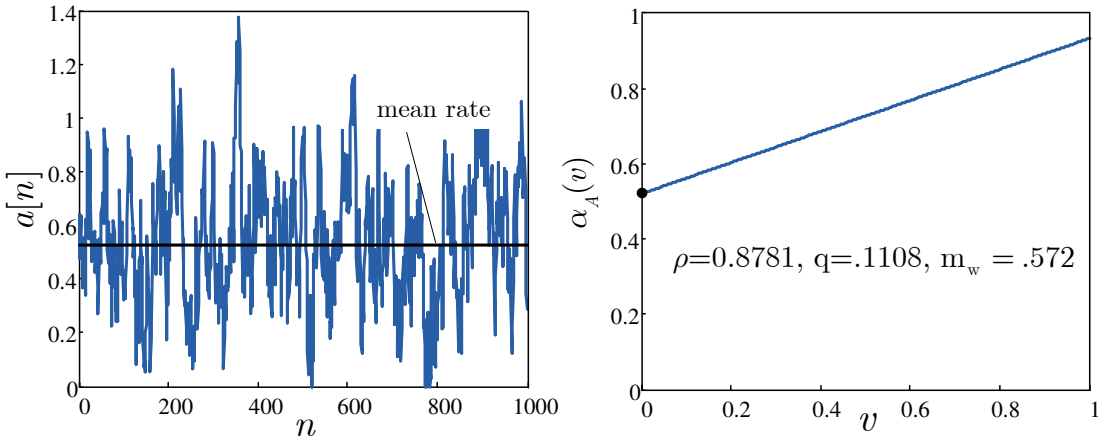
- $r = \frac{q}{1-\rho_A}$
- $m_A = r \cdot m_w$ is the mean of $a[n]$

Figure 3.5 illustrates the AR process. In Figure 3.5 (a) a realization of the process $a[n]$ is plot. In Figure 3.5 (b) the EBF of the autoregressive source is shown. In both cases, the parameters suggested in [Maglaris 1998] are considered (see Table 3.1). Notice that the EBF is here a straight line, starting in the mean rate and asymptotically approaching infinite, owing to the fact that the peak rate of the

source is infinite. If a source has a finite peak rate, then a constant rate channel at the peak rate will guarantee any QoS requirement, no matter how stringent it is. On the other hand, when the peak rate is infinite (as in the case of the AR source) it is not possible to ensure any QoS requirement. Thus, a deterministically guaranteed delay (corresponding to $\upsilon \to \infty$) would not be possible for the introduced AR source.

Table 3.1: Parameters of the AR(1) model in [Maglaris 1998]

ρ_A	0.8781
q	.1108
m_w	.572
$r = \frac{q}{1-\rho_A}$	.9889
$m_A = r \cdot m_w$	.5199



(a) (b)

Figure 3.5: Autoregressive traffic. (a) Realization of the process (b) Effective bandwidth function

3.2.4 The proposal of IEEE 802.16

There are two main problems for modeling Internet traffic. First of all, Internet is based upon a distributed architecture that makes it flexible and adaptable. Secondly, the growth of the Internet has been difficult to predict. In [802.16 2001b], the IEEE

802.16 Broadband Wireless Access Working Group proposed a set of traffic models suitable for MAC/PHY Simulations in 802.16 networks. The proposal provides not only the individual traffic models for each service but also the percentages necessary to define the mix of traffic arriving to an access point.

The standards and drafts from 802.16 technology are quite recent and therefore updated with the last and predicted changes in Internet. This ensures the relevance of the models. Besides, the proposed models are simple enough, making feasible the analytical evaluation of the effective bandwidth function. This simplicity must not be seen as a gap between the model and reality: as remarked in the AR source, simple traffic models have been demonstrated to be sufficient when the aim is the study of the queueing performance, such as the one we are referring to here.

The proposal includes three different services: voice, data (HTTP, TCP, FTP) and streaming. All of them are based on the superposition of several (up to four) generic IPP, so that the number of IPP's and the distribution involved in the ON state give rise to the different services.

Voice Traffic

One IDP represents one voice source, with the parameters shown in Table 3.2. h is expressed in packets per unit of time and λ and μ are transitions per unit of time. They are chosen to match the most cited voice model with ON period of 352ms and OFF period of 650ms, with the appropriate scaling.

Table 3.2: Parameters of the voice traffic in the proposal of 802.16.

source	h	μ	λ
IDP1	1.00	$5.682 \cdot 10^{-2}$	$3.076 \cdot 10^{-2}$

The instantaneous arrival rate is given by:

$$a[n] = \begin{cases} h & \text{if the process is in ON state} \\ 0 & \text{if the process is in OFF state} \end{cases} \tag{3.2.12}$$

The effective bandwidth function of an IDP process is the same presented in (3.2.2). The results with the parameters in Table 3.2 are shown in Figure 3.6. Figure 3.6 (a) shows a realization of $a[n]$ and the effective bandwidth function is plot in Figure 3.6 (b).

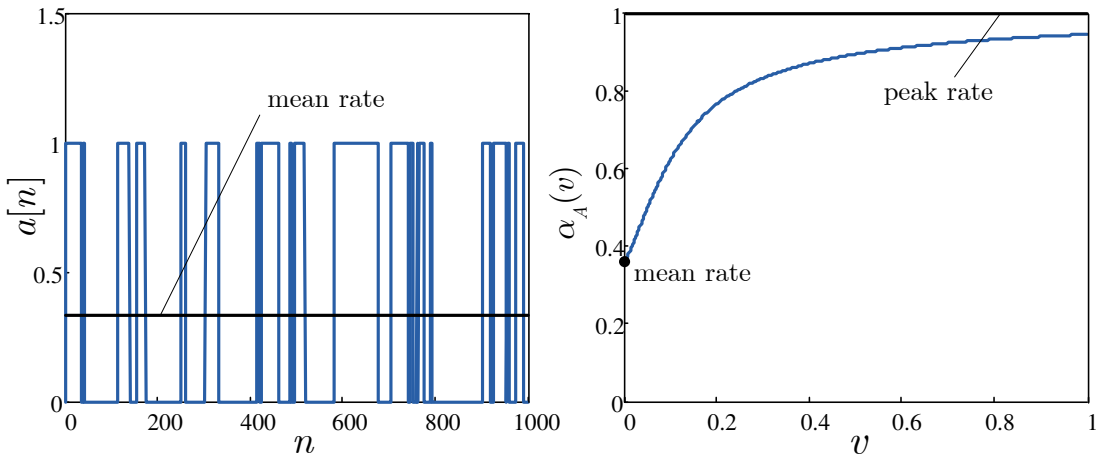


Figure 3.6: 802.16 voice traffic. (a) Realization of the process (b) Effective bandwidth function

Data Traffic

Data traffic is known to show self-similarity, i.e., autocorrelation coefficients with a slow or hyperbolic decay, associated with the existence of traffic variability in several time-scales. This self-similar nature is modeled with Long Range Dependent (LRD) traffic models [Leland 1994] [Adas 1997].

The generation of HTTP, TCP and FTP traffic is based on the superposition of 4 IPP processes that, as shown in [Brady 1968], can characterize self-similar traffic.

All four processes follow exponential distribution both for transitions among ON and OFF state and among packets during ON state, and different parameters $\mu^{(i)}$, $\lambda^{(i)}$ and $h^{(i)}$ for representing four different time scales, $i = 1..4$. Therefore, the 4IPP model superimposes four different time scales to generate an accurate representation of data traffic in Internet. Table 3.3 shows the parameters of data traffic, where $h^{(i)}$ is expressed in packets per unit of time and $\lambda^{(i)}$ and $\mu^{(i)}$ are transitions per unit of time. The mean rate of this traffic is: $m_A = \sum_{i=1}^{4} h^{(i)} \frac{\lambda^{(i)}}{\lambda^{(i)}+\mu^{(i)}}$.

The instantaneous arrival rate $a[n]$ now is the sum of the rates of the four processes:

$$a[n] = a^{(1)}[n] + a^{(2)}[n] + a^{(3)}[n] + a^{(4)}[n] \tag{3.2.13}$$

where:

$$a^{(i)}[n] = \begin{cases} H & \text{if the ith IPP process is in ON state} \\ 0 & \text{if the ith IPP process is in OFF state} \end{cases} \tag{3.2.14}$$

$$H \sim Poiss(h^{(i)}) \tag{3.2.15}$$

Table 3.3: Parameters of the data traffic in the proposal of 802.16.

source	$h^{(i)}$	$\mu^{(i)}$	$\lambda^{(i)}$
IPP1	2.679	$4.571 \cdot 10^{-1}$	$3.429 \cdot 10^{-1}$
IPP2	1.698	$1.445 \cdot 10^{-2}$	$1.084 \cdot 10^{-2}$
IPP3	1.388	$4.571 \cdot 10^{-4}$	$3.429 \cdot 10^{-4}$
IPP4	1.234	$4.571 \cdot 10^{-6}$	$3.429 \cdot 10^{-6}$

If the LRD nature is left out, the EBF can be directly evaluated as the sum of the effective bandwidth function of IPP processes. With exponentially distributed inter-arrival times and exponentially distributed sojourn times, it is just the superposition of 4 IPP with the EBF in (3.2.6), each of them with their specific parameters:

$$\alpha_A(\upsilon) = \sum_{i=1}^{4} \alpha^{(i)}(\upsilon) \tag{3.2.16}$$

$$\alpha^{(i)}(\upsilon) = \frac{1}{\upsilon}\log\left(\frac{\lambda^{(i)} + \mu^{(i)}\phi^{(i)}(\upsilon) + \sqrt{\left(\lambda^{(i)} + \mu^{(i)}\phi^{(i)}(\upsilon)\right)^2 - 4(\lambda^{(i)} + \mu^{(i)} - 1)\phi^{(i)}(\upsilon)}}{2}\right) \tag{3.2.17}$$

In this case, the moment generating function of the interarrival process during the active state, $\phi^{(i)}(\upsilon)$, is needed. It is that of an exponential distribution:

$$\phi^{(i)}(\upsilon) = \frac{h^{(i)}}{\upsilon - h^{(i)}} \tag{3.2.18}$$

The EBF in (3.2.17) has been validated by comparison with the measured EBF, by means of the methods explained in Section 3.3.

The results with the parameters in Table 3.3 are shown in Figure 3.7. A realization of $a[n]$ is plot in Figure 3.7 (a) and Figure 3.7 (b) presents the effective bandwidth curve.

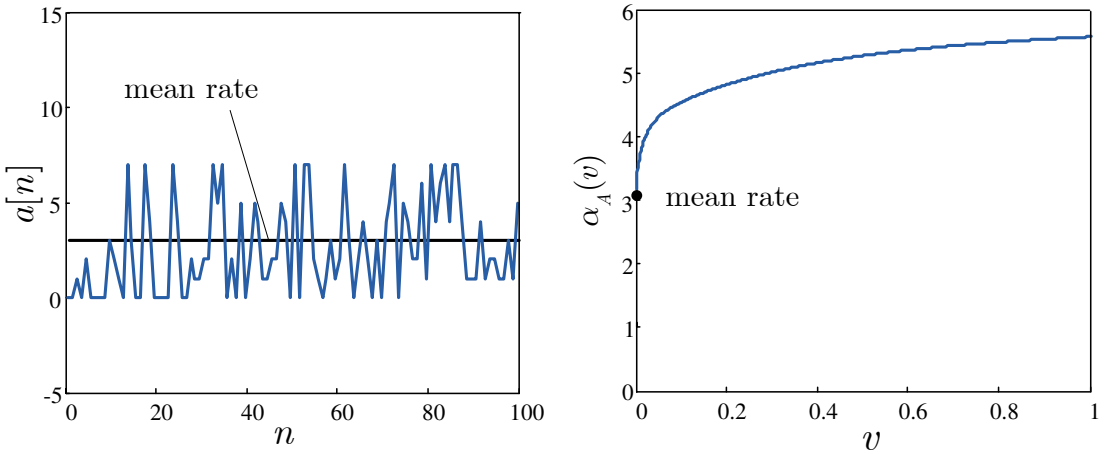


Figure 3.7: 802.16 data traffic. (a) Realization of the process (b) Effective bandwidth function

Video Traffic

Finally, a packet video source is modeled by means of two Interrupted Renewal Processes (2IRP) fitting the most cited video trace in past ten years, i.e., Star Wars movie. This kind of traffic also presents self-similarity. In the IRP the sojourn time is Pareto distributed rather than exponential and thereby it is not a Markov process anymore. The cumulative distribution function of a Pareto distribution is defined as:

$$F(x) = 1 - \left(\frac{b}{x}\right)^{\alpha} \quad x \geq 0 \tag{3.2.19}$$

with mean m_x and variance σ_x^2:

$$m_x = \frac{\alpha b}{\alpha - 1} \tag{3.2.20}$$

$$\sigma_x^2 = \frac{b^2 \alpha}{(\alpha - 1)^2(\alpha - 2)} \tag{3.2.21}$$

With the parameters of streaming traffic in Table 3.4, the model is appropriate, e.g., for MPEG packet video with 25 frames per second with local Hurst parameter ranging from 0.73 to 0.93. The parameter b in the Pareto distribution is set to 1, and α is λ and μ for the ON and OFF period, respectively. Once more, h is expressed in packets per unit of time and λ and μ are transitions per unit of time.

Table 3.4: Parameters of the streaming traffic in the proposal of 802.16.

source	h	μ	λ
IRP1	44.95	1.14	1.22
IRP2	61.90	1.54	1.28

In this case, two IRP processes are superposed with Pareto distributed sojourn time. The distribution among ON and OFF state is no longer memoryless and therefore the characterization with a Markov chain is not appropriate. In this case, we turn to estimation methods as the ones explained next.

3.3 Validation and numerical results

The effective bandwidth function of the sources presented in this chapter have been validated by comparison with measures. In particular, we have implemented two methods to estimate the effective bandwidth function of a source: the Dembo Point Estimator and the algorithm proposed in [Liu 2004], which is a measurement and simulation-based (MSB) approach to estimate the EBF.

Moreover, it happens very often that the EBF of a traffic source cannot be analytically obtained because the model is too complex. In other occasions, only a sample of the traffic is available and it is convenient to work with it. In any of these cases, the estimation methods presented here can be useful.

The Dembo Point Estimator

In [Dufield 1995] the authors propose the following estimator, originally suggested by Amir Dembo for a traffic stream $A[n]$. Consider a trace of duration n. Choose a block size k for which the block sums:

$$\widetilde{A}_1 = \sum_{n=1}^{k} A[n], \qquad \widetilde{A}_2 = \sum_{n=k+1}^{2k} A[n], ... \tag{3.3.1}$$

are approximately independent and identically distributed.

Then, use as estimator:

$$\widetilde{\alpha}(\upsilon) = \frac{1}{\upsilon k} \log \frac{1}{\lfloor n/k \rfloor} \sum_{i=1}^{\lfloor n/k \rfloor} e^{\upsilon \widetilde{A}_i} \tag{3.3.2}$$

The main difficulty is the selection of the block size k, which determines the performance of the estimation. k has to be large enough to assume independence of the blocks but, at the same time, the factor $e^{\upsilon \widetilde{A}_i}$ in the estimator grows with k giving rise to practical problems in the evaluation of $\widetilde{\alpha}(\upsilon)$. In Figure 3.8 the Dembo

estimator is applied to obtain the effective bandwidth function of the video traffic as defined in 802.16. The block size is 10000.

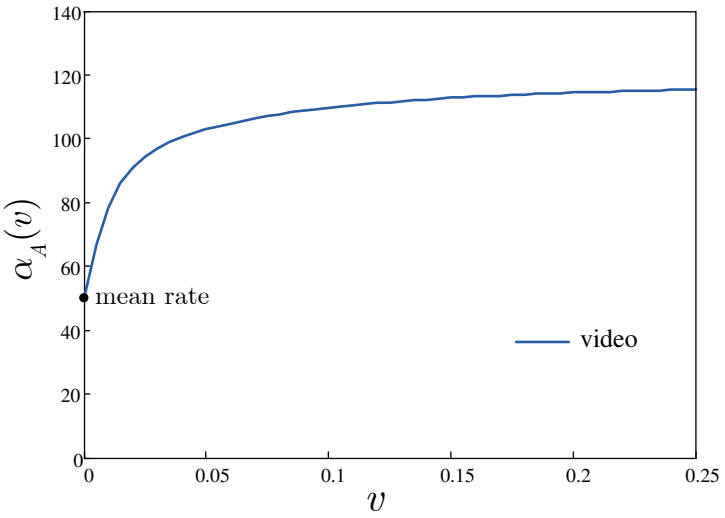


Figure 3.8: Dembo estimator for the video traffic defined in [802.16 2001b].

MSB estimation

In [Liu 2004] the authors propose a method based on measurement and simulation techniques to estimate the effective bandwidth of a source. This method is more practical than the Dembo estimator and it is the preferred option in this thesis.

A sample of the traffic is captured. The program simulates a single-server First Come First Served (FCFS) queue serving the traffic (Figure 3.9). The simulation is controlled by a search algorithm that decides whether the effective bandwidth satisfies the desired QoS requirement. The QoS requirement is specified through a delay constraint (D^t, ε), which states that the probability that the delay is beyond the target delay D^t should not be greater than ε.

The dynamics of the algorithm is as follows. An initial value of the service rate is selected, between the mean and the peak rate of the traffic source. The queue is

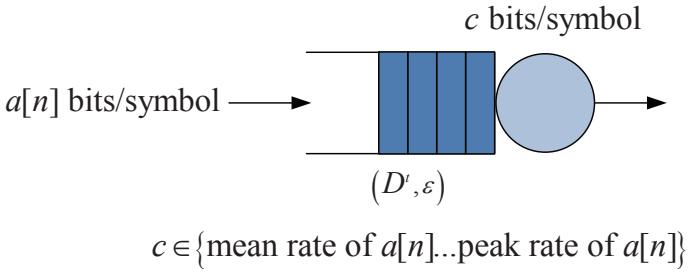


Figure 3.9: MSB estimation of the EBF.

simulated with this value. Then, the algorithm judges whether the QoS requirement is met or not by doing a hypothesis test. If the result is negative, a new value of the service rate is selected accordingly to the behaviour of the delay in current step (longer or shorter than expected) and the simulation is redone. Thus, the value of the service rate is adjusted in each step until it achieves the effective bandwidth of the traffic flow.

Figure 3.10 illustrates the MSB estimation. Traffic voice as defined in 802.16 is generated. The analytical EBF was shown before. The estimation is marked with squares. It can be checked that the simulation fits the expected curve.

3.4 Summary

In this chapter a literature review on traffic modeling has been done. First, the constant bit rate traffic, the classical Markov models and the autoregressive models are revised. Secondly, attention is directed to models that can characterize the mix of traffic arriving to a base station in a wireless system. In this respect, we have found interesting the proposal by the 802.16 Working Group, since a simple model with appropriate parameters is able to represent the set of services present in a wireless communications system, including HTTP, FTP, TCP, voice and streaming. The effective bandwidth function of the presented sources has been addressed. Finally,

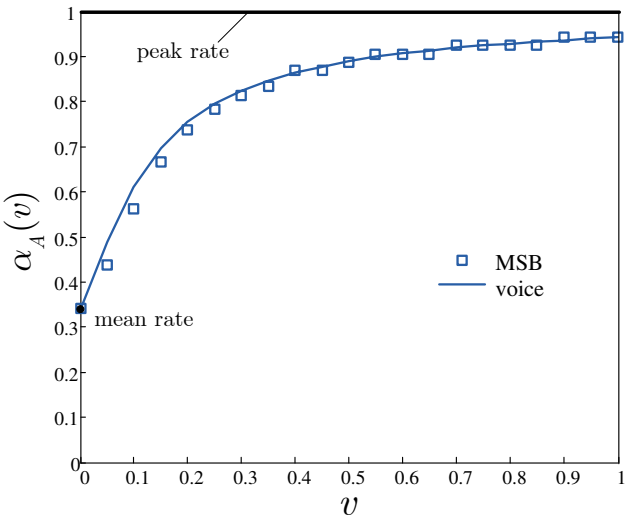


Figure 3.10: Comparison of the estimated EB with MSB and the analytical result for voice traffic.

two approaches to estimate the effective bandwidth function are detailed. They have been used to validate the analytical results presented in this chapter and can also be employed when no analytical solution is available.

Chapter 4

Effective Bandwidth Function of flat Rayleigh channels

In the previous Chapter, the EBF of several traffic sources was studied. In this Chapter the channel process is addressed. Closed-form expressions of the channel EBF of Rayleigh fading channels are obtained, introducing first the easiest case of a block fading channel and a generic time-correlated channel later on. It is worth highlighting that the EBF's are analytically derived from the physical channel. The result will be compared with the one provided by the Finite State Markov Chain (FSMC), a simple model for flat fading channels often found in the literature. The results of this Chapter have been presented in [Soret 2007a] (uncorrelated channel) and [Soret 2010b].

The Chapter is structured as follows. The channel process is described in Section 4.1. The analytical EBF of an uncorrelated Rayleigh channel is derived in Section 4.2. The extension to a generic time-correlated Rayleigh channel is addressed in Section 4.3. Section 4.4 details the Finite State Markov Chain model and its EBF. The obtained results are discussed and evaluated in Section 4.5. Finally, some concluding remarks are given in Section 4.6.

4.1 Channel process

4.1.1 The wireless channel

Rayleigh wireless channel

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 [Rappaport 2002].

Multipath channels are often characterized statistically. When there is not predominant direct line of sight between the transmitter and the receiver, the Rayleigh distribution approximates quite well the channel envelope and the fading is denoted Rayleigh fading. The effect of the channel over the transmitted signal in a flat channel can be represented through the low-pass equivalent of the channel gain, which is a Wide-Sense Stationary (WSS) random process and can be expressed as:

$$h[n] = h_{phase}[n] + jh_{quad}[n] \tag{4.1.1}$$

where the components in phase (h_{phase}) and in quadrature (h_{quad}) are independent of each other and Gaussian distributed.

The instantaneous Signal to Noise Ratio (SNR) at the receiver, $\gamma[n]$, 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 from the transmitter to the 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.

The instantaneous SNR is proportional to the square of $|h[n]|$:

$$\gamma[n] = |h[n]|^2 \frac{E_s}{N_0} \tag{4.1.2}$$

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), $\gamma[n]$ is exponentially distributed for Rayleigh channels, with pdf:

$$f(\gamma) = \frac{1}{\overline{\gamma}} e^{-\frac{\gamma}{\overline{\gamma}}} \tag{4.1.3}$$

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

More details on the Rayleigh channel and the model employed in this thesis can be found in Appendix A.

Autocorrelation function

The variability of the channel over time is usually reflected through the autocorrelation function (ACF) of the channel gain, denoted as $\mathcal{R}_z(m)$. This second-order statistic generally depends on the propagation geometry, the velocity of the mobile and the antenna characteristics. Two different autocorrelation functions have been considered in this thesis for flat Rayleigh channels.

- The first option, known as Jakes' model [Jakes 1989], assumes a uniform scattering environment. The discrete-time autocorrelation function of the channel response does not depend on the time n but just on the time difference m, and it is given by:

  $$\mathcal{R}_z(m) = J_0(2\pi f_D T_S m) \tag{4.1.4}$$

  where $J_0(.)$ is the zeroth order Bessel function of the first kind, f_D is the maximum Doppler frequency in Hertz and T_S is the symbol period.

- As second option, we employ a very simple model of correlation. The ACF is assumed to decay exponentially with a certain parameter ρ, $0 \leq \rho \leq 1$:

  $$\mathcal{R}_z(m) = \rho^m \tag{4.1.5}$$

The use of the exponential model simplifies and speeds up the simulations throughout this thesis without altering the conclusions. Anyway, the classical Jakes' model is also verified in many examples and any other correlation function could be used.

4.1.2 Rate adaptation

Adaptive modulation

Adaptive Modulation and Coding schemes (AMC) are widely employed in current wireless systems. The basic premise is to estimate the channel at the receiver and feed this estimate back to the transmitter [Goldsmith 2005]. Then, some transmission parameters such as constellation size and coding rate are modified dynamically, trying to adapt to the time-varying conditions of the channel. For the sake of simplicity, no encoding is assumed in this thesis, so that only the constellation size is adapted.

The range of received SNR is divided into M consecutive regions, each of which is associated to a constellation size. Thus, M different constellations are available, selecting a constellation of M_i symbols within the fading region (Γ_{i-1}, Γ_i), $i = 0, 1, .., M$ (defining $\Gamma_{-1} = 0$. We assume that Quadrature Amplitude Modulation (QAM) schemes are used. In particular, the set of constellations employed throughout this thesis is shown in Figure 4.1 where, as usual, BPSK (Binary Phase Shift Keying) is the constellation that can transmit 1 bit per symbol and QPSK (Quaternary Phase Shift Keying) in the case of $M = 4$. On their side, 16QAM can transmit 4 bits per symbol and 64QAM, the most dense constellation of this set, provides 6 bits per symbol.

One of the options to design of the SNR thresholds is to fulfill an instantaneous Bit Error Rate condition (I-BER) [Chung 2001]. In Figure 4.2, the curves of the BER of the set of constellations are shown as a function of the received SNR. Thus,

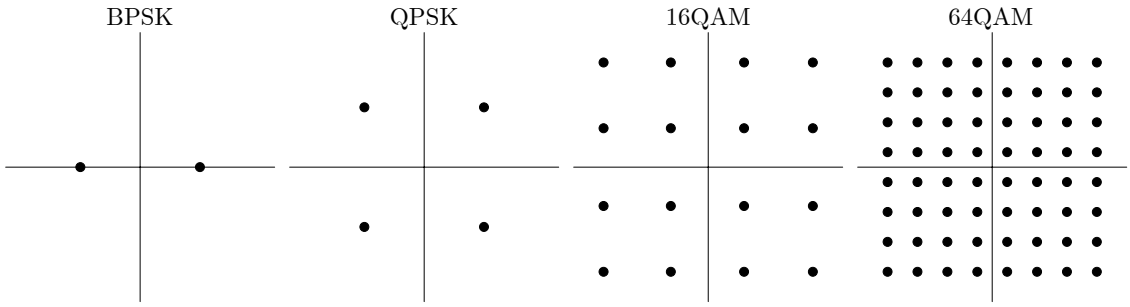


Figure 4.1: Set of constellations.

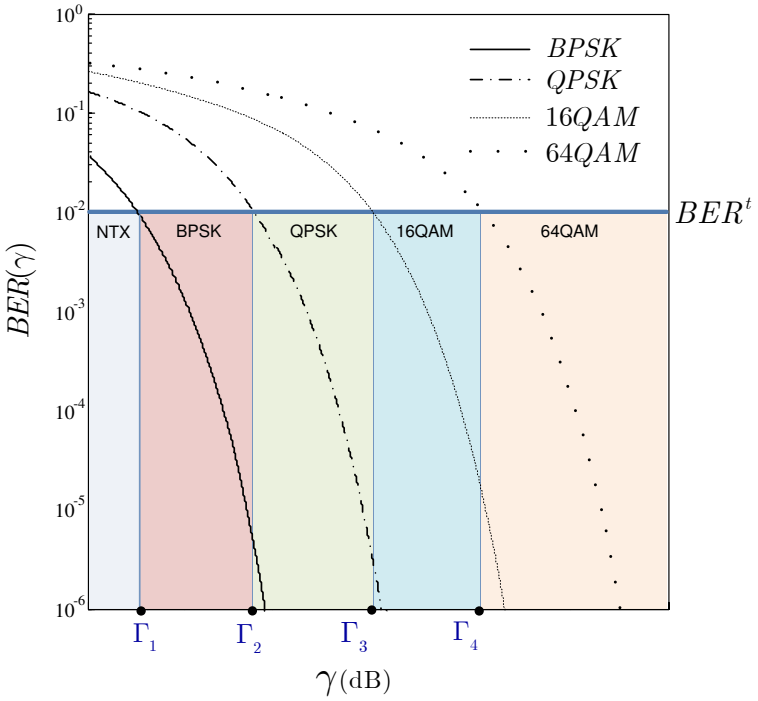


Figure 4.2: Adaptive Modulation.

the curve BER^{BPSK} is the instantaneous BER when using BPSK and likewise for the rest of the curves. A target BER, BER^t, is fixed (10^{-2} in the figure). Thus, 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. The thresholds for the constellations in the figure and target BER of 10^{-2}, 10^{-3} and 10^{-4} are shown in Table 4.1.

Table 4.1: Range of decision in dB.

	NTX	BPSK	QPSK	16QAM	64QAM
bits/symbol / BER^t	0	1	2	4	6
10^{-2}	(-∞..4.3)	(4.3..7.3)	(7.3..13.9)	(13.9..19.8)	(19.8..∞)
10^{-3}	(-∞..6.8)	(6.8..9.8)	(9.8..16.6)	(16.6..22.6)	(22.6..∞)
10^{-4}	(-∞..8.4)	(8.4..11.4)	(11.4..18.2)	(18.2..24.3)	(24.3..∞)

In short, the selection of region i means the utilization of certain transmission (or service) rate. Thus, the instantaneous channel rate $c[n]$ changes according to the channel gain, which determines the constellation size and the corresponding bits per symbol that are available for user transmission in that region.

Rate adaptation policies

Two different rate adaptation policies have been considered in this thesis: continuous and discrete.

- Discrete rate policy

In the discrete rate policy, $c[n]$ can only take a value from a discrete set corresponding to the constellation size available for user transmission in that region. The service

rate $c[n]$ is a function of the envelope of the complex channel gain, $z_n = |h[n]|$:

$$c[n] = c(z_n) = \log_2(M_i) = \mathrm{c}_i, \qquad \varrho^{i-1} \leqslant z_n < \varrho^i,\ i = 0..M, \quad \text{with } \varrho^k = \sqrt{\frac{\Gamma_k}{\bar{\gamma}}} \tag{4.1.6}$$

- Continuous rate policy

The continuous rate policy is a generalization of the discrete rate policy. In this case, the channel rate $c[n]$ can take any positive real number. This adaptation scheme does not have any practical applications in a real system, but it provides a theoretical limit that constitutes an upper bound of our results. With constant transmitted power, the instantaneous service rate can be expressed as:

$$c[n] = \log_2\left(1 + \beta\gamma[n]\right) \tag{4.1.7}$$

If $\beta = 1$ the Shannon-like capacity is evaluated. Under adaptive modulation, β is a constant related to the target BER, BER^t. Channel coding reduced the value of β in accordance to the coding gain. Its value for uncoded QAM [Chung 2001] is:

$$\beta \approx \frac{1.6}{-\log(5BER^t)} \tag{4.1.8}$$

The mean of $c[n]$ is the ergodic capacity of the channel [Biglieri 1998]:

$$\begin{aligned} m_c &= \mathrm{E}\left[\log_2(1+\beta\gamma)\right] = \int_0^\infty \log_2(1+\beta\gamma)\frac{1}{\bar{\gamma}}\exp(-\frac{\gamma}{\bar{\gamma}})d\gamma \\ &= \log_2(e)\left[-\exp(-\frac{\gamma}{\bar{\gamma}})\ln(1+\beta\gamma)\Big|_0^\infty + \int_0^\infty \frac{\beta}{1+\beta\gamma}\exp(-\frac{\gamma}{\bar{\gamma}})d\gamma\right] \\ &= \log_2(e)\exp(\frac{1}{\beta\bar{\gamma}})\mathrm{E}_1(\frac{1}{\beta\bar{\gamma}}) \end{aligned} \tag{4.1.9}$$

where $\mathrm{E}_1(\cdot)$ is the exponential integral function defined as:

$$\mathrm{E}_1(z) = \int_1^\infty \frac{1}{t}e^{-tz}dt \tag{4.1.10}$$

4.2 EBF of uncorrelated Rayleigh channels

The block fading channel (see [Ozarow 1994] [Biglieri 1998]) was introduced to model slowly-varying fading. It can be used when the channel response is the same over constant-sized blocks of channel uses so that it is modeled as a sequence of independent random variables, each of which is the fading gain in a block, with no correlation from one block to the next. This approximation is particularly relevant in wireless communication systems involving slow time-frequency hopping (e.g. GSM, EDGE). Despite its extreme simplification, the model has been shown to be useful in some particular cases and it is the starting point of our calculations, serving upper bounds on the QoS analysis performed later on. We refer to this channel hereafter as block fading or uncorrelated or time-independent channel.

The effective bandwidth function of the channel process is:

$$\alpha_C(\upsilon) = \lim_{n\to\infty} \frac{1}{n\upsilon} \log \mathrm{E}\left[e^{\upsilon C[n]}\right] = \lim_{n\to\infty} \frac{1}{n} \log \mathrm{E}\left[e^{\upsilon \sum_{m=0}^{n-1} c[m]}\right] \tag{4.2.1}$$

If there is no time-correlation among samples, then the accumulated transmission rate $C[n]$ is simply the addition of n uncorrelated and identically distributed random variables, c_m. As $n \to \infty$ the Central Limit Theorem (CLT) applies and $C[n]$ converges to a Gaussian random variable with average $n \cdot m_c$ and variance $n \cdot \sigma_c^2$, where m_c and σ_c^2 are the mean and the variance of the instantaneous channel rate $c[n]$. The pdf yields:

$$f(x) = \frac{1}{\sigma_c \sqrt{n 2\pi}} \exp\left(-\frac{(x - n m_c)^2}{2 n \sigma_c^2}\right) \tag{4.2.2}$$

With the considerations above, the EBF of the channel, $\alpha_C(\upsilon)$, is that of a Gaussian distribution. This function is easily computed:

$$\alpha_C(\upsilon) = \lim_{n\to\infty} \frac{1}{n\cdot\upsilon} \log \mathrm{E}\left[e^{\upsilon \sum_{m=0}^{n-1} c[m]}\right]$$
$$= \lim_{n\to\infty} \frac{1}{n\cdot\upsilon} \log\left[\exp\left(\frac{n(2m_c\sigma_c^2\upsilon + (\sigma_c^2\upsilon)^2}{2\sigma_c^2}\right)\right] = m_c + \frac{\upsilon}{2}\sigma_c^2 \tag{4.2.3}$$

It can be observed that the computation of the EBF comes down to obtaining the mean m_c and the variance σ_c^2 of the channel process $c[n]$ under the specific rate adaptation policy (continuous or discrete as described in Section 4.1.2). Next, these statistics are particularized for both options. Moreover, the EBF is valid for any uncorrelated channel process. We particularize next for a Rayleigh channel, but any other channel can be tackled by only calculating the mean and the variance of $c[n]$.

Continuous rate policy

From Section 4.1.2, if constant transmitted power is assumed the instantaneous service rate can be expressed:

$$c[n] = \log_2\left(1 + \beta\gamma[n]\right) \tag{4.2.4}$$

When $\beta = 1$ the upper bound of the transmission rate is evaluated.

The mean of $c[n]$ is the ergodic capacity of the channel [Biglieri 1998] already described in (4.1.9); we reproduce it here:

$$m_c = \log_2(e)\exp(\frac{1}{\beta\bar{\gamma}})\mathrm{E}_1(\frac{1}{\beta\bar{\gamma}}) \tag{4.2.5}$$

To obtain the variance, $\sigma_c^2 = \mathrm{E}\left[c^2[n]\right] - \mathrm{E}^2\left[c[n]\right]$, the second moment of $c[n]$ needs

to be computed:

$$\begin{aligned}
\mathrm{E}\left[(\log_2(1+\beta\gamma))^2\right] &= \log_2(e)\int_0^\infty (\ln(1+\beta\gamma))^2\,\frac{1}{\bar{\gamma}}e^{-\frac{\gamma}{\bar{\gamma}}}d\gamma \\
&= \log_2(e)\left[-\exp(-\frac{\gamma}{\bar{\gamma}})\,(\ln(1+\beta\gamma))^2\Big|_0^\infty + 2\int_0^\infty \frac{\beta\ln(1+\beta\gamma)}{1+\beta\gamma}\exp(-\frac{\gamma}{\bar{\gamma}})d\gamma\right] \\
&= 2\log_2(e)\left[-\ln(1+\beta\gamma)\exp(\frac{1}{\beta\bar{\gamma}})\mathrm{E}_1\left(\frac{\gamma}{\bar{\gamma}}+\frac{1}{\beta\bar{\gamma}}\right)\Big|_0^\infty\right. \\
&\qquad \left. +\exp(\frac{1}{\beta\bar{\gamma}})\int_0^\infty \frac{\beta}{1+\beta\gamma}\mathrm{E}_1\left(\frac{\gamma}{\bar{\gamma}}+\frac{1}{\beta\bar{\gamma}}\right)d\gamma\right] \\
&= 2\log_2(e)\exp(\frac{1}{\beta\bar{\gamma}})\left[\int_0^\infty \frac{\beta}{1+\beta\gamma}\mathrm{E}_1\left(\frac{\gamma}{\bar{\gamma}}+\frac{1}{\beta\bar{\gamma}}\right)d\gamma\right] \qquad (4.2.6)
\end{aligned}$$

The following result will be employed to solve the integral in the last step:

$$\int_a^\infty \frac{1}{z}\mathrm{E}_1(z)dz = \frac{1}{2}\left(\frac{\pi^2}{6}+g^2+2g\ln(a)+\ln^2(a)\right) - a\;{}_3\mathbf{F}_3\left([1,1,1],[2,2,2],-a\right) \qquad (4.2.7)$$

where:

- g is the Euler constant:

$$g = \lim_{n\longrightarrow\infty}\sum_{i=1} n\frac{1}{i} - \ln(n) \approx 0.5772$$

- ${}_p\mathbf{F}_q([a_1,...,a_p];[b_1,...,b_q];z)$ is the Hypergeometric function:

$${}_p\mathbf{F}_q([a_1,...,a_p];[b_1,...,b_q];z) = \sum_{n=0}^{\infty}\frac{(a_1)_n...(a_p)_n}{(b_1)_n...(b_q)_n}\cdot\frac{z^n}{n!}$$

- $(p)_n$ is the Pochhammer symbol:

$$(p)_n = p(p+1)(p+2)...(p+n-1)$$

The inclusion of the obtained second moment in the variance of $c[n]$ yields:

$$\begin{aligned}
\sigma_c^2 &= \mathrm{E}\left[(\log_2(1+\beta\gamma))^2\right] - m_c^2 \\
&= (\log_2(e))^2 e^{\frac{1}{\beta\bar{\gamma}}}\left[\frac{\pi^2}{6}+g^2+2g\log\left(\frac{1}{\beta\bar{\gamma}}\right)+\log^2\left(\frac{1}{\beta\bar{\gamma}}\right)\right. \\
&\quad \left. -2\left(\frac{1}{\beta\bar{\gamma}}\right){}_3\mathbf{F}_3\left([1,1,1],[2,2,2],-\frac{1}{\beta\bar{\gamma}}\right) - e^{\frac{1}{\beta\bar{\gamma}}}\mathrm{E}_1^2(\frac{1}{\beta\bar{\gamma}})\right] \qquad (4.2.8)
\end{aligned}$$

Discrete rate policy

The other rate policy consists of a transmission adaptation scheme following a discrete rate policy with constant transmitted power and $c[n]$ taking only a finite number of integer values.

For Rayleigh channels, the probability of using the ith constellation is given by:

$$p_i = \exp(-\frac{\gamma_{i-1}}{\bar{\gamma}}) - \exp(-\frac{\gamma_i}{\bar{\gamma}}) \tag{4.2.9}$$

and the mean and the variance of $c[n]$ are quite easily computed for the discrete random variable:

$$m_c = \sum_{i=0}^{M} p_i \cdot \mathrm{c}_i \tag{4.2.10}$$

$$\sigma_c^2 = \sum_{i=0}^{M} p_i \cdot (\mathrm{c}_i)^2 - m_c^2 \tag{4.2.11}$$

The set of constellations and thresholds Γ_k to be used throughout this thesis were detailed in Section 4.1.

4.3 EBF of time-correlated Rayleigh channels

After studying the uncorrelated channel, we are ready to deal with the EBF of a generic time-correlated Rayleigh channel. With the purpose of applying again the CLT, the cumulative transmission rate $C[n]$ in the interval $[0..n]$ is split into B blocks of size k, like shown in Figure 4.3.

With that division $C[n]$ is expressed:

$$C[n] = \sum_{b=0}^{B-1} C_b[k] = \sum_{b=0}^{B-1} \sum_{m=0}^{k-1} c[k \cdot b + m] \tag{4.3.1}$$

The correlation between the elements in the block is considered but, with the proper selection of the block's length, i.e. k large enough, independence between

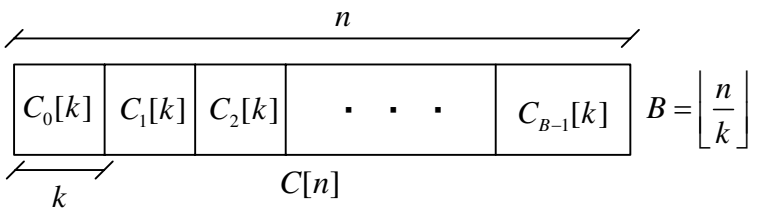


Figure 4.3: Block division of C[n].

elements of different blocks may be assumed. Thus, $C[n]$ can be considered as the sum of independent random variables, the blocks $C_b[k]$, and the CLT can be applied over the blocks. To justify this decision, several considerations have to be pointed out:

- The autocorrelation function is a decreasing function (Bessel decay, exponential decay or any other ACF).
- The choice of k will be closely related to the correlation of the channel. If the channel is strongly correlated, longer blocks have to be defined in order to assume independent blocks. Whatever the value of k is, there is a residual value of correlation between the last elements of one block and the first elements of next one. Nevertheless, this *border* correlation is negligible when the value of k is large enough.
- The EBF is defined as the limit when n tends to infinite, so that n and k can be selected indeed as large as wanted.
- In statistics, the Lilliefors test [Sheskin 2004] is an adaptation of the Kolmogorov - Smirnov test. It is used to test whether an observed sample distribution is consistent with normality. The statistic measures the maximum distance between the observed distribution and a Gaussian distribution with the same mean and standard deviation as the sample, and assesses if this distance is greater than might be accounted for by chance. In the numerical

results and simulations conducted throughout this dissertation, the Gaussian approximation for $C[n]$ has been validated by testing for normality with the Lilliefors test for the selected values of k and n.

Under these conditions (sufficiently long k and n), $C[n]$ is the sum of a sufficiently large number of independent random variables and, as stated by the Central Limit Theorem, it will be approximately normally distributed. The EBF for the resulting Gaussian distribution of $C[n]$ is computed similarly to the uncorrelated channel. Now the mean is $B \cdot m_k$, whereas the variance is $B \cdot \sigma_k^2$, being m_k and σ_k^2 the mean and the variance of a block of size k under the adaptation rate policy:

$$\begin{aligned} m_k &= \mathrm{E}[C_b[k]] \\ \sigma_k^2 &= \mathrm{E}[C_b^2[k]] - m_k^2 \end{aligned} \tag{4.3.2}$$

The effective bandwidth function yields:

$$\begin{aligned} \alpha_C(\upsilon) &= \lim_{n\to\infty} \frac{1}{n \cdot \upsilon} \log \mathrm{E}\left[e^{\upsilon C[n]}\right] \\ &= \lim_{n\to\infty} \frac{1}{n \cdot \upsilon} \left(\frac{n}{k} m_k \upsilon + \frac{n}{2k} \upsilon^2 \sigma_k^2\right) = \frac{m_k}{k} + \frac{\upsilon}{2} \frac{\sigma_k^2}{k} \end{aligned} \tag{4.3.3}$$

The evaluation of the EBF comes down again to the computation of two statistics related to the channel rate, but this time it is about the mean and the variances of the blocks.

Like in the uncorrelated channel, it is worth noting that the argument up to here is valid for any channel process, since the split of the accumulated transmission rate into blocks as long as necessary does not depend on the process. The mean and the variance of the blocks in the case of a Rayleigh channel is derived next.

Being $C[n]$ a stationary and ergodic process, the mean of each block equals the mean value of the process without correlation:

$$m_k = \mathrm{E}[C_b[k]] = k \cdot m_c \tag{4.3.4}$$

where m_c is the mean of $c[n]$, equations (4.2.5) and (4.2.10) for continuous and discrete rate policy, respectively.

However, the evaluation of the variance of the blocks is not straightforward as it results from a multivariate Rayleigh distribution.

The following property will be employed to derive σ_k^2 ([Papoulis 2002]):

$$\mathrm{var}\left(\sum_{p=0}^{k-1} X_p\right) = \sum_{q=0}^{k-1}\sum_{r=0}^{k-1} \mathrm{cov}(X_q, X_r) \tag{4.3.5}$$

with the autocovariance function

$$\mathcal{K}_X(m) = \mathrm{cov}(X_n, X_{n+m}) = \mathrm{E}[X_n X_{n+m}] - \mathrm{E}[X_n]\mathrm{E}[X_{n+m}] \tag{4.3.6}$$

Applying the expression above to $C[n]$ we obtain:

$$\begin{aligned}\sigma_k^2 = \mathrm{var}(\sum_{p=0}^{k-1} c[p]) &= \sum_{q=0}^{k-1} \mathrm{var}(c[q]) + 2\sum_{q=0}^{k-2}\sum_{r=q+1}^{k-1} \mathcal{K}_c(r-q) \\ &= k\cdot\sigma_c^2 + 2\sum_{q=0}^{k-2}\sum_{r=q+1}^{k-1} \mathcal{K}_c(r-q)\end{aligned} \tag{4.3.7}$$

with σ_c^2 the variance of $c[n]$, and the autocovariance function

$$\mathcal{K}_c(m) = \mathrm{E}[c[n]c[n+m]] - m_c^2 \tag{4.3.8}$$

The value of σ_c^2 for the instantaneous service rate has been obtained in Section 4.2 for continuous and discrete rate policy, equations (4.2.8) and (4.2.11) respectively. Thus, only the evaluation of $\mathrm{E}[c[n]c[n+m]]$ in the autocovariance function in (4.3.8) is needed to complete the analysis. This expectation is calculated next for the two adaptation policies.

Covariance for a continuous rate policy

To derive $\mathrm{E}[c[n]c[n+m]]$, we need the following expression of the bivariate probability density function (pdf) for Rayleigh distributed variables in terms of the instantaneous SNR [Simon 2005] (p. 142, eq. 6.2.):

$$f_\gamma(\gamma_n, \gamma_{n+m}) = \frac{1}{(1-\mathcal{R}_z^2(m))\bar{\gamma}^2} \exp\left(-\frac{(\gamma_n+\gamma_{n+m})}{(1-\mathcal{R}_z^2(m))\bar{\gamma}}\right) \cdot I_0\left(\frac{2\mathcal{R}_z(m)\sqrt{\gamma_n\gamma_{n+m}}}{(1-\mathcal{R}_z^2(m))\bar{\gamma}}\right) \tag{4.3.9}$$

where

- $I_0(u)$ the modified Bessel function of 0th order
- $\mathcal{R}_z(m)$ is the value of the autocorrelation function of $z_n = |h[n]|$ for a time lag m:

$$\mathcal{R}_z(m) = \frac{\mathrm{cov}(|h_n|, |h_{n+m}|)}{\sqrt{\mathrm{var}(|h_n|)\,\mathrm{var}(|h_{n+m}|)}} \tag{4.3.10}$$

We proceed to evaluate $\mathrm{E}[c[n]c[n+m]]$ explicitly:

$$\begin{aligned}
&\mathrm{E}\left[c[n]c[n+m]\right] = \mathrm{E}\left[c(\gamma_n)c(\gamma_{n+m})\right] \\
&= \int_{\gamma_n=0}^{\infty}\int_{\gamma_{n+m}=0}^{\infty} c(\gamma_n)c(\gamma_{n+m})f_\gamma(\gamma_n,\gamma_{n+m})d\gamma_n d\gamma_{n+m} \\
&= \int_{\gamma_n=0}^{\infty}\int_{\gamma_{n+m}=0}^{\infty} \log_2(1+\beta\gamma_n)\log_2(1+\beta\gamma_{n+m}) \\
&\cdot \frac{1}{(1-\mathcal{R}_z^2(m))\bar{\gamma}^2} \cdot \exp\left(-\frac{(\gamma_n+\gamma_{n+m})}{(1-\mathcal{R}_z^2(m))\bar{\gamma}}\right) \cdot I_0\left(\frac{2\mathcal{R}_z(m)\sqrt{\gamma_n\gamma_{n+m}}}{(1-\mathcal{R}_z^2)\bar{\gamma}}\right) d\gamma_n d\gamma_{n+m}
\end{aligned} \tag{4.3.11}$$

The modified Bessel function in (4.3.11) can be expanded as an infinite series:

$$I_0(\mathrm{u}) = \sum_{\mathrm{k}\,=\,0}^{\infty} \frac{1}{(k!)^2}\left(\frac{\mathrm{u}}{2}\right)^{2\mathrm{k}} \tag{4.3.12}$$

Some algebraic manipulations can be done then in the integral that permit to split the parts in γ_n and in γ_{n+m}, reaching the following expression:

$$\begin{aligned}
\mathrm{E}\left[c[n]c[n+m]\right] = & \int_{\gamma_n=0}^{\infty}\int_{\gamma_{n+m}=0}^{\infty} \log_2(1+\beta\gamma_n)\log_2(1+\beta\gamma_{n+m}) \\
& \cdot \sum_{\mathrm{k}\,=\,0}^{\infty} \frac{1}{(k!)^2}\left(\frac{\mathcal{R}_z(m)}{(1-\mathcal{R}_z^2(m))\bar{\gamma}}\frac{\mathcal{R}_z(m)}{(1-\mathcal{R}_z^2(m))\bar{\gamma}}\gamma_n\gamma_{n+m}\right)^{\mathrm{k}} \\
& \cdot \frac{1}{(1-\mathcal{R}_z^2(m))\bar{\gamma}^2}\cdot\exp\left(-\frac{(\gamma_n+\gamma_{n+m})}{(1-\mathcal{R}_z^2(m))\bar{\gamma}}\right)d\gamma_n d\gamma_{n+m} \\
= & \frac{a}{\bar{\gamma}}\sum_{\mathrm{k}\,=\,0}^{\infty}\frac{1}{k!k!}\int_{\gamma_n=0}^{\infty}\int_{\gamma_{n+m}=0}^{\infty} \log_2(1+\beta\gamma_n)\log_2(1+\beta\gamma_{n+m}) \\
& \cdot \mathrm{e}^{-a\gamma_n}\mathrm{e}^{-a\gamma_{n+m}}\left(a\mathcal{R}_z(m)\gamma_n\right)^k\cdot\left(a\mathcal{R}_z(m)\gamma_{n+m}\right)^k d\gamma_n d\gamma_{n+m} \\
= & \frac{a}{\bar{\gamma}}\sum_{k=0}^{\infty}\left(I_k(\beta,\mathcal{R}_z(m),a)\right)^2
\end{aligned} \tag{4.3.13}$$

where:

- The parameter a is:

$$a = \frac{1}{(1-\mathcal{R}_z^2(m))\cdot\bar{\gamma}} \tag{4.3.14}$$

- The integral $I_k(\beta,\mathcal{R}_z(m),a)$ is defined as:

$$I_k(\beta,\mathcal{R}_z(m),a) \triangleq \frac{\left(a\mathcal{R}_z(m)\right)^k}{k!\log(2)}\int_{x=0}^{\infty}\log(1+\beta x)\cdot\mathrm{e}^{-ax}x^k dx \tag{4.3.15}$$

The solution to the integral $I_k(\beta,\mathcal{R}_z(m),a)$ can be obtained from a generic result in [Prudnikov 1981] (p. 530, sec. 2.6.23), and taking into account that k is integer. The final result is:

$$I_k(\beta,\mathcal{R}_z(m),a) = \frac{\mathcal{R}_z^k(m)}{k\cdot\log(2)}\left(-\frac{k}{a}\left(-\psi(1+k)+\log(a)\right)+{}_2\mathbf{F}_2([1,1],[2,1-k],a)\right) \tag{4.3.16}$$

where:

- $\psi(x)$ is the digamma function:

$$\psi(x) = \frac{d(\log(\Gamma(x)))}{dx}$$

- $\Gamma(x)$ is the gamma function:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$

Covariance for a discrete rate policy

In this case, we work with the envelope of the complex response of the channel $z_n = |h_n|$. We proceed to evaluate $\mathrm{E}[c[n]c[n+m]]$ explicitly. First, the integral in z_n is split into several addends:

$$\begin{aligned}
\mathrm{E}[c[n]c[n+m]] &= \int_{z_n=0}^{\infty} \int_{z_{n+m}=0}^{\infty} c(z_n)c(z_{n+m}) f_{\mathbf{z}}(z_n, z_{n+m}) dz_n dz_{n+m} \\
&= \int_{z_n=0}^{\infty} c(z_n) \int_{z_{n+m}=0}^{\infty} c(z_{n+m}) f_{\mathbf{z}}(z_n, z_{n+m}) dz_n dz_{n+m} \\
&= \int_{z_n=\varrho^0}^{\varrho^1} c(z_n) \int_{z_{n+m}=0}^{\infty} c(z_{n+m}) f_{\mathbf{z}}(z_n, z_{n+m}) dz_n dz_{n+m} \\
&+ \int_{z_n=\varrho^1}^{\varrho^2} c(z_n) \int_{z_{n+m}=0}^{\infty} c(z_{n+m}) f_{\mathbf{z}}(z_n, z_{n+m}) dz_n dz_{n+m} \\
&+ \ldots \\
&+ \int_{z_n=\varrho^{M-1}}^{\varrho^M} c(z_n) \int_{z_{n+m}=0}^{\infty} c(z_{n+m}) f_{\mathbf{z}}(z_n, z_{n+m}) dz_n dz_{n+m} = \qquad (4.3.17) \\
&= \sum_{i=1}^{M} \mathrm{c}_i \int_{z_n=\varrho^{i-1}}^{\varrho^i} \int_{z_{n+m}=0}^{\infty} c(z_{n+m}) f_{\mathbf{z}}(z_n, z_{n+m}) dz_n dz_{n+m} \qquad (4.3.18)
\end{aligned}$$

And splitting now over the integral in z_{n+m} we finally obtain:

$$\mathrm{E}[c[n]c[n+m]] = \sum_{i=1}^{M} \mathrm{c}_i \sum_{j=1}^{M} \mathrm{c}_j \int_{z_n=\varrho^{i-1}}^{\varrho^i} \int_{z_{n+m}=\varrho^{j-1}}^{\varrho^j} f_{\mathbf{z}}(z_n, z_{n+m}) dz_n dz_{n+m} \qquad (4.3.19)$$

The following expression is the definition of bivariate cumulative distribution function (CDF) for Rayleigh distributed variables in terms of the probability density function $f_{\mathbf{z}}(z_n, z_{n+m})$:

$$F_{\mathbf{z}}(\varrho^i, \varrho^j, m) = \int_{z_n=0}^{\varrho^i} \int_{z_{n+m}=0}^{\varrho^j} f_{\mathbf{z}}(z_n, z_{n+m}) dz_n dz_{n+m} \tag{4.3.20}$$

In [Simon 2005] the authors provide an integral formula for the bivariate CDF (p. 144, ec. 6.7). In this case, with normalized power, the expression reduces to:

$$F_{\mathbf{z}}(u, v, m) = 1 - g(u, v, m) +$$
$$+\frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\left(-\frac{u^2 + v^2 + 2\sqrt{\mathcal{R}_{z^2}(m)}uv\sin\theta}{1 - \mathcal{R}_{z^2}(m)}\right)$$
$$\cdot \left[\frac{\left(1 - \mathcal{R}_{z^2}(m)\right) u^2v^2 + \sqrt{\mathcal{R}_{z^2}(m)}\left(1 - \mathcal{R}_{z^2}(m)\right) uv\left(u^2 + v^2\right)\sin\theta}{\left(\mathcal{R}_{z^2}(m)u^2 + 2\sqrt{\mathcal{R}_{z^2}(m)}uv\sin\theta + v^2\right)\left(u^2 + 2\sqrt{\mathcal{R}_{z^2}(m)}uv\sin\theta + \mathcal{R}_{z^2}(m)v^2\right)}\right] d\theta,$$

$$g(u, v, m) = \begin{cases} \exp(-v^2) & 0 \leqslant v < \sqrt{\mathcal{R}_{z^2}(m)}u \\ \frac{1}{2}\exp(-u^2) + \exp(-u^2\mathcal{R}_{z^2}(m)) & v = \sqrt{\mathcal{R}_{z^2}(m)}u \\ \exp(-u^2) + \exp(-v^2) & \sqrt{\mathcal{R}_{z^2}(m)}u \leqslant v < u/\sqrt{\mathcal{R}_{z^2}(m)} \\ \frac{1}{2}\exp(-v^2) + \exp(-v^2\mathcal{R}_{z^2}(m)) & v = u/\sqrt{\mathcal{R}_{z^2}(m)} \\ \exp(-u^2) & u/\sqrt{\mathcal{R}_{z^2}(m)} < v \end{cases} \tag{4.3.21}$$

$\mathcal{R}_{z^2}(m)$ is the value of the autocorrelation function of the square of z_n for a time lag m:

$$\mathcal{R}_{z^2}(m) = \frac{\operatorname{cov}(z_n^2, z_{n+m}^2)}{\sqrt{\operatorname{var}(z_n^2)\operatorname{var}(z_{n+m}^2)}} \tag{4.3.22}$$

Define the function:

$$C_2^m(\varrho^i, \varrho^j) \triangleq \int_{z_n=\varrho^{i-1}}^{\varrho^i} \int_{z_{n+m}=\varrho^{j-1}}^{\varrho^j} f_{\mathbf{z}}(z_n, z_{n+m}) dz_n dz_{n+m}$$
$$= F_{\mathbf{z}}(\varrho^i, \varrho^j, m) + F_{\mathbf{z}}(\varrho^{i-1}, \varrho^{j-1}, m) - F_{\mathbf{z}}(\varrho^{i-1}, \varrho^j, m) - F_{\mathbf{z}}(\varrho^i, \varrho^{j-1}, m) \tag{4.3.23}$$

With the definition above, the final result is written as:

$$\mathrm{E}[c[n]c[n+m]] = \sum_{i=1}^{M} \mathrm{c}_i \sum_{j=1}^{M} \mathrm{c}_j \; C_2^m(\varrho^i, \varrho^j) \tag{4.3.24}$$

Validation of the variance results

The results of the obtained analytical variances are validated, on account of its main significance in the calculation of the effective bandwidth function. To that end, the sampled variance is obtained with a program in MATLAB that splits the accumulated transmission rate into blocks of length k and measures the variance over a long realization. This measured variance is compared with the analytical one. Several numerical examples are shown here, for continuous and discrete policies. Both the analytical and measured variances are normalized with the block length k in the figures. Thus, the Gaussian approximation can be validated: if the normalized variance is independent of k, then the Gaussian approximation is valid for these values of channel and block length k. On the other hand, if the normalized variance increases with k, then a larger value of the block length k should be considered for that channel (i.e. for that value of time-correlation).

In the expressions in (4.3.13) and (4.3.24) there is a dependence on the autocorrelation function of the channel response, $\mathcal{R}_z(m)$. The ACF's in equations (4.1.4) and (4.1.5) are considered here[1]. Besides, for calculating the EBF of the radio channel the transmission rate has been split into blocks of length k. It is equivalent to say that the original process is filtered by a rectangular window of length k. As a consequence, the ACF of the resulting process is the original autocorrelation function of a Rayleigh channel multiplied by a triangular function, which is the result of the convolution of the rectangular window with itself [Papoulis 2002].

On the other hand, the application of the CLT over the accumulated transmission

[1]Recall that the procedure is generic and any other autocorrelation function can be considered.

rate demands that the length of the blocks is high enough to assume independence among blocks, so that the effect of windowing the original autocorrelation becomes negligible as the size of the blocks increases. Nevertheless, the ACF $\mathcal{R}_z(m)$ used hereinafter is the original one modified with a triangular function, for the computation of the autocovariance in (4.3.13) and (4.3.24).

Firstly, the exponential model for the ACF is employed (equation (4.1.5)). Figures 4.4 and 4.5 show the evaluation of the variance in (4.3.7) for a continuous rate policy and the ACF following an exponential rule with parameter ρ. The normalized variance is plot as a function of the block length k.

In Figure 4.4 a fixed value of $\rho = 0.7$ is set, and different values of the mean SNR (5, 10 and 15dB) are represented, whereas Figure 4.5 presents the influence of the correlation through different values of ρ ($\rho = 0.7, 0.8$ and 0.9). The average SNR is set to 5dB.

As expected, the normalized variance is constant with regard to the block length, both the analytical and the measured one. It is also an indirect sign that the Gaussian approximation is valid for the represented block lengths (from 500 to 2000), meaning that those values of k are big enough for the selected parameters of the channel (SNR and ρ).

Furthermore, the variance increases with the instantaneous SNR and with ρ. Thus, the variance is the parameter capturing the autocorrelation of the channel response in the following way: higher time-correlation leads to higher values of the variance σ_k^2 and viceversa.

Finally, it can be checked that for higher values of the correlation, a bigger value of the block length is needed in order to ensure the independence among blocks. Thus, if we observe the variance for $\rho = 0.9$, the measure and the analysis slightly differ for the smaller values of k plot in the Figure. As the block length increases, these differences become negligible.

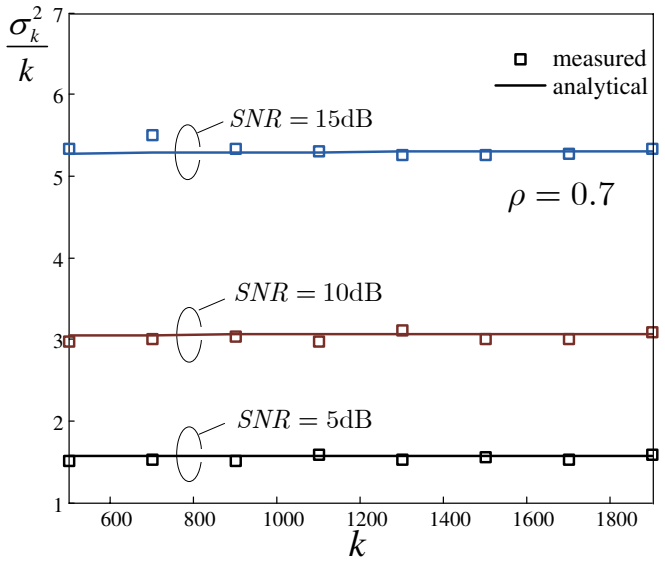


Figure 4.4: Variance of the blocks for a continuous rate policy. Exponential ACF. Influence of the SNR

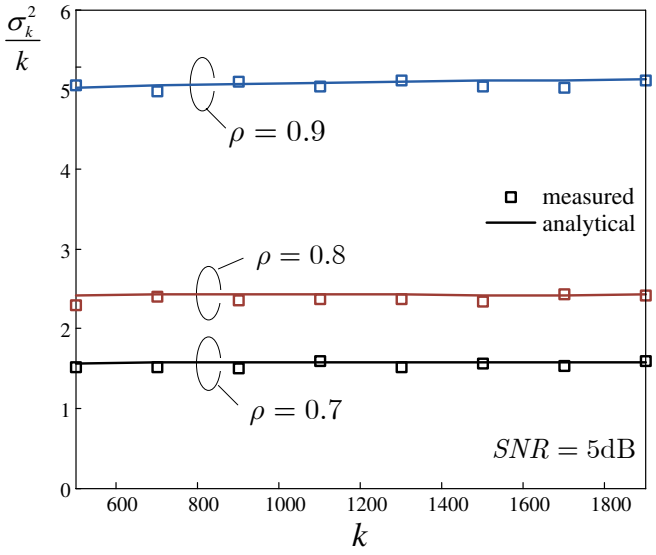


Figure 4.5: Variance of the blocks for a continuous rate policy. Exponential ACF. Influence of the correlation

The results for a discrete rate policy are also satisfactory, as shown in Figures 4.6 and 4.7, with the same conditions of the continuous rate policy: exponential ACF with parameter ρ and the same three values of instantaneous SNR. Notice that lower values of the block length k are needed here to accomplish the Gaussian assumption.

Lastly, the variance of a channel with the ACF from Jakes' model (equation (4.1.4)) is illustrated in Figures 4.8 and 4.9, for a discrete rate policy. In Figure 4.8 the correlation product $f_D \cdot T_S$ is set to 0.1, and different values of the average SNR are presented (5, 10 and 15dB). In Figure 4.9 different values of the product $f_D \cdot T_S$ are shown ($f_D \cdot T_S = 0.1, 0.01$ and 0.001) and the SNR is fixed to $=$ 5dB. As expected, the variance increases with the value of average SNR and with $f_D \cdot T_S$. Moreover, the values of k needed to fulfill the Lilliefors test are much higher in this evaluation. This is due to the oscillating nature of the Bessel function, which demands higher values of k, in contrast to the exponential ACF. For the highest value of correlation shown in this Figure, 0.001, the variance did not pass the Lilliefors test for the plot values of k. Indeed, it could have been predicted from the observation of the figure, since the normalized variance keeps growing with the parameter k. In this case, a minimum block length of $k = 15000$ is required.

It is worth highlighting the convenience of measuring the variance of $c[n]$. After some calculations, we have obtained it for a Rayleigh channel with a generic Doppler spectrum. Nevertheless it can be difficult to get the analytical expression of σ_c^2 for other channel models or in other scenarios. For example, in Chapter 7 we will see the difficulty in obtaining it in a multi-user system. In these cases, a semi-analytical approach consists of measuring the variance as in the previous figures and then using it in the expression of $\alpha_C(-\upsilon)$, as it will be explained in Section 4.5.

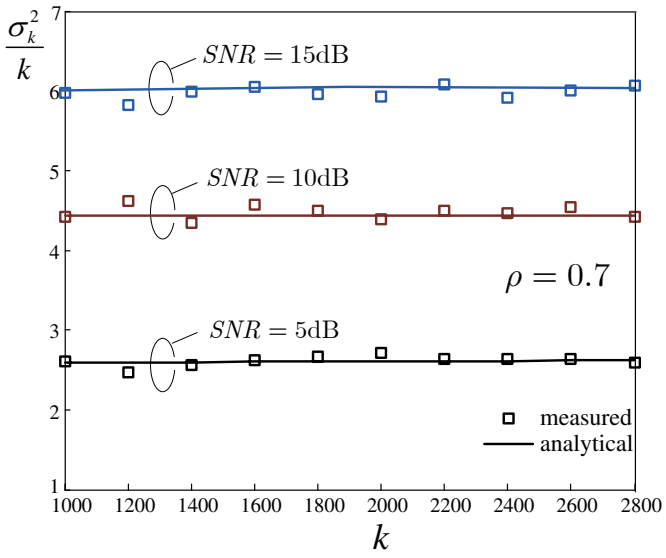


Figure 4.6: Variance of the blocks for a discrete rate policy. Exponential ACF. Influence of the SNR

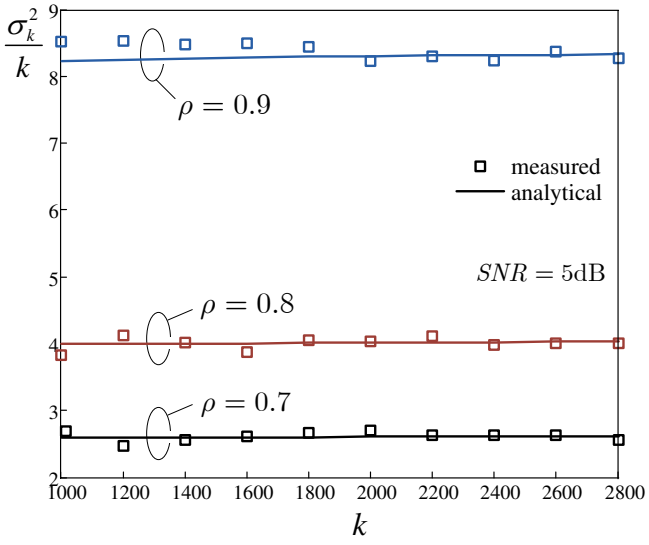


Figure 4.7: Variance of the blocks for a discrete rate policy. Exponential ACF. Influence of the correlation

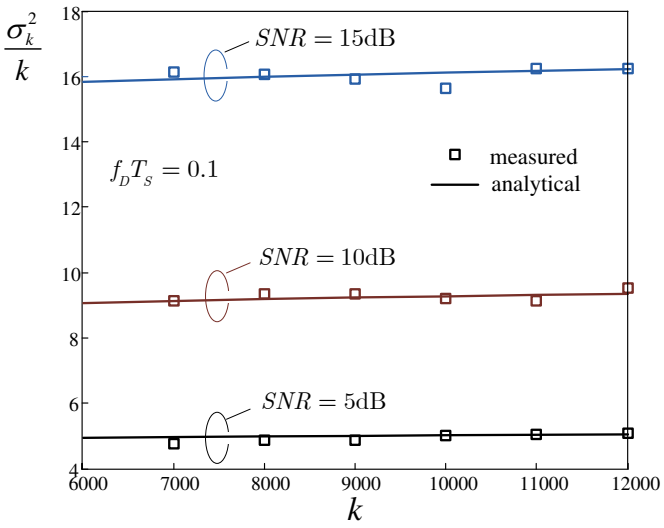


Figure 4.8: Variance of the blocks for a discrete rate policy. Bessel ACF. Influence of the SNR

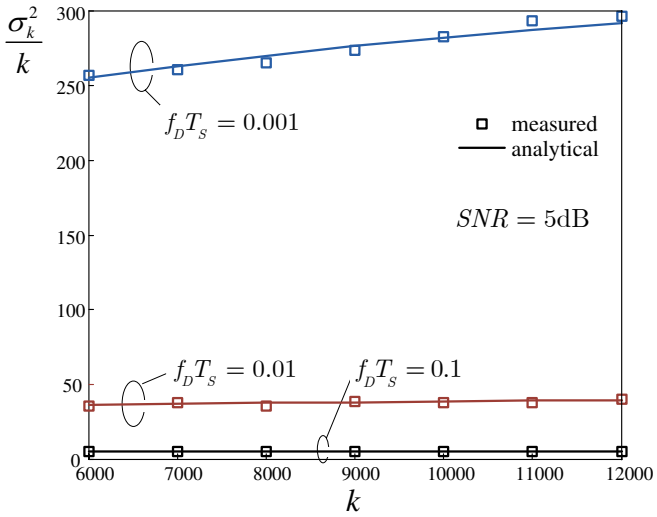


Figure 4.9: Variance of the blocks for a discrete rate policy. Bessel ACF. Influence of the autocorrelation

4.4 EBF based on FSMC models

The analytical effective bandwidth of the Rayleigh channel has been derived from the physical channel model described in Section 4.1 and Appendix C. However, simpler models have been often preferred in the literature. One of these models is the Finite State Markov Chain (FSMC) [Zhang 1999].

FSMC models are versatile, and with suitable choices of model parameters, can capture the essence of time-varying fading channels while avoiding complex mathematical formulations. It is a suitable model for many upper-layer procedures but, as it will be shown further on, not for the objectives of this thesis. Nevertheless, the study of the effective bandwidth function of a channel described through a FSMC is done here since the model has been employed by several authors [Hassan 2004] [Park 2006] [Tang 2006] [Tang 2007b] to analyze some aspects of the QoS in a wireless system. The comparison with our analytical results will demonstrate the limitations of the model. A study of the parameters and performance of the FSMC is done in Appendix B.

In a FSMC, 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}. Thus, each state of the chain represents one fading region (Figure 4.10).

The steady state probability for state i is directly the probability that the exponentially distributed SNR is between the thresholds of the region:

$$\pi_i = \int_{\Gamma_i}^{\Gamma_{i+1}} p(\gamma) d\gamma = e^{-\frac{\Gamma_i}{\bar{\gamma}}} - e^{-\frac{\Gamma_{i+1}}{\bar{\gamma}}} \tag{4.4.1}$$

An adaptive scheme over a fading channel can be directly mapped into a FSMC, so that each fading region is characterized by a constellation size. In short, the selection of region i means the utilization of a transmission rate r_i, the number of bits per second available for user transmission in that region, i.e., discrete rates are assumed.

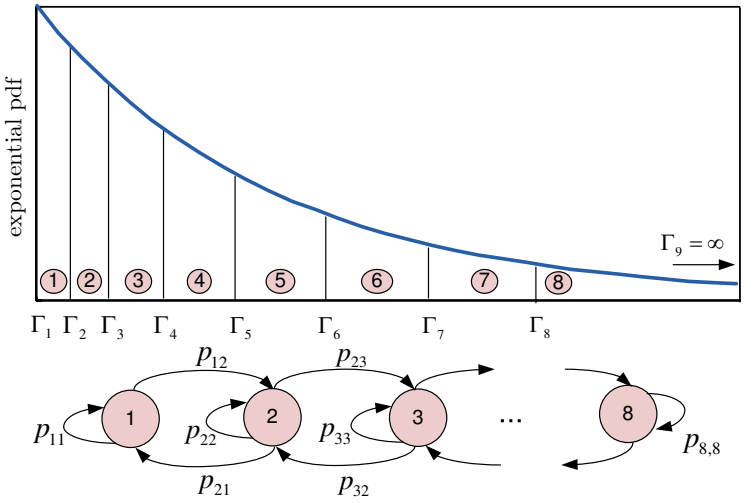


Figure 4.10: Finite State Markov Chain.

The Level Crossing Rate (LCR) of the instantaneous SNR 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:

$$N(\Gamma) = \sqrt{\frac{2\pi \cdot \Gamma}{\overline{\gamma}}} f_D \cdot \exp(-\frac{\Gamma}{\overline{\gamma}}) \tag{4.4.2}$$

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 bits per second staying in state i. Similarly, the transition probability $p_{i,i-1}$ is approximately the ratio of the LCR at threshold Γ and the average number of bits per second staying in state i:

$$p_{i,i+1} \approx \frac{N(\Gamma_{i+1}) \cdot T_S}{\pi_i} \tag{4.4.3}$$

$$p_{i,i-1} \approx \frac{N(\Gamma_i) \cdot T_S}{\pi_i} \tag{4.4.4}$$

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

The EBF of a FSMC is a well-known result [Kessidis 1996]:

$$\alpha_C(\upsilon) = \frac{\Lambda_C(\upsilon)}{\upsilon} = \frac{1}{\upsilon}\log(\Omega(\mathbf{P}\cdot\mathbf{C}(\upsilon))) \tag{4.4.5}$$

where

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

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

$$\mathbf{P}\cdot\mathbf{C}(\upsilon) = \begin{bmatrix} p_{00}e^{r_0\upsilon} & p_{01}e^{r_1\upsilon} & 0 & 0 & \cdots \\ p_{10}e^{r_0\upsilon} & p_{11}e^{r_1\upsilon} & p_{12}e^{r_2\upsilon} & 0 & \\ 0 & p_{21}e^{r_1\upsilon} & p_{22}e^{r_2\upsilon} & p_{23}e^{r_3\upsilon} & \\ \vdots & & \vdots & & \ddots \\ & & & & \end{bmatrix} \tag{4.4.6}$$

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

$$\mathbf{P}\cdot\mathbf{C}(\upsilon) = \begin{bmatrix} p_{00}e^{r_0\upsilon} & p_{01}\sqrt{\frac{\pi_0}{\pi_1}}e^{\upsilon\frac{r_0+r_1}{2}} & 0 & 0 & \cdots \\ p_{10}\sqrt{\frac{\pi_1}{\pi_0}}e^{\upsilon\frac{r_0+r_1}{2}} & p_{11}e^{r_1\upsilon} & p_{12}\sqrt{\frac{\pi_1}{\pi_2}}e^{\upsilon\frac{r_1+r_2}{2}} & 0 & \\ 0 & p_{21}\sqrt{\frac{\pi_2}{\pi_1}}e^{\upsilon\frac{r_1+r_2}{2}} & p_{22}e^{r_2\upsilon} & p_{23}\sqrt{\frac{\pi_2}{\pi_3}}e^{\upsilon\frac{r_2+r_3}{2}} & \\ 0 & 0 & p_{32}\sqrt{\frac{\pi_3}{\pi_2}}e^{\upsilon\frac{r_2+r_3}{2}} & p_{33}e^{r_3\upsilon} & \\ \vdots & & \vdots & & \ddots \\ & & & & \end{bmatrix} \tag{4.4.7}$$

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

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

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

Unfortunately, for a number of regions above three, the resultant polynomial to be solved has not explicit solution and the eigenvalues of the matrix cannot be derived in a closed form. The usual number of regions employed in AMC schemes is between five and eight. Thus, we conclude that the result of the EBF of the FSMC has to be given in terms of the spectral radius as in (4.4.5).

Tan and Beaulieu [Tan 2000] showed the limitations of a first order FSMC in the modeling of the fading process. More specifically, they proposed a stochastic analysis of the problem, concluding that it is only appropriate for applications that require analysis over a short period of time but not for those requiring a large number of consecutive samples. The analysis was done, without loss of generality, for equal state probabilities in the chain. Nevertheless, similar results are obtained for other state partitions [Ruiz 2009]. By applying it to our effective bandwidth analysis, we can expect the FSMC model to be valid when the QoS requirements are relaxed, and not appropriate as the delay constraint becomes tighter, which would correspond to longer periods of observation.

4.5 Validation and numerical results

In this Section the evaluation of EBF of a Rayleigh channel is presented. The curve $\alpha_C(-\upsilon)$ is plot under different channel conditions.

The continuous rate policy is first evaluated assuming an exponential decay in the ACF (Figure 4.11). The EBF of the corresponding uncorrelated channel is also shown with solid line. The average SNR is set to 10dB and different values of the parameter ρ are evaluated. The analytical solution is compared with semi-analytical methods. The semi-analytical EBF is plot by means of the mean and variance measured over a long realization of the channel process. Notice that this semi-analytical technique can be useful for other purposes, when working with channels whose EBF is not known or cannot be analytically derived. In these cases, we just need to measure the variance of the channel process and to include it in equation (4.3.3).

The curve of the EBF exhibits the behaviour stated by the EBT: it starts in the ergodic capacity of the channel when $\upsilon = 0$, i.e. when no QoS requirements are imposed. Further, it decreases as υ increases, reaching zero at a certain point υ_{MAX}, above which the EBF of the channel is zero. This point indicates that stricter QoS requirements are not achievable by this channel. As expected, the value of υ_{max} decreases when the correlation increases.

In [Wu 2003b] the authors provide another solution of the effective bandwidth function assuming low average SNR and the instantaneous Shannon-like channel capacity (continuous rate policy with $\beta = 1$). The expression for the effective bandwidth function obtained there is:

$$\alpha_C(\upsilon) = \frac{\int \log(\upsilon S(f) + 1) df}{\upsilon} \tag{4.5.1}$$

where $S(f)$ is the Power Spectral Density of the Rayleigh channel, for Clarke's model:

$$S(f) = \frac{Pr}{2\pi f_D} \frac{1}{\sqrt{1 - (f/f_D)^2}} \qquad |f| \leq f_D \tag{4.5.2}$$

We have compared this solution with our continuous rate policy following a Bessel ACF and with the parameter β set to 1. The result is shown in Figure 4.12.

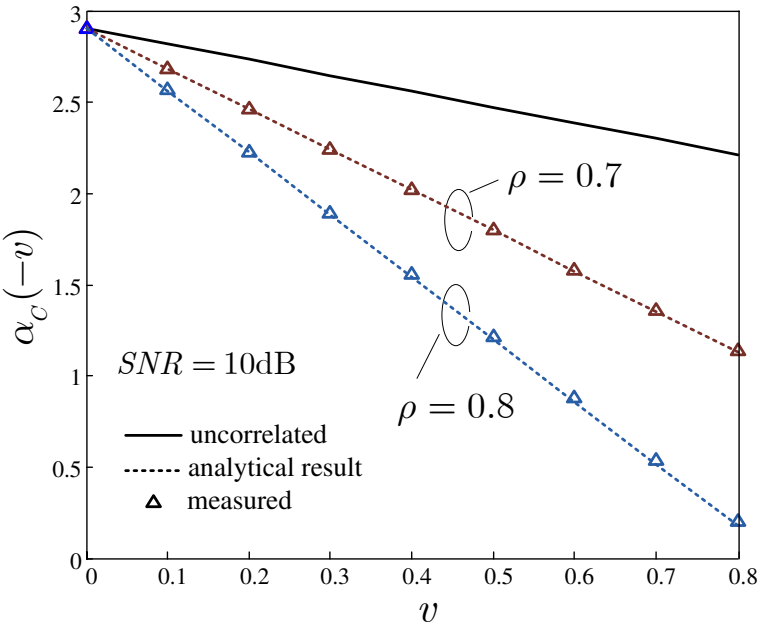


Figure 4.11: EBF for continuous rate policy and exponential ACF

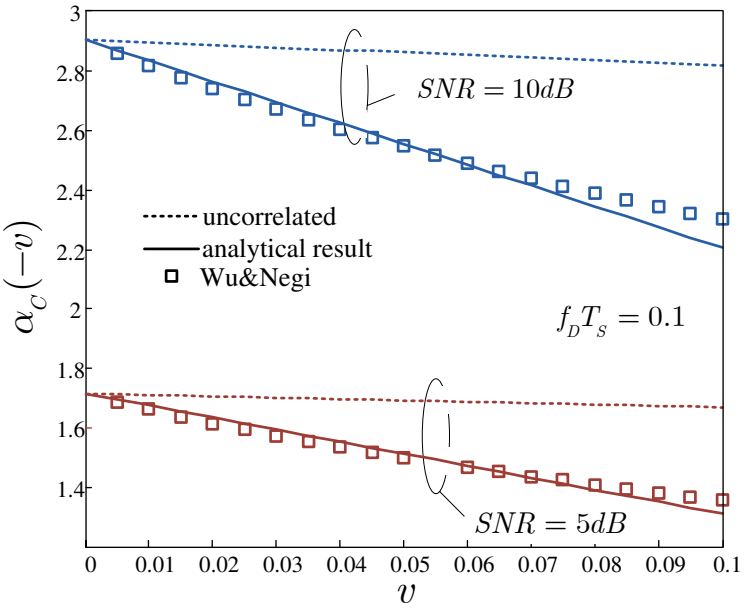


Figure 4.12: EBF for continuous rate policy and Bessel ACF

Two different values of the average SNR are plotted, 5dB and 10dB and the block length is 10000. The product $f_D \cdot T_S$ is set to 0.1. As expected, the curves diverge for higher values of the SNR (10dB) since the authors assumed low SNR regime.

Figures 4.13 and 4.14 show the EBF for discrete rate policy and Bessel autocorrelation function. In this case the analytical result is compared with the result provided by the FSMC model. In Figure 4.13, the product $f_D \cdot T_S$ is set to 0.01 and different average SNR are shown. In Figure 4.14, the average SNR is set to 5dB and different values of the product $f_D \cdot T_S$ are shown. As it can be observed, the FSMC overestimates the effective bandwidth of the channel, and that excess becomes more noticeable with stricter QoS requirements and stronger time-correlation.

4.6 Summary

The effective bandwidth function of the channel indicates the maximum arrival rate that a given service process can support by fulfilling certain QoS requirements. In this chapter, we have evaluated it for flat Rayleigh channels. First, the block fading channel (uncorrelated channel) is addressed. Then, the EBF is calculated for a generic time-correlated channel. To tackled this case, we propose the application of the Central Limit Theorem under the appropriate conditions. This assumption is shown to be valid by means of the Lilliefors test, which checks if an observed sample distribution is consistent with normality, and with the confrontation of the results to simulations.

At the end, the evaluation of the EBF comes down to the computation of two statistics, the mean and the variance of the channel rate. The expressions of both parameters are obtained for two rate policies: continuous and discrete. The results are evaluated under different conditions and compared to a widely employed model for the channel, the FSMC. The limitations of the FSMC for the purposes of this thesis are demonstrated in the numerical results, which show the overestimation of

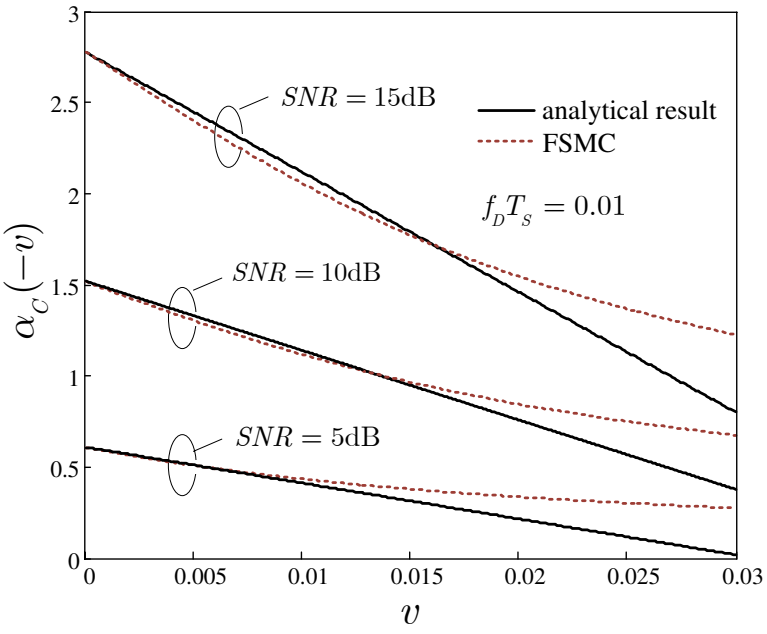


Figure 4.13: EBF for discrete rate policy. Influence of the SNR

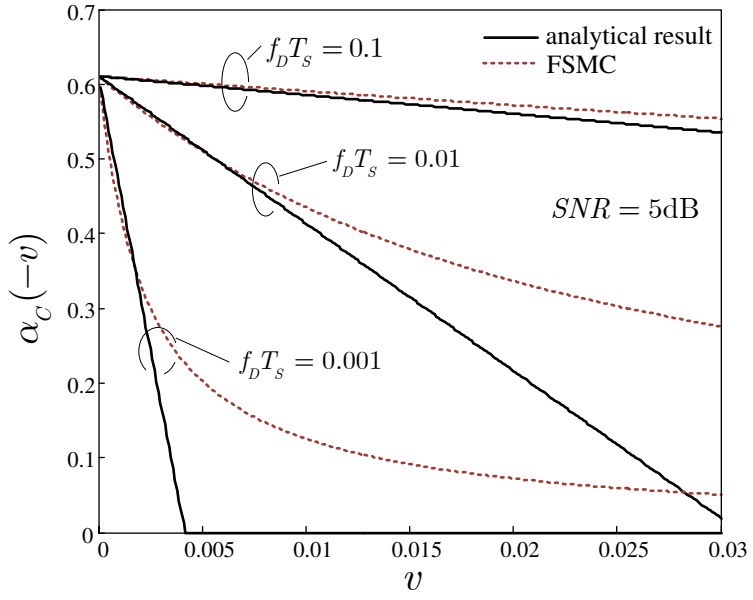


Figure 4.14: EBF for discrete rate policy. Influence of the autocorrelation

the model with regard to our analytical results. This overestimation is more noticeable as the QoS requirements become tighter. Besides, a semi-analytical method to calculate the EBF is proposed and confronted to our results. This semi-analytical technique can be also useful in those cases in which the EBF of a channel is not known.

Chapter 5

Delay constrained communications over flat Rayleigh channels

So far, the EBF of the outsourcing and the channel process have been obtained (in Chapters 3 and 4, respectively). In this chapter, we use both functions to analyze the QoS in a wireless system over flat Rayleigh channels. Particularly, emphasis is given on the delay suffered by the information flow. The P-percentile of the delay, the density function and the tradeoff between delay and source rate are analyzed by means of the effective bandwidth theory. Moreover, we propose a new concept, the Capacity with Probabilistic Delay Constraint $C_{D^t,\varepsilon}$, which expresses the maximum source rate a wireless channel can support while accomplishing a target Bit Error Rate (BER) and a probabilistic delay constraint.

Part of the results in this chapter have been published in [Soret 2007b], [Soret 2007a], [Soret 2009b] and [Soret 2010b].

The Chapter is organized as follows. In Section 5.1 we analyze the delay in flat Rayleigh channels. Next, in Section 5.2, we proceed to present the capacity $\mathrm{C}_{D^t,\varepsilon}$, first for constant rate sources and later its extension to variable rate sources. The results are validated by confrontation with simulations as explained in Section 5.3. Finally, some concluding remarks are done in Section 5.4.

5.1 Analysis of the delay in flat Rayleigh channels

5.1.1 The QoS exponent

Figure 5.1 shows the single-user single-channel system model from Chapter 2. The source process characterizes the incoming user traffic and the server represents the information transmitted to the wireless channel.

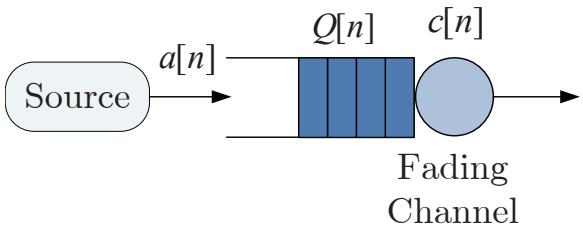


Figure 5.1: Queueing system.

As it was explained in Chapter 2, if the source and the channel processes are stationary and the steady state queue length exists, then the probability of exceeding a target delay D^t can be written:

$$\varepsilon = Pr\{D(\infty) > D^t\} \asymp e^{-\theta\alpha_A(\theta)D^t} \qquad D^t \to \infty \tag{5.1.1}$$

Including the probability that the queue is not empty, η, the expression yields:

$$\varepsilon = Pr\{D(\infty) > D^t\} \approx \eta \cdot e^{-\theta\cdot\alpha_A(\theta)D^t} = \eta \cdot e^{-\theta\cdot\alpha_C(-\theta)D^t} \tag{5.1.2}$$

where the QoS exponent θ is the solution to:

$$\alpha_A(\theta) - \alpha_C(-\theta) = 0 \tag{5.1.3}$$

Graphically, θ is the intersection of the curves of the effective bandwidth of the source and the channel process.

Let us consider constant rate traffic at the rate λ, with constant EBF:

$$\alpha_A(\upsilon) = \lambda \tag{5.1.4}$$

Figure 5.2 illustrates the intersection of the two curves in this case, when the EBF of the source is a straight line.

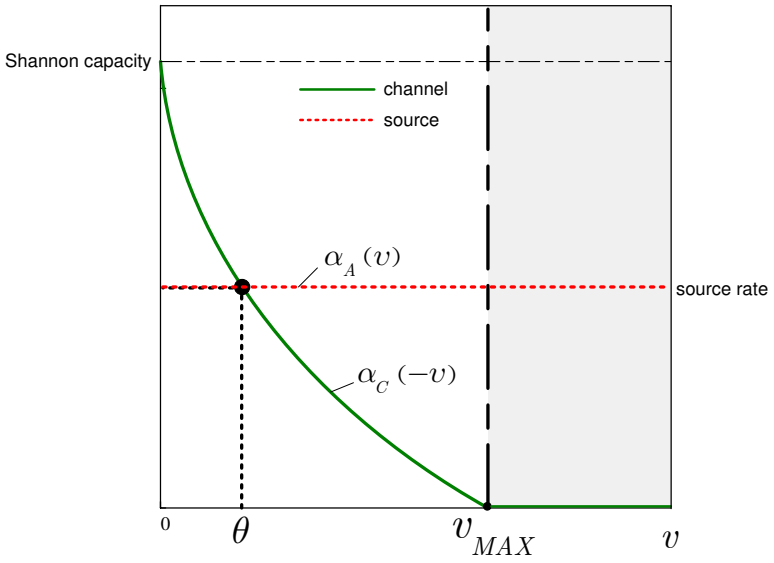


Figure 5.2: Intersection of the EBF curves for CBR.

Uncorrelated channel

With a fixed target BER, the parameters of the uncorrelated channel, m_c and σ_c^2, depend uniquely on the average SNR, $\bar{\gamma}$. Figure 5.3 shows the QoS exponent as a function of the source rate λ and two different examples of uncorrelated channel: $\bar{\gamma} = 10dB$ and $\bar{\gamma} = 20dB$. β is set to 1 for the continuous rate policy (i.e. the upper bound is considered) and $BER^t = 10^{-2}$ for the discrete case. It is observed that as the traffic rate increases the QoS exponent that may be set has to be necessarily smaller, i.e., less stringent QoS requirements can be demanded. Naturally, a channel

with a higher value of average SNR will support lower values of θ for the same source rate. Moreover, the continuous rate policy obtains better results with regard to the discrete rate policy, which sets a constraint on the target BER. On the other hand, there is a limit in the QoS metrics that a given channel can offer, corresponding to the region delimited by the maximum value of υ, υ_{MAX}. In the Figure, υ_{MAX} is the value $\theta(\lambda = 0)$. For example, in the case of continuous rate with $\bar{\gamma} = 10dB$, $\upsilon_{MAX} = 3.5$.

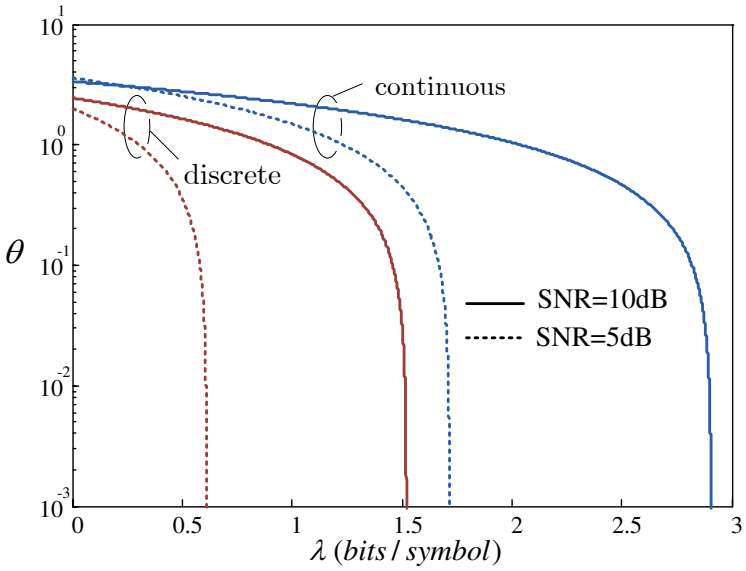


Figure 5.3: QoS exponent vs. source rate for an uncorrelated channel. Influence of the average SNR.

Time-correlated channel

When the response of the channel is time-correlated, the ACF fixes the shape of the correlation among samples. Figures 5.4 and 5.5 illustrate this case with an ACF that decays exponentially with a parameter ρ.

In Figure 5.4 ρ is fixed to 0.90 and two values of $\bar{\gamma}$ are plot: 5 and 10 dB. Figure 5.5 fixes the value of $\bar{\gamma}$ to 10dB and shows the influence of the correlation, with

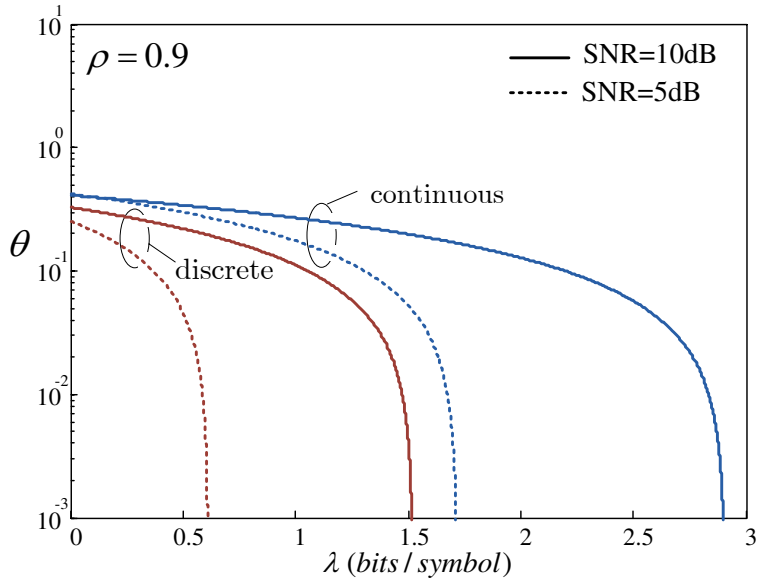


Figure 5.4: QoS exponent vs. source rate for a time-correlated channel. Exponential ACF. Influence of the average SNR.

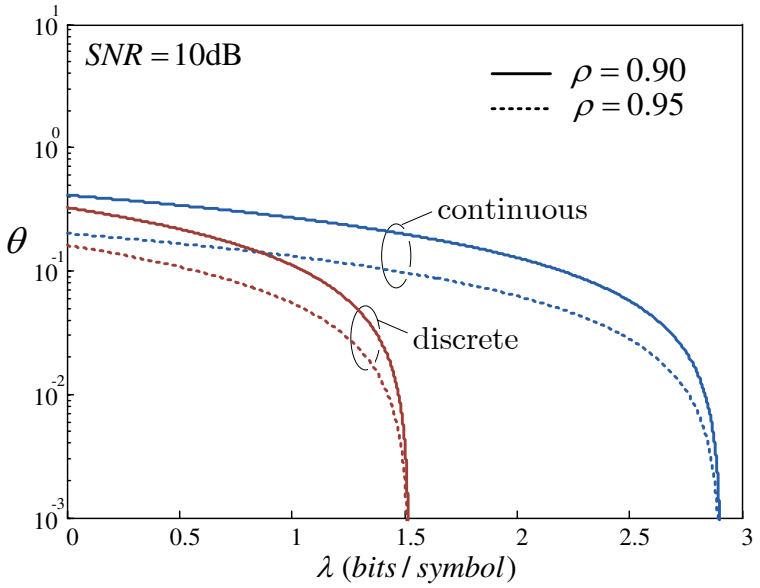


Figure 5.5: QoS exponent vs. source rate for a time-correlated channel. Exponential ACF. Influence of the correlation.

two different values of ρ. Thus, it can be observed that, for a given source rate, a higher value of correlation in the channel leads to lower values of θ, i.e., the QoS requirements have to be relaxed. This is our first observation of the influence of the correlation in the delay: time-correlation in the channel process is harmful to the delay behaviour.

5.1.2 P-percentile of the delay

The QoS requirement for delay is often defined in terms of its percentile, i.e., the value of the delay below which a certain percent of observations fall. Thus, the 90^{th}-percentile is the value below which 90 percent of the observations may be found. Together with the average delay it is a very common metric of QoS for delay-sensitive applications. Typical values of percentiles are 99^{th} and 95^{th}.

Probability of non-empty queue

First of all, the parameter η in equation (5.1.2) is discussed. In a GI/GI/1 queue, which refers to a queue with i.i.d. arrival and service processes, the probability that the buffer is not empty is approximated by the ratio of the mean arrival rate to the mean channel rate [Kleinrock 1975]:

$$\eta \approx \frac{\lambda}{m_c} \tag{5.1.5}$$

Particularly, when the source rate is approximately half of the mean of the instantaneous transmission rate, η approaches 0.5:

$$\lambda \approx \frac{m_c}{2} \Rightarrow \eta \approx 0.5 \tag{5.1.6}$$

This approximation of η is not suitable to our correlated system. It has been checked via simulation that the approximation is quite accurate in the case of CBR and uncorrelated Rayleigh channel, which represents the worst case. Any other

variable source or correlated channel will lead to higher values of η. An example is plot in Figure 5.6. A Rayleigh channel with average SNR of 10dB has been simulated. The rate policy is discrete and the target BER is 10^{-3}, leading to a mean channel rate $m_c = 1.01$. The uncorrelated Rayleigh channel and two different Bessel channels ($f_D T_S = 0.10$ and $f_D T_S = 0.01$) are plot. Moreover, the approximation in (5.1.5) is shown with solid line. The arrival process is constant at a rate varying from $m_c/2$ to m_c. As expected, the uncorrelated channel is well approximated with equation (5.1.5). As the correlation increases, the probability of non-empty queue increases. Similar results are obtained with continuous rate policy.

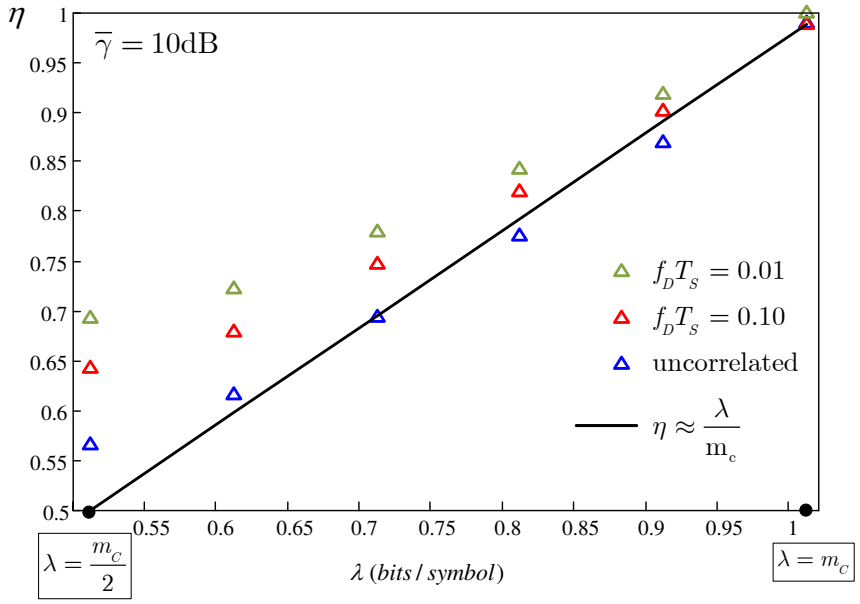


Figure 5.6: Probability of non-empty queue vs. source rate. Discrete rate policy with average SNR = 10dB and $BER^t = 10^{-3}$. Uncorrelated channel and Bessel ACF.

Constant Rate Traffic

Let us calculate the percentile of the delay for a constant rate traffic at the rate λ.

The percentile of the delay is directly obtained from equation (5.1.2). Typical

values of the percentile $(95^{th}, 99^{th})$ lead to small values of the violation probability ε, corresponding to the tail of the distribution where the term η is not needed (equation (5.1.1)). It means that we work in a high load scenario and the probability that the buffer is not empty approaches one, i.e. $\eta \to 1$.

With constant rate traffic, the percentile of the delay yields:

$$1 - \varepsilon = 1 - Pr\{D(\infty) > D^t\} = 1 - e^{-\theta \cdot \lambda D^t} \tag{5.1.7}$$

Let us work out the QoS exponent. Replacing the EBF of a flat Rayleigh channel, θ is given by:

$$\lambda - \alpha_C(-\theta) = 0 \Rightarrow \theta(\lambda) \triangleq \theta(m_c, \sigma_c^2, \lambda) = \frac{2(m_c - \lambda)}{\sigma_c^2} \tag{5.1.8}$$

If the channel is uncorrelated, the mean m_c and the variance σ_c^2 are those in equations (4.2.5) and (4.2.8) (continuous rate policy), and (4.2.10) and (4.2.11) (discrete rate policy). If the Rayleigh channel is time-correlated, we work with the mean and the variance of the blocks normalized with k, m_k/k and σ_k^2/k, obtained with (4.3.4) and (4.3.7).

With θ replaced in (5.1.7), the percentile of the delay is obtained as:

$$1 - \varepsilon = 1 - Pr\{D(\infty) > D^t\} = 1 - e^{-\frac{2(m_c - \lambda)}{\sigma_c^2} \cdot \lambda D^t} \tag{5.1.9}$$

In Figure 5.7, the percentile of the delay is plot as a function of the source rate λ, for a given target delay. A continuous rate policy with exponential ACF is considered, with $\rho = 0.8$ and $\rho = 0.9$. The uncorrelated channel is also plot. The target BER is 10^{-2} and D^t is fixed to 20 symbols. The channel has average SNR 5, 10 and 15dB. It can be checked that the percentile of the delay approaches 100% for small values of the source rate, that is, all bits will suffer a delay lower than the target delay. When the source rate increases, the percentile decreases until 0%. It is worth highlighting the drop down to zero in the case of an uncorrelated

channel, so that very small changes in the source rate will lead to drastic changes in the percentile of the delay. It is the expected behaviour with no variability in the source and no correlation in the channel. When the time correlated of the channel is considered, the curve of the percentile decays more slowly.

In Figure 5.8, the result for a discrete rate policy is shown. The ACF follows a Bessel function with $f_D T_S = 0.10$ and $f_D T_S = 0.01$. The average SNR of the channel is set to 5, 10 and 15dB. The target BER is 10^{-2} and D^t is fixed to 150 symbols. The effect of the correlation in the decay of the curve is more noticeable than in the exponential ACF.

In Figures 5.9 and 5.10 the influence of the target delay in the uncorrelated channel is illustrated, for a continuous and a discrete rate policy, respectively. Three different values are fixed ($D^t = 3$, 7 and 11 symbols). The average SNR is 20dB. $\beta = 1$ in the continuous rate policy and $BER^t = 10^{-2}$ in the discrete rate policy. A zoom of the x-axis (source rate) has been done in order to have a better sight of the decay of the percentile. As expected, the curves decay earlier for smaller values of the target delay (tighter QoS requirements).

Variable rate traffic

As example of variable bit rate, the autoregressive traffic described in Section 3.2.3 modeling streaming services is considered.

A video source is approximated by a continuous fluid flow model that assumes that the output bit rate within a symbol period is constant and changes from symbol to symbol according to the following AR(1) model [Maglaris 1998]:

$$a[n] = \rho_A \cdot a[n-1] + q \cdot w[n] \tag{5.1.10}$$

$w = \mathcal{N}(m_w, 1)$ is Gaussian white noise of mean m_w and variance 1.

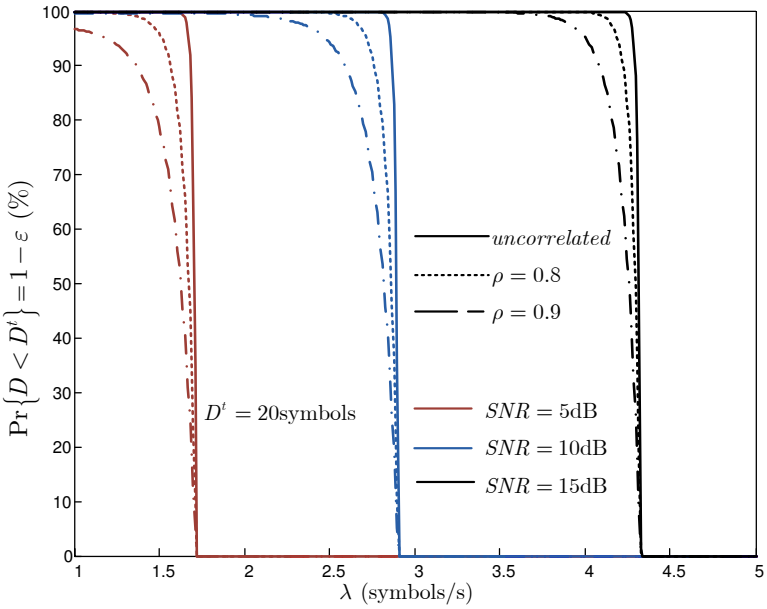


Figure 5.7: Percentile of the delay for constant rate traffic and continuous rate policy. Exponential ACF. $D^t = 20$ symbols. $BER^t = 10^{-2}$.

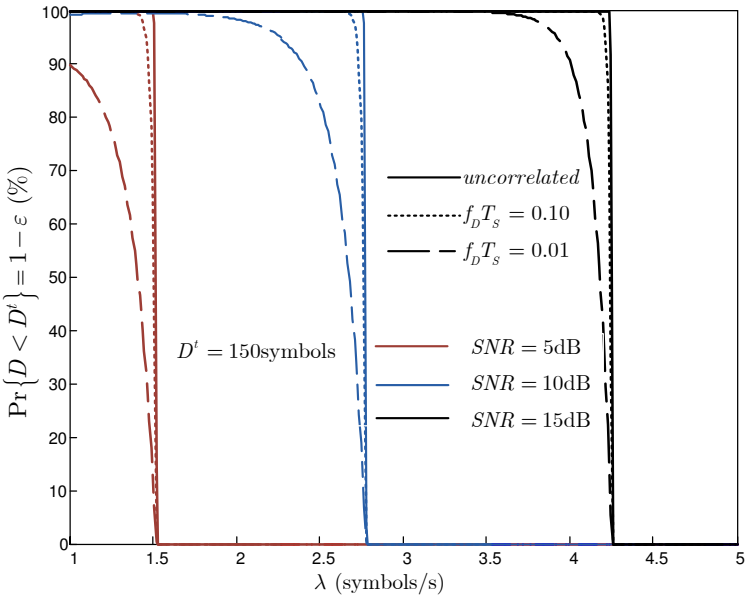


Figure 5.8: Percentile of the delay for constant rate traffic and discrete rate policy. Bessel ACF. $D^t = 150$ symbols. $BER^t = 10^{-2}$.

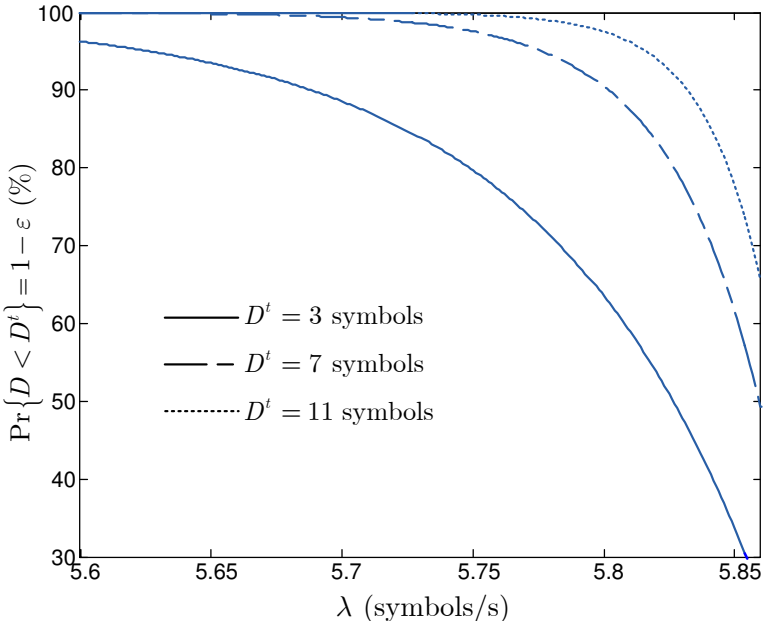


Figure 5.9: Percentile of the delay for constant rate traffic. Influence of the target delay. Uncorrelated Rayleigh channel and continuous rate policy. Average SNR= 20dB. $\beta = 1$.

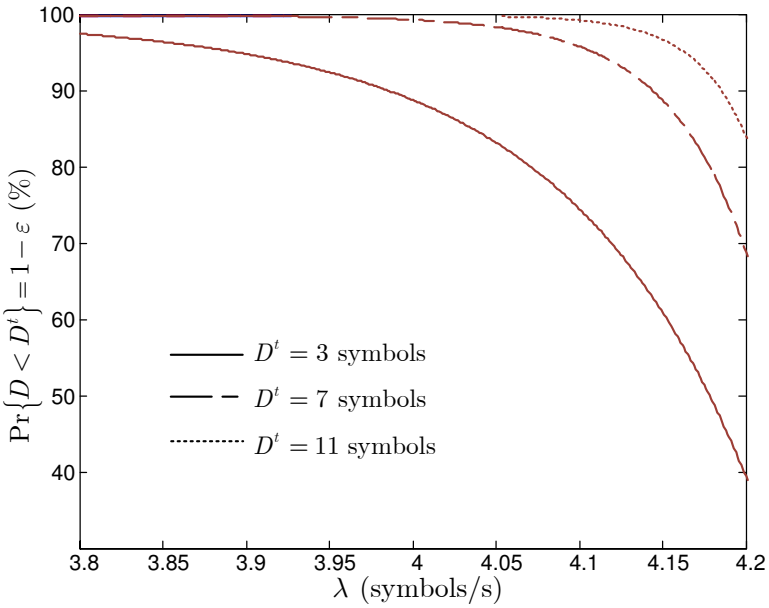


Figure 5.10: Percentile of the delay for constant rate traffic. Influence of the target delay. Uncorrelated Rayleigh channel and discrete rate policy. Average SNR= 20dB. $BER^t = 10^{-2}$.

The effective bandwidth of this AR(1) process was [Courcoubetis 1994]:

$$\alpha_A(u) = r \cdot m_w + \frac{r^2}{2} \cdot u \tag{5.1.11}$$

where $r = \frac{q}{1-\rho_A}$.

The mean of $a[n]$ is denoted by m_A and is obtained as:

$$m_A = r \cdot m_w \tag{5.1.12}$$

The parameters of the model were detailed in Table 3.1.

To work out θ, the approximation $\eta \to 1$ is made. The value of the QoS exponent for the uncorrelated channel is derived by replacing (5.1.3) with (5.1.11) and (4.3.3):

$$\theta(r, m_w) = \frac{2(m_c - r \cdot m_w)}{r^2 + \sigma_c^2} = \frac{2(m_c - m_A)}{r^2 + \sigma_c^2} \tag{5.1.13}$$

The percentile of the delay yields:

$$\begin{aligned} 1 - \varepsilon &= 1 - Pr\{D(\infty) > D^t\} \\ &= 1 - \exp\left[-\frac{2(m_c - m_A)}{r^2 + \sigma_c^2}\left(m_A + \frac{r^2}{2} \cdot \frac{2(m_c - m_A)}{r^2 + \sigma_c^2}\right) \cdot D^t\right] \end{aligned} \tag{5.1.14}$$

The percentile is plot in Figure 5.11. In the source, the correlation parameter ρ_A varies from 0.95 to 0.99, which leads to mean source rates m_A from 1.05 to 6.33 (abscissa). A continuous rate policy is considered, with $\beta = 1$. The target delay is fixed to 7 symbols. The channel is uncorrelated, with average SNR 10, 15 and 20dB. As expected, with a variable source the decay of the percentile is slower than in the case of a constant rate source.

5.1.3 PDF and pdf of the delay

From the tail probability in (5.1.1), it is observed that it provides an expression of the tail of the probability distribution of the delay (PDF) as $1 - \epsilon$. In the case of CBR:

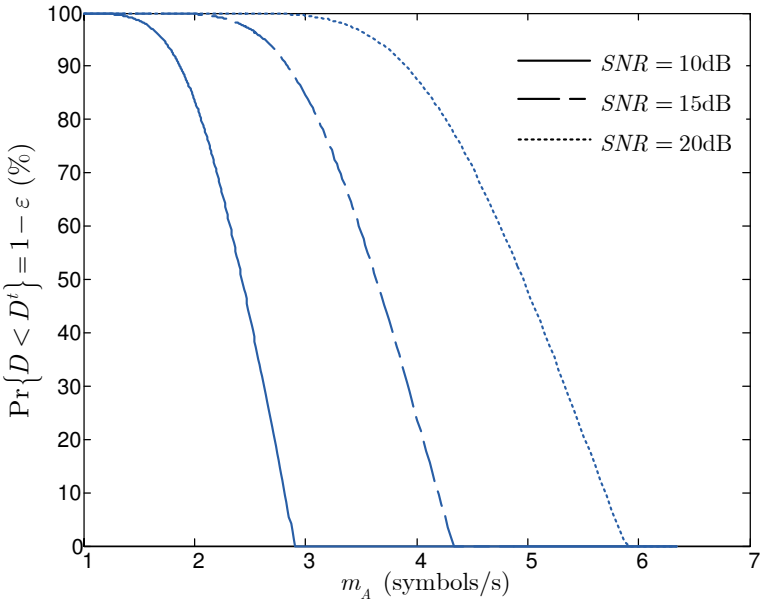


Figure 5.11: Percentile of the delay for AR traffic and uncorrelated Rayleigh channel. Influence of the SNR. $D^t = 7$ symbols. Continuous rate policy. $\beta = 1$.

$$F(D^t) = Pr\{D(\infty) > D^t\} \approx 1 - e^{-\theta \cdot \lambda D^t} \qquad D^t \to \infty \tag{5.1.15}$$

Moreover, the tail of the probability density function (pdf) is obtained by taking the derivative of the PDF, which yields:

$$f(D^t) = \frac{d(F(D^t))}{dD^t} \approx \theta \cdot \lambda \cdot e^{-\theta \lambda D^t} \qquad D^t \to \infty \tag{5.1.16}$$

The theoretical tail of the pdf has been checked by comparison with simulations, for constant rate traffic. Figures 5.12 and 5.13 plot the results. The average SNR has been set to $\bar{\gamma} = 15$dB and the violation probability is $\varepsilon = 0.1$. In the first one, Figure 5.12, the channel is uncorrelated. In Figure 5.13 the channel is time-correlated following a Bessel function with $f_D \cdot T_S = 10^{-3}$. It can be observed that the pdf tail approximates precisely the measured histograms for large values of the target delay in both cases. In fact, when the channel is uncorrelated the approximation is even exact for small values of D^t.

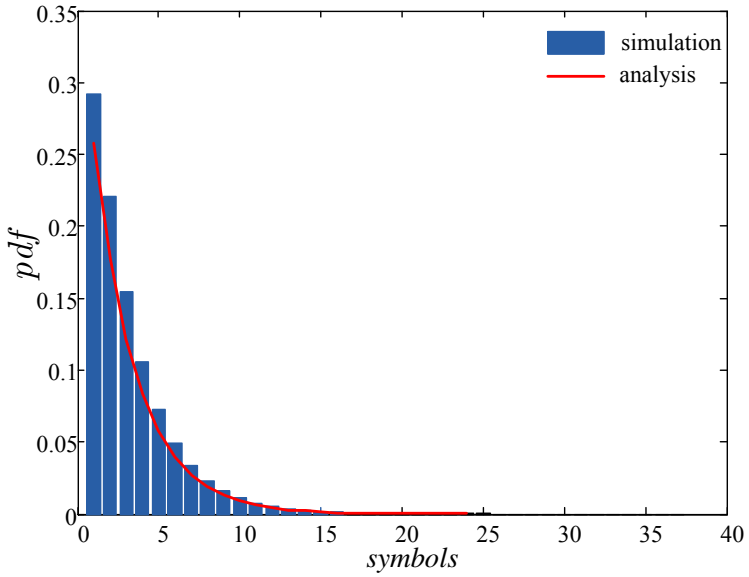


Figure 5.12: Tail of the pdf of the delay for an uncorrelated channel.

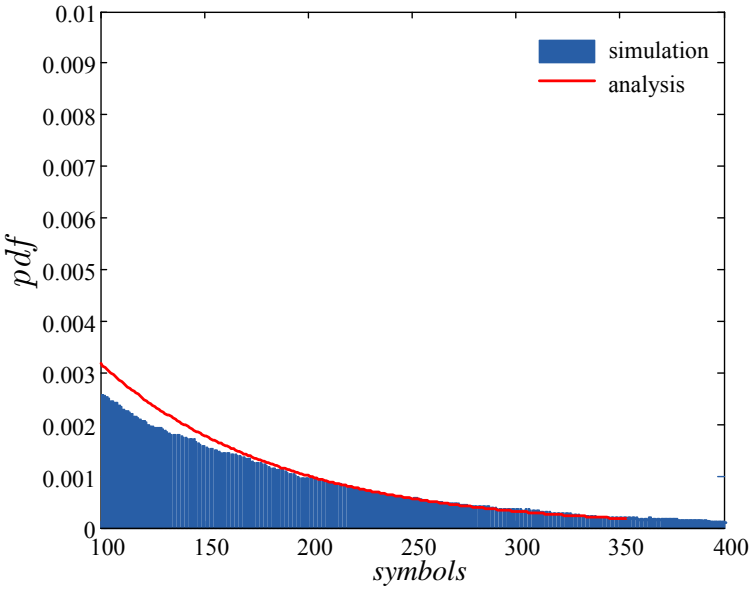


Figure 5.13: Tail of the pdf of the delay for a time-correlated channel.

5.1.4 Delay QoS factor

In a Rayleigh channel an attempt to provide deterministic QoS will most likely result in extremely conservative guarantees. Thus, a requirement of delay will be always given in terms of a delay constraint (D^t, ε), where D^t is the target delay and ε is the probability of exceeding D^t.

Under these circumstances, let us define $\frac{-log(\varepsilon)}{D^t}$ as the Delay QoS Factor (DQF). A low value of the DQF (DQF approaching zero) means a poor level of QoS, i.e., either the target delay tends to infinity or the delay is not guaranteed (Shannon conditions). On the other hand, the DQF approaching infinity means that either the target delay is zero or the delay is deterministically guaranteed (unattainable for Rayleigh channels).

If the traffic has constant rate and $\eta \to 1$, then the delay QoS factor gives:

$$\text{DQF} = \frac{-\log \varepsilon}{D^t} = \frac{2 \cdot (m_c - \lambda)}{\sigma_c^2} \lambda \tag{5.1.17}$$

The evaluation of (5.1.17) is shown in Figure 5.14, for an uncorrelated channel whose average SNR is 10dB and under a continuous rate policy with $\beta = 1$. The ratio $\frac{-\log(\varepsilon)}{D^t}$ is represented as a function of the average SNR $\bar{\gamma}$ varying from 5dB to 20dB, and the source rate λ varying between $\frac{mc}{2}$ and m_c. Obviously, a high value of the DQF means that stringent QoS requirements are supported. Thus, the DQF increases when λ decreases or $\bar{\gamma}$ increases.

Consider now AR traffic as defined in Section 5.1.2. Substituting (5.1.13) into (5.1.2), the following expression for the delay QoS factor is obtained:

$$\text{DQF} = \frac{-\log \varepsilon}{D^t} = \frac{2 \cdot r(m_c - r \cdot m_w)}{r^2 + \sigma_c^2} \cdot \left(m_w + \frac{r(m_c - r \cdot m_w)}{r^2 + \sigma_c^2} \right) \tag{5.1.18}$$

The evaluation of (5.1.18) is shown in Figure 5.15, for a channel whose average SNR is 5dB, under a continuous rate policy and exponential fall for the autocorrelation with $\rho = 0.95$, which leads to a mean $m_c = 1.72$ and $\sigma_c^2 = 18.12$. The DQF

is represented as a function of the parameters r and m_w, varying between 0.5 and 1.5, and 0 and 1, respectively. One can see in the Figure that sources with the same mean, $r \cdot m_w$, do not necessarily correspond to the same QoS guarantees expressed through the ratio $-\log(\varepsilon)/D^t$.

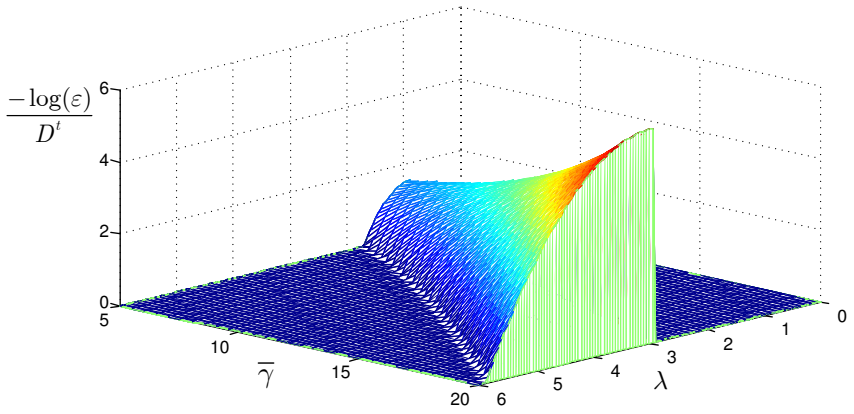


Figure 5.14: Delay QoS factor for CBR.

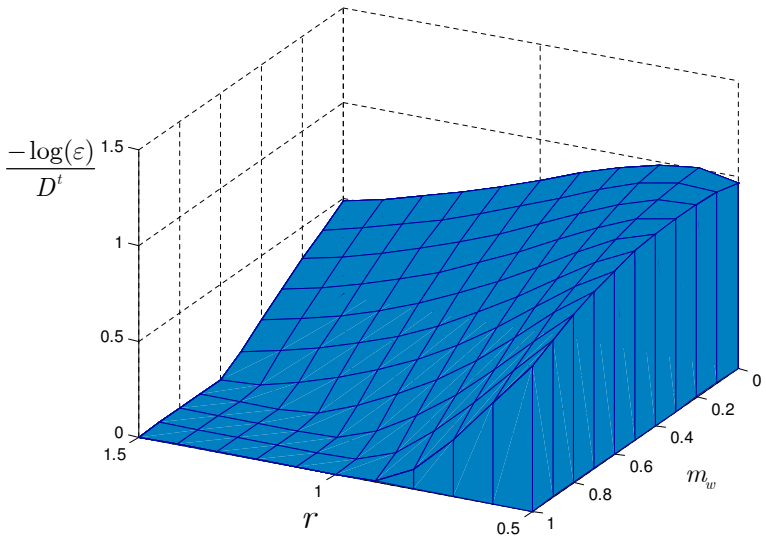


Figure 5.15: Delay QoS factor for AR traffic.

5.1.5 Delay vs. source rate

Let us assume a source transmitting at a rate λ under a delay constraint (D^t, ε). From (5.1.2), the target delay yields:

$$D^t = -\frac{\log(\varepsilon) \cdot \sigma_c^2}{2m_c\lambda - 2\lambda^2} \tag{5.1.19}$$

Now, imagine a source transmitting above or below λ, at a rate $\lambda' = P\lambda$, where P is a proportionality constant. It corresponds to moving to the left or to the right, depending on P, in the curve of the EBF of the channel, obtaining a smaller or a higher value of the parameter v, respectively. This leads to a loss or an improvement in the QoS guarantees. The case of moving to the left is illustrated in Figure 5.16.

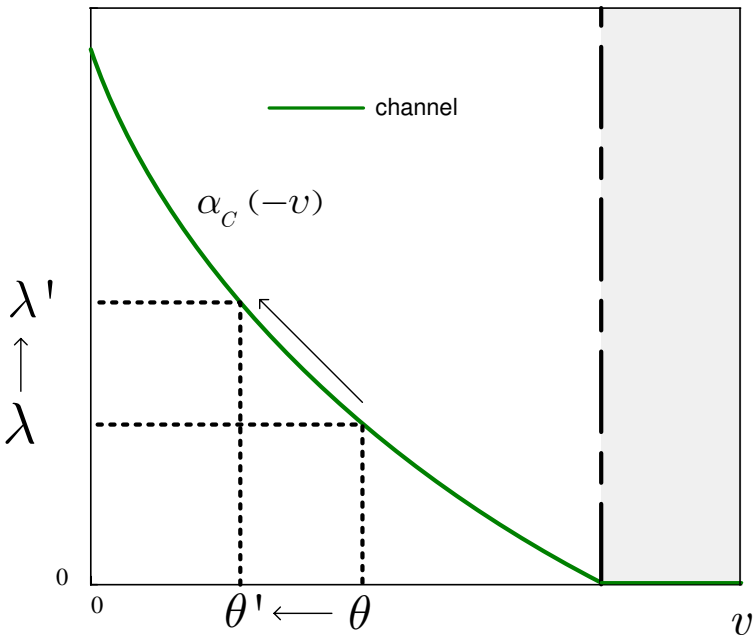


Figure 5.16: Effect of increasing the source rate.

We want to quantify the QoS metrics in the new conditions. If the probability of violation is to be preserved, then the new delay is:

$$D = -\frac{\log(\varepsilon) \cdot \sigma_c^2}{2m_cP\lambda - 2(P\lambda)^2} \tag{5.1.20}$$

Observe that the denominator changes from positive values to negative values exactly at the rate m_c. Obviously, we do not want to transmit at a higher rate than the ergodic capacity of the channel. In terms of the target delay D^t, equation (5.1.20) can be expressed:

$$D = D^t \frac{m_c - \lambda}{P(m_c - 2P\lambda)} \tag{5.1.21}$$

Figure 5.17 shows an example for uncorrelated channel. The average SNR $\bar{\gamma}$ is 20dB and the violation probability ε is set to 0.1. The mean of the channel rate is $m_c = 5.88$. The figure shows the delay as a function of the source rate, for values of λ between $\frac{m_c}{2}$ and m_c. As expected, the delay increases asymptotically when approaching m_c. It is also observed that very small reductions in the source rate make the delay to approach 0 quickly, owing to the uncorrelated channel.

If we turn to a correlated channel, we can see larger differences in the delay as the source rate changes. This is illustrated in Figure 5.18. The rate policy is discrete. The ACF follows a Bessel function with the product $f_D T_S = 0.01$. The average SNR is 10dB and $\varepsilon = 0.1$. The mean of the channel is $m_c = 1.52$. As expected, the delay increases with the source rate, starting with less than 200 symbols for a source rate of 0.9 bits per symbol, and increasing up to 600 symbols for source rates of 1.4 bits per symbol.

5.2 Capacity with Probabilistic Delay Constraint $\mathrm{C}_{D^t,\varepsilon}$

5.2.1 Constant rate traffic

In view of Figure 5.2 and the graphics of the QoS exponent as a function of the source rate, we propose the definition of a new capacity in a wireless system, the **Capacity with Probabilistic Delay Constraint** $\mathrm{C}_{D^t,\varepsilon}$. Its derivation and meaning is explained next.

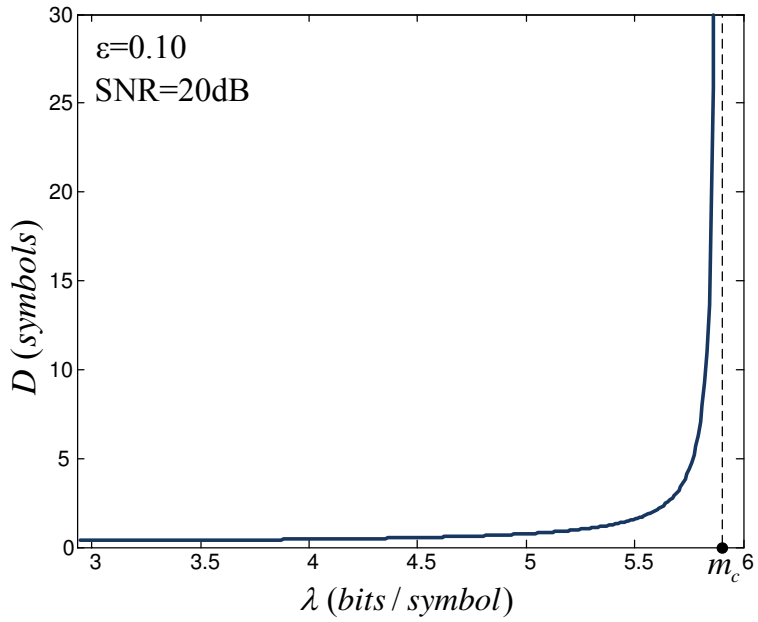


Figure 5.17: Delay vs. transmission rate for uncorrelated channel with continuous rate policy.

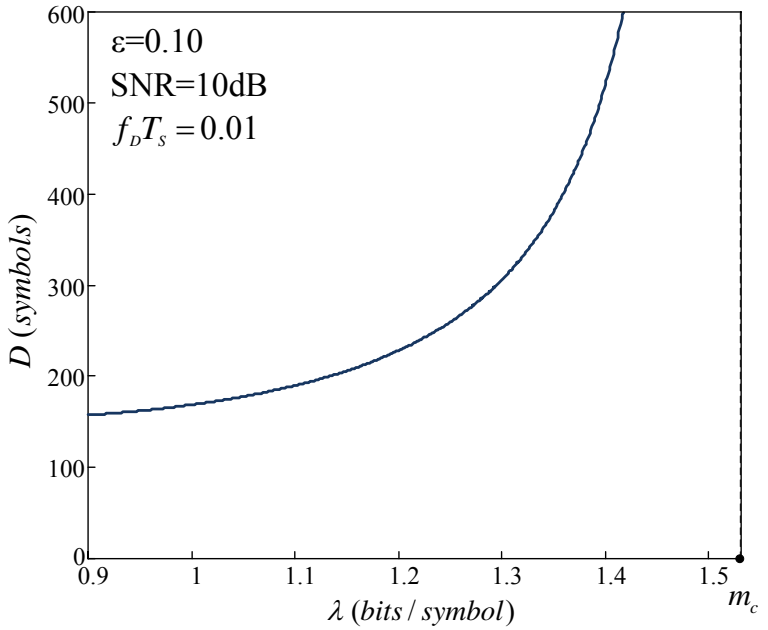


Figure 5.18: Delay vs. transmission rate for time-correlated channel with discrete rate policy.

The capacity of a channel dictates the maximum data rate that can be transmitted over it. In wireless communications, several capacities have been defined in the literature under a variety of constraints and models. The well known ergodic capacity [Biglieri 1998] is the maximum error-free digital communication on a wireless noisy channel in Shannon's sense, i.e. with no restrictions to its complexity and delay. On the other hand, delay-limited capacity is associated to the maximum *fixed-rate* that can be deterministically guaranteed, which implicitly ensures delay.

Delay-limited capacity is zero for Rayleigh fading channels. In such cases, we propose the definition of its probabilistic version, the Capacity with Probabilistic Delay Constraint $C_{D^t,\varepsilon}$. $C_{D^t,\varepsilon}$ is defined as the maximum source rate supported by the channel under a target BER and keeping delay under the bound D^t except for a given probability ε. Reducing ε ensures D^t for a higher percentage of time and the specific case with assurance ($\varepsilon = 0$) represents delay-limited capacity.

To derive $C_{D^t,\varepsilon}$, the delay constraint is obtained from (5.1.2), under high load conditions:

$$-\frac{\log(\varepsilon)}{D^t} = \theta \cdot \alpha_A(\theta) \tag{5.2.1}$$

With the QoS exponent (5.1.8) substituted into (5.2.1), the value of λ that matches the equation is the **Capacity with Probabilistic Delay Constraint** $\mathrm{C}_{D^t,\varepsilon}$ and is written in very simple terms as:

$$\mathrm{C}_{D^t,\varepsilon} = \frac{m_c}{2} + \frac{1}{2}\sqrt{m_c^2 - 2\sigma_c^2\frac{(-\log\varepsilon)}{D^t}} = \frac{m_c}{2} + \frac{1}{2}\sqrt{m_c^2 - 2\sigma_c^2 \cdot \mathrm{DQF}} \tag{5.2.2}$$

Further analysis of equation (5.2.2) reveals interesting points. Two foreseen limits can be checked. As the DQF decreases (high D^t values or $\varepsilon \to 1$), the QoS requirement relaxes and $\mathrm{C}_{D^t,\varepsilon}$ approaches m_c, which is the ergodic capacity of the channel for the corresponding rate policy and under Shannon conditions. On the other hand, as the DQF increases (the target delay D^t or ε become lower), the wireless channel tolerates lower traffic arrival rates in order to guarantee the delay

constraints. Obviously, $\mathrm{C}_{D^t,\varepsilon}$ always stays below the ergodic capacity. The influence of the target BER is captured in the mean and the variance (through the parameter β in the continuous rate policy and the thresholds in the discrete case). The lower the tolerated target BER, the lower the capacity. Moreover, the variance also captures the influence of the correlation among samples when the channel is time-correlated. Thus, as correlation increases (σ_c^2 increases and $f_D \cdot T_S$ decreases), the second addend in the square root increases in absolute value and $\mathrm{C}_{D^t,\varepsilon}$ diminishes. Naturally, source rates below $\mathrm{C}_{D^t,\varepsilon}$ will carry out with the delay QoS requirements.

For certain values of ε and D^t, the evaluation of (5.2.2) results in a complex number (when the expression under the square root becomes negative). It happens for values of the delay QoS factor above $\mathrm{DQF}_{max} = \frac{1}{2}\left(\frac{m_c}{\sigma_c}\right)^2$. For this limiting value of DQF, the capacity is $\mathrm{C}_{D^t,\varepsilon} = \frac{m_c}{2}$. Values of DQF above DQF_{max} are not part of the domain of the function $\mathrm{C}_{D^t,\varepsilon}$ and the capacity is plot as zero in the figures.

The DQF could be redefined to include the probability of non-empty queue:

$$\mathrm{DQF} = \frac{-\log\frac{\varepsilon}{\eta}}{D^t} \tag{5.2.3}$$

In this case, a tighter expression of the capacity should be given:

$$\mathrm{C}_{D^t,\varepsilon} = \frac{m_c}{2} + \frac{1}{2}\sqrt{m_c^2 - 2\sigma_c^2\frac{(-\log\frac{\varepsilon}{\eta})}{D^t}} \tag{5.2.4}$$

Let us observe the tradeoff between the delay constraint and the capacity, illustrated in Figure 5.19.

In Figure 5.19 (a), the capacity normalized by the mean m_c is plot as a function of the target delay. There is a minimum value of D^t denoted as D^t_{min} in the figure, corresponding to the target delay that makes the expression under the square root negative (given a violation probability and some channel conditions). Thus, D^t_{min} expresses the tighter value of target delay that can be supported by the channel

with a certain ε. The influence of the violation probability and the variance of the channel is also indicated. If ε decreases (approaches zero), D^t_{min} moves to the right and the capacity decreases. On the other hand, if the correlation of the channel decreases or the target BER increases (σ_c^2 decreases), the capacity increases and the curve moves to the left.

In Figure 5.19 (b) $C_{D^t,\varepsilon}/m_c$ is plot as a function of the probability of exceeding the target delay. In this case, the value ε_{min} is represented in the figure, expressing the minimum value of violation probability supported by the channel for a given target delay. If the QoS requirements are tight (small value of the target delay), ε_{min} moves to the right and the normalized capacity decreases. Finally, when the correlation of the channel diminishes or the target BER increases, the curve of capacity moves to the left.

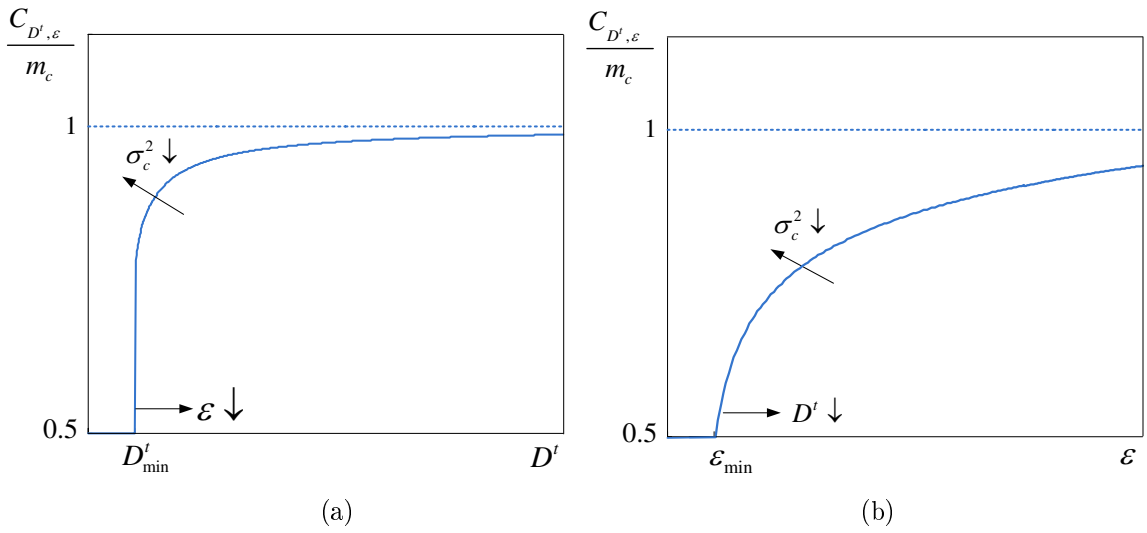


Figure 5.19: Normalized Capacity with Probabilistic Delay Constraint. (a) Capacity vs. target delay. (b) Capacity vs. violation probability

Evaluation of $C_{D^t,\varepsilon}$

The evaluation of $C_{D^t,\varepsilon}$ under different channel conditions is done next. The capacity is plot as a function of the target delay. Values of capacity below $m_c/2$ are not part of the domain of the function and are plot as zero in the figures.

Figure 5.20 presents the evaluation of $C_{D^t,\varepsilon}$ for an uncorrelated channel and with both rate mechanisms, continuous and discrete. Two values of $\overline{\gamma}$ are shown, 10dB and 20dB, and two violation probabilities, ε=0.1 (percentile 90%) and ε=0.05 (percentile 95%). $C_{D^t,\varepsilon}$ is drawn as a function of D^t for the different values of $\overline{\gamma}$ and ε. The target BER is set to $BER^t = 10^{-3}$. As presumed, for the same $\overline{\gamma}$ value, the discrete rate capacity is always below the continuous rate case. The behavior of both schemes (continuous and discrete) is similar, but the reduction in the maximum rate for the discrete case is more noticeable.

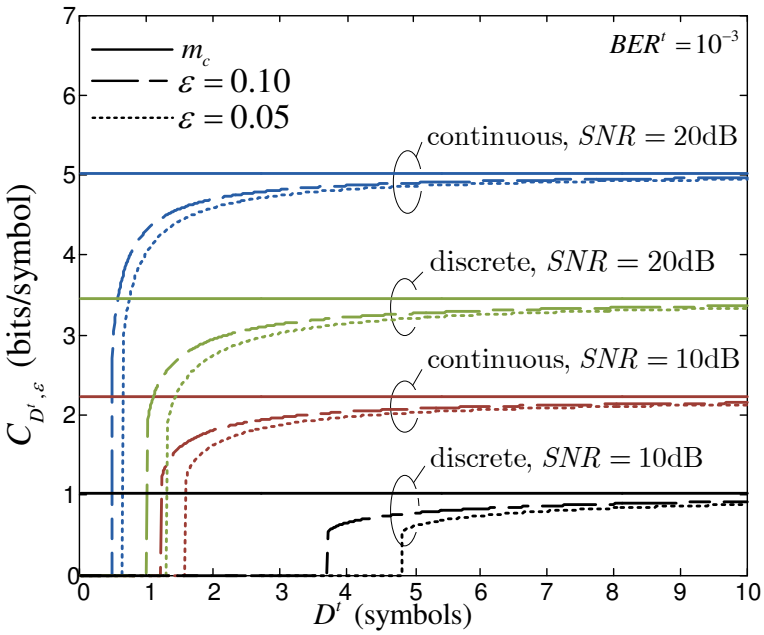


Figure 5.20: Capacity with Probabilistic Delay Constraint. Uncorrelated channel.

In Figures 5.21 and 5.22 the influence of the target BER is illustrated for the

continuous and the discrete rate policy, respectively. In the continuous case the value $\beta = 1$, corresponding to the upper bound, is also shown. The average SNR is set to 20dB and the capacity is plot as a function of the target delay. The demand of a stricter target BER results in a lower value of capacity. For any value of the target BER the continuous rate policy obtains higher values of capacity.

The evaluation of a time-correlated Rayleigh channel is shown in Figures 5.23 and 5.24 for continuous and discrete rate policy, respectively.

In the case of a continuous rate mechanism, we present two different values of $\overline{\gamma}$, 5dB and 10dB. The parameter β is 1 and the target BER is set to 10^{-2} in the discrete rate policy. $\varepsilon = 0.10$ and the correlation follows an exponential decay with the parameter ρ set to 0.95 and 0.99. The result of the uncorrelated case (marked with triangles) is also represented to perform the comparison with the time-correlated channel. We observe that the correlation in the channel response is harmful to the delay behaviour and thus the capacity is lower for stronger correlations, i.e. higher values of ρ. Consequently, the time-correlated channel capacity is always below the uncorrelated case.

In the discrete rate mechanism, Figure 5.24, the average SNR is set to 15dB. The correlation is now a first-kind Bessel function and the product $f_D \cdot T_S$ takes on two values, $50 \cdot 10^{-3}$ and $10 \cdot 10^{-3}$. The capacity is plot as a function of the target delay. The result of the uncorrelated case is again marked with triangles. Once more, for the same value of $\overline{\gamma}$ the time-correlated capacity is always below the uncorrelated case, which is an upper bound that is very close to the ergodic capacity for the values of D^t drawn in the figure. The uncorrelated case is practically the same as the ergodic capacity m_c for the values of D^t drawn in the figure. Notice that for the same violation probability (percentile 90%), the target delays D^t (x-axis) to be represented in the correlated channel in order to obtain and draw $\mathrm{C}_{D^t,\varepsilon} > 0$ are much higher than in the uncorrelated case, even for the values of correlation shown in the figure, which are not very high.

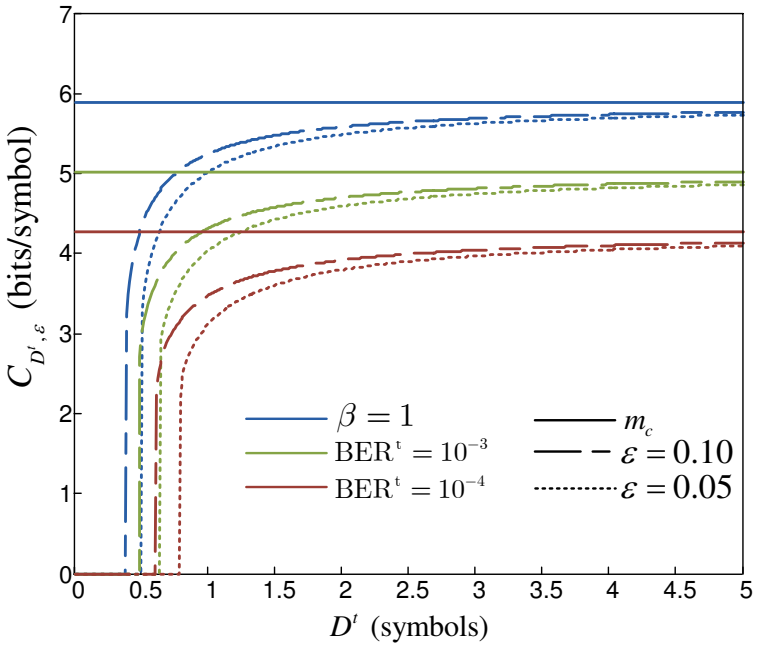


Figure 5.21: Capacity with Probabilistic Delay Constraint. Uncorrelated channel and continuous rate policy. Influence of the target BER.

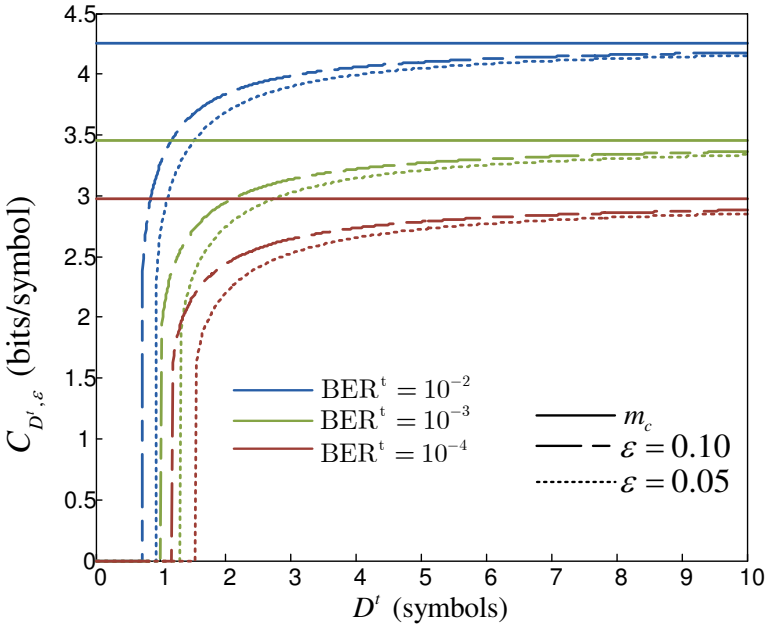


Figure 5.22: Capacity with Probabilistic Delay Constraint. Uncorrelated channel and discrete rate policy. Influence of the target BER.

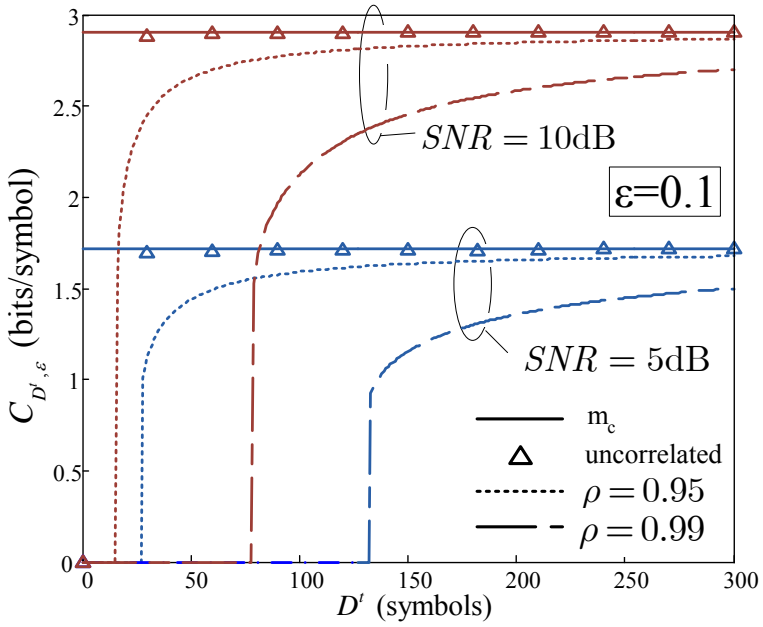


Figure 5.23: Capacity with Probabilistic Delay Constraint. Time-correlated channel. Continuous rate policy and exponential ACF

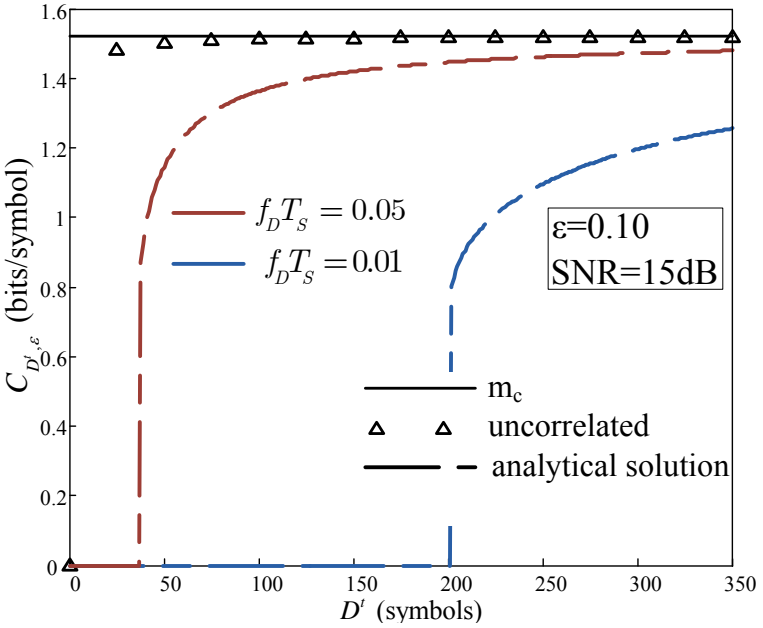


Figure 5.24: Capacity with Probabilistic Delay Constraint. Time-correlated channel. Discrete rate policy and Bessel ACF

Finite State Markov Chain

We repeat the calculations of the capacity with a channel modeled through a FSMC, which will be compared to ours later.

The effective bandwidth of a channel modeled with a FSMC was (equation (4.4.5)):

$$\alpha_C(\upsilon) = \frac{1}{\upsilon}\log(\Omega(\mathbf{P}\cdot\mathbf{C}(\upsilon))) \tag{5.2.5}$$

where $\mathbf{P}$ is the transition probabilities matrix, $\mathbf{C}(\upsilon)$ is a diagonal matrix whose diagonal elements are of the form $c_{ii} = \exp(r_i\upsilon)$, 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.

With a constant rate source, the value of λ is worked out from (5.2.1) :

$$\lambda \approx -\frac{\log\frac{\varepsilon}{\eta}}{\theta\cdot D^t} \tag{5.2.6}$$

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

The value of the QoS exponent θ is obtained by replacing (5.1.3) with (5.2.6) and (5.2.5):

$$\lambda\cdot\theta + \log(\Omega(\mathbf{P}\cdot\mathbf{C}(-\theta))) = 0 \tag{5.2.7}$$

The computation of the rate λ provides the limiting source rate that accomplishes the QoS constraint, i.e., the Capacity with Probabilistic Delay Constraint $\mathrm{C}_{D^t,\varepsilon}$. In this case, it cannot be explicitly derived:

$$C_{D^t,\varepsilon} = -\frac{\log\varepsilon}{\theta\cdot D^t} \tag{5.2.8}$$

where θ is the solution to:

$$-\frac{\log \varepsilon}{D^t} - \log(\Omega(\mathbf{P} \cdot \mathbf{C}(-\theta))) = 0 \tag{5.2.9}$$

Figure 5.25 compares our result (marked with dashes) to that provided by the FSMC (marked with dots), for a discrete rate policy. The average SNR is set to 15dB, the correlation is now a first-kind Bessel function and the product $f_D \cdot T_S$ takes on two values, $50 \cdot 10^{-3}$ and $10 \cdot 10^{-3}$. The result of the uncorrelated case is again marked with triangles. As expected, the result provided by the FSMC overestimates the capacity for strict delay constraints, corresponding here to small values of the target delay.

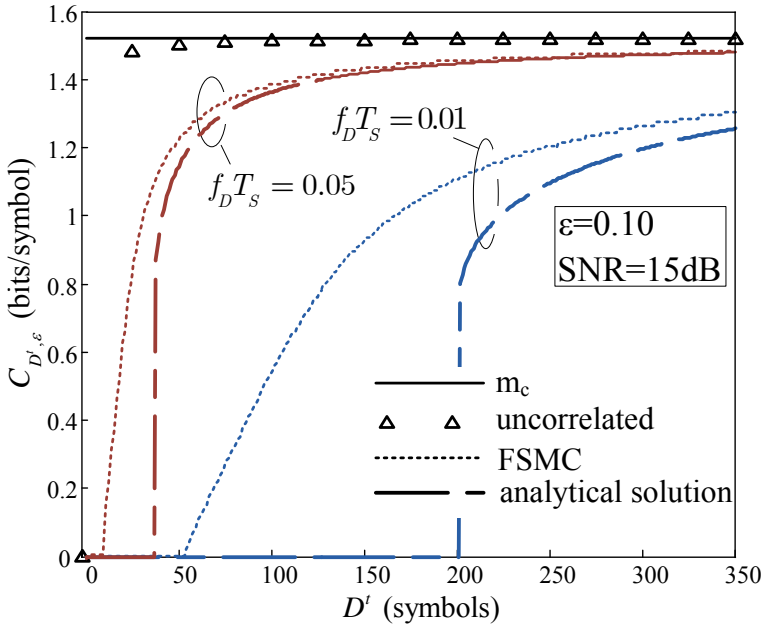


Figure 5.25: Capacity with Probabilistic Delay Constraint. Time-correlated channel. Bessel ACF. Comparison with the FSMC

5.2.2 Variable rate traffic

It is well known that if sources are more bursty, queueing performance is degraded, and to meet QoS the service rate has to be increased. In particular, delay suffered by an information flow depends not only on the transmission rate but also on the distribution and self-correlation of the information rate. Even in wired systems (constant rate channels) different distributions of the information process having the same average rate will cause different delays. As an example, it is known that ON-OFF sources with longer on-off periods and the same average rate suffer higher delays than less variable rate sources.

Under certain channel conditions and QoS metrics, the maximum $C_{D^t,\varepsilon}$ has been obtained when the incoming user traffic has constant rate. For any other source process, the maximum average source rate with supported delay will be lower. Thus, the influence of the parameters of the outsourcing process can be investigated. The derivation of the capacity for variable bit rate processes will depend on the model of the traffic source. As an example, the study of the capacity is carried out for two of the traffic models presented in Chapter 3: ON-OFF traffic modeling voice traffic, and an autoregressive source rate which models streaming services. Finally, the procedure to be applied when the EBF of the source cannot be obtained analytically is explained.

ON-OFF traffic

Let us consider the ON-OFF traffic detailed in Section 3.2.2, modeling a voice user. We reproduce here the EBF of this source, found in [Kelly 1996]:

$$\alpha_A(u) = \frac{1}{2u}\left[h \cdot u - \mu - \lambda + \sqrt{(h \cdot u - \mu + \lambda)^2 + 4 \cdot \lambda \cdot \mu}\right] \tag{5.2.10}$$

Since three parameters (h, λ and μ) are now involved in the definition of the arrival process, the formulation of $\mathrm{C}_{D^t,\varepsilon}$ gets complicated. The maximum *mean*

source rate m_A that can be supported fulfilling the desired value of ε can be obtained.

The QoS exponent is obtained by solving (5.1.3) for an ON-OFF process in the source and a channel of parameters m_c and σ_c^2. With (5.2.10) and (4.3.3) substituted into equation (5.1.3), the QoS exponent θ is found to be the solution of the following cubic equation:

$$\begin{aligned}(\sigma_c^2)^2\theta^3 - 2\sigma_c^2(2m_c - h)\theta^2 + (4m_c^2 - 4hm_c - 2\mu\sigma_c^2 \\ - 2\lambda\sigma_c^2)\theta + 4m_c\lambda + 4m_c\mu - 4h\lambda = 0\end{aligned} \tag{5.2.11}$$

The solution of (5.2.11) is employed in (5.2.1) to obtain this mean $m_A(D^t, \varepsilon)$:

$$m_A(D^t, \varepsilon) = -\frac{1}{2} \cdot \frac{\lambda}{\lambda + \mu} \cdot \frac{(\sigma_c^2)^2 s^3 - 4\sigma_c^2 m_c s^2 + 4m_c^2 s - 2\sigma_c^2(\lambda + \mu)s + 4m_c(\lambda + \mu)}{\sigma_c^2 s^2 - 2m_c s - 2\lambda} \tag{5.2.12}$$

with s:

$$s = \frac{-m_c D^t - \sqrt{(m_c D^t)^2 - 2\sigma_c^2 D^t \log \varepsilon}}{\sigma_c^2 D^t} \tag{5.2.13}$$

The analysis of the limits in (5.2.12) is not so obvious as in (5.2.2) but the dependence with D^t and ε is expected to be similar. The effect of the parameters of the channel process (m_c and σ_c^2) and the source process (λ and μ) are discussed next.

Figures 5.26 and 5.27 show the evaluation of $m_A(D^t, \varepsilon)$ as a function of the target delay D^t for $\varepsilon = 0.1$. A discrete rate policy with the constellations and thresholds defined before is employed. Additionally, the discrete-time ACF of the channel response follows a Bessel function with f_D the maximum Doppler frequency in Hertz and T_S the symbol period.

In Figure 5.26 the influence of the channel parameters can be evaluated. No changes with regard to the case of constant traffic are expected. The transition

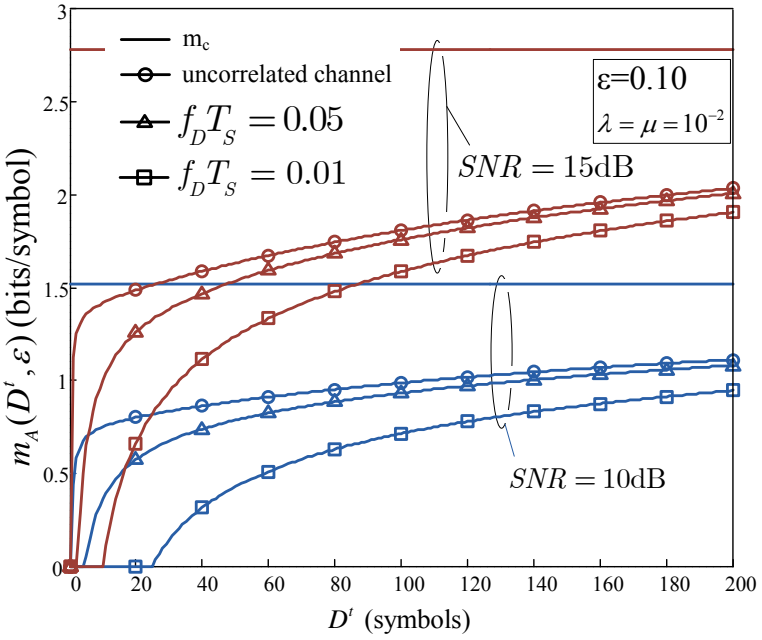


Figure 5.26: Capacity with probabilistic delay constraint for voice traffic. Influence of the channel parameters

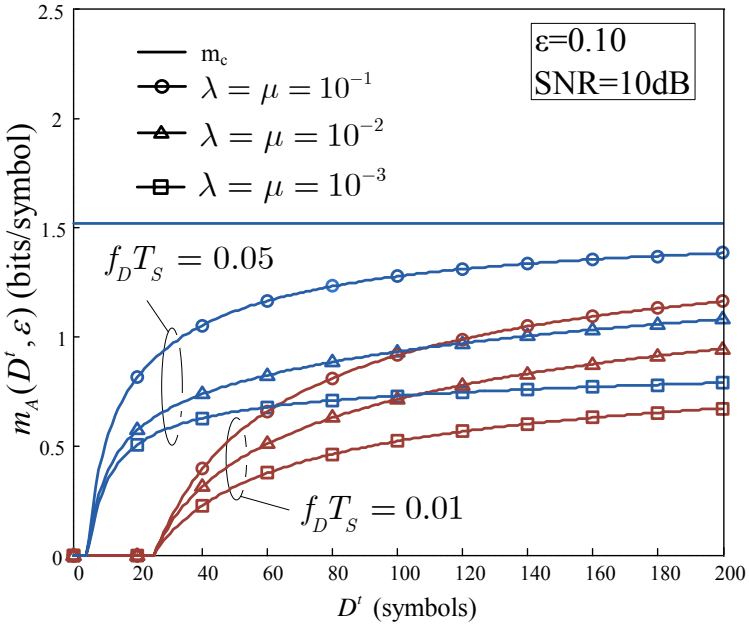


Figure 5.27: Capacity with probabilistic delay constraint for voice traffic. Influence of the source parameters

probability rates λ and μ are equal and set to 0.01. The average SNR takes two values, 10dB and 15dB. Two different values of the product $f_D \cdot T_S$ are represented, $5 \cdot 10^{-2}$ and $1 \cdot 10^{-2}$. The result of the uncorrelated case is also printed in order to compare with the time-correlated channel. As we have already observed in the constant traffic, the correlation in the channel response is harmful to the delay and the capacity diminishes with stronger correlation. Consequently, the uncorrelated channel capacity is always above the time-correlated case. In any case, both of them keep below the ergodic capacity.

In Figure 5.27 we can go deeper in the joint influence of the channel and the source parameters. The transition probability rates are equal with three different values (0.1, 0.01 and 0.001). The product $f_D \cdot T_S$ is set to $5 \cdot 10^{-2}$ and $1 \cdot 10^{-2}$. As expected, shorter ON and OFF states in the source process are beneficial to the capacity, and therefore the capacity increases for higher values of transition probability rates. Furthermore, it can be observed that the curves of the two different correlation values cross, showing the combined effect of the channel and the source parameters. Thus, in order to statistically accomplish certain delay, less frequent transitions between ON and OFF states can be carried over channels with weaker time correlation.

AR traffic

In the case of AR traffic, it is also feasible to obtain the maximum *mean* source rate m_A that can be supported fulfilling the desired value of ε. This mean $m_A(D^t, \varepsilon)$ is derived from (5.1.11) by replacing the EBF of the Rayleigh channel and as a function of the QoS exponent:

$$\alpha_A(\theta) = m_A + \frac{r^2}{2}\theta = \alpha_C(-\theta) \Rightarrow m_A(D^t, \varepsilon) = m_c - \left(\frac{\sigma_c^2}{2} + \frac{r^2}{2}\right)\theta \qquad (5.2.14)$$

where the QoS exponent is derived from (5.1.2):

$$\log(\varepsilon) = -\theta\alpha_C(-\theta)D^t \Rightarrow \theta = \frac{m_c}{\sigma_c^2} - \sqrt{\left(\frac{m_c}{\sigma_c^2}\right)^2 + \frac{2\log\varepsilon}{\sigma_c^2 D^t}} \qquad (5.2.15)$$

Figures 5.28 and 5.29 evaluate (5.2.14) as a function of the target delay D^t. In Figure 5.28 the source parameter r is set to 0.8781 as in [Maglaris 1998]. Two values of violation probability, $\varepsilon = 0.1$ and $\varepsilon = 0.01$ are plot. The channel is uncorrelated with average SNR 15dB and 25dB. The discrete and continuous rate policies are shown both with a target BER of 10^{-3}. The same conclusions obtained for the constant traffic apply here.

In Figure 5.29 ε is set to 0.1 and the influence of the source captured in r is illustrated. When r increases the capacity decreases and viceversa. Thus, if ρ_A (in the denominator of r) increases, i.e. the traffic is more correlated, the capacity decreases. In the figure, the parameter ρ_A is set to 0.7, 0.8 and 0.9. Obviously, the influence of the parameter q is the opposite: when q increases the weight of the white noise increases and so does the capacity.

Other traffic sources

From Chapter 3 we know that in many occasions the effective bandwidth of the source either has no analytical expression or the analytical formula is highly complex. In these cases, it is more convenient to work with the measured function. In other occasions the number of parameters involved in the EBF is too high and it does not make sense to work out the capacity with probabilistic delay constraint. Anyway it is feasible to study the behaviour of the delay and to determine if the delay constraint can or can not be accomplished.

The iterative procedure to study the delay in this situation is as follows.

- Consider a flat Rayleigh channel with parameters m_c and σ_c^2 (m_k/k and σ_k^2/k if it is time-correlated).
- Assume that the delay constraint to be fulfilled is given by (D^t, ε).
- Evaluate the QoS exponent with (5.2.15).

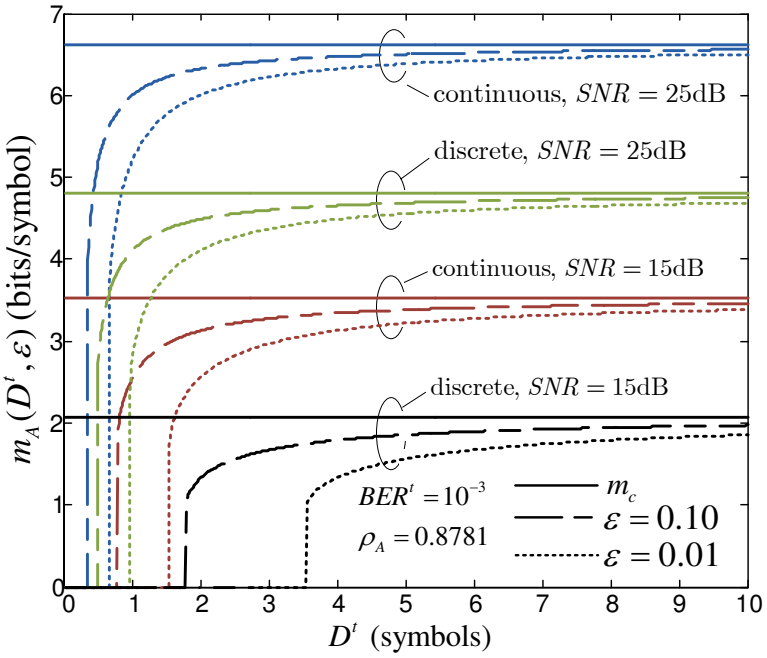


Figure 5.28: Capacity with probabilistic delay constraint for AR traffic. Influence of the channel parameters

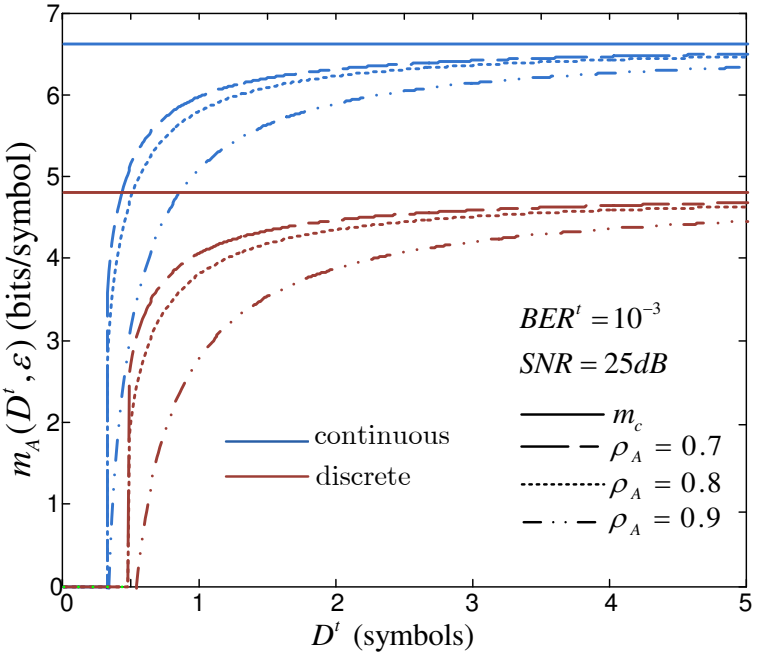


Figure 5.29: Capacity with probabilistic delay constraint for AR traffic. Influence of the source parameters

- Evaluate $\alpha_A(\theta)$ from the measured or estimated function.
- Evaluate $\alpha_C(\theta) = m_c + \frac{\sigma_c^2}{2}$
- If $\alpha_A(\theta)$ is less than or equal to $\alpha_C(\theta)$, then the delay constraint can be accomplished.
- Otherwise, the delay constraint can not be fulfilled. Try with a higher value of D^t or ε.

5.3 Simulation comparison

A queueing system has been simulated and compared with the analytical results in order to validate our analysis.

Constant rate traffic

First of all, constant rate traffic is simulated, with the scenario shown in Figure 5.30.

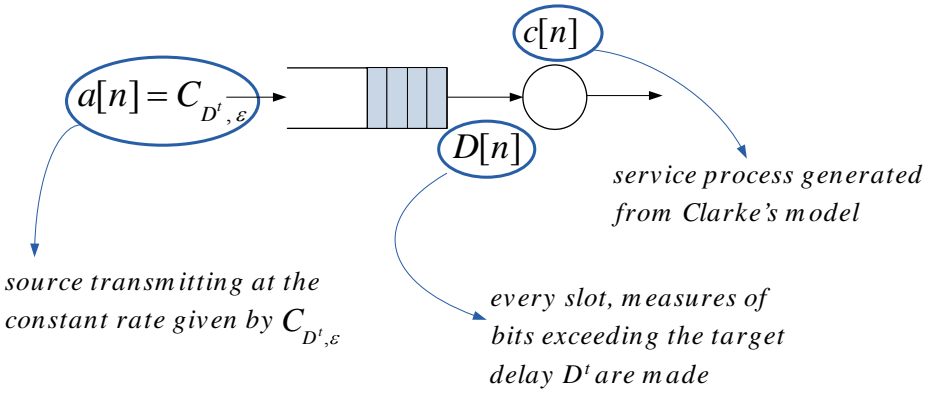


Figure 5.30: Simulation system

For certain values of D^t and ε, the *Capacity with Probabilistic Delay Constraint*

is evaluated with (5.2.2). The arrival process generates source data at the (constant) resulting limiting rate, $C_{D^t,\varepsilon}$. Bits are sent to the buffer of queue length $Q[n]$ in the nth sample interval. The bits in the queue are transmitted over the fading channel according to a FCFS discipline and at service rate $c[n]$. The description of the queue with Kendall notation is D/G/1, with service times that are correlated in the case of generic Rayleigh channel and time-independent when the block fading scenario is considered. The tail probability of exceeding the delay target is measured based upon the measures of the delay suffered by bits leaving the queue. Notice that the expected value of this tail probability, $Pr\{D(\infty) > D^t\}$, is ε.

The result of the simulations for an uncorrelated and time-correlated channel are shown in Figure 5.31 and Figure 5.32, respectively. In Figure 5.31, $Pr\{D(\infty) > D^t\}$ is represented for a target delay of 10 symbols and an expected violation probability $\varepsilon = 0.1$, for the continuous case, and $\varepsilon = 0.05$, for the discrete case. 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 since the capacity was obtained for $\eta = 1$, η approaching 1 as D^t or $\overline{\gamma}$ increases. Note that the QoS requirements are very accurately reached with the simple result for $C_{D^t,\varepsilon}$ in (5.2.2).

Figure 5.32 represents $Pr\{D(\infty) > D^t\}$ for the simulations of the time-correlated channel with a target delay of 140 symbols. Two different values of the product $f_D \cdot T_s$ are presented. For the first one, $f_D \cdot T_s = 10 \cdot 10^{-4}$, the rate is adapted following the continuous policy and the violation probability is fixed to $\varepsilon = 0.1$. For the second value of $f_D \cdot T_s$, $10 \cdot 10^{-3}$, the discrete rate policy is employed and $\varepsilon = 0.05$. Once more, the curve including the term η approximates more precisely the simulated results in both cases. Nevertheless, the QoS requirements are fulfilled, achieving a very accurate approximation in the continuous rate policy.

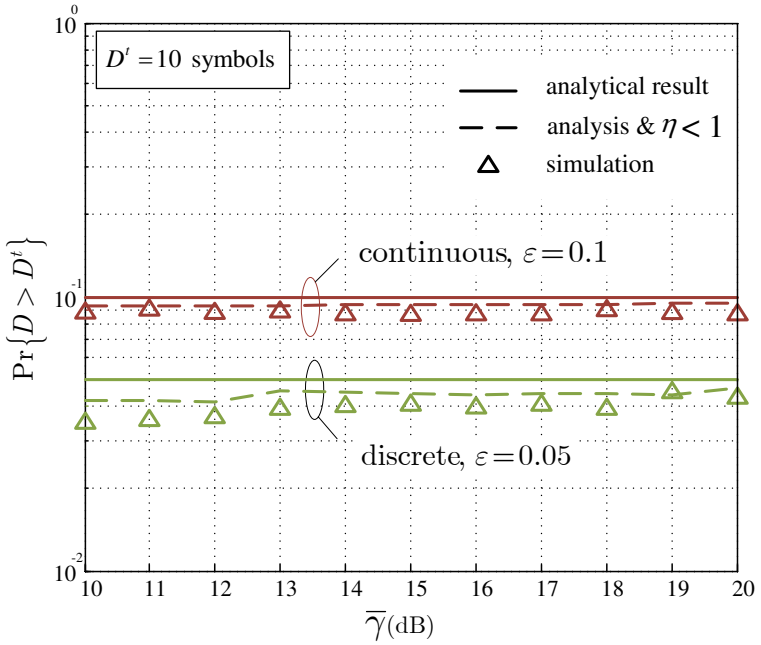


Figure 5.31: Simulated probability of delay violation versus analytical bound. Uncorrelated channel.

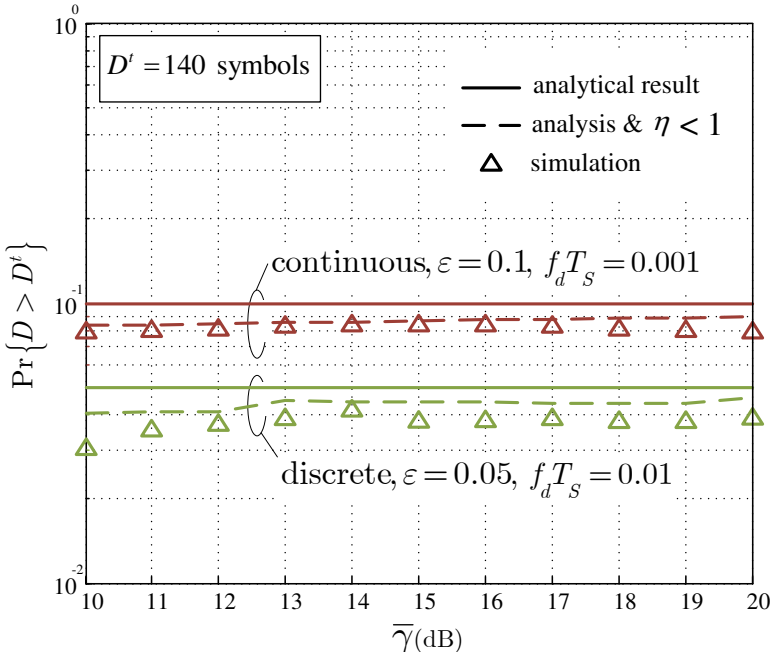


Figure 5.32: Simulated probability of delay violation versus analytical bound. Time-correlated channel. Bessel ACF

Variable rate traffic

- ON-OFF traffic

The arrival process is generated according to an ON-OFF process. Regarding the service process, the service times are correlated when the time-correlation of the Rayleigh channel is considered and time-independent if block fading is assumed.

Certain values of the delay constraint D^t, ε are fixed, as well as the channel and the source parameters. Then, the *Capacity with Probabilistic Delay Constraint* is evaluated with (5.2.12). With the arrival process generating the source data at the mean resulting limiting rate, $C_{D^t,\varepsilon}$, the tail probability of exceeding the delay target is measured, with expected value ε.

Figure 5.33 shows the results of the simulations. $Pr\{D(\infty) > D^t\}$ is represented

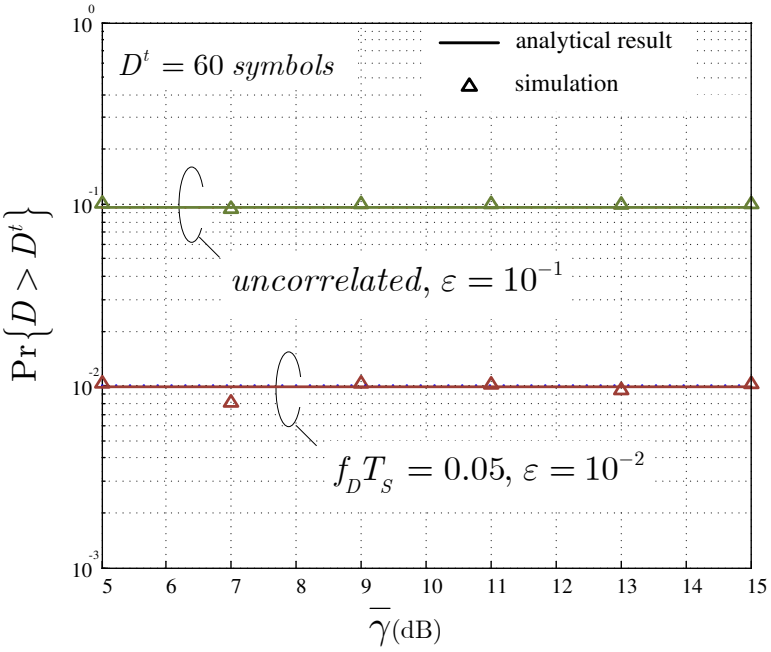


Figure 5.33: ON-OFF traffic

as a function of the mean SNR, in the range from 5dB to 15dB, for a target delay

of 60 symbols and violation probability $\varepsilon = 0.10$ and $\varepsilon = 0.01$. The service rate $c[n]$ has been generated for the two options considered here: uncorrelated and correlated Rayleigh wireless channel. In the case of the correlated one, the product $f_D \cdot T_S$ has been set to 0.05. The discrete rate policy with adaptive modulation works with the same parameters specified in Chapter 2: for each SNR value the transmitter selects the most dense constellation that keeps the instantaneous BER under the maximum $BER^t = 10^{-2}$ between the four different uncoded constellations (BPSK, QPSK, 16QAM and 64QAM) and the option of no transmitting if the instantaneous SNR is below a certain threshold. In the source process, the mean ON and OFF rate, λ and μ, are equal and set to 10^{-2}. Once again the QoS requirements are very accurately reached with the result for $C_{D^t,\varepsilon}$ in (5.2.12): the measured violation probability is the expected ε. The capacity was obtained under the assumption that $\eta = 1$, a high load scenario that constitutes an upper bound. It has been checked in the simulations that the measured probability of non-empty queue η is one.

- AR traffic

The simulation of the AR traffic is shown in Figure 5.34. The arrival process now generates source data according to an autoregressive model of order 1. The parameters of the source have been taken from [Maglaris 1998].

The rate policy is assumed to be continuous, with $\beta = 1$, average SNR 5dB and a discrete-time ACF following an exponential fall with rate $\rho = 0.95$, which leads to a mean $m_c = 1.72$ and a variance $\sigma_c^2 = 18.12$. With the expression in (5.1.18) it is feasible to predict the ratio $log(\varepsilon)/D^t$. In the simulations, the target delay has been varied from 0 to 10 symbols and the probability of exceeding this target delay is measured and compared to the predicted ε. Once more, the predicted values of delay behaviour are accurately reached, obtaining a better approximation when the curve includes the measured term η (dashed line).

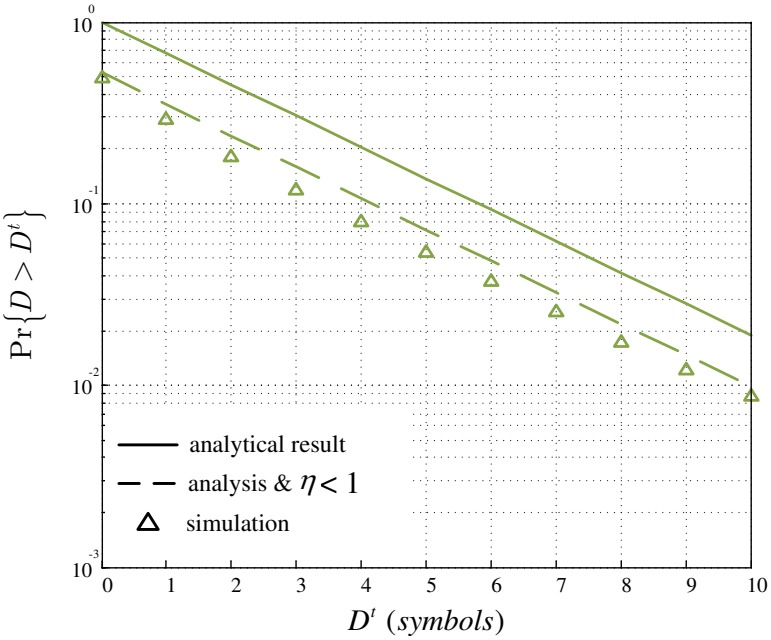


Figure 5.34: AR traffic

5.4 Summary

The effective bandwidth function of the source and the channel process make feasible the analysis of the queueing system. The EBT provides a solution of the tail probability function of the delay. The analysis of the delay in a flat Rayleigh channel can be done by means of the EBT. The percentile of the delay or the tradeoff between delay and source rate are obtained. In addition, we propose the definition of a new concept of capacity in a wireless system, the Capacity with Probabilistic Delay Constraint, $C_{D^t,\varepsilon}$. This capacity represents the maximum allowable rate that the channel can transport under a delay constraint, expressed in terms of a target delay D^t and the probability of exceeding it, ε. In contrast to the delay-limited capacity, which turns to be zero in Rayleigh channels, $C_{D^t,\varepsilon}$ is related to statistical QoS guarantees. With the numerical evaluation, the potential of the Capacity with Probabilistic Delay Constraint has been shown, giving explicit information of the limits on the information that can be transmitted under certain conditions in the

channel, the outsourcing process and the QoS metrics. It is also worth mentioning the mathematical tractability of the result, which permits its exact numerical evaluation with explicit values of the QoS parameters.

Chapter 6

Delay constrained communications over frequency selective channels

---

The analysis of the delay for flat Rayleigh channels was addressed in the precedent chapter. Here, the results are generalized to a multi-channel system employing OFDM modulation over a frequency-selective Rayleigh channel. With this purpose, the effective bandwidth function of the frequency selective channel is first obtained. Then, it is feasible to redo the analysis of the delay for the new system model. Moreover, the Capacity with Probabilistic Delay Constraint $C_{D^t,\varepsilon}$ in this new scenario is also examined.

The results presented in this chapter have been partially published in [Soret 2009a].

The chapter is organized as follows. First of all, the system model is presented in Section 6.1. Closed-form expressions of the new channel effective bandwidth function are obtained in Section 6.2. This result makes it possible to obtain the percentile of the delay in a OFDM system, as detailed in Section 6.3. The Capacity with Probabilistic Delay Constraint in a OFDM system is presented in Section 6.4. Section 6.5 shows the comparison of our results with simulations. Finally, some concluding remarks are given in Section 6.6.

6.1 System model for the OFDM system

6.1.1 OFDM

The origins of the Orthogonal Frequency Division Multiplexing (OFDM) date back to the 1960s. In [Doeltz 1957] the authors introduce the principle of transmitting data by dividing it into several interleaved bit streams, and using these to modulate several carriers. Nevertheless, it was not really considered an interesting option due to technological limitations until recently, where it has been adopted as the main transmission strategy in many standards. Among others, it has been selected for digital audio and video broadcasting (DAB/DVB), wireless Local Area Networks (LAN) (HIPERLAN/2 in Europe, IEEE802.11a in North America) and Metropolitan Area Networks (MAN) (IEEE802.16) and the fourth generation of cellular wireless (LTE).

OFDM combats frequency selective fading by decomposing the wideband channel into a set of flat fading narrowband orthogonal subchannels that can be independently modulated. Correlation between subcarriers is closely related to the delay spread of the channel. Moreover, OFDM multicarrier nature allows the use of adaptive modulation [Chung 2001] to enhance its performance significantly. In particular, constellation modulating each subcarrier can be selected suitably to the frequency and time varying conditions of the physical channel.

Over frequency selective channels, OFDM decomposes the total bandwidth into F equally spaced subcarriers, each of them relatively narrowband (Figure 6.1). An Inverse Fast Fourier Transform (IFFT) is used at the transmitter to modulate the comb of subcarriers, and a cyclic prefix is added to eliminate intersymbol interference. At the receiver, the dual process is performed. At each subcarrier, the perceived time variant channel response at instant n is denoted as $H[n, f]$, where the index f indexes the subcarriers.

Assuming the employment of adaptive modulation at each subchannel, subcarrier

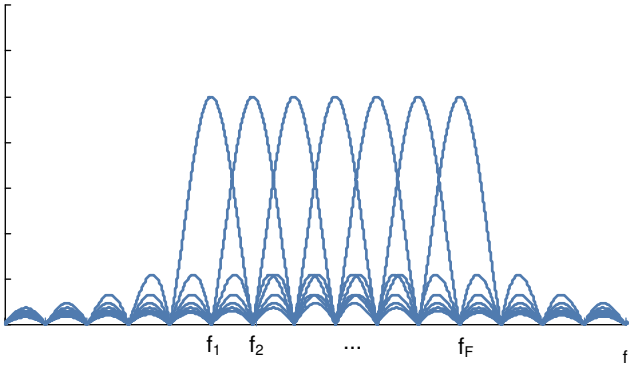


Figure 6.1: OFDM spectrum.

f can transmit at an instantaneous rate $c^f[n]$, which changes according to the time varying channel state. A discrete rate policy with constant transmitted power is considered (see Section 4.1.2). At each subcarrier the range of received SNR is divided into M consecutive regions, each of them associated with a constellation size. A constellation of M_i symbols is selected within the fading region (Γ_{i-1}, Γ_i), $i = 0, 1, .., M$ (defining $\Gamma_{-1} = 0$). Therefore, $c^f[n]$ can be expressed as a function of the envelope of the channel gain, $z_{n,f} = |H[n, f]|$:

$$\begin{aligned} & c^f[n] = c(z_{n,f}) = \log_2(M_i) = \mathrm{c}_i, \\ & \varrho^{i-1} \leqslant z_{n,f} < \varrho^i, \ i = 0..M, \quad \text{with} \ \ \varrho^k = \sqrt{\frac{\Gamma_k}{\overline{\gamma}}} \end{aligned} \tag{6.1.1}$$

6.1.2 Channel model

The frequency selective channel is modeled as a set of L taps, each of them delayed τ_l from the first one and with a time varying gain $h_l[n]$ [Rappaport 2002]:

$$h[n, \tau] = \sum_{l=0}^{L-1} h_l[n]\delta[\tau - \tau_l] \tag{6.1.2}$$

The time varying gain of each tap, $h_l[n]$ is:

$$h_l[n] = p_l h_{norm}[n] \tag{6.1.3}$$

where the constant p_l determines the power of the tap and $h_{norm}[n]$ is a zero-mean complex-valued Gaussian process.

The channel perceived at each subcarrier f is obtained by means of the Fourier transform of (6.1.2) respect to the delay parameter τ:

$$H[n, f] = \underset{\tau}{\mathfrak{F}}\{h[n, \tau]\} \tag{6.1.4}$$

The power delay profile (PDP) or multipath intensity profile gives the intensity of a signal received through a multipath channel as a function of the time delay. The time delay is the difference in travel time between multipath arrivals. The abscissa is in units of time and the ordinate is the powers p_l. It is easily measured empirically and can be used to extract certain channel's parameters such as the delay spread.

We can also characterize the time-varying multipath channel in the frequency domain. In this case, the coherence bandwidth B_c is a parameter commonly used to provide a measure of the range of frequencies over which the channel can be considered *frequency-flat*. It is closely related to the PDP through the rms delay spread:

$$B_c \approx \frac{1}{2\pi D} \tag{6.1.5}$$

where D is typically taken to be the rms delay spread in seconds.

Two examples of frequency selective channel are employed in this chapter. Both of them belong to the recommendation of ITU for IMT-2000 systems [ITU-R 1997]. In particular, the pedestrian test environment channel A and the vehicular test environment channel A have been implemented. The tapped delay line parameters of the channels are shown in Table 6.1. For each tap of the channels three parameters are given: the time delay relative to the first tap, the average power relative to the strongest tap, and the Doppler spectrum of each tap [1].

[1] Recall that the term classic in the Doppler spectrum refers to Jakes' spectrum.

Table 6.1: Parameters of the pedestrian and vehicular channels.

	Pedestrian Channel A		Vehicular Channel A		
Tap	Relative Delay (ns)	Average Power (dB)	Relative Delay (ns)	Average Power (dB)	Doppler Spectrum
1	0	0	0	0	Classic
2	110	-9.7	310	-1	Classic
3	190	-19.2	710	-9	Classic
4	410	-22.8	1090	-10	Classic
5	-	-	1730	-15	Classic
6	-	-	2510	-20	Classic

6.1.3 Queueing model

The queueing system (see Figure 6.2) is an extension of that applied for the flat Rayleigh fading. It consists of a source process that characterizes the incoming user traffic and F servers in parallel representing the F subcarriers.

With the same discrete-time fluid model applied in flat channels, the incoming source traffic has an instantaneous rate $a[n]$. On the other hand, subcarrier f can transmit at an instantaneous rate $c^f[n]$, so that the total instantaneous rate of the wireless channel is $c[n] = \sum_{f=1}^{F} c^f[n]$, where F is the number of subcarriers. The definition of accumulated source rate $A[n]$ is the same considered up to now, $A[n] = \sum_{m=0}^{n-1} a[m]$, and for the channel process:

$$C[n] = \sum_{m=0}^{n-1} c[m] = \sum_{m=0}^{n-1} \sum_{f=1}^{F} c^f[m] \tag{6.1.6}$$

With the considerations above, the rest of the analysis of the queueing system is analogous to the one in Chapter 2, with the difference that now the transmission rate is the sum of the instantaneous transmission rate in every subcarrier. The expression of the probability of exceeding a target delay D^t has not changed:

$$\varepsilon = Pr\{D(\infty) > D^t\} \approx \eta \cdot e^{-\theta \cdot \alpha_A(\theta) D^t} = \eta \cdot e^{-\theta \cdot \alpha_C(-\theta) D^t} \tag{6.1.7}$$

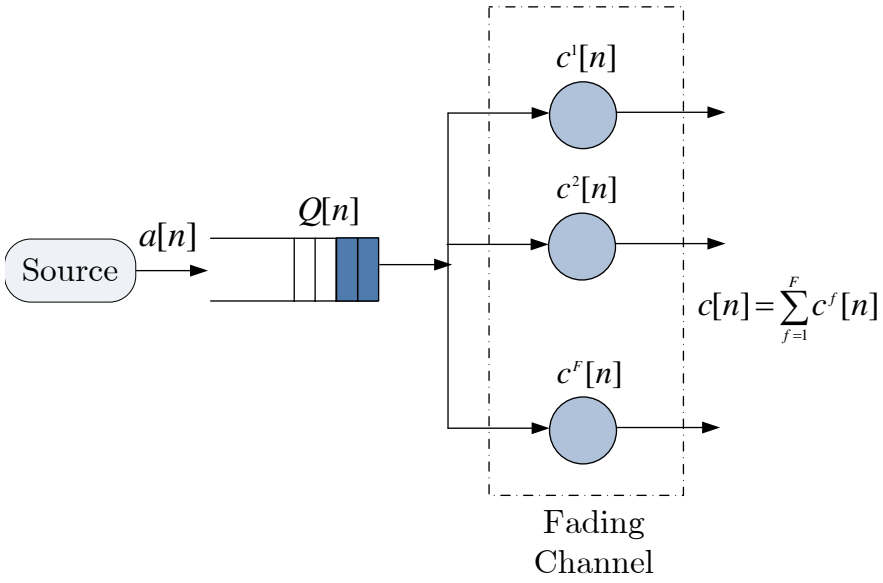


Figure 6.2: Queueing model used for OFDM systems.

where ε is the probability of exceeding the target delay D^t. The EBF of the frequency selective channel has to be calculated.

6.2 Effective Bandwidth Function of a frequency selective channel

Similarly to the case of flat Rayleigh channels, the analysis of the EBF of the channel $\alpha_C(u)$ is faced by dividing the cumulative channel rate $C[n]$ in the interval $[0..n]$ into $B = \lfloor \frac{n}{k} \rfloor$ blocks of size k, as shown in Figure 6.3:

The cumulative transmission rate is written:

$$C[n] = \sum_{b=0}^{B-1} C_b[k] \tag{6.2.1}$$

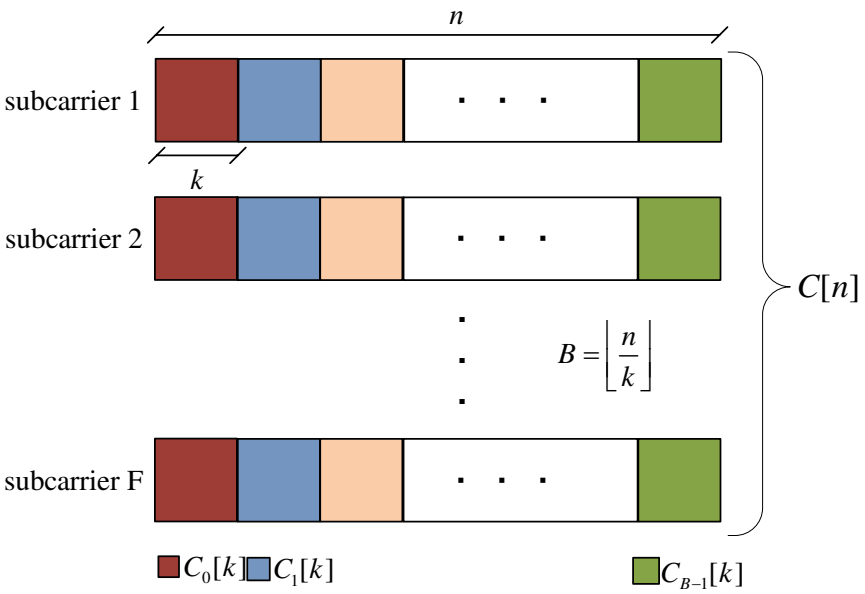


Figure 6.3: Block division of C[n].

where $C_b[k]$ is the accumulated channel rate in block b:

$$C_b[k] = \sum_{p=0}^{k-1} c[k \cdot b + p] = \sum_{p=0}^{k-1} \sum_{f=1}^{F} c^f[k \cdot b + p] \tag{6.2.2}$$

The same arguments considered in the flat channel apply here. Not only the time correlation between the elements in the block is considered but also the frequency correlation. Nevertheless, independence between elements of different blocks can be assumed if a proper selection of the block's length k is done. This way $C[n]$ is the sum of independent and identically distributed random variables, the blocks $C_b[k]$, and the Central Limit Theorem is applicable.

The choice of k has to fit in with the same considerations stated in the flat channel. Therefore, it is closely related to the correlation of the channel, now not only on time but also on frequency. If the channel is strongly time- or frequency-correlated, longer blocks have to be defined in order to assume independent blocks. In the numerical evaluations and simulations conducted throughout this chapter,

the validity of this Gaussian approximation for $C[n]$ has been validated by testing for normality (Lilliefors test) on the result for the selected values of k.

Under these conditions (sufficiently long k and n), the effective bandwidth function of the resulting Gaussian distribution of $C[n]$ is computed as:

$$\alpha_C(\upsilon) = \lim_{n\to\infty} \frac{1}{n \cdot \upsilon} \log \mathrm{E}\left[e^{\upsilon C[n]}\right] = \frac{m_k}{kF} + \frac{u}{2}\frac{\sigma_k^2}{kF} \tag{6.2.3}$$

where m_k and σ_k^2 are the mean and the variance of the blocks $C_b[k]$ normalized with the block length.

The mean m_k is readily computed from the mean m_c of the discrete random variable $c^f[n]$:

$$m_k = F \cdot k \cdot m_c = F \cdot \sum_{i=0}^{M} p_i \cdot \mathrm{c}_i \tag{6.2.4}$$

where the probability of using the ith constellation is given by:

$$p_i = \exp(-\frac{\Gamma_{i-1}}{\bar{\gamma}}) - \exp(-\frac{\Gamma_i}{\bar{\gamma}}) \tag{6.2.5}$$

However, the evaluation of the variance of the blocks is not straightforward. The following two properties will be used to derive σ_k^2 ([Papoulis 2002]):

$$\begin{aligned} \mathrm{var}(\sum_{p=0}^{k-1} X_p) &= \sum_{p=0}^{k-1}\sum_{q=0}^{k-1} \mathrm{cov}(X_p, X_q) \\ \mathrm{cov}(\sum_{i=0}^{k-1} X_i, \sum_{j=0}^{k-1} Y_i) &= \sum_{i=0}^{k-1}\sum_{j=0}^{k-1} \mathrm{cov}(X_i, Y_j) \end{aligned} \tag{6.2.6}$$

with autocovariance function

$$\mathrm{cov}(X_n, X_{n+m}) = \mathrm{E}[X_n X_{n+m}] - \mathrm{E}[X_n]\mathrm{E}[X_{n+m}] \tag{6.2.7}$$

Applying the expression above to the variance of the blocks we obtain:

$$\sigma_k^2 = \mathrm{var}(\sum_{p=0}^{k-1}\sum_{f=1}^{F} c^f[p]) = \sum_{p=0}^{k-1}\sum_{q=1}^{k-1}\sum_{f=1}^{F}\sum_{g=1}^{F}\left(\mathrm{E}\left[c^f[p]c^g[q]\right] - m_c^2\right) \tag{6.2.8}$$

The mean m_c is known from the flat channel. Thus, the evaluation of the variance of the blocks comes down to evaluating $\mathrm{E}[c^f[n]c^g[m]]$. To do it the instantaneous service rate $c^f[n]$ is expressed as a function of the envelope of the response of the channel $z_{n,f} = |H[n,f]|$, as in (6.1.1). The expectation is solved by splitting the integral in $z_{n,f}$ in several addends and the same is done for the integral in $z_{m,g}$. The expectation can then be written as:

$$\mathrm{E}[c^f[n]c^g[m]] = \sum_{i=1}^{M} \mathrm{c}_i \sum_{j=1}^{M} \mathrm{c}_j \; \mathcal{F}_{m-n,g-f}(\varrho^i, \varrho^j) \tag{6.2.9}$$

with:

$$\begin{aligned} &\mathcal{F}_{\Delta n,\Delta f}(\varrho^i, \varrho^j) \\ &= F_{\mathbf{z}}(\varrho^i, \varrho^j, \Delta n, \Delta f) + F_{\mathbf{z}}(\varrho^{i-1}, \varrho^{j-1}, \Delta n, \Delta f) \\ &-F_{\mathbf{z}}(\varrho^{i-1}, \varrho^j, \Delta n, \Delta f) - F_{\mathbf{z}}(\varrho^i, \varrho^{j-1}, \Delta n, \Delta f) \end{aligned} \tag{6.2.10}$$

$F_{\mathbf{z}}(\varrho^i, \varrho^j, \Delta n, \Delta f)$ is the bivariate cumulative distribution function (CDF) for

Rayleigh distributed variables [Simon 2005] (pp. 172, eq. 6.7):

$$
\begin{aligned}
&F_{\mathbf{z}}(u,v,m,g) = 1 - \mathbf{g}(u,v,m,g) + \\
&+\frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left(-\frac{u^2+v^2+2\sqrt{\mathcal{R}_{z^2}(m,g)}uv\sin\theta}{1-\mathcal{R}_{z^2}(m,g)}\right) \\
&\cdot\left[\frac{\left(1-\mathcal{R}_{z^2}(m,g)\right)u^2v^2+\sqrt{\mathcal{R}_{z^2}(m,g)}\left(1-\mathcal{R}_{z^2}(m,g)\right)uv\left(u^2+v^2\right)\sin\theta}{\left(\mathcal{R}_{z^2}(m,g)u^2+2\sqrt{\mathcal{R}_{z^2}(m,g)}uv\sin\theta+v^2\right)}\cdot\right. \\
&\left.\frac{1}{\left(u^2+2\sqrt{\mathcal{R}_{z^2}(m,g)}uv\sin\theta+\mathcal{R}_{z^2}(m,g)v^2\right)}\right]d\theta,
\end{aligned}
$$

$$
\mathbf{g}(u,v,m,g) = \begin{cases}
\exp(-v^2) & 0 \leqslant v < \sqrt{\mathcal{R}_{z^2}(m,g)}u \\
\frac{1}{2}\exp(-u^2) + \exp(-u^2\mathcal{R}_{z^2}(m,g)) & v = \sqrt{\mathcal{R}_{z^2}(m,g)}u \\
\exp(-u^2) + \exp(-v^2) & \sqrt{\mathcal{R}_{z^2}(m,g)}u \leqslant v < u/\sqrt{\mathcal{R}_{z^2}(m,g)} \\
\frac{1}{2}\exp(-v^2) + \exp(-v^2\mathcal{R}_{z^2}(m,g)) & v = u/\sqrt{\mathcal{R}_{z^2}(m,g)} \\
\exp(-u^2) & u/\sqrt{\mathcal{R}_{z^2}(m,g)} < v
\end{cases} \tag{6.2.11}
$$

The correlation parameter $\mathcal{R}_{z^2}(\Delta n, \Delta f)$ is:

$$
\mathcal{R}_{z^2}(\Delta n, \Delta f) = \frac{\mathrm{cov}(z^2_{n,f}, z^2_{m,g})}{\sqrt{\mathrm{var}(z^2_{n,f})\mathrm{var}(z^2_{m,g})}} \tag{6.2.12}
$$

which in turns depends on the expectation $\mathrm{E}[|H(n,f)|\,|H(m,g)|]$.

The expectation $\mathrm{E}[|H(n,f)|\,|H(m,g)|]$ expresses the correlation of the channel

gain at different frequencies and time instants. This expectation yields:

$$\begin{aligned}
&\mathrm{E}[|H(n,f)|\,|H(m,g)|] \\
&= \mathrm{E}[\sum_{l=0}^{L-1} |h_l[n]| e^{-j\frac{2\pi}{M}fl} \sum_{l'=0}^{L-1} |h'_l[m]| e^{-j\frac{2\pi}{M}gl'}] \\
&= \mathrm{E}[\sum_{l=0}^{L-1} |h_l[n]||h'_l[m]| e^{-j\frac{2\pi}{M}(f-g)l}] \\
&= \mathrm{E}[|h_{norm}[n,f]|\,|h_{norm}[m,g]|] \cdot \left(\sum_{l=0}^{L-1} p_l e^{-j\frac{2\pi}{M}(f-g)l} \right)
\end{aligned} \tag{6.2.13}$$

where the second equality comes from the independence among taps and $|h_{norm}[n,f]|$ and $|h_{norm}[m,g]|$ are Rayleigh distributed.

Observing (6.2.13), the expression consists of two independent factors, corresponding to the time and frequency correlation, respectively. The first term is the time-correlation function. If the Jakes' model is assumed, then it is the well-known result of the Bessel function already employed in the flat channel:

$$\mathrm{E}[|h_{norm}(n,f)|\,|h_{norm}(m,f)|] = J_0(2\pi f_D T_S(m-n)) \tag{6.2.14}$$

with $J_0(.)$ the zeroth order Bessel function of the first kind, f_D the maximum Doppler frequency and T_S the symbol period. This is the time-correlation function used along this chapter. Nevertheless, any other correlation function can be considered.

The second factor in (6.2.13) is the Fourier transform of the Power Delay Profile, p_l, which describes the number L and delay τ_l of the multipath components, as well as the average power associated with each multipath delay of the considered channel.

This section about the effective bandwidth function of a frequency selective channel concludes by presenting the numerical evaluation of $\alpha_C(-\upsilon)$ with the detailed mean and variance for a frequency-selective channel. To check the validity of the outcomes, the mean and the variance have been measured over a long realization of the instantaneous transmission rate process and compared to the analytical results, following the same method as described in Chapter 5.

Figure 6.4 shows the result of computing (6.2.3) with an average SNR of 10dB and 14dB and 16 subcarriers. The PDP corresponds to a vehicular channel A as defined in [ITU-R 1997], which leads to a coherence bandwidth of $B_c \approx 430$KHz. The length of the blocks k is set to 10000. Notice that if the choice of k is appropriate (high enough) and the Central Limit Theorem can be applied, there is no dependence of $\alpha_C(-\upsilon)$ with the block length k. At each subcarrier, the usual discrete rate policy parameters are assumed. As expected, $\alpha_C(-\upsilon)$ decreases with υ, as a stringent QoS constraint implies a high value of the parameter υ.

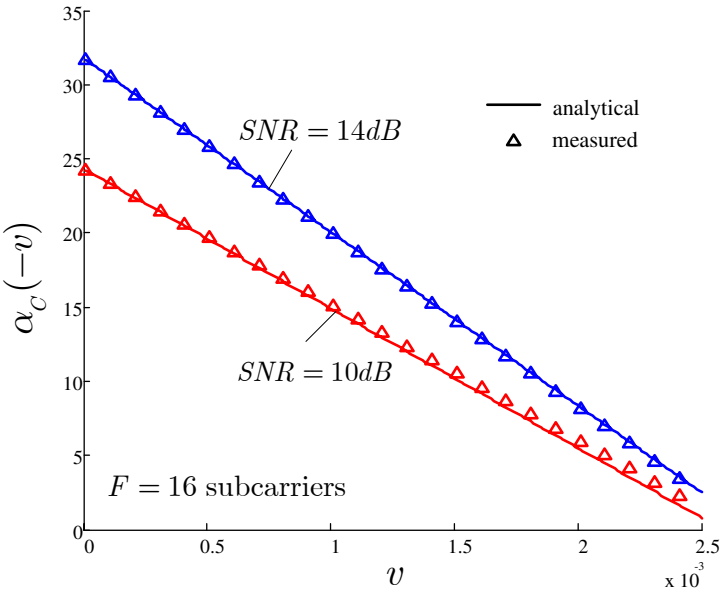


Figure 6.4: Effective bandwidth function of a frequency selective channel

6.3 Percentile of the delay in frequency selective channels

For simplicity, we rename $m_c = \frac{m_k}{kF}$ and $\sigma_c^2 = \frac{\sigma_k^2}{kF}$ and work with this notation for the mean and the variance of the channel hereinafter in the chapter.

Likewise in the flat Rayleigh channel, the percentile of the delay for constant rate traffic yields:

$$1-\varepsilon = 1 - Pr\{D(\infty) > D^t\} = 1 - e^{-\theta \cdot \lambda D^t} \tag{6.3.1}$$

with θ:

$$\lambda - \alpha_C(-\theta) = 0 \Rightarrow \theta(\lambda) \triangleq \theta(m_c, \sigma_c^2, \lambda) = \frac{2(m_c - \lambda)}{\sigma_c^2} \tag{6.3.2}$$

With a frequency selective Rayleigh channel, the mean and the variance are obtained with (6.2.4) and (6.2.8), respectively.

In Figure 6.5, the percentile of the delay is plot as a function of the source rate λ, for a given target delay and a discrete rate policy. The channel is the pedestrian channel A in Table 6.1, with a coherence bandwidth of approximately 3.4MHz. The Bessel ACF takes values $f_D T_S = 0.10$ and $f_D T_S = 0.01$. The channel with no time-correlation is also plot. The target BER is 10^{-2} and D^t is fixed to 200 symbols. The channel has average SNR 10 and 15dB. As in the flat channel, when the time-correlation increases (the frequency correlation is fixed in this Figure), the curve of the percentile decays more slowly.

The effect of the frequency correlation is illustrate in Figure 6.6. Two different PDP's are represented: the pedestrian channel of the previous Figure and a vehicular channel A as defined in Table 6.1. The Bessel ACF is set to $f_D T_S = 0.10$. The target BER is 10^{-2} and D^t is fixed to 50 symbols. The channel has average SNR 10 and 15dB. As expected, the frequency-correlation is also harmful to the delay performance and the curve of the percentile decays more slowly as the frequency-correlation increases.

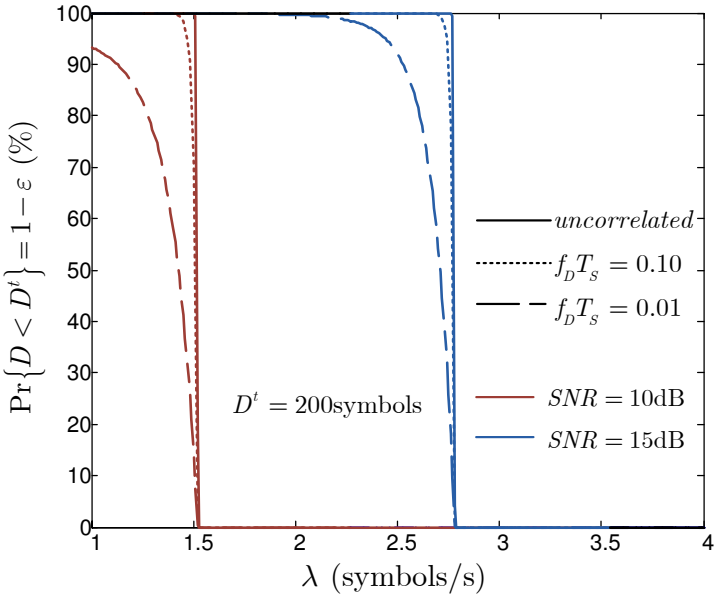


Figure 6.5: Percentile of the delay for a pedestrian channel A with Bessel ACF. Constant rate traffic and discrete rate policy. $D^t = 200$ symbols. $BER^t = 10^{-2}$.

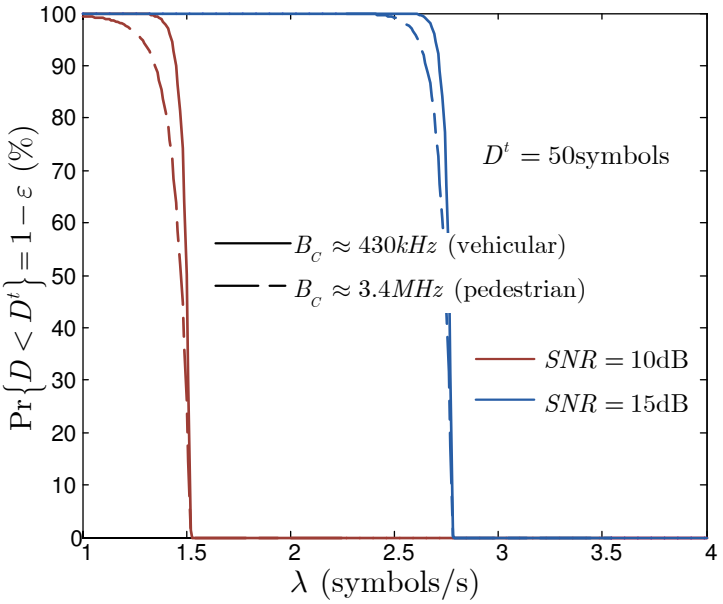


Figure 6.6: Percentile of the delay for a pedestrian channel A and a vehicular channel A. Bessel ACF with $f_D T_S = 0.10$. Constant rate traffic and discrete rate policy. $D^t = 50$ symbols. $BER^t = 10^{-2}$.

6.4 Capacity with Probabilistic Delay Constraint $\mathrm{C}_{D^t,\varepsilon}$

In this section, we derive the capacity with probabilistic delay constraint, $\mathrm{C}_{D^t,\varepsilon}$ for an OFDM system.

For simplicity, the source traffic arrives to the buffer at a constant rate, i.e. $\alpha_A(u) = \lambda$. In a high load scenario, the probability that the buffer is not empty approaches one, i.e. $\eta \to 1$. The QoS exponent is obtained by solving for a constant source and the effective bandwidth of the channel in (6.2.3):

$$\lambda - \alpha_C(-\theta) = 0 \Rightarrow \theta(\lambda) \triangleq \theta(m_c, \sigma_c^2, \lambda) = \frac{2(m_c - \lambda)}{\sigma_c^2} \tag{6.4.1}$$

The value of λ is worked out by substituting (6.4.1) into (6.1.7), and it represents the maximum source rate that the OFDM system may support with a probability ε of exceeding a delay bound D^t. The expression of the capacity with probabilistic delay constraint $\mathrm{C}_{D^t,\varepsilon}$ is:

$$\mathrm{C}_{D^t,\varepsilon} = \frac{m_c}{2} + \frac{1}{2}\sqrt{m_c^2 - 2\sigma_c^2 \frac{(-\log \varepsilon)}{D^t}} \tag{6.4.2}$$

Notice that we have obtained the same result as in the flat channel with differences in the involved parameters: the mean m_c is the average of the instantaneous OFDM rate and the variance σ_c^2 includes now not only the time-frequency but also the frequency-correlation of the channel via the expression in (6.2.8). Two foreseen limits can be checked in the capacity $\mathrm{C}_{D^t,\varepsilon}$. For high D^t values or $\varepsilon \to 1$, the QoS requirement relaxes and $\mathrm{C}_{D^t,\varepsilon}$ approaches m_c. On the other hand, as the target delay D^t or ε become lower, the wireless channel tolerates lower traffic arrival rates in order to guarantee the delay constraints. Now the dependence on the variance captures the influence of both time and frequency correlation of the channel. Thus, as correlation increases, the second addend in the square root increases in absolute value and $\mathrm{C}_{D^t,\varepsilon}$ diminishes. Finally, it is worth pointing out that the capacity

has been obtained for a constant rate source. Nevertheless, the procedure can be generalized to any other source process as it was done in Chapter 5.

Figures 6.7 and 6.8 show the evaluation of $C_{D^t,\varepsilon}$ for different values of the time correlation of the channel.

In Figure 6.7 the channel is the vehicular channel A defined in Table 6.1, with a coherence bandwidth of approximately 430kHz. The autocorrelation function of the different echoes is a first-kind Bessel function, with values of the product $f_D \cdot T_S$ of 0.05 and 0.005. The average SNR is set to 10dB and 15dB and F is set to 16. One can see that a higher value of time-correlation (smaller value of the product $f_D \cdot T_S$) implies a lower value of the capacity, as expected.

Likewise, Figure 6.8 presents the influence of the frequency correlation, by means of different values of the coherence bandwidth. In this case two different PDP are represented: the vehicular channel of the previous figure and a pedestrian channel A as defined in Table 6.1, which leads to a coherence bandwidth of $B_c \approx 430$kHz and $B_c \approx 3.4$MHz, respectively. There are 16 subcarriers and the average SNR is again set to 10dB and 15dB. The time-correlation is fixed with a product $f_D \cdot T_S$ of 0.05 in both cases. As expected, a smaller value of the coherence bandwidth (vehicular channel) corresponds to a higher value of the capacity, and the reduction is more noticeable for smaller values of the average SNR.

6.5 Simulation comparison

In this section, the results of this chapter are validated by comparison with simulations. Bits are sent to the buffer of queue length $Q[n]$ in the nth symbol. The bits in the queue are transmitted over the fading channel on a First Come First Served basis and at a service rate $c^f[n]$ at each subcarrier. Certain values of D^t and ε are fixed and the *Capacity with Probabilistic Delay Constraint* is evaluated

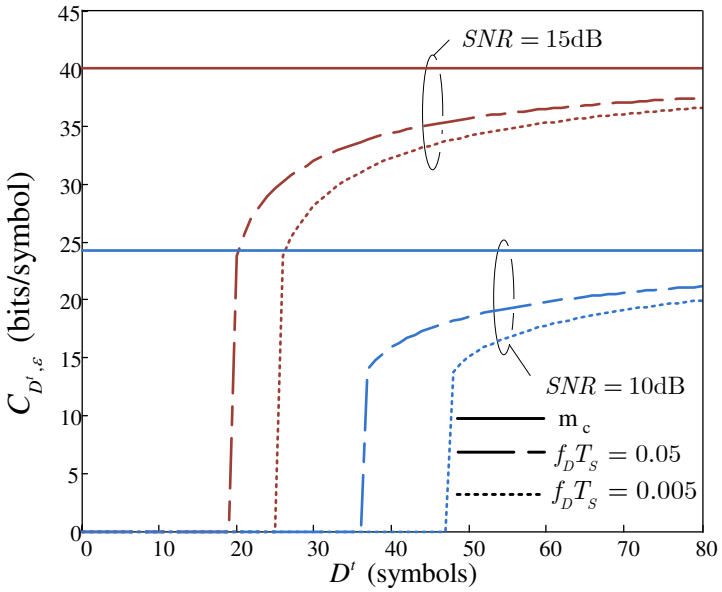


Figure 6.7: Capacity with probabilistic delay constraint for a constant rate source and 16 subcarriers. Time correlation

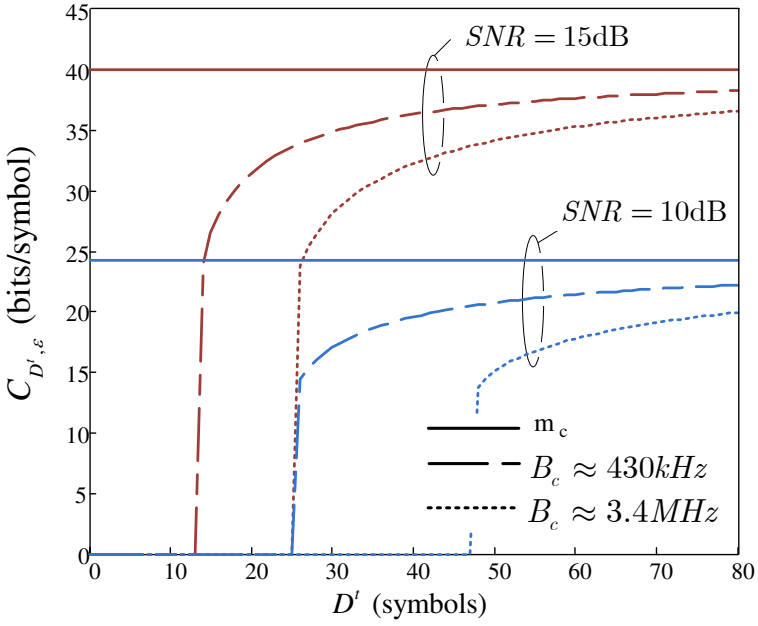


Figure 6.8: Capacity with probabilistic delay constraint for a constant rate source and 16 subcarriers. Frequency correlation

with (6.4.2). Then, the result is given in the simulator to the arrival process to generate source data at the (constant) rate $\mathrm{C}_{D^t,\varepsilon}$. The simulation is run and the tail probability of exceeding the target delay is measured based on the measurements of the delay suffered by bits leaving the queue. Notice that the expected value of this tail probability, $Pr\{D(\infty) > D^t\}$, is ε.

Figure 6.9 shows the results of the simulations. $Pr\{D(\infty) > D^t\}$ is represented for a target delay of 60 symbols. Two different simulations are presented, with different parameters of the channel and the violation probability. In the first one, the vehicular channel A with $B_c \approx 430$kHz and product $f_D \cdot T_S = 0.05$ is simulated. The violation probability is $\varepsilon = 0.1$. For the second simulation, the pedestrian channel A of $B_c \approx 3.4$MHz and $f_D \cdot T_S = 0.05$ is employed, with $\varepsilon = 0.05$. It is worth mentioning that the QoS requirements are accurately reached with the simple result for $\mathrm{C}_{D^t,\varepsilon}$ in (6.4.2): the measured violation probability is always lower than ε. The capacity was obtained under the assumption that $\eta = 1$, a high load scenario that constitutes an upper bound. It has been checked in the simulations that the measured probability of a non-empty queue η approaches one as D^t or $\overline{\gamma}$ increases. Nevertheless, the analytic result modified by the measured η is also shown in the figure with a dashed line, leading to a closer approximation to the simulations, as could be expected. It can also be checked that the results of the simulation of the frequency selective channel are not so closed to the analysis as in the case of flat channels. This is because in this case there is not only time-correlation but also frequency correlation. Thus, the length of the blocks k necessary to apply the CLT is larger and longer values of k would had been desirable in these simulations.

6.6 Summary

During the last decade, OFDM has been selected in many wireless standards as modulation scheme. In this chapter we have extended our previous analysis of the

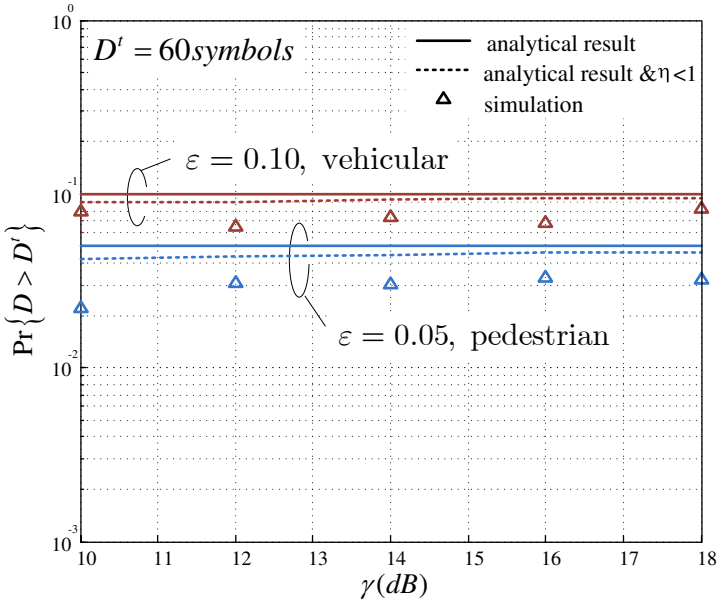


Figure 6.9: Simulation results

QoS metrics in a flat Rayleigh channel to a frequency selective channel employing OFDM as modulation scheme. Few changes have been done in the system model. Observing the new multi-channel model, it is noticed that the extension can be done by calculating the effective bandwidth function of the frequency selective channel. The result is analytically derived for a constant rate source with similar techniques as employed in the flat channel. The main difference is that in this case not only the time correlation of the channel influences the result but also de frequency correlation among subcarriers. The P-percentile of the delay and the capacity with probabilistic delay constraint are calculated under these new conditions. Figures show that the delay performance behaves as expected with regard to the frequency correlation of the channel: the higher the frequency correlation, the lower the P-percentile of the delay and the capacity with probabilistic delay constraint. Two examples of frequency selective channel are employed, corresponding to a pedestrian and a vehicular channel. Nevertheless, any other channel may be studied. Finally, the system is simulated and the results are compared to the analytical ones.

Chapter 7

Delay constrained multiuser communications

In Chapter 5 we discussed the advisability of the ergodic capacity as information-theoretic measure for delay sensitive applications in single user systems. On the other hand, delay-limited capacity becomes zero for Rayleigh channels. In this case, the Capacity with Probabilistic Delay Constraint becomes useful.

Nevertheless, in multiuser communications the capacity of the channel is no longer fully characterized by a single number. Instead, the capacity should be redefined to consider each user's data rate separately. Thus, in a system with U users, the capacity takes the form of a U dimensional vector representing rates allocated to the U users [Cover 1991]. A capacity region is then defined as the set of all U dimensional rate vectors that are achievable in the channel.

In this chapter, we propose a multi-user formulation of the previous results. Rather than the preceding Capacity with Probabilistic Delay Constraint, a closed-form expression of the vector of users' data rates is obtained. A new factor comes out in the redefinition of the system model: the scheduling strategy. Now, the attainable rate of each user depends not only on the channel and source statistics and the QoS metrics but also on the scheduling algorithm at the MAC layer. $\mathrm{R}^{u}{}_{D^t,\varepsilon}$

represents the maximum rate that the uth user can transmit by fulfilling the delay constraint (D^t, ε) and under the selected discipline. The total system capacity will be the sum of the individual users' rates, where each user can have a different delay constraint and can experience a different channel.

The results of this chapter have been partially published in [Soret 2009c] and [Soret 2010a].

The remainder of the Chapter is organized as follows. Section 7.1 describes the multiuser system model. Section 7.2 first details the derivation of the maximum users' rates subject to a delay constraint for an uncorrelated flat Rayleigh channel. Later on, the expressions are particularized to three widely employed disciplines: Round Robin, Best Channel and Proportional Fair. The procedure is repeated for a correlated single channel in Section 7.3. In Section 7.4, a multiple shared fading channel is addressed. In particular, we study an OFDMA system, where different subcarriers are assigned to different users. Finally, simulation comparison and concluding remarks are presented in Sections 7.5 and 7.6, respectively.

7.1 Multiuser system model

In multiuser communications, the channel resources can be allocated to different users in an infinite number of ways. The multiuser capacity is defined by a region that contains the set of all U dimensional rate vectors that are achievable in the channel. Each vector element represents a set of user rates that can be simultaneously supported by the channel with arbitrarily small error probability (Shannon conditions).

We consider the multiuser system model depicted in Figure 7.1, which is an extension of the system model presented in Chapter 2. A single channel is shared among U users, whose incoming traffics are characterized by U source processes.

Each user has its own queue where the data are stored before being transmitted.

Physical time is divided into units referred to as symbol periods, representing the transmission discrete time unit, n. The channel gain of each user is assumed to be constant over the symbol. The scheduler allocates resources to users in a symbol per symbol basis: every new symbol, a user is selected for transmission.

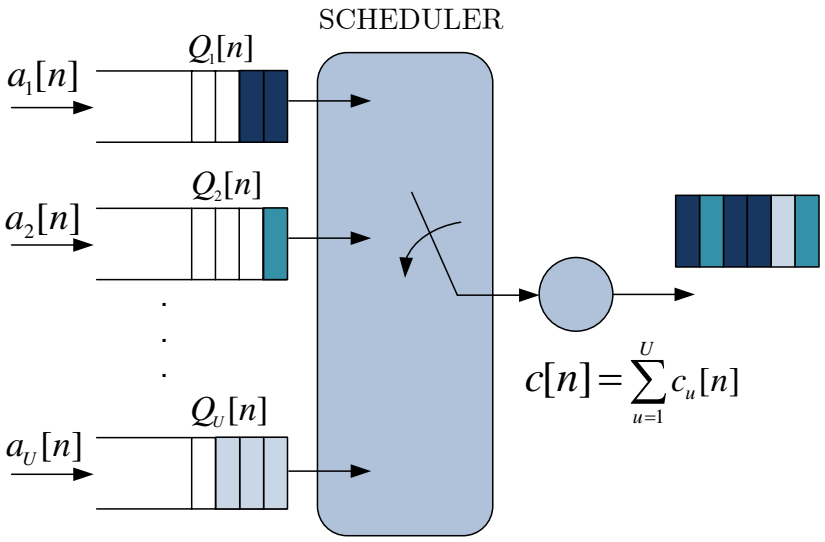


Figure 7.1: Multiuser system model

Each incoming user traffic has an instantaneous rate $a_u[n]$. On his side, the wireless channel can transmit at an instantaneous rate $c[n]$. For simplicity, the analysis is done for continuous rate policy. Each user has a potential rate $r_u[n]$, which represents the channel rate that he may use if the channel is assigned to him, and which depends on his channel conditions. Moreover, the instantaneous channel rate of user uth, $c_u[n]$, is given by:

$$c_u[n] = \begin{cases} r_u[n] \text{ if channel is assigned to user u} \\ 0 \text{ in other case} \end{cases} \tag{7.1.1}$$

Since the channel is shared among U users, $c[n]$ can be expressed:

$$c[n] = \sum_{u=1}^{U} c_u[n] \tag{7.1.2}$$

Notice that in the sum above only one of the terms is non-zero, corresponding to the user allocated to the channel.

The processes $a_u[n]$ and $c_u[n]$ are not necessarily white and represent the amount of bits per symbol generated by user u and the amount of bits per symbol of uth user transmitted by the server, respectively. In addition, the accumulated source rate $A_u[n]$ is the amount of bits generated by user u from 0 to instant $n-1$:

$$A_u[n] = \sum_{m=0}^{n-1} a_u[m] \tag{7.1.3}$$

And similarly the accumulated channel process of uth user is:

$$C_u[n] = \sum_{m=0}^{n-1} c_u[m] \tag{7.1.4}$$

The queue size is assumed to be infinite and $Q_u[n]$ denotes the length of the uth user queue at time n. The dynamics of the queueing system seen by user u is characterized by the equation $Q_u[n] = (Q_u[n-1] + a_u[n] - c_u[n])^+$. The asymptotic log-moment generating function of $Q_u[n]$ is:

$$\Lambda_u(\upsilon) = \lim_{n\to\infty} \frac{1}{n} \log \mathrm{E}\left[e^{\upsilon Q_u[n]}\right] \tag{7.1.5}$$

Since $a_u[n]$ and $c_u[n]$ are independent of each other, $\Lambda(\upsilon)$ may be decomposed into two terms, $\Lambda(\upsilon) = \Lambda_{A_u}(\upsilon) + \Lambda_{C_u}(-\upsilon)$, where $\Lambda_{A_u}(\upsilon)$ and $\Lambda_{C_u}(\upsilon)$ are the log-moment generating functions of the accumulated source process $A_u[n]$ and the accumulated channel process of user u $C_u[n]$, respectively.

When the source and the channel processes are stationary and the steady state queue length exists, then $Q_u[n]$ satisfies a LDP and the following asymptotic decay is satisfied:

$$Pr\{Q_u(\infty) > B\} \asymp e^{-\theta B} \qquad B \to \infty \tag{7.1.6}$$

with θ the solution to: $\Lambda_{A_u}(\upsilon) + \Lambda_{C_u}(-\upsilon) \mid_{\upsilon=\theta} = 0$

Defining the EBF of $A_u[n]$ and $C_u[n]$ the equation to obtain θ can be expressed as:

$$\alpha_u(\upsilon) = \alpha_{A_u}(\upsilon) - \alpha_{C_u}(-\upsilon) \mid_{\upsilon=\theta} = 0 \tag{7.1.7}$$

$\eta_u = Pr\{Q_u[n] > 0\}$ is the probability that the queue of user u is not empty. By including it in the analysis the tail probability of the queue is:

$$Pr\{Q_u(\infty) > B\} \approx \eta_u \cdot e^{-\theta B} \tag{7.1.8}$$

The delay of the bits leaving the queue of user u at symbol n is denoted as $D_u[n]$. The probability of exceeding the target delayD^t is given by:

$$\varepsilon = Pr\{D_u(\infty) > D^t\} \approx \eta_u \cdot e^{-\theta \cdot \alpha_{A_u}(\theta) D^t} = \eta_u \cdot e^{-\theta \cdot \alpha_{C_u}(-\theta) D^t} \tag{7.1.9}$$

7.2 Uncorrelated channel

We start the analysis with the case of users experiencing an uncorrelated Rayleigh channel.

7.2.1 Achievable users' rates with a delay constraint

The Capacity with Probabilistic Delay Constraint in a single-user system under uncorrelated Rayleigh channel was detailed in Chapter 5. On the basis of that result, the extension to a multiuser system is done here. Now, each user can have his own delay constraint[1].

Like in the single user system, two functions are necessary: the effective bandwidth of the source, $\alpha_{A_u}(\upsilon)$, and the channel, $\alpha_{C_u}(\upsilon)$.

[1]Nevertheless, for the sake of simplicity we maintain the pair (D^t, ε) to denote the delay constraint of user uth.

For simplicity, assume that the source traffic from the uth user arrives to the buffer at a constant rate, $a_u[n] = \lambda_u$, leading to a constant EBF: $\alpha_{A_u}(\upsilon) = \lambda_u$.

Every user experiences a flat Rayleigh channel. Furthermore users are independent among them, i.e. the channel gain seen by one user is independent of the rest. Let us define $\gamma_u[n]$ as the instantaneous Signal to Noise Ratio of user u at the receiver. With Additive White Gaussian Noise, $\gamma_u[n]$ is exponentially distributed:

$$f(\gamma) = \frac{1}{\overline{\gamma}_u} e^{-\frac{\gamma_u}{\overline{\gamma}_u}} \tag{7.2.1}$$

where $\overline{\gamma}_u$ is the average Signal to Noise Ratio of user u.

$\gamma_u[n]$ is proportional to the square of $|h_u[n]|$, the envelope of the complex channel gain of user u, $\gamma_u[n] = |h_u[n]|^2 \frac{E_s}{N_0}$, with E_s the average energy per symbol and N_0 the noise power spectral density.

The analysis of the multiuser system will be done assuming constant transmitted power and a continuous rate policy, i.e. the channel rate takes any positive real number.

Then, the potential channel rate of user u, $r_u[n]$, is a function of $\gamma_u[n]$:

$$r_u[n] = \log_2 \left(1 + \beta_u \gamma_u[n]\right) \tag{7.2.2}$$

where β_u, under adaptive modulation, is a constant related to the target BER. Its value for uncoded QAM is [Chung 2001] : $\beta_u \approx \frac{1.6}{-\log(5BER^t)}$, where BER^t is the target BER. Furthermore, the value $\beta_u = 1$ represents the upper bound corresponding to the evaluation of the AWGN channel capacity (in Shannon's sense).

With no time-correlation among samples, the accumulated transmission rate for the uth user $C_u[n]$ is simply the addition of n uncorrelated and identically distributed random variables. As $n \to \infty$ the CLT can be applied and $C_u[n]$ is a Gaussian random variable with average $n \cdot m_u$ and variance $n \cdot \sigma_u^2$, where m_u and σ_u^2 are the mean and the variance of the instantaneous channel rate for the uth user $c_u[n]$.

Then, the effective bandwidth function for the resulting Gaussian distribution of $C_u[n]$ is computed as:

$$\alpha_{C_u}(\upsilon) = \lim_{n\to\infty} \frac{1}{n \cdot u} \log \mathrm{E}\left[e^{\upsilon C_u[n]}\right] = m_u + \frac{u}{2}\sigma_u^2 \tag{7.2.3}$$

In a high load scenario, the probability that the buffer is not empty approaches one, i.e. $\eta_u \to 1$. Under this assumption, the delay constraint is worked out from (7.1.9):

$$-\frac{log(\varepsilon)}{D^t} = \theta \cdot \alpha_{A_u}(\theta) \tag{7.2.4}$$

Assume that the uncorrelated channel has parameters m_u and σ_u^2, then the QoS exponent is obtained by solving (7.1.7) for a constant source:

$$\lambda_u - \alpha_{C_u}(-\theta) = 0 \Rightarrow \theta(\lambda_u) \triangleq \theta(m_u, \sigma_u^2, \lambda_u) = \frac{2(m_u - \lambda_u)}{\sigma_u^2} \tag{7.2.5}$$

With (7.2.5) substituted into (7.2.4), the value of λ_u is worked out and it is the achievable user rate that we were seeking. It represents the maximum source rate of user u that may be supported with a probability ε of exceeding a certain delay bound D^t. That's the rate of uth user $\mathrm{R}^u{}_{D^t,\varepsilon}$:

$$\mathrm{R}^u{}_{D^t,\varepsilon} = \frac{m_u}{2} + \frac{1}{2}\sqrt{m_u^2 - 2\sigma_u^2 \frac{(-\log \varepsilon)}{D^t}} \tag{7.2.6}$$

The vector of achievable users' rates is:

$$\mathbf{R}_{\mathbf{D}^t,\boldsymbol{\epsilon}} = (\mathrm{R}^1{}_{D^t,\varepsilon}, ..., \mathrm{R}^U{}_{D^t,\varepsilon}) \tag{7.2.7}$$

with $\mathbf{D}^t$ and $\boldsymbol{\epsilon}$ the vectors with the target delays and probabilities of violation of each user, respectively.

From (7.2.6) it can be observed that the evaluation of the user rate comes down to obtaining the mean and the variance of the channel process seen by user u:

$$m_u = \mathrm{E}\left[c_u[n]\right] \tag{7.2.8}$$

$$\sigma_u^2 = \mathrm{E}\left[c_u^2[n]\right] - m_u^2 \tag{7.2.9}$$

It should be emphasized that these statistics depend on the distribution of $c_u[n]$ which in turn depends on the scheduling algorithm. Thus, the discipline determines the maximum users' rates, as expected. These maximum rates will be detailed in next sections for several scheduling disciplines.

It can be observed from (7.2.6) that for high D^t values or $\varepsilon \to 1$, the QoS requirement relaxes and $\mathrm{R}^u{}_{D^t,\varepsilon}$ approaches m_u. On the other hand, as the target delay D^t or ε become lower, the user has to transmit at a lower rate in order to guarantee his own delay constraint.

Finally, the total system capacity $\mathbf{C}_{\mathbf{D}^t,\boldsymbol{\epsilon}}$ is obtained as the sum of the individual user rates, each of them with his own delay constraint:

$$\mathbf{C}_{\mathbf{D}^t,\boldsymbol{\epsilon}} = \sum_{u=1}^{U} \mathrm{R}^u{}_{D^t,\varepsilon} \tag{7.2.10}$$

7.2.2 Round Robin

First of all, the mean and the variance to compute $\mathbf{R}_{\mathbf{D}^t,\boldsymbol{\epsilon}}$ under a Round Robin strategy are calculated.

Round Robin (RR) is a fixed cyclic algorithm without priorities, which dispenses the channel equally among the different flows independently of their priorities or radio channel conditions. Transmission at symbol n is assigned to the following user in a cyclic order and therefore:

$$c_u[n] = \begin{cases} \log_2(1+\beta_u\gamma_u[n]) & \text{if} \quad \mod(n,U) = u \\ 0 \text{ in other case} \end{cases} \tag{7.2.11}$$

It is known that this strategy does not work well over varying channels and a low efficiency in terms of system capacity and QoS differentiation is expected.

The mean and the variance of $c_u[n]$ are required. With only one user ($U = 1$) and continuous rate policy, the mean $c_u[n]$ matches up with the mean given in Section 2, equation (4.1.9). We denote it by m_1:

$$m_1 = \mathrm{E}\left[\log_2(1+\beta\gamma)\right] = \log_2(e)\exp(\frac{1}{\beta\bar{\gamma}})E_1(\frac{1}{\beta\bar{\gamma}}) \tag{7.2.12}$$

where $E_1(x)$ is the exponential integral and $\overline{\gamma}$ is the average Signal to Noise Ratio of the single user.

Likewise, the expression of the variance σ_1^2 with only one user is (equation 4.2.8):

$$\begin{aligned}\sigma_1^2 &= \mathrm{E}\left[(\log_2(1+\beta\gamma))^2\right] - m_1^2 \\ &= (\log_2(e))^2 e^{\frac{1}{\beta\bar{\gamma}}}\left[\frac{\pi^2}{6} + g^2 + 2g\log\left(\frac{1}{\beta\bar{\gamma}}\right) + \log^2\left(\frac{1}{\beta\bar{\gamma}}\right)\right. \\ &\left. - 2\left(\frac{1}{\beta\bar{\gamma}}\right){}_3\mathbf{F}_3\left([1,1,1],[2,2,2],-\frac{1}{\beta\bar{\gamma}}\right) - e^{\frac{1}{\beta\bar{\gamma}}}\mathrm{E}_1^2(\frac{1}{\beta\bar{\gamma}})\right]\end{aligned} \tag{7.2.13}$$

where g is the Euler constant and ${}_p\mathrm{F}_q(\boldsymbol{n}, \boldsymbol{d}, z)$ is the hypergeometric function.

When U users share the channel under a RR discipline, we only need to take into account that the channel is equally divided among users. Thus, the expressions of m_u and σ_u^2 are written directly from m_1 and σ_1^2, by just replacing with the average Signal to Noise Ratio of each user and dividing by the number of users:

$$m_u = \mathrm{E}\left[\log_2(1+\beta_u\gamma_u)\right] = \frac{1}{U}\log_2(e)\exp(\frac{1}{\beta_u\bar{\gamma}_u})E_1(\frac{1}{\beta_u\bar{\gamma}_u}) \tag{7.2.14}$$

$$\begin{aligned}\sigma_u^2 &= \mathrm{E}\left[(\log_2(1+\beta_u\gamma_u))^2\right] - m_u^2 \\ &= \frac{1}{U^2}(\log_2(e))^2 e^{\frac{1}{\beta_u\bar{\gamma}_u}}\left[\frac{\pi^2}{6} + g^2 + 2g\log\left(\frac{1}{\beta_u\bar{\gamma}_u}\right) + \log^2\left(\frac{1}{\beta_u\bar{\gamma}_u}\right)\right. \\ &\left. - 2\left(\frac{1}{\beta_u\bar{\gamma}_u}\right){}_3\mathbf{F}_3\left([1,1,1],[2,2,2],-\frac{1}{\beta_u\bar{\gamma}_u}\right) - e^{\frac{1}{\beta_u\bar{\gamma}_u}}\mathrm{E}_1^2(\frac{1}{\beta_u\bar{\gamma}_u})\right]\end{aligned} \tag{7.2.15}$$

The expressions above make it possible to evaluate the vector of users' rates in (7.2.7) under a Round Robin discipline.

An example is shown in Figure 7.2. Three users have been considered, with average SNR 5, 7 and 12 dB respectively. The individual rates $\mathrm{R}^{u}{}_{D^t,\varepsilon}$ are plot as a function of the target delay D^t. The other parameter in the delay constraint, the violation probability ε, has been set to 0.1 for all users. β_u is set to 1[2]. The mean m_u of each user is represented with solid line, whereas the dashed line is $\mathrm{R}^{u}{}_{D^t,\varepsilon}$. Both m_u and $\mathrm{R}^{u}{}_{D^t,\varepsilon}$ are plotted for each user (users marked with triangles, squares and circles). Moreover, the system capacity normalized with the number of users, corresponding to the *average* rate, is represented with no marks and thicker line.

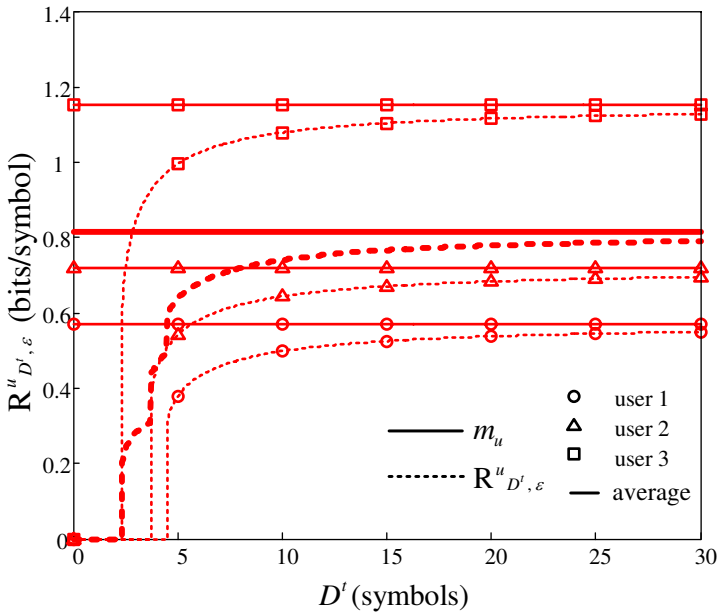


Figure 7.2: Achievable users' rates with Round Robin scheduling in an uncorrelated channel.

The behaviour of $\mathrm{R}^{u}{}_{D^t,\varepsilon}$ for all the user is analogous to the behaviour of the Capacity with Probabilistic Delay Constraint in a single user system. Observing

[2] β_u is set to 1 in all the numerical results hereinafter in the chapter, i.e., the upper bound of the users' rates are evaluated.

the differences among users, those with better channel conditions obtain higher rates and can demand stringent QoS conditions, as it was expected. Thus, the best user in this example could fix a delay constraint with a target delay of 2 symbols in contrast to the 4 symbols of the worst user. On the other hand, if the same target delay is fixed for the three users the rate to be employed increases for *better* users (users with better channel conditions). For example, for $D^t = 10$ symbols user 1 can transmit 0.5 bits/symbol, user 2, 0.7 bits/symbol and user 3, 1.1 bits/symbol.

7.2.3 Best Channel

Best Channel (BC) strategy is adaptive to the channel state, giving priority to those users with higher potential transmission rate. The channel is assigned to the user that may transmit with the highest number of bits per symbol:

$$c_u[n] = \begin{cases} \log_2(1+\beta_u\gamma_u[n]) & \text{if } \gamma_u[n] > \gamma_k[n] \ \ \forall k \neq u \\ 0 & \text{in other case} \end{cases} \tag{7.2.16}$$

This algorithm maximizes the total system efficiency. However, under this strategy good average SNR users get more average throughput than low SNR users.

Let us define γ_{max}:

$$\gamma_{max} = \max_u \{\gamma_u\} \tag{7.2.17}$$

The CDF of γ_{max} can be written:

$$F_{\gamma_{max}}(\gamma) = Pr(\gamma_{max} < \gamma) = Pr(\gamma_1 < \gamma, \gamma_2 < \gamma, ..., \gamma_U < \gamma) \tag{7.2.18}$$

Since the users are i.i.d. it comes down to:

$$F_{\gamma_{max}}(\gamma) = \prod_{u=1}^{U} (1 - \exp(-\gamma/\bar{\gamma}_u)) \tag{7.2.19}$$

Consider the following effective SNR for the uth user:

$$\gamma_u^* = \begin{cases} \gamma_u, & \gamma_u > \gamma_{-u} \\ 0, & \gamma_u < \gamma_{-u} \end{cases} \tag{7.2.20}$$

where $\gamma_{-u} = \max_{k \neq u} \{\gamma_k\}$ is the maximum of the average SNR of all the users except u.

The pdf of the effective SNR γ_u^* can be expressed as follows [Pérez 2007]:

$$f_u^*(\gamma_u^*) = \text{Prob}\{\gamma_u < \gamma_{-u}\}\delta(\gamma_u^*) + f_u(\gamma_u^*)F_{-u}(\gamma_u^*) \tag{7.2.21}$$

where:

- $\delta(x)$ is the Dirac delta function
- $f_u(x)$ is the exponential pdf in (7.2.1)
- $F_{-u}(x)$ is the CDF of γ_{-u}, which can be written:

$$F_{-u}(x) = \prod_{k \neq u}^{U} [1 - \exp(-\frac{x}{\bar{\gamma}_k})] = \sum_{\mathbf{i} \in \mathfrak{U}} (-1)^{\mathbf{i} \cdot \mathbf{1}} (1 - i_u) \exp(-x\mathbf{b} \cdot \mathbf{i}) \tag{7.2.22}$$

- $\mathbf{b} = \left[\frac{1}{\bar{\gamma}_1} \frac{1}{\bar{\gamma}_2} ... \frac{1}{\bar{\gamma}_U}\right]$
- $\mathbf{1}$ denotes the all-ones U-dimensional vector
- $\mathfrak{U}$ is the set of all U-dimensional vectors with entries taking values 0 of 1, i.e., $\mathfrak{U}$ contains the 2^U binary words of length U
- i_u is the uth component of $\mathbf{i}$

The mean m_u to be computed for BC is:

$$m_u = \text{E}[c(\gamma_u)] = \int_0^\infty c(\gamma_u^*) f_u^*(\gamma_u^*) d\gamma_u^* \tag{7.2.23}$$

From (7.2.21) it can be observed that the first addend will be zero in the required expectation and only the second term needs to been integrated:

$$f_u(\gamma_u^*)F_{-u}(\gamma_u^*) = -\sum_{\mathbf{i} \in \mathfrak{U}} (-1)^{\mathbf{i} \cdot \mathbf{1}} \frac{i_u}{\bar{\gamma}_u} \exp(-x\mathbf{b} \cdot \mathbf{i}) \tag{7.2.24}$$

Substituting into the mean it yields:

$$m_u = -\int_0^{\infty} \log_2(1+\beta_u\gamma_u^*) \sum_{\mathbf{i}\in\mathfrak{U}} (-1)^{\mathbf{i}\cdot\mathbf{1}} \frac{i_u}{\overline{\gamma}_u} \exp(-\gamma_u^* \mathbf{b}\cdot\mathbf{i}) d\gamma_u^* \tag{7.2.25}$$

This integral is analogous to the single user case by simply defining $\frac{1}{\bar{\gamma}} = \mathbf{b}\cdot\mathbf{i}$. The result is then:

$$m_u = -\sum_{\mathbf{i}\in\mathfrak{U}} (-1)^{\mathbf{i}\cdot\mathbf{1}} \frac{i_u}{\overline{\gamma}_u \mathbf{b}\cdot\mathbf{i}} \log_2(e) \exp\left(\frac{\mathbf{b}\cdot\mathbf{i}}{\beta_u}\right) E_1\left(\frac{\mathbf{b}\cdot\mathbf{i}}{\beta_u}\right) \tag{7.2.26}$$

Likewise, the calculation of the variance is similar to the single user case, obtaining:

$$\begin{aligned}
\sigma_u^2 &= \mathrm{E}\left[\log_2^2(1+\beta_u\gamma)\right] - m_u^2 \\
&= -\sum_{\mathbf{i}\in S} (-1)^{\mathbf{i}\cdot\mathbf{1}} \frac{i_s}{\bar{\gamma}_u \mathbf{b}\cdot\mathbf{i}} (\log_2(e))^2 e^{\frac{1}{\beta_u}\mathbf{b}\cdot\mathbf{i}} \\
&\left[\frac{\pi^2}{6} + g^2 + 2g\ln\left(\frac{1}{\beta_u}\mathbf{b}\cdot i\right) + \ln^2\left(\frac{1}{\beta_u}\mathbf{b}\cdot\mathbf{i}\right)\right. \\
&\left. -2\left(\frac{1}{\beta_u}\mathbf{b}\cdot\mathbf{i}\right) {}_3\mathbf{F}_3\left([1,1,1],[2,2,2],-\frac{1}{\beta_u}\mathbf{b}\cdot\mathbf{i}\right)\right]
\end{aligned} \tag{7.2.27}$$

The same evaluation example presented for RR is shown in Figure 7.3, now for BC allocation. It is then three users with average SNR $5, 7$ and 12 dB respectively.

The differences among users are much more noticeable than for RR. Thus, the best user is better off with the change to BC allocation at the expenses of users with lower average SNR. Notice that not only the differences in the mean m_u are remarkable (the asymptotic behaviour when relaxing the QoS constraint) but also the minimum target delays of each user move away. For example, the worst user cannot demand a target delay below 45 symbols for these channel conditions and scheduling, in contrast to the 2 symbols of the best user. As expected, the average rate is higher than for RR, since this algorithm maximizes the total system efficiency.

It is well-known that by exploiting the multiuser diversity one can achieve higher system capacity as the number of users increases. This multiuser diversity gain is

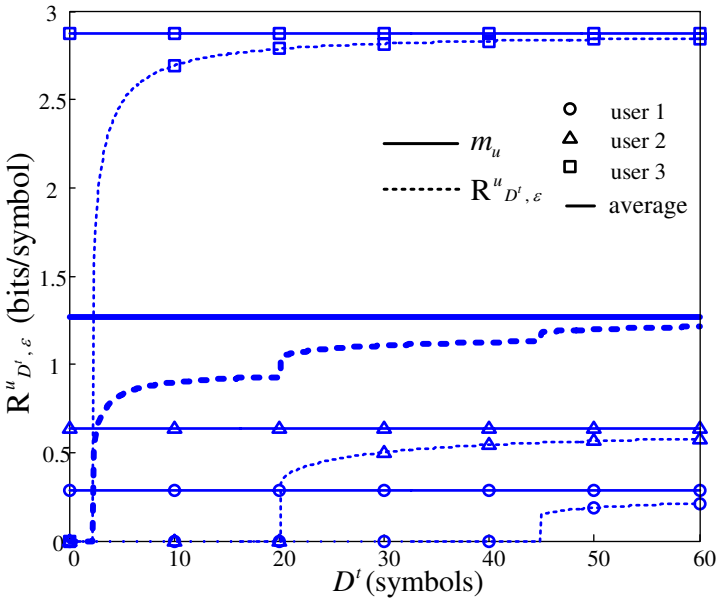


Figure 7.3: Achievable users' rates with Best Channel scheduling in an uncorrelated channel.

illustrated in Figure 7.4. Lognormal shadowing is considered, so that the average SNR of users follows a lognormal distribution, with average 10dB and standard deviation 4dB. The violation probability is 0.1 for all users. The maximum achievable rate of the median user is plot, for 4, 7 and 10 users. It can be observed that as the number of users increases, the maximum achievable rate of the median user increases, due to multiuser diversity.

7.2.4 Proportional Fair

Proportional Fair (PF) is a compromise based scheduling algorithm. It is intended to improve Best Channel by maintaining a balance between two competing interests: trying to maximize the total throughput while allowing a minimum level of service to all users. Fair sharing will lower the total throughput over the maximum possible, but it will provide more acceptable levels to users with poorer SNR. Instead of using the instantaneous potential transmission rate of BC, PF uses as metrics the ratio

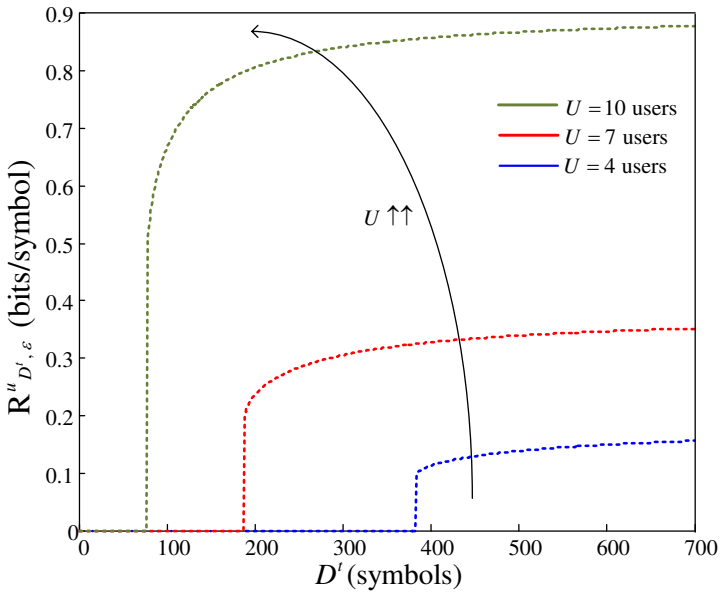


Figure 7.4: Multiuser diversity for BC and uncorrelated channel: maximum achievable rate of the median user. Average SNR following lognormal shadowing with mean 10dB and standard deviation 4dB. $\varepsilon = 0.1$. $\beta_u = 1$

$\gamma_u[n]/\bar{\gamma}_u$:

$$c_u[n] = \begin{cases} \log_2(1+\beta_u\gamma_u[n]) & \text{if } \gamma_u[n]/\bar{\gamma}_u > \gamma_k[n]/\bar{\gamma}_k \ \ \forall k \neq u \\ 0 & \text{in other case} \end{cases} \tag{7.2.28}$$

Let us define Γ_{max}:

$$\Gamma_{max} = \max_u \left\{ \frac{\gamma_u}{\bar{\gamma}_u} \right\} \tag{7.2.29}$$

The effective SNR for the uth user is:

$$\Gamma_u^* = \begin{cases} \gamma_u, & \frac{\gamma_u}{\bar{\gamma}_u} > \frac{\gamma_{-u}}{\bar{\gamma}_{-u}} \\ 0, & \frac{\gamma_u}{\bar{\gamma}_u} < \frac{\gamma_{-u}}{\bar{\gamma}_{-u}} \end{cases} \tag{7.2.30}$$

With the change of variable, now the second term of the pdf of Γ_u^* is expressed:

$$\bar{\gamma}_u f_u(x) F_{-u}(x\bar{\gamma}_u) = -\bar{\gamma}_u \cdot \sum_{\mathbf{i}\in\mathfrak{U}} (-1)^{\mathbf{i}\cdot\mathbf{1}} \frac{i_u}{\bar{\gamma}_u} \exp(-\frac{x}{\bar{\gamma}_u}\mathbf{1}) \tag{7.2.31}$$

with $\mathbf{1}$ the all-ones U-dimensional vector.

The result of the mean m_u is analogous to the BC:

$$m_u = -\sum_{\mathbf{i}\in\mathfrak{U}} (-1)^{\mathbf{i}\cdot\mathbf{1}} \frac{i_u}{\mathbf{1}} \log_2(e) \exp(\mathbf{q})\, E_1(\mathbf{q}) \tag{7.2.32}$$

Likewise, the next result is obtained for the variance:

$$\begin{aligned}\sigma_u^2 &= \mathrm{E}\left[\log_2^2(1+\beta_u\gamma)\right] - m_u^2 \\ &= -\sum_{\mathbf{i}\in S} (-1)^{\mathbf{i}\cdot\mathbf{1}} \frac{i_u}{\mathbf{q}} (\log_2(e))^2 e^{\mathbf{q}} \\ &\left[\frac{\pi^2}{6} + g^2 + 2g\ln(\mathbf{q}) + \ln^2(\mathbf{q})\right. \\ &\left. -2(\mathbf{q})\, F([1,1,1],[2,2,2],-\mathbf{q})\right] \qquad (7.2.33) \\ & \qquad (7.2.34)\end{aligned}$$

where $\mathbf{q}$ is a row vector of length U whose entries take the form $\frac{1}{\bar{\gamma}_u\beta_u}$

In Figure 7.5, the maximum achievable users' rates are evaluated under the same conditions as it was done with RR and BC, i.e., three users with average SNR $5, 7$ and 12 dB.

It can be observed that the differences among users reduce if we compare with the BC strategy. That is exactly the goal of this discipline: to maintain a balance between the total throughput and the level of service of all users. Obviously, the average rate reduces to increase the fairness. The achieved fairness is specially noticeable in the behaviour of the target delay, which is 10 symbols for the three users.

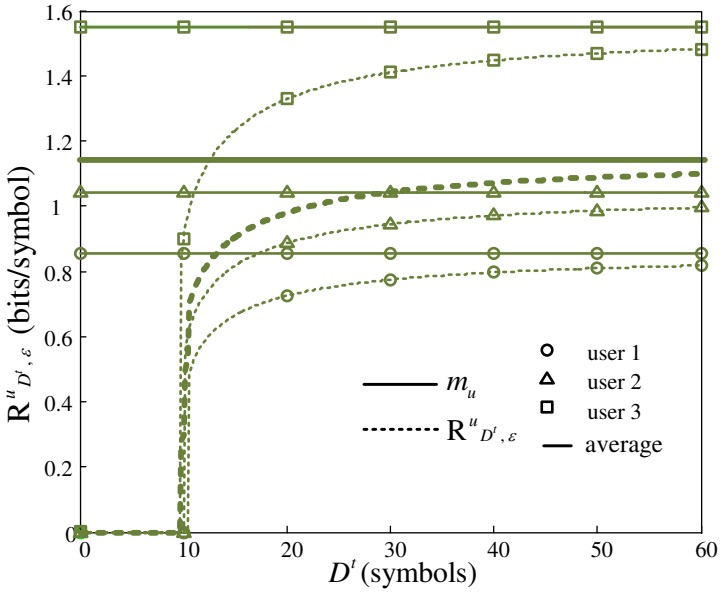


Figure 7.5: Achievable users' rates with Proportional Fair scheduling in an uncorrelated channel.

7.3 Time-correlated channel

7.3.1 Achievable users' rates with a delay constraint

To face the time-correlated Rayleigh channel, we follow the same strategy as in the single user system: to split the accumulated transmission rate for the uth user into b blocks of length k, with k large enough to assume independence among blocks.

Thus, $C_u[n]$ can be considered a Gaussian random variable with average $b \cdot m_{k_u}$ and variance $b \cdot \sigma^2_{k_u}$, where m_{k_u} and $\sigma^2_{k_u}$ are the mean and the variance of a block of size k of the uth user. The effective bandwidth function of the Gaussian distribution of $C_u[n]$ yields:

$$\alpha_{C_u}(\upsilon) = \lim_{n\to\infty} \frac{1}{n \cdot \upsilon} \log \mathrm{E}\left[e^{\upsilon C_u[n]}\right] = \frac{m_{k_u}}{k} + \frac{\upsilon}{2}\frac{\sigma^2_{k_u}}{k} \tag{7.3.1}$$

And the achievable rate of uth user is:

$$\mathrm{R}^{u}{}_{D^t,\varepsilon} = \frac{m_{k_u}}{2k} + \frac{1}{2}\sqrt{\frac{m_{k_u}^2}{k^2} - 2\frac{\sigma_{k_u}^2}{k}\frac{(-\log\varepsilon)}{D^t}} \tag{7.3.2}$$

7.3.2 Round Robin

The extension of the results of the RR discipline to a time-correlated channel is straightforward.

With only one user ($U = 1$) and continuous rate policy, the mean and variance of the blocks were given in Chapter 4. For the sake of simplicity, we reproduce them here and rename them as m_{k_1} and $\sigma_{k_1}^2$:

$$m_{k_1} = \mathrm{E}[C_i[k]] = k \cdot m_1 \tag{7.3.3}$$

$$\sigma_{k_1}^2 = \sum_{q=0}^{k-1}\sum_{r=0}^{k-1} \mathcal{K}_c(r-q) \tag{7.3.4}$$

$$\mathcal{K}_c(m) = \mathrm{E}[c[n]c[n+m]] - m_c^2 \tag{7.3.5}$$

$$\mathrm{E}\left[c[n]c[n+m]\right] = \frac{a}{\bar{\gamma}}\sum_{\mathrm{p}\,=\,0}^{\infty}\left(I_p(\beta, \mathcal{R}_z(m), a)\right)^2 \tag{7.3.6}$$

$$a = \frac{1}{(1-\mathcal{R}_z^2(m))\cdot\overline{\gamma}} \tag{7.3.7}$$

$$I_p(\beta, \mathcal{R}_z(m), a) = \frac{\mathcal{R}_z^p(m)}{p\cdot\log(2)}\left(-\frac{p}{a}\left(-\psi(1+p) + \log(a)\right) + {}_2\mathbf{F}_2([1,1],[2,1-p],a)\right) \tag{7.3.8}$$

When U users share the channel under a RR discipline, the channel is equally divided among users. Like in the uncorrelated channel, the expressions of m_{k_u} and $\sigma^2_{k_u}$ are written directly as a function of m_{k_1} and $\sigma^2_{k_1}$, by just replacing $\bar{\gamma}$ with $\bar{\gamma}_u$, β with β_u and dividing by the number of users:

$$m_{k_u} = \frac{m_{k_1}(\bar{\gamma}_u, \beta_u)}{U} \tag{7.3.9}$$

$$\sigma_{k_u} = \frac{1}{U}\sigma_{k_1}(\bar{\gamma}_u, \beta_u) \tag{7.3.10}$$

The evaluation of RR in a correlated channel is presented in Figure 7.6. There are three users with average SNR 5, 7 and 12 dB, $\varepsilon = 0.01$ and $\beta_u = 1$ for the three users. The correlation follows an exponential decay and two values of ρ are evaluated, 0.8 and 0.9. The predecent result for uncorrelated channel is also shown with black line. The qualitative behaviour is the same as in the uncorrelated channel. It is remarkable the influence of the parameter ρ: as expected, the correlation is harmful to the delay performance, so that when the correlation increases, the achievable user rate decrease. Moreover, it is observed that the differences among users increases with the time-correlation.

7.3.3 Best Channel

Similarly as done in RR, the calculation of the maximum attainable rates in the case of Best Channel strategy leads to the computation of the variance of the blocks (the evaluation of the mean of the blocks is straightforward), which comes down to the evaluation of $E[c_u[n]c_u[n+m]]$:

$$\sigma^2_{k_u} = \sum_{q=0}^{k-1}\sum_{r=0}^{k-1}\mathcal{K}_{c_u}(r-q) \tag{7.3.11}$$

$$\mathcal{K}_{c_u}(m) = \mathrm{E}[c_u[n]c_u[n+m]] - m_u^2 \tag{7.3.12}$$

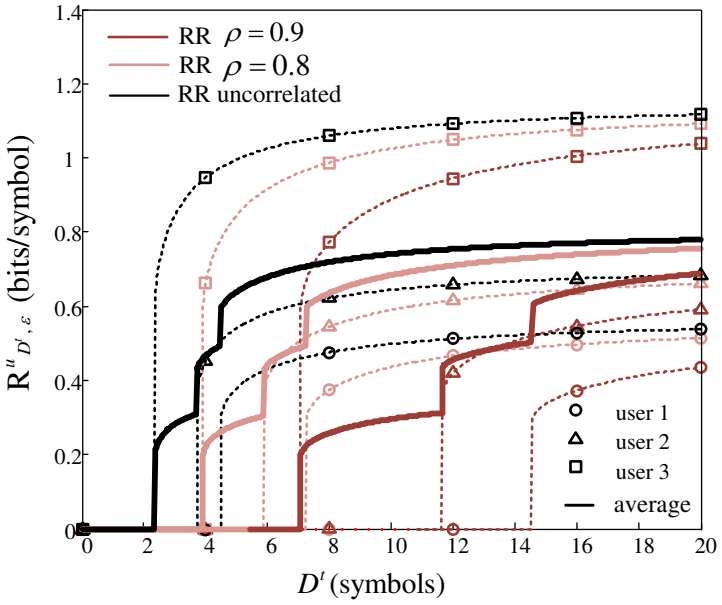


Figure 7.6: Achievable users' rates with Round Robin scheduling in a time-correlated channel.

with m_u as in (7.2.23).

The joint pdf and CDF of two correlated Rayleigh variates are needed. We have the expressions in terms of the envelope of the channel response $z_n = |h[n]|$ [Simon 2005].

For normalized power, the pfd is given by:

$$f_{\mathbf{z}}(z_n, z_{n+m}) = \frac{4 z_n z_{n+m}}{(1 - \mathcal{R}_z(m))} \exp\left(-\frac{z_n^2 + z_{n+m}^2}{(1 - \mathcal{R}_z(m))}\right) \cdot I_0\left(\frac{2\sqrt{\mathcal{R}_z(m)} z_n z_{n+m}}{(1 - \mathcal{R}_z(m))}\right) \tag{7.3.13}$$

And the CDF of the envelope of the channel is:

$$\begin{aligned} F_{\mathbf{z}}(z_n, z_m) = 1 &- \exp\left(-z_n^2\right) Q_1\left(\sqrt{\frac{2}{(1 - \mathcal{R}_z(m))}} z_m, \sqrt{\frac{2\mathcal{R}_z(m)}{(1 - \mathcal{R}_z(m))}} z_n\right) \\ &- \exp\left(-z_m^2\right)\left[1 - Q_1\left(\sqrt{\frac{2\mathcal{R}_z(m)}{(1 - \mathcal{R}_z(m))}} z_m, \sqrt{\frac{2}{(1 - \mathcal{R}_z(m))}} z_n\right)\right] \end{aligned} \tag{7.3.14}$$

where $Q_1(a,b)$ is the Marcum Q function.

Let the effective envelope of the uth user be:

$$z^*_{n_u} = \begin{cases} z_u[n], & z_u[n]\cdot\bar{\gamma}_u > z_{-u}[n]\cdot\bar{\gamma}_{-u} \\ 0, & z_u[n]\cdot\bar{\gamma}_u > z_{-u}[n]\cdot\bar{\gamma}_{-u} \end{cases} \tag{7.3.15}$$

where $z_{-u}[n] = \max_{k\neq u}\{z_u[n]\}$.

This random variable, equivalent to the effective SNR defined in the uncorrelated channel, indicates the fact that the user only gets the channel if his instantaneous SNR is the highest among all the users. In contrast to the uncorrelated channel, we include the time through the subindex n as we need to calculate the expectation evaluated in two different symbols.

Consider the following vector of decision:

$$(z^*_{n_u}, z^*_{m_u}) = \begin{cases} (0,0) & z_u[n]\bar{\gamma}_u < z_{-u}[n]\bar{\gamma}_{-u} \;\; and \;\; z_u[m]\bar{\gamma}_u < z_{-u}[m]\bar{\gamma}_{-u} \\ (z_u[n],0) & z_u[n]\bar{\gamma}_u > z_{-u}[n]\bar{\gamma}_{-u} \;\; and \;\; z_u[m]\bar{\gamma}_u < z_{-u}[m]\bar{\gamma}_{-u} \\ (0,z_u[m]) & z_u[n]\bar{\gamma}_u < z_{-u}[n]\bar{\gamma}_{-u} \;\; and \;\; z_u[m]\bar{\gamma}_u > z_{-u}[m]\bar{\gamma}_{-u} \\ (z_u[n],z_u[m]) & z_u[n]\bar{\gamma}_u > z_{-u}[n]\bar{\gamma}_{-u} \;\; and \;\; z_u[m]\bar{\gamma}_u > z_{-u}[m]\bar{\gamma}_{-u} \end{cases} \tag{7.3.16}$$

Notice that to calculate the expectation $E[c_u[n]c_u[n+m]]$ only the last case is needed, as the other three options will result in zero in the evaluation of $E[c_u[n]c_u[n+m]]$. Therefore, only the case in which the channel is assigned to user u in both symbols n and m is required. With the same change of variable done in the uncorrelated PF, the joint pdf is:

$$\bar{\gamma}_u f_u(z^*_{n_u}, z^*_{m_u}) F_{-u}(z^*_{n_u}\cdot\bar{\gamma}_u, z^*_{m_u}\cdot\bar{\gamma}_u) \tag{7.3.17}$$

where $f_u(z^*_{n_u}, z^*_{m_u})$ is the pdf in (7.3.13) and the CDF:

$$F_{-u}(z^*_{n_u}, z^*_{m_u}) = \prod_{k\neq u}^{U} F_k(z^*_{n_u}, z^*_{m_u}) \tag{7.3.18}$$

Gathering together the previous expressions, the expectation to be calculated is:

$$
\begin{aligned}
&\mathrm{E}\left[c_u[n]c_u[m]\right] = \mathrm{E}\left[c(z^*_{n_u})c(z^*_{m_u})\right] \\
&= \int_{z^*_{n_u}=0}^{\infty}\int_{z^*_{m_u}=0}^{\infty} \bar{\gamma}_u c(z^*_{n_u})c(z^*_{m_u}) f_u(z^*_{n_u}, z^*_{m_u}) F_{-u}(z^*_{n_u}\cdot\bar{\gamma}_u, z^*_{m_u}\cdot\bar{\gamma}_u) dz^*_{n_u} dz^*_{m_u} \\
&= \left\{x = z^*_{n_u}; y = z^*_{m_u}; p = m-n\right\} \\
&= \int_{x=0}^{\infty}\int_{y=0}^{\infty} \bar{\gamma}_u \log_2(1+\beta_u x^2)\log_2(1+\beta_u y^2) \\
&\cdot \frac{4xy}{(1-\mathcal{R}_z(p))}\cdot\exp\left(-\frac{(x^2+y^2)}{(1-\mathcal{R}_z(p))}\right)\cdot I_0\left(\frac{2\sqrt{\mathcal{R}_z(p)}xy}{(1-\mathcal{R}_z(p))}\right) \\
&\left\{1-\exp\left(-\bar{\gamma}_u^2 x^2\right) Q_1\left(\sqrt{\frac{2}{(1-\mathcal{R}_z(p))}}\bar{\gamma}_u y, \sqrt{\frac{2\mathcal{R}_z(p)}{(1-\mathcal{R}_z(p))}}\bar{\gamma}_u x\right)\right. \\
&\left.-\exp\left(-\bar{\gamma}_u^2 y^2\right)\left[1-Q_1\left(\sqrt{\frac{2\mathcal{R}_z(p)}{(1-\mathcal{R}_z(p))}}\bar{\gamma}_u y, \sqrt{\frac{2}{(1-\mathcal{R}_z(p))}}\bar{\gamma}_u x\right)\right]\right\}^{U-1} dxdy
\end{aligned}
\tag{7.3.19}
$$

The evaluation of the users' rates is presented in Figure 7.7 for the same conditions as in RR. The variance in (7.3.19) has been obtained by simulation methods. A long trace of the instantaneous transmission rate process is generated and the sample variance is get from it. The qualitative behaviour already observed in the uncorrelated channel is highlighted here: the differences among users increase significantly with the time correlation of the channel.

7.3.4 Proportional Fair

The calculation of the variance in the PF discipline is very similar to the BC algorithm. It is necessary the evaluation of $\mathrm{E}[c_u[n]c_u[n+m]]$:

$$\sigma^2_{k_u} = \sum_{q=0}^{k-1}\sum_{r=0}^{k-1}\mathcal{K}_{c_u}(r-q) \tag{7.3.20}$$

$$\mathcal{K}_{c_u}(m) = \mathrm{E}[c_u[n]c_u[n+m]] - m_u^2 \tag{7.3.21}$$

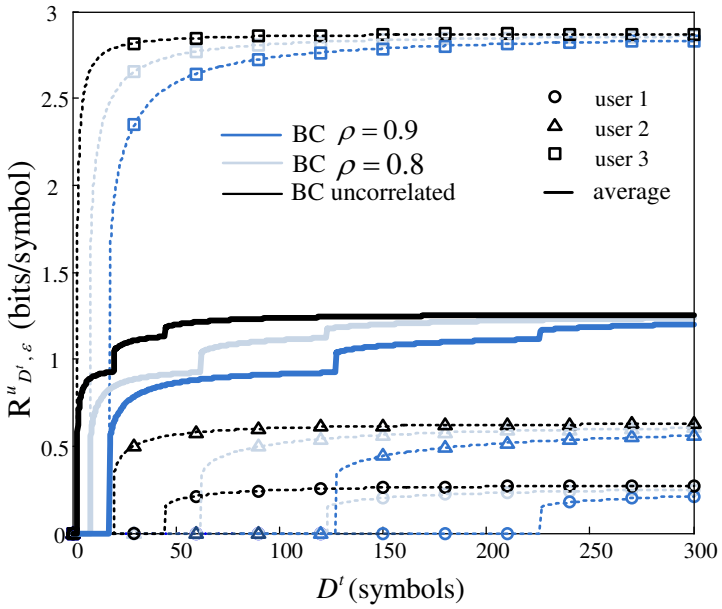


Figure 7.7: Achievable users' rates with Best Channel scheduling in a time-correlated channel.

with m_u as in (7.2.32).

The effective envelope of the uth user yields:

$$z^*_{n_u} = \begin{cases} z_u[n], & z_u[n] > z_{-u}[n] \\ 0, & z_u[n] < z_{-u}[n] \end{cases} \tag{7.3.22}$$

Now the vector of decision simplifies:

$$(z^*_{n_u}, z^*_{m_u}) = \begin{cases} (0,0) & z_u[n] < z_{-u}[n] \;\; and \; z_u[m] < z_{-u}[m] \\ (z_u[n], 0) & z_u[n] > z_{-u}[n] \;\; and \;\; z_u[m] < z_{-u}[m] \\ (0, z_u[m]) & z_u[n] < z_{-u}[n] \;\; and \;\; z_u[m] > z_{-u}[m] \\ (z_u[n], z_u[m]) & z_u[n] > z_{-u}[n] \;\; and \;\; z_u[m] > z_{-u}[m] \end{cases} \tag{7.3.23}$$

Like in the BC discipline, only the joint pdf of the last case is needed, as the other three options will result in zero in the expression of $E[c_u[n]c_u[n+m]]$. This joint pdf is:

$$f_u(z^*_{n_u}, z^*_{m_u}) F_{-u}(z^*_{n_u}, z^*_{m_u}) \tag{7.3.24}$$

where $f_u(z^*_{n_u}, z^*_{m_u})$ is the pdf in (7.3.13) and the CDF $F_{-u}(z^*_{n_u}, z^*_{m_u})$ is:

$$F_{-u}(z^*_{n_u}, z^*_{m_u}) = \prod_{k \neq u}^{U} F_k(z^*_{n_u}, z^*_{m_u}) \tag{7.3.25}$$

Finally, the expression of the expectation is:

$$\begin{aligned}
&\mathrm{E}\left[c_u[n]c_u[m]\right] = \mathrm{E}\left[c(z^*_{n_u})c(z^*_{m_u})\right] \\
&= \int_{x=0}^{\infty}\int_{y=0}^{\infty} \log_2(1+\beta_u x^2)\log_2(1+\beta_u y^2) \\
&\cdot \frac{4xy}{(1-\mathcal{R}_z(p))} \cdot \exp\left(-\frac{(x^2+y^2)}{(1-\mathcal{R}_z(p))}\right) \cdot I_0\left(\frac{2\sqrt{\mathcal{R}_z(p)}xy}{(1-\mathcal{R}_z(p))}\right) \\
&\left\{1-\exp\left(-x^2\right)Q_1\left(\sqrt{\frac{2}{(1-\mathcal{R}_z(p))}}y, \sqrt{\frac{2\mathcal{R}_z(p)}{(1-\mathcal{R}_z(p))}}x\right)\right. \\
&\left.-\exp\left(-y^2\right)\left[1-Q_1\left(\sqrt{\frac{2\mathcal{R}_z(p)}{(1-\mathcal{R}_z(p))}}y, \sqrt{\frac{2}{(1-\mathcal{R}_z(p))}}x\right)\right]\right\}^{U-1} dxdy
\end{aligned} \tag{7.3.26}$$

Figure 7.8 shows the evaluation of the users' rates for a PF discipline and with the same parameters defined before. The variance has been obtained by simulation methods. In the uncorrelated channel, the algorithm was able to equal the users in terms of minimum target delay. When the time-correlation of the channel comes on, the algorithm cannot maintain the fairness anymore and differences among users can be observed. Like in the other two algorithms, the minimum target delay that each user can demand is related to the quality of his channel. In spite of not maintaining the fairness among users anymore, it is still the more fair in terms of target delay.

7.4 OFDMA

System model

Up to now we have studied a system with users sharing a single channel. To complete the multi-user scenario, the last step is to consider a multi-user wireless

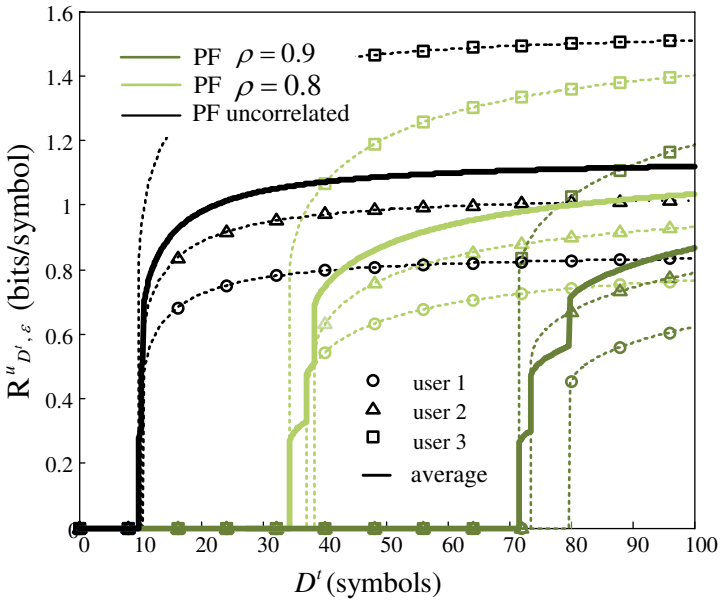


Figure 7.8: Achievable users' rates with Proportional Fair scheduling in a time-correlated channel.

system over a multiple shared fading channel. In OFDMA (Orthogonal Frequency Division Multiple Access), multiuser diversity and time diversity are exploited by allocating different users to distinct subcarriers dynamically. The results obtained for OFDM (Chapter 6) can be incorporated to the multiuser system. Figure 7.9 illustrates the system model.

F servers model the F subcarriers of the OFDM system. The scheduler assigns resources to users. The instantaneous transmission rate seen by user uth is:

$$c_u[n] = \sum_{f=1}^{F} c_u^f[n],$$
$$c_u^f[n] = \begin{cases} r_u^f[n] \text{ if subcarrier } f \text{ is assigned to user u} \\ 0 \text{ in other case} \end{cases} \tag{7.4.1}$$

The effective bandwidth function of the channel depends on the scheduling discipline and on the time- and frequency- correlation of the channel. Splitting the total transmission rate into blocks of length k, the EBF of user uth yields:

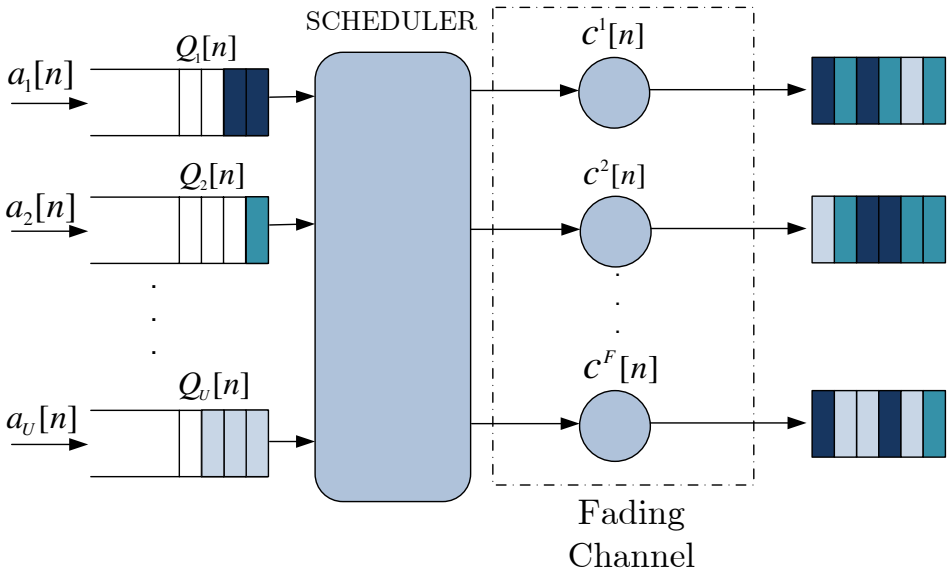


Figure 7.9: OFDMA system model

$$\alpha_{C_u}(\upsilon) = \lim_{n\to\infty} \frac{1}{n\cdot\upsilon} \log \mathrm{E}\left[e^{\upsilon C_u[n]}\right] = \frac{m_{k_u}}{kF} + \frac{\upsilon}{2}\frac{\sigma^2_{k_u}}{kF} \tag{7.4.2}$$

And the achievable rate of uth user is:

$$\mathrm{R}^u{}_{D^t,\varepsilon} = \frac{m_{k_u}}{2kF} + \frac{1}{2}\sqrt{\frac{m^2_{k_u}}{(kF)^2} - 2\frac{\sigma^2_{k_u}}{kF}\frac{(-\log\varepsilon)}{D^t}} \tag{7.4.3}$$

The only difference with the single channel case is that now the mean and the variance are obtained as in Chapter 6:

- The mean is readily computed from the mean m_u of the uncorrelated single channel:

$$m_{k_u} = F\cdot k\cdot m_u \tag{7.4.4}$$

- The variance of user uth is:

$$\sigma^2_{k_u} = \mathrm{var}(\sum_{p=0}^{k-1}\sum_{f=1}^{F} c_u^f[p]) = \sum_{p=0}^{k-1}\sum_{q=1}^{k-1}\sum_{f=1}^{F}\sum_{g=1}^{F}\left(\mathrm{E}\left[c_u^f[p]c_u^g[q]\right] - m_u^2\right) \tag{7.4.5}$$

Thus, the variances needed for the three disciplines are the result of computing (7.4.5) with the corresponding expectations $\mathrm{E}[c_u^f[n]c_u^g[m]]$. Next, the procedure is summarized for the three disciplines.

Round Robin

- Mean. Compute (7.4.4) with m_u in equation (7.2.14).
- Variance. Compute the variance:

$$\sigma_{k_u}^2 = \frac{1}{U^2}\sigma_{k_1}^2(\bar{\gamma}_u, \beta_u) \tag{7.4.6}$$

  with $\sigma_{k_1}^2$ from Section 6.2.

Best Channel

- Mean. Compute (7.4.4) with m_u in equation (7.2.23).
- Variance. Compute (7.4.5) with (7.3.19) and m_u in equation (7.2.23), and replace the autocorrelation function $\mathcal{R}_z(\Delta n)$ by:

$$\mathcal{R}_z(\Delta n, \Delta f) = \frac{\mathrm{cov}(z_{n,f}, z_{m,g})}{\sqrt{\mathrm{var}(z_{n,f})\mathrm{var}(z_{m,g})}} \tag{7.4.7}$$

Proportional Fair

- Mean. Compute (7.4.4) with m_u in equation (7.2.32).
- Variance. Compute (7.4.5) with (7.3.26) and m_u in equation (7.2.32), and replace the autocorrelation function $\mathcal{R}_z(\Delta n)$ by:

$$\mathcal{R}_z(\Delta n, \Delta f) = \frac{\mathrm{cov}(z_{n,f}, z_{m,g})}{\sqrt{\mathrm{var}(z_{n,f})\mathrm{var}(z_{m,g})}} \tag{7.4.8}$$

Numerical results

Figures 7.10, 7.11 and 7.12 show the evaluation of the users' rates in an OFDMA system and for the three disciplines: RR, BC and PF. The variance of the BC and the PF has been obtained by simulation methods. A pedestrian channel with the parameters in Table 6.1 is considered. The ACF follows a Bessel function with parameter $f_D T_S$. The coherence bandwidth is approximately $3.4MHz$ and the product $f_D T_S$ is set to 0.1 and 0.01. There are 5 users with average SNR 5, 6, 7, 8 and 9 dB and 4 subcarriers. The maximum achievable rate $\mathrm{R}^u{}_{D^t,\varepsilon}$ is plot for the best (9dB), the medium (7dB) and the worst (5dB) user.

The same behaviour observed in the single channel is emphasized with the inclusion of the frequency correlation, in the same way as we observed in a single user OFDM system with regard to the flat channel. Thus, the differences among users increases with the correlation of the channel, not just the time correlation but also the frequency correlation. It is worth noting that even adding the frequency correlation, PF is still able to maintain a good degree of fairness among users in terms of target delay.

7.5 Simulation comparison

7.5.1 Uncorrelated channel

A queueing system has been simulated to validate the results presented in this chapter. Each user sends bits to its buffer of queue length $Q_u[n]$ in the nth symbol. The selected user depends on the scheduling algorithm: Round Robin or Best Channel. Certain values of D^t and ε are fixed and the users' rates are evaluated with (7.2.6). Then, the result is given in the simulator to the arrival process of each user to generate source data at the (constant) rate $\mathrm{R}^u{}_{D^t,\varepsilon}$. The simulation is run and the tail probability of exceeding the target delay is measured based on the measurements of

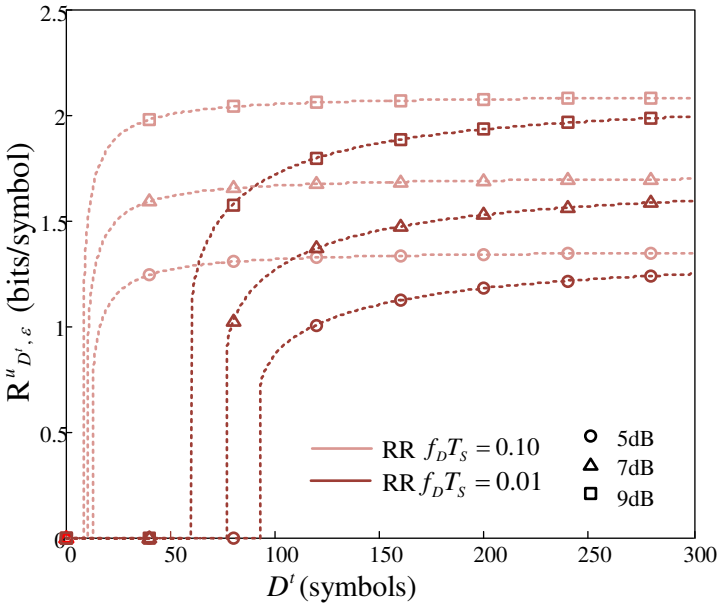


Figure 7.10: Achievable users' rates in a OFDMA system with Round Robin discipline.

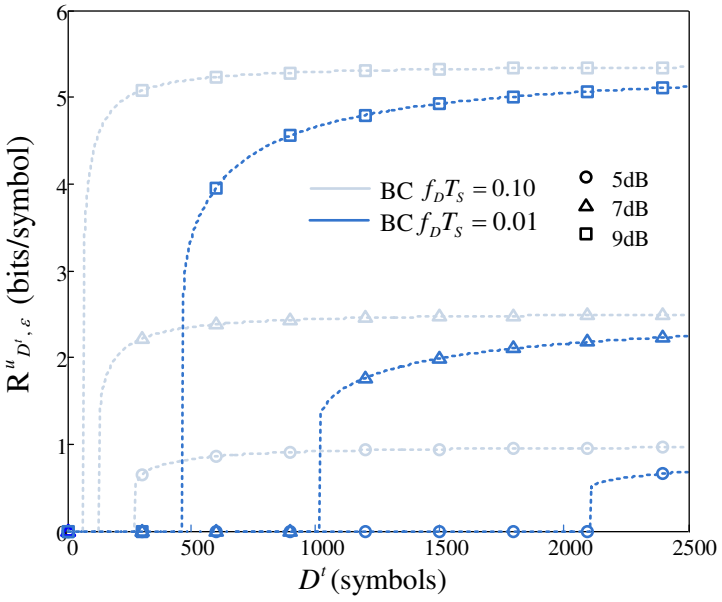


Figure 7.11: Achievable users' rates in a OFDMA system with Best Channel discipline.

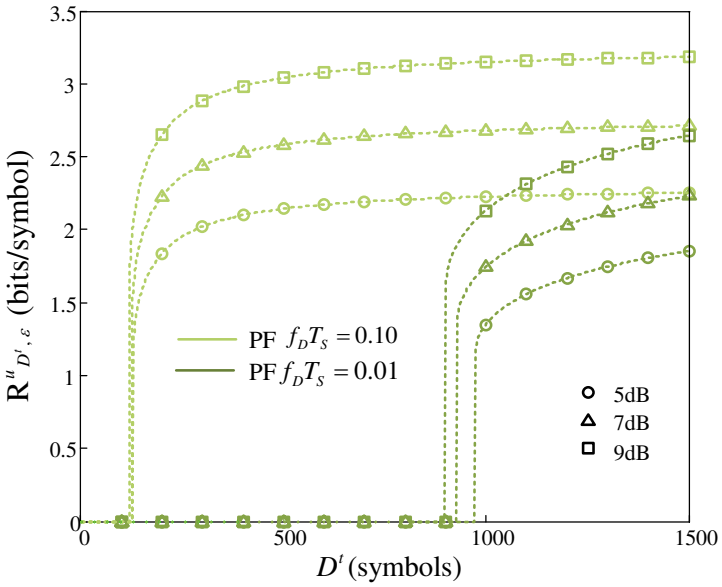


Figure 7.12: Achievable users' rates in a OFDMA channel with Proportional Fair discipline.

the delay suffered by bits leaving the queue. Notice that the expected value of this tail probability, $Pr\{D(\infty) > D^t\}$, is ε.

Figure 7.13 shows the results of the simulations. Five users with average SNR's $8, 9, 10, 11, 12$dB are simulated. $Pr\{D(\infty) > D^t\}$ is measured for different values of the average SNR and a target delay of 60 symbols, and compared to the analytical ε. The probability of exceeding the target delay, ε, has been set to 0.10 (BC), 0.05 (PF) and 0.01 (RR). The parameters of the simulation are shown in Table 7.1. The results show that the QoS requirements are accurately reached with the result for $\mathrm{R}^u{}_{D^t,\varepsilon}$ in (7.2.6). The users' rates were obtained under the assumption that $\eta_u = 1$, a high load scenario that constitutes an upper bound. It has been checked in the simulations that the measured probability of a non-empty queue η_u is very close to one in our simulations.

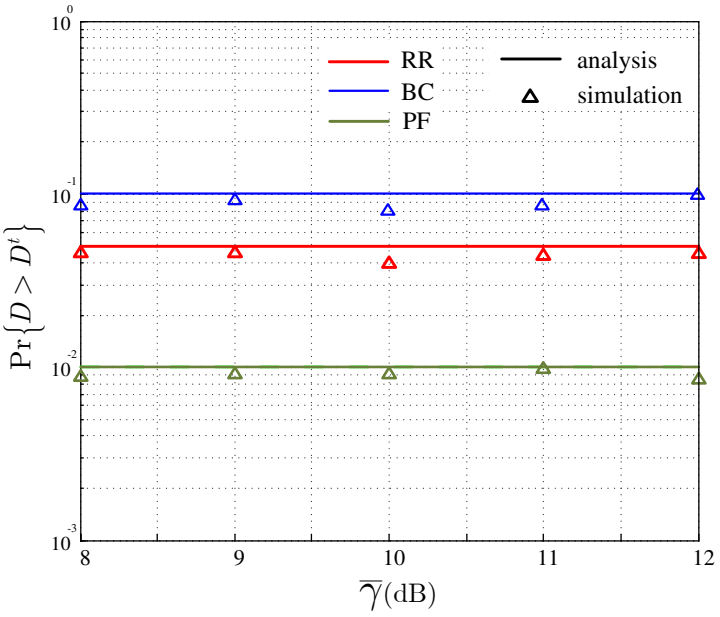


Figure 7.13: Simulation comparison for uncorrelated channel. 5 users with $\bar{\gamma} = 8, 9, 10, 11, 12$dB respectively.

Table 7.1: Achievable users' rates per each algorithm in the simulations.

		RR			BC			PF		
	$\bar{\gamma}_u$	ε	D^t	$\mathrm{R}^u{}_{D^t,\varepsilon}$	ε	D^t	$\mathrm{R}^u{}_{D^t,\varepsilon}$	ε	D^t	$\mathrm{R}^u{}_{D^t,\varepsilon}$
user 1	8dB	0.05	60	4.6	0.10	60	2.1	0.01	60	6.0
user 2	9dB	0.05	60	5.1	0.10	60	4.3	0.01	60	6.5
user 3	10dB	0.05	60	5.6	0.10	60	7.3	0.01	60	7.1
user 4	11dB	0.05	60	6.2	0.10	60	1.3	0.01	60	7.6
user 5	12dB	0.05	60	6.8	0.10	60	6.1	0.01	60	8.1

7.5.2 Correlated channel

As example of correlated channel, a simple time-correlated channel has been simulated under the same conditions of Section 7.5.1. Figure 7.14 shows the results of the simulations. The ACF follows an exponential rule with correlation parameter $\rho = 0.8$ and the target delay is 250 symbols. The conclusions from the uncorrelated channel apply also here. Once more, the simulation results are satisfactory, approaching accurately the analytical results.

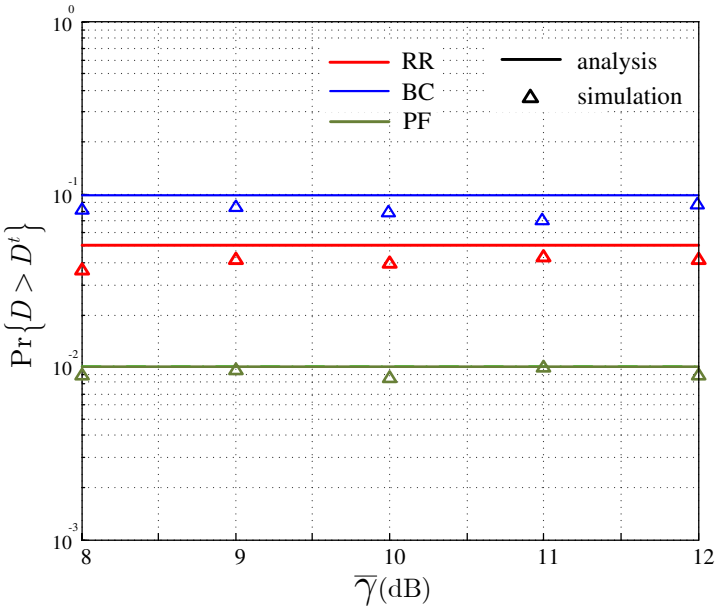


Figure 7.14: Simulation comparison for time correlated channel. 5 users with $\bar{\gamma} = 8, 9, 10, 11, 12$dB respectively.

7.6 Summary

In this chapter, we have obtained analytical expressions of the achievable users' rates and the system capacity under a selected scheduling discipline and a QoS constraint given in terms of a BER and a delay constraint. Each user can have its own delay constraint. Three widely employed disciplines have been analyzed (Round Robin,

Best Channel and Proportional Fair), but the procedure can be applied to any other algorithm. The results have been obtained first for a single shared channel, starting with the uncorrelated fading channel and with a time-correlated channel later on. Finally, a multiple shared fading channel has been addressed by means of an OFDMA system where different subcarriers are assigned to distinct users. In the case of correlated channels, on account of the difficulty to obtain the EBF of the channel, a semi-analytical strategy has been applied to the Best Channel and Proportional Fair algorithms.

The evaluation of the individual rates and the total capacity confirms the expected qualitative behaviour of the algorithms: Round Robin obtains the lower total throughput whereas Best Channel maximizes it at the expenses of fairness among users. On its side, Proportional Fair provides a balance among throughput and fairness, although this balance deteriorates as the time- of frequency- correlation of the channel increases. There are two interesting parameters in the observation of the fairness among users. First the asymptotic maximum rate that a user can achieve, when no QoS is guaranteed. The other one is the minimum target delay that he can demand under his channel conditions. Finally, simulations of the algorithms were conducted to validate our outcomes.

Chapter 8

Conclusions

This dissertation addressed the provisioning of statistical QoS in a wireless system. In particular, the tradeoff among the channel fading, the data outsourcing process and the QoS metrics has been analyzed. For this goal, we have proposed the application of the effective bandwidth theory.

The main conclusions drawn from the realization of the thesis and some issues to be tackled in future work are presented in this chapter. Section 8.1 summarizes the contents of the thesis. Several topics to be addressed in the future are discussed in Section 8.2.

8.1 Synthesis of the dissertation

- In Chapter 1 we have introduced and motivated the problem.
- In Chapter 2 a review on related bibliography has bee done. Next, the proposed system model has been detailed. In contrast to most of the related literature, we have included the variability of the outsourcing process. Additionally, the variability of the physical transmission rate has also been considered, representing the dynamic adaptation to the channel conditions so common in

current wireless systems. An effective bandwidth analysis of the model has been addressed, and the meaning of the effective bandwidth functions involved in the system has been explained. Finally, the generalization to multi-user and multi-channel systems has been detailed.

- In Chapter 3 we have investigated the effective bandwidth function of several traffic sources. First, a review on traffic modeling has been presented. The aim is to collect a set of representative models for traffic in wireless systems. Moreover, the corresponding effective bandwidth function of each model has been studied. Two estimators of the EBF found in the literature have been used to validate the results of the chapter.

- In Chapter 4 the effective bandwidth of a flat Rayleigh channel has been analytically derived, first for block fading channels and secondly for time-correlated Rayleigh channels. In contrast to existing results which resort to measures or simple models of the channel, the effective bandwidth function has been analytically derived from the radio channel. The methodology to face the problem consists of splitting the transmission rate into blocks in such a way that they can be considered independent, making feasible the application of the Central Limit Theorem. Despite the fact that the evaluation has been done for Rayleigh channels and two examples of Doppler spectrum, the procedure is generic and provides a framework under which any other channel model or Doppler spectrum can be studied. For the cases in which the analytical EBF of the channel is not available, a semi-analytical approach has also been proposed.

- In Chapter 5 the analysis of the single-user single-channel system has been tackled by means of the effective bandwidth functions of the source and the channel. The effective bandwidth approach provides a solution of the tail probability of the delay. The percentile of the delay has been analyzed for flat Rayleigh channels. The results show the harmful influence of the correlation

of the source or the channel in the delay performance. Since the effective bandwidth of the channel has been calculated by means of a Gaussian distribution, only two parameters define the channel: the mean and the variance of the transmission rate. The parameter that captures the influence of the time-correlation of the channel is the variance. Thus, a higher value of the correlation leads to a higher value of variance and hence a smaller value of the percentile of the delay. Furthermore, a couple of examples of variable sources were illustrated, representing the procedure to generalize the results of this thesis to any desired traffic model.

In addition, a new concept has been defined, the capacity with probabilistic delay constraint, which guarantees statistical QoS in contrast to the well-known delay-limited capacity. The capacity has been derived for constant traffic sources, when the maximum capacity is obtained. This upper bound is readily evaluated and it is of great use to understand the trade-off between the channel conditions and the QoS metrics. Figures on how much the capacity decreases for more correlated channels or shorter allowable delays have been shown. It is expected that any kind of variability in the source process is harmful to the delay QoS performance leading to lower values of capacity. The predicted behaviour with the burstiness of the source has been observed: the higher the correlation of the source, the lower the maximum allowable rate. Finally, the results have been validated by comparison with simulations.

- In Chapter 6 the analysis of the delay and the capacity with probabilistic delay constraint has been extended to a multi-channel system employing OFDM multiplexing over a frequency selective Rayleigh channel. First, the effective bandwidth function of a frequency selective channel had to be calculated, with the same techniques applied in the case of flat channels. Secondly, the percentile of the delay has been calculated. Moreover, the proposed capacity has been evaluated. Similarly to flat channels, the channel is represented through two statistics, mean and variance. Now, the variance captures not

only the time-correlation but also the frequency correlation among subcarriers. As it could be expected, the effects of the frequency correlation on the capacity are similar to the time-correlation: it is harmful to the delay behaviour. Thus, when the time- or frequency- correlation increases, the capacity under a delay constraint decreases. Two examples of frequency selective channel have been considered, corresponding to a pedestrian and a vehicular channel. The results have been validated by comparison with simulations.

- In Chapter 7 a multiuser formulation of the system model has been proposed. A new factor comes out in the scenario: the scheduling discipline. The same techniques presented before has been applied to analyze the maximum rate achievable by each user under a delay constraint and a queueing discipline. Thus, the capacity with probabilistic delay constraint becomes a vector containing the maximum available rate per user, where each user can have his own delay constraint and channel conditions. Three examples of discipline were analyzed: Round Robin, Best Channel and Proportional Fair, first for a single link and secondly for a multiple channel system employing OFDM as multiplexing mechanism. The expected behaviour of the algorithms (observed via simulations) has been confirmed by our analytical results. Thus, Round Robin obtains the lower total throughput whereas Best Channel maximizes it at the expenses of fairness among users. Proportional Fair provides a balance among throughput and fairness, although this balance deteriorates as the time- or frequency- correlation of the channel increases. There are two interesting parameters in the observation of the fairness among users. First the asymptotic maximum rate a user can achieve, when no QoS is guaranteed. The other one is the minimum target delay he can demand under his channel conditions. The presented procedure can be extended to other algorithms. For the cases in which the analytical EBF of the channel is not available, a semi-analytical approach has also been proposed.

In summary, this research analyzes the delay in a wireless system. The tradeoff among the traffic source, the channel and the QoS requirements has been investigated. The known qualitative behaviour of the delay is asserted by the analytical results obtained throughout the dissertation. Thus, the time and frequency variability of the channel as well as the burstiness of the traffic source are harmful to the delay performance. Moreover, the discipline turns to be a significant factor on the behaviour of the delay.

8.2 Future work

In this section, we point out future research directions.

- The classical Clarke's model is starting to be replaced in new standards by specific models that take into account the peculiarities of the communication. For example, IEEE 802.16 [802.16 2001a] proposes a different Doppler spectrum suitable for its application in fixed wireless communications. The approach to include other models can be the same proposed in this thesis. The key is to calculate the effective bandwidth function of the channel process. If the analytical solution is not feasible, we can always turn to semianalytical methods.

- QoS-aware multiplexing algorithms could be addressed. We refer to disciplines that include some QoS indicators in the dynamic allocation of the channel, such as Longest Delay First (which decides based on the delay of the queue) or Exponential Rule [Shakkottai 2001] (which uses both the channel state and the delay of the queue). In the same way as we have done with BC and PF, the key is the computation of the effective pdf seen by user uth. If the pdf cannot be analytically obtained, semianalytical methods measuring the required variance can be applied.

- In this thesis, we have considered the time- and frequency- correlation. More and more, wireless standards are adding a third coordinate, the space. Techniques that exploit the spatial diversity are the multiples antennas schemes (Multiple Input Multiple Output, MIMO) and cooperative communications, including relays and network MIMO. The results of this work can be extended to MIMO systems similarly as we have done from flat channels to frequency selective channels. Thus, the space correlation among antennas has to be included and the main difficulty is once more to get an expression of the joint pdf. On the other hand, cooperative communications are becoming increasingly popular thanks to their ability to provide spatial diversity without packing multiple antennas physically into small-size mobile nodes. They form a virtual antenna array by exploiting the broadcast nature of wireless transmissions and utilize distributed antennas that belong to different transmitters. The extension of the work to cooperative communications opens a broad research direction. A wide range of scenarios arise depending on the kind of cooperation. The challenge is to provide a framework under which the effective bandwidth analysis of the network can be addressed for different configurations of the cooperative communication.

Bibliography

[3GPP 2009] 3GPP. *TS23.107 Quality of Service (QoS) concept and architecture v 9.0.0.* 2009.

[802.16 2001a] IEEE 802.16. *Channel Models for Fixed Wireless Applications.* IEEE 802.16 Broadband Wireless Access Working Group, July 2001.

[802.16 2001b] IEEE 802.16. *Traffic Model for 802.16 TG3 MAC/PHY Simulations.* IEEE 802.16 Broadband Wireless Access Working Group, March 2001.

[Adas 1997] A. Adas. *Traffic Models in Broadband Networks.* IEEE Communications Magazine, vol. 35, no. 7, pages 82–89, 1997.

[Arauz 2004] J. Arauz, P. Krishnamurthy and M. A. Labrador. *Discrete Rayleigh fading channel modeling.* Wireless Communications & Mobile Computing, vol. 4, pages 413–425, 2004.

[Baddour 2005] K. E. Baddour and N. C. Beaulieu. *Autoregressive modeling for fading channel simulation.* IEEE Transactions on Wireless Communications, vol. 4, no. 4, pages 1650–1662, 2005.

[Bergamo 2002] P. Bergamo, D. Meniezzo, A. Giovanardi, G. Mazini and M. Zorzi. *Improved Markov model for Rayleigh fading envelope.* IEEE Electronic Letters, vol. 38, pages 477–478, May 2002.

[Berry 2002] R. Berry and R. Gallager. *Communication over fading channels with delay constraints.* IEEE Transactions on Information Theory, vol. 48, pages 1135–1149, May 2002.

[Biglieri 1998] E. Biglieri, J. Proakis and S. Shamai. *Fading channels: information-theoretic and communications aspects.* IEEE Transactions on Information Theory, vol. 44, no. 6, pages 2619–2692, 1998.

[Botvich 1995] D. D. Botvich and N. G. Duffield. *Large Deviations, the Shape of the Loss Curve, and Economies of Scale in Large Multiplexers*, 1995.

[Brady 1968] P. T. Brady. *A Statistical Analysis of ON-OFF Patterns in 16 Conversations.* Bell Syst. Tech. J., vol. 48, no. 1, January 1968.

[Chang 1994] C. S. Chang. *Stability, queue length and delay of deterministic and stochastic queueing networks.* IEEE Transactions on Automatic Control, vol. 39, no. 5, pages 913–931, May 1994.

[Chang 1995a] C. S. Chang and J. A. Thomas. *Effective Bandwidth in High-Speed Digital Networks.* IEEE Journal on Selected Areas in Communications, vol. 13, August 1995.

[Chang 1995b] C. S. Chang and T. Zajic. *Effective Bandwidths of Departure Processes from Queues with Time Varying Capacities.* In Proc. IEEE INFOCOM'95, pages 1001–1009, Boston, USA, April 1995.

[Chang 2000] C. S. Chang. *Performance Guarantees in Communication Networks.* Springer, 1 ed., 2000.

[Chen 2006] S. Chen and J. Cobb. *Wireless Quality-of-Service Support.* ACM/Kluwer Journal of Wireless Networks, vol. 12, no. 4, pages 409–410, August 2006.

[Chou 1996] L. S. Chou and C. S. Chang. *Experiments of the theory of effective bandwidth for Markov sources and video traces.* In Proc. IEEE INFOCOM'96, pages 497–504, San Francisco, USA, Mar 1996.

[Choudhury 1996] G. L. Choudhury and D. M. Lucantoni. *Squeezing the Most Out of ATM.* IEEE Transactions on Communications, vol. 44, pages 203–217, February 1996.

[Chung 2001] S. T. Chung and A. J. Goldsmith. *Degrees of freedom in adaptive modulation: A unified view.* IEEE Transactions on Communications, vol. 49, pages 1561–1571, September 2001.

[Clarke 1968] R. H. Clarke. *A statistical theory of mobile radio reception.* Bell Systm. Tech. J., pages 957–1000, July-August 1968.

[Courcoubetis 1994] C. Courcoubetis, G. Fouskas and R. Weber. *On the performance of an effective bandwidths formula.* Proc. 14th Int. Teletraffic Cong., June 1994.

[Courcoubetis 1999] C. Courcoubetis, V. A. Siris and G. D. Stamoulis. *Application of the Many Sources Asymptotic and Effective Bandwidths to Traffic Engineering.* Telecommunication Systems, vol. 12, pages 167–191, November 1999.

[Cover 1991] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley Series in Telecommunications, 1st ed., 1991.

[Cruz 1991] R. Cruz. *A calculus for network delay: parts I and II.* IEEE Transactions on Information Theory, vol. 37, no. 1, pages 114–141, January 1991.

[Doeltz 1957] M. L. Doeltz, E. T. Heald and D. L. Martin. *Binary Data Transmission Techniques for Linear Systems.* Proc. IRE, vol. 45, pages 656–661, May 1957.

[Dufield 1995] N. G. Dufield, J. T. Lewis, N. O'Connell, R. Russell and F. Toomey. *Entropy of ATM Traffic Streams: A Tool for Estimating QoS Parameters.* IEEE Journal on Selected Areas in Communications, vol. 13, pages 981–990, August 1995.

[Entrambasaguas 2007] J. T. Entrambasaguas, M. C. Aguayo-Torres, G. Gómez and J. F. Paris. *Multiuser Capacity and Fairness Evaluation of Channel/QoS-Aware Multiplexing Algorithms.* IEEE Network, vol. 21, no. 3, pages 24–30, May 2007.

[Femenias 2009] G. Femenias, Jaume Ramis and Loren Carrasco. *Using Two-Dimensional Markov Models and the Effective-Capacity Approach for Cross-Layer Design in AMC/ARQ-Based Wireless Networks.* IEEE Transactions on Vehicular Technology, vol. 58, no. 8, pages 4193–4203, October 2009.

[Georgiadis 2006] L. Georgiadis, M. J. Neely and L. Tassiulas. *Resource allocation and cross-layer control in wireless networks.* Foundations and Trends in Networking, vol. 1, no. 1, pages 1–149, 2006.

[Goldsmith 2005] A. Goldsmith. *Wireless Communications.* Cambridge University Press, 1 ed., 2005.

[Hanssen 2004] M. R. Hanssen. *Opportunistic Scheduling in Wireless Networks.* Term proj. rep., Dept. of Elect. and Telecommun., NTNU, December 2004.

[Hassan 2004] M. Hassan, M. M. Krunz and I. Matta. *Markov-based channel characterization for tractable performance analysis in wireless packet networks.* IEEE Transactions on Wireless Communications, vol. 3, no. 3, pages 821–831, May 2004.

[Heyman 1996] D. P. Heyman and T. V. Lakshman. *What are the implications of long-range dependence for vbr-video traffic engineering?* IEEE/ACM Transactions on Networking, vol. 4, no. 3, pages 301–317, June 1996.

[ITU-R 1997] ITU-R. *Guidelines for evaluation of radio transmission for IMT-2000.* Recommendation ITU-R M.1225, 1997.

[Jakes 1989] W. C. Jakes. *Microwave Mobile Communications.* McGraw Hill, 1 ed., 1989.

[Kelly 1996] F. P Kelly. *Notes on Effective Bandwidth.* in F. P. Kelly, S. Zachary, and I. Zeidins (editors) Stochastic Networks: Theory and Applications, vol. 4, pages 141–168, 1996.

[Kessidis 1996] G. Kessidis, J. Walrand and C. S. Chang. *Effective Bandwidth for multiclass Markov fluids and other ATM sources.* IEEE/ACM Transactions on Networking, vol. 1, no. 4, pages 424–428, August 1996.

[Kleinrock 1975] L. Kleinrock. *Queueing Systems. Volume I: Theory.* Wiley Interscience, 1 ed., 1975.

[Knopp 1995] R. Knopp and P. A. Humblet. *Information Capacity and Power Control in Single-Cell multiuser Communications.* Proc. IEEE ICC'95, vol. 1, pages 331–335, June 1995.

[Kurose 1992] J. Kurose. *On computing per-session performance bounds in high-speed multi-hop computer networks.* Proc. of ACM SIGMETRKS'92, pages 128–139, June 1992.

[LeBoudec 2001] J. Y. LeBoudec and P. Thiran. *Network calculus.* Springer Verlag, Lecture Notes in Computer Science, LNCS 2050, 2001.

[Leland 1994] W. E. Leland, M. S. Taqqu, W. R. Willinger and D. V. Wilson. *On the self-similar nature of Ethernet traffic (extended version.* IEEE/ACM Transactions on Networking, vol. 2, pages 1–15, February 1994.

[Liu 2004] N. X. Liu and J. S. Baras. *Measurement and Simulation Based Effective Bandwidth Estimation.* Proc. IEEE GLOBECOM'04, November 2004.

[Liu 2007] L. Liu, P. Parag, J. Tang, W. Y. Chen and J. F. Chamberland. *Resource Allocation and Quality of Service Evaluation for Wireless Communication Systems Using Fluid Models.* IEEE Transactions on Information Theory, vol. 53, no. 5, pages 1767–1777, May 2007.

[Maglaris 1998] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson and J. D. Robbins. *Performance Models of Statistical Multiplexing in Packet Video Communications.* IEEE Transactions on Communications, vol. 36, no. 7, pages 834–844, July 1998.

[Neely 2003] M. J. Neely, E. Modiano and C. E. Rohrs. *Power allocation and routing in multi-beam satellites with time varying channels.* IEEE Transactions on Networking, vol. 11, no. 1, pages 138–152, February 2003.

[Neely 2009] M. J. Neely. *Delay Analysis for Max Weight Opportunistic Scheduling in Wirelss Systems.* IEEE Transactions on Automatic Control, vol. 54, no. 9, pages 2137–2150, September 2009.

[O'Connell 1999] Neil O'Connell. *Large Deviations with Application to Telecommunications*, 1999.

[Ozarow 1994] L. H. Ozarow, S. Shamai and A. D. Wyner. *Information theoretic considerations for cellular mobile radio.* IEEE Transactions on Vehicular Technology, vol. 43, no. 2, pages 359–378, May 1994.

[Papoulis 2002] A. Papoulis and S. U. Pillai. *Probability, Random Variables and Stochastic Processes.* McGraw-Hill, 4th ed., 2002.

[Park 2006] D. Park and B. G. Lee. *QoS Support by Using CDF-Based Wireless Packet Scheduling in Fading Channels.* IEEE Trans. on Communications, vol. 54, pages 2051–2061, November 2006.

[Pérez 2007] J. Pérez, J. Ibá nez and I. Santamaría. *Exact closed-form expressions for the sum capacity and individual users' rates in broadcast ergodic Rayleigh fading channels.* IEEE Workshop on Signal Processing Advances in Wireless Communications, SPAWCŠ07, June 2007.

[Pop 2002] M.F. Pop and N.C. Beaulieu. *Design of Wide-Sense Stationary Sum-of-Sinusoids Fading Channel Simulators.* Proceedings of IEEE International Conference on Communications, ICC'02, vol. 2, pages 709–716, April 2002.

[Prudnikov 1981] A. P. Prudnikov. *Integrals and series, Vol. 1 Elementary Functions.* Science, Moscow, Russia, 1981.

[Quimi 2005] A. M. A. Quimi, A. Saiadian and A. Mirzaee. *Explicit statistical QoS guarantees for video stream over wireless fading channel.* In Proceedings of IEEE International Conference on Wireless, Networking and Mobile Computing, pages 1245–1248, September 2005.

[Rappaport 2002] Theodore S. Rappaport. *Wireless Communications: Principles and Practice.* Prentice-Hall, New-Delhi, Indian, 2002.

[Ren 2009] S. Ren and K. B. Letaief. *Maximizing the Effective Capacity for Wireless Cooperative Relay Networks with QoS Guarantees".* IEEE Transactions on Communications, vol. 57, no. 7, pages 2148–2159, July 2009.

[Ruiz 2009] J. G. Ruiz, B. Soret, M. C. Aguayo-Torres and J. T. Entrambasaguas. *On Finite State Markov Chains for Rayleigh channel modeling.* Proc. Wireless Vitae'09, May 2009.

[Ryu 1996] B. K. Ryu and A. Elwalid. *The importance of long-range dependence of vbr video traffic in atm traffic engineering: Myths and realities.* ACM SIGCOMM Conf, vol. 26, no. 4, pages 3–14, August 1996.

[Shakkottai 2001] S. Shakkottai and A. L. Stolyar. *Scheduling algorithms for a mixture of real-time and non-real-time data in HDR.* Proc. ITC 17, September 2001.

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

[Sheskin 2004] D. J. Sheskin. *Handbook of parametric and non-parametric statistical procedures.* Chapman and Hall/CRC Statistics, 3 ed., 2004.

[Simon 2005] M. K. Simon and M. S. Alouini. *Digital Communications over Fading Channels.* Wiley Interscience, 2nd ed., 2005.

[Smith 1975] J. I. Smith. *A computer generated multipath fading simulation for mobile radio.* IEEE Transaction on Vehicular Technology, vol. VT-24, pages 39–40, August 1975.

[Soret 2007a] B. Soret, M. C. Aguayo-Torres and J. T. Entrambasaguas. *Capacity with Probabilistic Delay Constraint for Rayleigh Channels.* Proc. IEEE Globecom'07, November 2007.

[Soret 2007b] B. Soret, M. C. Aguayo-Torres and J. T. Entrambasaguas. *Maximum Delay-Constrained Source Rate over a Wireless Channel.* Proc. Valuetools'07, October 2007.

[Soret 2008] B. Soret, M. C. Aguayo-Torres and J. T. Entrambasaguas. *On the effective bandwidth function of 802.16 traffic models.* Proc. I International Interdisciplinary Technical Conference of Young Scientists Intertech'08, April 2008.

[Soret 2009a] B. Soret, M. C. Aguayo-Torres and J. T. Entrambasaguas. *Analysis of delay constrained communications over OFDM systems.* Proc. IEEE Globecom'09, November 2009.

[Soret 2009b] B. Soret, M. C. Aguayo-Torres and J. T. Entrambasaguas. *Capacity with Probabilistic Delay Constraint for Voice Traffic in a Rayleigh channel.* Proc. IEEE ICC'09, June 2009.

[Soret 2009c] B. Soret, M. C. Aguayo-Torres and J. T. Entrambasaguas. *Multiuser capacity for heterogeneous QoS constraints in uncorrelated Rayleigh channels.* Proc. Information Theory Workshop ITW'09, October 2009.

[Soret 2010a] B. Soret, M. C. Aguayo-Torres and J. T. Entrambasaguas. *Analysis of the tradeoff between delay and source rate in multiuser wireless systems.* EURASIP Journal on Wireless Communications and Networking, vol. 2010, Article ID 726750, 13 pages, doi: 10.1155/2010/72675, 2010.

[Soret 2010b] B. Soret, M. C. Aguayo-Torres and J. T. Entrambasaguas. *Capacity with explicit delay guarantees for generic sources over correlated Rayleigh channel.* IEEE Transaction on Wireless Communications, vol. 9, no. 6, pages 1901–1911, June 2010.

[Tan 2000] C. C. Tan and N. C. Beaulieu. *On first-order Markov modeling for the Rayleigh fading channel.* IEEE Transactions on Communications, vol. 48, no. 12, pages 2032–2040, December 2000.

[Tang 2006] J. Tang and X. Zhang. *Quality of service driven power and rate control in mobile wireless networks.* Proc. IEEE ICC'06, June 2006.

[Tang 2007a] J. Tang and X. Zhang. *Cross-Layer-Model Based Adaptive Resource Allocation for Statistical QoS Guarantees in Mobile Wireless Networks.* IEEE Transactions on Wireless Communications, vol. 6, no. 12, December 2007.

[Tang 2007b] J. Tang and X. Zhang. *Quality of service driven power and rate adaptation over wireless links.* IEEE Transactions on Wireless Communications, vol. 6, no. 8, pages 3058–3068, August 2007.

[Wrege 1996] Dallas Wrege, Edward W. Knightly, Hui Zhang and Jorg Liebeherr. *Deterministic Delay Bounds for VBR Video in Packet-Switching Networks: Fundamental Limits and Practical Tradeoffs.* IEEE/ACM Transactions on Networking, vol. 4, pages 352–362, 1996.

[Wu 2003a] D. Wu. *Providing Quality of Service Guarantees in Wireless Networks.* PhD thesis, Department of Electrical and Computer Engineering, Carnegie Mellon University, August 2003.

[Wu 2003b] D. Wu and R. Negi. *Effective capacity: A wireless link model for support of Quality of Service.* IEEE Transactions on Wireless Communications, vol. 2, pages 640–643, July 2003.

[Ying 2006] L. Ying, R. Srikant, A. Eryilmaz and G. E. Dullerud. *A large deviation analysis of scheduling in wireless networks.* IEEE Transactions on Information Theory, vol. 52, pages 5088–5098, November 2006.

[Young 2000] D. J. Young and N. C. Beaulieu. *The Generation of Correlated Rayleigh Random Variates by Inverse Discrete Fourier Transform.* IEEE Transactions on Communications, vol. 48, pages 1114–1127, July 2000.

[Young 2003] D. J. Young and N. C. Beaulieu. *Power Margin Quality Measures for correlated random variates derived from the normal distribution.* IEEE Transactions on Information Theory, vol. 49, pages 241–256, June 2003.

[Zhang 1999] Q. Zhang and S. A. Kassam. *Finite-State Markov Model for Rayleigh Fading Channels.* IEEE Transactions on Communications, vol. 47, October 1999.

[Zhang 2006] X. Zhang, J. Tang, H. H. Chen, S. Ci and M. Guizani. *Cross-layer-based modeling for quality of service gurantees in mobile wireless networks.* IEEE Communications Magazine, pages 100–106, January 2006.

[Zheng 2002] Y. R. Zheng and C. Xiao. *Improved Models for the Generation of Multiple Uncorrelated Rayleigh Fading Waveforms.* IEEE Communications Letters, vol. 6, no. 6, pages 256–258, June 2002.

Appendix A

Fundamentals of the Large Deviations Theory

Roughly speaking, the Large Deviations Theory (LDT) is the theory of *rare events*. At the present time, it is one of the most active topics in probability theory. One of its applications is to the analysis of the tails of probability distributions.

One indicator of QoS is loss probability, which can be analyzed using the LDT, by a transformation of the large deviations rate function called the *effective bandwidth*. Of course, the loss probability could be also explicitly calculated for traffic models that are sufficiently simple. The problem is that for many traffic models the calculations are extremely difficult. Thus, having a generic result for the loss probability makes the Effective Bandwidth Theory (EBT) very appealing for its application in the field of queueing systems. Besides, the tail of the probability of the delay in the queue can be directly obtained from the loss probability.

In this Appendix the fundamentals of the LDT and its connection with the EBT are revised.

A.1 Large Deviations Theory

Log-moment generating function

There is a close relation between the cumulants of a distribution and its moments. The first cumulant is simply the mean, m; the second cumulant is the variance σ^2 (the second moment less the square of the first moment). The relationship between the higher cumulants and the moments gets more complicated, but in general the kth cumulant can be written in terms of the first k moments.

The generating functions simplify the relationship between the moments and the cumulants.

- The function $M(\theta) = \mathrm{E}[e^{\theta X}]$ is the Moment Generating Function (MGF) of X. The kth moment is found by taking the kth derivative of ϕ and evaluating it at $\theta = 0$:

$$\begin{aligned} &\frac{d^k}{d\theta^k} M(\theta) = \mathrm{E}[X^k e^{\theta X}] \\ &\frac{d^k M(\theta)}{d\theta^k} \mid_{\theta=0} = \mathrm{E}[X^k] = k\text{th moment} \end{aligned} \tag{A.1.1}$$

- The log-moment generating function (L-MGF) or cumulant generating function is defined to be the logarithm of the MGF:

$$\Lambda(\theta) = \log \mathrm{E}[e^{\theta X}] \tag{A.1.2}$$

And the cumulants are just the derivatives of $\Lambda(\theta)$:

$$\begin{aligned} &\frac{d}{d\theta} \Lambda(\theta) = m \\ &\frac{d^2}{d\theta^2} \Lambda(\theta) \mid_{\theta=0} = \sigma^2 \ \ ... \end{aligned} \tag{A.1.3}$$

Legendre transform

For a function $\Lambda(\theta) : \mathbb{R} \mapsto \mathbb{R}$, the Legendre transform of $\Lambda(\theta)$ is defined [Chang 2000]:

$$\Lambda^*(a) = \sup_{\theta} [\theta a - \Lambda(\theta)] \tag{A.1.4}$$

The Legendre transform is a convex function, and thereby it is also called the convex transform. It attains its minimum at the mean $\mathrm{E}[X]$ (Figure A.1).

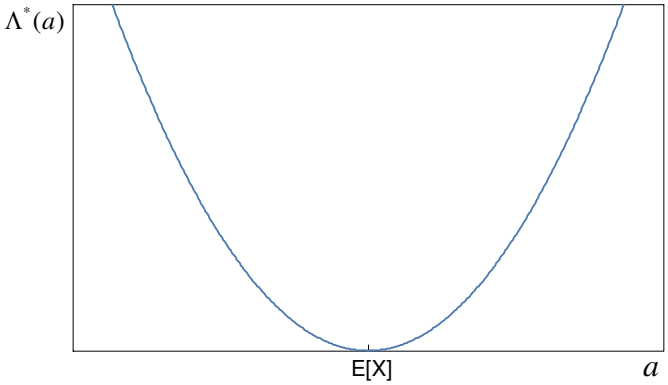


Figure A.1: Legendre Transform.

The Central Limit Theorem

The Central Limit Theorem (CLT) [Papoulis 2002] gives information about the behaviour of a probability distribution near its mean.

Let $X_1, X_2, ..., X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables with mean μ and finite variance σ^2. The CLT states that as the sample size n increases the distribution of the sample average of these random variables approaches the normal distribution with mean μ and variance σ^2/n, irrespective of the shape of the common distribution of the individual terms X_i. Therefore, the probability density function converges to:

$$f(x) = \frac{1}{\sqrt{\frac{2\pi\sigma^2}{n}}} e^{-\frac{n}{2}\frac{(x-\mu)^2}{\sigma^2}} \tag{A.1.5}$$

The CLT gives only an asymptotic distribution. As an approximation for a finite number of observations, it provides a reasonable approximation only when close to the peak of the normal distribution; it requires a very large number of observations to stretch into the tails. The approximation is only valid for x within about $\sigma/\sqrt{n}$ of μ.

Cramér's theorem

In many situations we are concerned with rare events out on the tail of a probability distribution. Cramér was looking for a refinement of the CLT and here what he proved [Chang 2000]:

Theorem. *Let X_n be a sequence of i.i.d. random variables and set $S_n = X_1 + ... + X_n$.*

Define the log-moment generating function:

$$\Lambda(\theta) = \log E[e^{\theta X_n}] \tag{A.1.6}$$

Let $\Lambda^*(a)$ be the Legendre transform of $\Lambda(\theta)$:

$$\Lambda^*(a) = \sup_{\theta}[\theta a - \Lambda(\theta)] \tag{A.1.7}$$

Then, the sequence of random variables S_n/n satisfies a Large Deviation Principle with rate function Λ^*:

- *for all closed sets F:*

$$\limsup_{n\to\infty} \frac{1}{n} \log Pr(S_n \in F) \leq - \inf_{a\in F} \Lambda^*(a) \tag{A.1.8}$$

- *and for all open sets G:*

$$\liminf_{n\to\infty} \frac{1}{n} \log Pr(S_n \in G) \geq - \inf_{a\in G} \Lambda^*(a) \tag{A.1.9}$$

The inequality (A.1.8) is usually referred to as the large deviation upper bound, and (A.1.9) is the large deviation lower bound. If both hold we say that the sequence S_n/n satisfies a Large Deviation Principle with rate function $\Lambda^*(a)$.

Let G be an open set G and $\overline{G}$ its closure. If:

$$\inf_{a\in G} \Lambda^*(a) = \inf_{a\in \overline{G}} \Lambda^*(a) \tag{A.1.10}$$

then from Cramér's theorem:

$$\lim_{n\to\infty} \frac{1}{n} \log Pr(S_n \in G) = \lim_{n\to\infty} \frac{1}{n} \log Pr(S_n \in \overline{G}) = - \inf_{a\in G} \Lambda^*(a) \tag{A.1.11}$$

In particular, if $G = (a, \infty)$ and $\overline{G} = [a, \infty)$, with $a \geq E[X]$, then:

$$\lim_{n\to\infty} \frac{1}{n} \log Pr(S_n > a) = \lim_{n\to\infty} \frac{1}{n} \log Pr(S_n \geq a) = -\Lambda^*(a) \tag{A.1.12}$$

Remark that $\Lambda^*(a)$ is increasing for $a \geq \mathrm{E}[X]$ since $\Lambda^*(a)$ is convex and attains its minimum at $a = \mathrm{E}[X]$.

In summary, under appropriate conditions Cramér's theorem provides the asymptotic behaviour:

$$Pr(S_n > a) \asymp e^{-n\Lambda^*(a)} \tag{A.1.13}$$

where $f(x) \asymp g(x)$ means that $\lim\limits_{x\to\infty} f(x)/g(x) = 1$

Gärtner-Ellis theorem

The extension of Cramér's theorem to sequences of not necessarily independent random variables is known as the Gärtner-Ellis theorem [Chang 2000]. The main difference lies in the assumptions of the theorem:

Consider a sequence of random variables $Y_1, Y_2, ..., Y_n$.

$\Lambda_n(\theta)$ is the log-moment generating function:

$$\Lambda_n(\theta) = \frac{1}{n} \log \mathrm{E}[e^{\theta Y_n}] \tag{A.1.14}$$

Theorem. *Assume:*

- $\limsup \frac{1}{n}\Lambda_n(\theta) = \Lambda(\theta) < \infty$ *for all* $\theta \in \mathbb{R}$
- $\Lambda(\theta)$ *is differentiable for all* $\theta \in \mathbb{R}$
- $\Lambda^*(a)$ *is the Legendre transform of* $\Lambda(\theta)$

Then, for all closed sets F:

$$\limsup_{n\to\infty} \frac{1}{n} \log Pr\left(\frac{Y_n}{n} \in F\right) \leq -\inf_{a\in F} \Lambda^*(a) \tag{A.1.15}$$

and for all open sets G:

$$\liminf_{n\to\infty} \frac{1}{n} \log Pr\left(\frac{Y_n}{n} \in G\right) \geq -\inf_{a\in G} \Lambda^*(a) \tag{A.1.16}$$

A.2 Large deviations in queueing systems

The results of Cramér's and Gärtner-Ellis theorems can be applied to a queueing system [O'Connell 1999]. Consider a single-server queue with the discipline to be First Come First Served (FCFS). Let $a[n]$ be the amount of work brought by customers arriving at time n, and $c[n]$ the amount of work the server can do at time

n. Further, $x[n]$ is the difference between $a[n]$ and $c[n]$. The queue length (i.e. the amount of work remaining in the queue) $Q[n]$ is given by Lindey's recursion:

$$Q[n] = (Q[n-1] + x[n])^+ \tag{A.2.1}$$

In a queueing system, it is interesting the study of the equilibrium behaviour of the queue: what the queue-length is when the system has been running for a very long time and the initial queue-length has no influence. If the work process and the service process are stationary, then the equilibrium queue length is given by:

$$Q(\infty) = \max_{n \geq 0} W[n]^+ \tag{A.2.2}$$

where $W[n] = x[1] + x[2] + ... + x[n]$ is the workload process.

For many purposes, we may be interested in the behaviour of the tail of the probability distribution, i.e. the probability that the queue length Q exceeds a certain value B. Let $\Lambda_n(\upsilon) = \log \mathrm{E}[e^{\upsilon W[n]}]$ be the log-moment generating function of the workload process. Define the limiting scaled cumulant generating function to be:

$$\Lambda(\upsilon) = \lim_{n \to \infty} \Lambda_n(\upsilon)/n = \lim_{n \to \infty} \frac{1}{n} \log \mathrm{E}[e^{\upsilon W[n]}] \tag{A.2.3}$$

It can be proved [Botvich 1995] [O'Connell 1999] that if the workload process satisfies a LDP:

$$Pr(X[n]/n > x) \asymp e^{-n\Lambda^*(x)}, \tag{A.2.4}$$

with $\Lambda^*(x)$ the Legendre transform of $\Lambda(\upsilon)$, then, for large values of B:

$$Pr(Q(\infty) > B) \asymp e^{-\theta B} \tag{A.2.5}$$

and the decay-constant θ, known as the *QoS exponent*, is the solution to:

$$\theta = \min_x \frac{\Lambda^*(x)}{x} \tag{A.2.6}$$

More interesting is to characterize θ in terms of the L-MGF:

$$\begin{aligned} \upsilon \leq \min_x I(x)/x \text{ if and only if } & \upsilon \leq I(x)/x \ \forall x \\ \text{if and only if } & \upsilon x - I(x) \leq 0 \ \forall x \\ \text{if and only if } & \upsilon x - I(x) \leq 0 \ \forall x \\ \text{if and only if } & \max_x \{\upsilon x - I(x)\} \leq 0 \end{aligned} \tag{A.2.7}$$

Finally, $\upsilon \leq \theta$ if and only if $\Lambda(\upsilon) \leq 0$ and so θ is the solution to:

$$\Lambda(\upsilon) \mid_{\upsilon=\theta} = 0 \tag{A.2.8}$$

Recall that $x[n]$ is the difference $a[n] - c[n]$. If we assume that the arrival process is independent of the service process, then $\Lambda(\upsilon) = \Lambda_A(\upsilon) + \Lambda_C(\upsilon)$, with:

$$\Lambda_A(\upsilon) = \lim_{n\to\infty} \frac{1}{n} \log \mathrm{E}\left[\exp\left(\upsilon \sum_{k=1}^{n} a[k]\right)\right] \tag{A.2.9}$$

$$\Lambda_C(\upsilon) = \lim_{n\to\infty} \frac{1}{n} \log \mathrm{E}\left[\exp\left(\upsilon \sum_{k=1}^{n} c[k]\right)\right] \tag{A.2.10}$$

Effective bandwidth

The concept of effective bandwidth of an arrival process comes from the asymptotic tail probability in (A.2.5). For a single server queue with arrival process $a[n]$ and constant service capacity c we may ask which service capacity will ensure:

$$Pr(Q(\infty) > B) \leq e^{-\theta B} \tag{A.2.11}$$

for large B and some specified value of θ. Define the effective bandwidth of the arrival process to be:

$$\alpha_A(\upsilon) = \frac{\Lambda_A(\upsilon)}{\upsilon} \tag{A.2.12}$$

Then, under appropriate conditions, the service capacity c that ensures (A.2.11) for a given QoS exponent is just the effective bandwidth of the source evaluated in θ:

$$c = \alpha_A(\theta) \tag{A.2.13}$$

Similarly, the effective bandwidth function of the service process is defined as:

$$\alpha_C(\upsilon) = \frac{\Lambda_C(\upsilon)}{\upsilon} \tag{A.2.14}$$

Appendix B

Channel Simulation

In the literature, many methods to generate realizations of the wireless channel are available. A typical wireless channel is a time-varying system in which the parameters are random and liable to change with time. The final aim is to obtain an ergodic method able to approximate as much as possible the statistics of the fading channel. In general, the goodness of a generator is mainly evaluated by the accuracy of the approximation of the first and second order statistics.

Results obtained in this thesis have been validated by confrontation with computer simulations, including a model of the Rayleigh channel. Thus, an adequate simulation of the communication channel is essential. Several Rayleigh generators have been evaluated in this thesis, on account of its vital importance in the validation of the results.

The rest of the appendix is organized as follows. In Section B.1 a brief description of the wireless channel is given. The examined generators are detailed in Sections B.2, B.3, B.4 and B.5. Finally, Section B.6 compares the performance of the different methods.

The work presented in this Appendix has been partially published in [Ruiz 2009].

B.1 Multipath propagation

The wireless channel is much more hostile than the AWGN channel which is usually the channel model assumed for most wired communication systems. 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 [Rappaport 2002] [Goldsmith 2005]. 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. 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.

B.1.1 Representation of the fading process

In the case of flat fading, the effect of the channel over the transmitted signal can be represented through the low-pass equivalent as in Figure B.1.

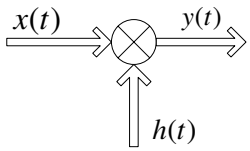


Figure B.1: Low-pass equivalent of the flat fading channel.

$x(t)$ 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 channel gain:

$$h(t) = h_{phase}(t) + jh_{quad}(t) \tag{B.1.1}$$

The phase (h_{phase}) and quadrature (h_{quad}) components are independent of each other and Gaussian distributed, with probability density function (pdf) $f_G(h)$, mean zero and variance b_0:

$$f_G(x) = \frac{1}{\sqrt{2\pi b_0}} e^{-\frac{x^2}{2b_0}} \tag{B.1.2}$$

The cumulative distribution function (CDF) can be expressed:

$$F_G(x) = \int_{-\infty}^{x} f_G(u)du = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x}{\sqrt{2b_0}}\right)\right] \tag{B.1.3}$$

with the error function:

$$\text{erf}(y) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-t^2} dt \tag{B.1.4}$$

The envelope of the fading channel, calculated from the phase and quadrature components as $r(t) = \sqrt{h_{phase}^2(t) + h_{quad}^2(t)}$, is Rayleigh distributed, due to the

underlying Gaussian components. We denote its pdf $f_R(r)$:

$$f_R(r) = \frac{r}{b_0} e^{-\frac{r^2}{2b_0}} \qquad r \geq 0 \tag{B.1.5}$$

It is a uniparametric function whose mean, variance and root mean square can be expressed in terms of b_0:

$$\begin{aligned} \mathrm{E}[r] &= \sqrt{\frac{b_0 \pi}{2}} \\ \sigma_r^2 &= \frac{4-\pi}{2} b_0 \\ \mathrm{E}[r^2] &= 2b_0 \end{aligned} \tag{B.1.6}$$

The CDF is obtained as:

$$F_G(r) = \int_{-\infty}^{r} f_R(u)du = 1 - e^{-\frac{r^2}{2b_0}} \tag{B.1.7}$$

Finally, 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 from the transmitter to the 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.

The instantaneous SNR is proportional to the square of $|h(t)|$:

$$\gamma(t) = |h(t)|^2 \frac{E_s}{N_0} \tag{B.1.8}$$

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 pdf:

$$f_E(\gamma) = \frac{1}{\overline{\gamma}} e^{-\frac{\gamma}{\overline{\gamma}}} \tag{B.1.9}$$

where $\overline{\gamma}$ is the average Signal to Noise Ratio. The mean, variance and root mean

square are:

$$\begin{aligned} \mathrm{E}[\gamma] &= \bar{\gamma} \\ \sigma_\gamma^2 &= \bar{\gamma}^2 \\ \mathrm{E}[\gamma^2] &= 2\bar{\gamma}^2 \end{aligned} \tag{B.1.10}$$

The CDF yields:

$$F_E(\gamma) = \int_{-\infty}^{\gamma} f_E(u)du = 1 - e^{-\frac{\gamma}{\bar{\gamma}}} \tag{B.1.11}$$

Autocorrelation Function

An important 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 gain 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:

$$T_C \approx \frac{1}{f_D} \tag{B.1.12}$$

The variability of the 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. The isotropic scattering is a common assumption that means that the channel consists of many

scatterers densely packed with respect to angle. This approach was introduced by Clarke [Clarke 1968] and further developed by Jakes [Jakes 1989], so it is usually referred to as Clarke's model, Jakes' model or *classical* model. The channel gain is a wide-sense stationary random process and the continuous-time ACF of the phase and quadrature components does not depend on the time t but just on the time difference τ:

$$R_{h_{phase}}(\tau) = R_{h_{quad}}(\tau) = b_0 J_0(2\pi f_D \tau); \tag{B.1.13}$$

where $J_0(.)$ is the zeroth order Bessel function of the first kind and f_D is the maximum Doppler frequency in Hertz. The Bessel functions are canonic solutions of the Bessel differential equation. In the case of zero-order first kind functions:

$$J_0(x) = \sum_{i=0}^{\infty} (-1)^i \frac{\frac{x}{2}^2 i}{(i!)^2} \tag{B.1.14}$$

Bessel functions are oscillating functions that decay proportionally with $1/\sqrt{x}$, although the zeros of the functions are not periodic, except asymptotically when $x \to \infty$.

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 ($f_D \cdot T_S$):

$$R_{h_{phase}}[m] = R_{h_{quad}}[m] = b_0 J_0(2\pi f_D T_S m); \tag{B.1.15}$$

A low value of the product $f_D \cdot T_S$ implies high correlation in the signal and, likewise, a high value of $f_D \cdot T_S$ means low correlation. In the limit ($f_D \cdot T_S \to \infty$) there is no correlation and the samples are independent of each others.

Figure B.2 shows the ACF function of the in-phase and quadrature components for a product $f_D \cdot T_S$ equal to 0.01 and $b_0 = 1$.

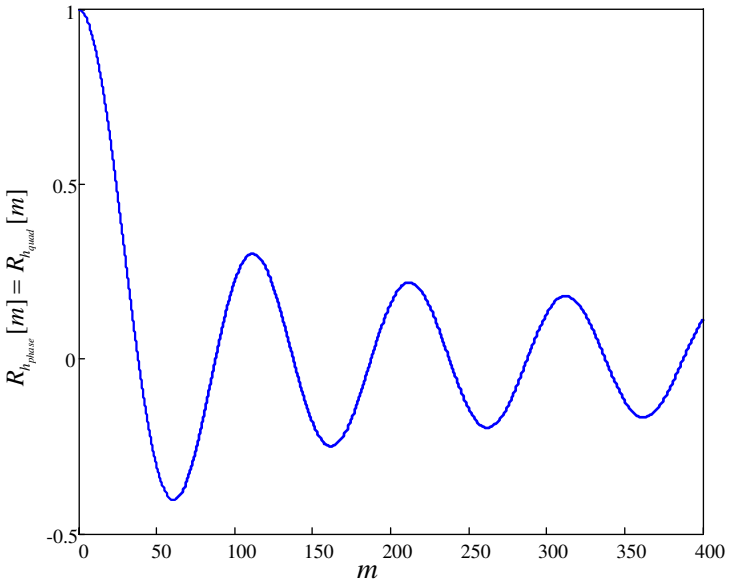


Figure B.2: ACF of the in-phase and quadrature components.

If we are working with the Rayleigh envelope, the expression of the ACF, calculated from (B.1.13), is as follows [Tan 2000]:

$$R_r[m] = \frac{\pi b_0}{2} {}_2\mathbf{F}_1\left(-\frac{1}{2}, -\frac{1}{2}; 1; (J_0(2\pi f_D m))^2\right); \tag{B.1.16}$$

where ${}_p\mathbf{F}_q(\boldsymbol{n}, \boldsymbol{d}, z)$ is the hypergeometric function. The ACF of the envelope is represented in Figure B.1.1, for a product $f_D \cdot T_S$ equal to 0.01 and $b_0 = 1$.

The hypergeometric function can be expanded as an infinite series. Thus, the ACF of the envelope can be rewritten:

$$R_r[m] = \frac{\pi b_0}{2} \sum_{i=0}^{\infty} \frac{(J_0(2\pi f_D m))^2 i}{2^{(i+1)!}}; \tag{B.1.17}$$

A common approximation is to consider the first two terms of the series. It yields:

$$R_r[m] \approx \frac{\pi b_0}{2}\left[1 + \frac{1}{4}(J_0(2\pi f_D m))^2\right]; \tag{B.1.18}$$

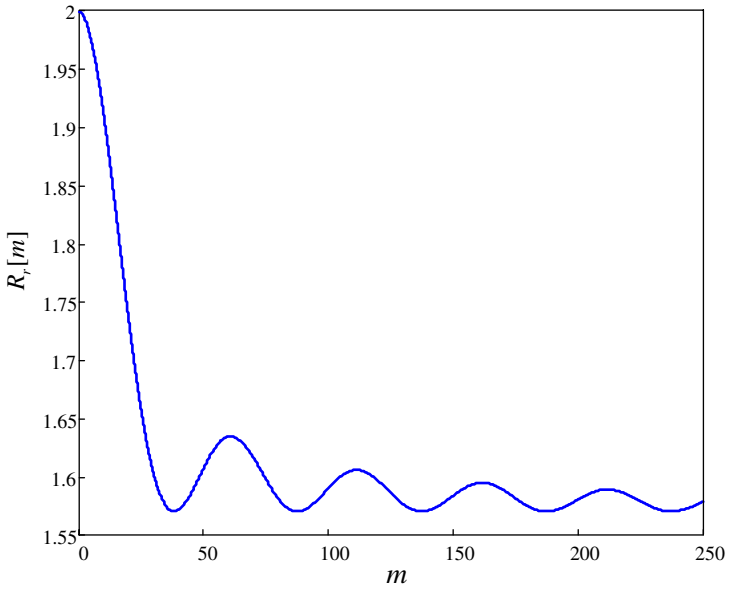


Figure B.3: ACF of the envelope.

Finally, the ACF of the Signal to Noise Ratio (Figure B.1.1, average SNR of 5dB and $f_D \cdot T_S = 0.01$) is:

$$R_\gamma(\tau) = \gamma_0^2 + (\gamma_0 J_0(2\pi f_D \tau))^2; \tag{B.1.19}$$

The autocorrelation function is often given in the frequency domain. The power spectral density (PSD) is obtained by taking the Fourier transform of $R(\tau)$ relative to the time parameter τ. In the case of the in-phase and quadrature components (Figure B.1.1):

$$S_{h_phase}(f) = S_{h_quad}(f) = \begin{cases} \frac{1}{\pi f_d \sqrt{1-\left(\frac{f}{f_d}\right)^2}}, & |f| \leq |f_d| \\ 0 & elsewhere \end{cases} \tag{B.1.20}$$

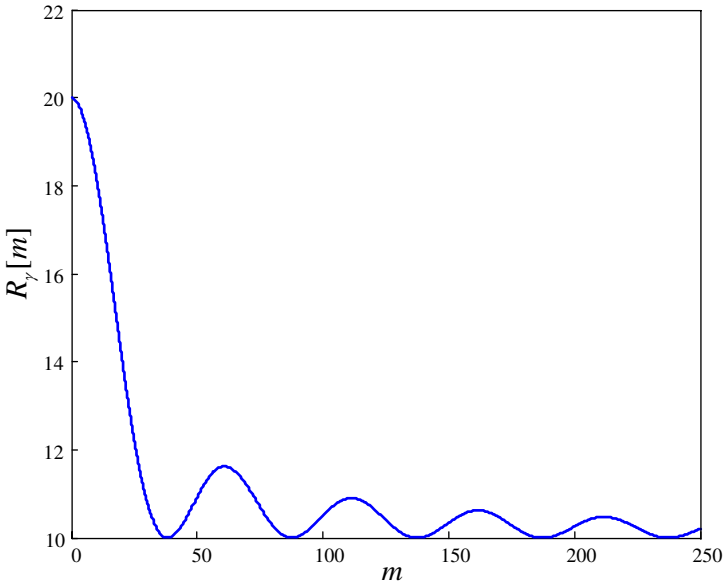


Figure B.4: ACF of the Signal to Noise Ratio.

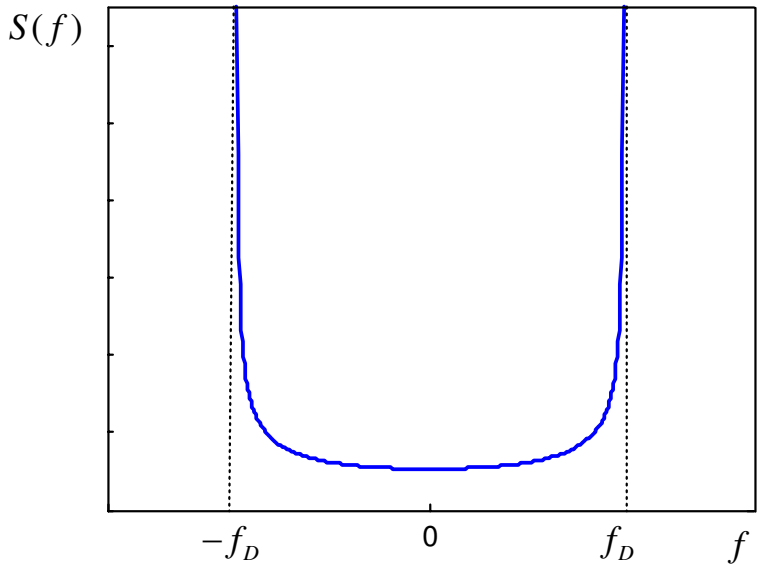


Figure B.5: In-Phase and Quadrature PSD.

B.1.2 Generation of correlated Rayleigh random variates

A typical wireless channel is a time-varying system in which the parameters are random and liable to change with time. When simulating a mobile wireless channel for communication systems, it is usually assumed that the fading process is a random variate with Rayleigh distribution. Moreover, the fading process should also have other properties, like autocorrelation, spectrum, etc., and with an execution time (time necessary to get a realization of X samples) as small as possible. A generator that provides a realization of the random process able to accurately approximate the main statistics with a small computational effort is a challenging task. There is a huge literature on generators of the Rayleigh fading process. Due to this diversity, different simulations of the same communication system can yield different results depending on the selected Rayleigh generator.

The recent literature on Rayleigh generators can be classified into four approaches:

- Generators based on FIR filtering of Gaussian noise through the Doppler spectrum filtering. The Rayleigh variate is generated by filtering two zero-mean independent white Gaussian processes and then adding the outputs in quadrature. Here, rational transfer function approximations of the PSD are typically used to shape the spectrum.

- Sum of Sinusoids generators. The in-phase and quadrature components are generated as the sum of N complex sinusoidal generators. In practice, the generated sequence closely approximates a complex Gaussian process provided a sufficient number of sinusoids are used. With proper choice of the distribution of the sinusoid frequencies, the ACF approaches the desired function.

- IDFT generators. The IDFT operation is applied to sequences of uncorrelated complex Gaussian variates, each sequence weighted by appropriate filter

coefficients to shape the PSD.

- AR generators. An autoregressive model is applied to approximate the random processes.

B.2 FIR filtering

The first group of methods are the generators based on FIR (Finite Impulse Response) filtering of Gaussian noise. The generator implemented by MATLAB is one of them. Gaussian noise is filtered through the Doppler filter [Rappaport 2002] [Jakes 1989], following the block diagram in Figure B.6. Two independent Gaussian low-pass noise sources are used to produce the in-phase and quadrature components. Then, they are introduced in the Doppler filter defined in (B.1.20).

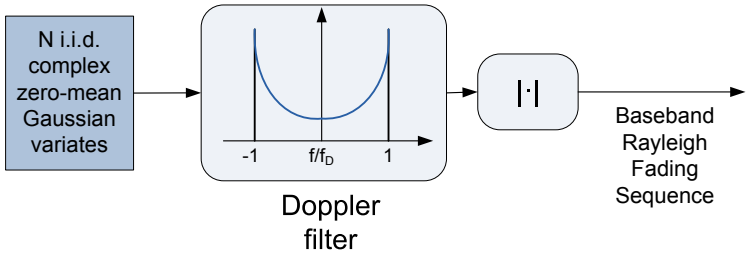


Figure B.6: MATLAB generator.

B.3 Sum Of Sinusoids generator

One of the most common procedures for the generation of Rayleigh distributed variates is the Sum of Sinusoids [Pop 2002] [Zheng 2002]. The principle behind is to express the fading process as the superposition of an infinite number of weighted

sinusoids with equidistant frequencies and random phases. Using this principle, the underlying discrete-time Gaussian process for the in-phase and quadrature components can be modeled as:

$$h_{phase}[n] = \sum_{i=1}^{N} C_i \cos(2\pi f_D T_S n \cos\alpha_i + \phi_i) \tag{B.3.1}$$

$$h_{quad}[n] = \sum_{i=1}^{N} C_i \sin(2\pi f_D T_S n \cos\alpha_i + \varphi_i) \tag{B.3.2}$$

where b_0 is the mean power and C_i, ϕ_i, φ_i and α_i are mutually independent random variables, representing the amplitude, initial phase and angle of incoming wave associated with the ith propagation path, respectively.

The main difficulty in implementing Clarke's model is not on having infinite sinusoids but in the randomness of each of them. Thus, the best method would be that with less sinusoids but able to approximate the randomness of the model.

A way to noticeably reduce the complexity of the model is to consider some of the random variables as constants. A simple and efficient simulator design results from setting:

$$\begin{aligned} N &= 4M \\ C_i &= \frac{1}{\sqrt{N}} \\ \alpha_i &= \frac{2\pi i - \pi + \theta_i}{4M} \end{aligned} \tag{B.3.3}$$

where M is the number of sinusoids of the model and θ_i are random variables uniformly distributed on $[-\pi, \pi)$.

Including the expressions above and normalizing for unit power, $g(t)$ becomes:

$$h_{phase}[n] = \sqrt{\frac{2b_0}{M}} \sum_{i=1}^{N} \cos(2\pi f_D T_S n \cos \frac{2\pi i - \pi + \theta_i}{4M} + \phi_i) \tag{B.3.4}$$

$$h_{quad}[n] = \sqrt{\frac{2b_0}{M}} \sum_{i=1}^{N} \cos(2\pi f_D T_S n \sin \frac{2\pi i - \pi + \theta_i}{4M} + \varphi_i) \tag{B.3.5}$$

B.4 IDFT generator

A method based on the Inverse Discrete Fourier Transform (IDFT) was first presented by Smith as FORTRAN code in [Smith 1975]. Since then, it has been widely utilized in simulations of wireless systems together with the SoS generators. Young and Beaulieu [Young 2000] modified Smith's algorithm for greater computational efficiency and provided a statistical analysis of the method. This second version is the one implemented in this thesis.

The block diagram of the algorithm is shown in Figure B.7. The IDFT operation is applied to complex sequences of independent, normally distributed random numbers, each sequence weighted by appropriate filter coefficients F_k. Therefore, these coefficients are responsible of setting the statistical properties to the final sequence.

To get a realization of N samples, N coefficients of the filter are required, with the following expression:

$$F[k] = \begin{cases} 0, & k = 0, \\ \sqrt{\dfrac{1}{2\sqrt{1-\left(\frac{k}{Nf_D}\right)^2}}}, & k = 1, 2, ..., K_m - 1 \\ \sqrt{\dfrac{K_m}{2}\left[\dfrac{\pi}{2} - \arctan\left(\dfrac{K_m - 1}{\sqrt{2K_m - 1}}\right)\right]}, & k = K_m \\ 0, & k = K_m + 1, ..., N - K_m - 1, \\ \sqrt{\dfrac{K_m}{2}\left[\dfrac{\pi}{2} - \arctan\left(\dfrac{K_m - 1}{\sqrt{2K_m - 1}}\right)\right]}, & k = N - K_m \\ \sqrt{\dfrac{1}{2\sqrt{1-\left(\frac{N-k}{Nf_D}\right)^2}}}, & k = N - K_m + 1, ..., N - 1 \end{cases} \tag{B.4.1}$$

where $K_m = \lfloor Nf_D \rfloor$.

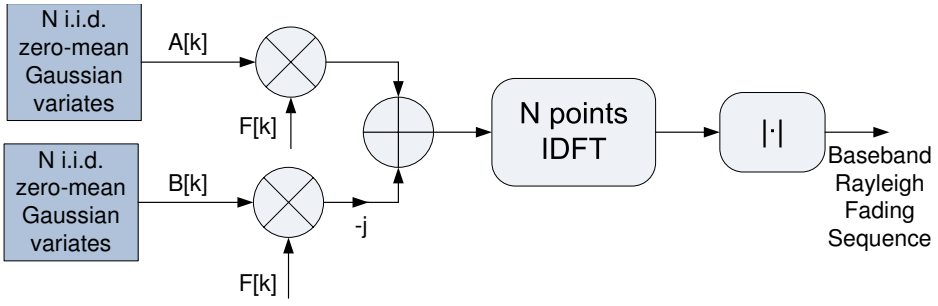


Figure B.7: IDFT generator.

B.5 AR generator

Auto-Regressive (AR) models [Baddour 2005] are commonly used to approximate discrete-time random processes, due due to the simplicity with which their parameters can be computed and their correlation matching property. In the time domain, a complex AR process of order p, AR(p), can be generated via the recursion:

$$x[n] = -\sum_{k=1}^{p} a_k x[n-k] + w[n] \tag{B.5.1}$$

where $w[n]$ is a complex white Gaussian noise process with uncorrelated real and imaginary components and zero mean. The AR model parameters consist of the filter coefficients $a_1, a_2, ..., a_p$ and the variance σ_p^2 of the noise process. These parameters are related with the desired ACF for the output process through the Yule-Walker equations:

$$R_{xx}[m] = \begin{cases} -\sum_{k=1}^{p} a_k R_{xx}[m-k] & m > 0 \\ -\sum_{k=1}^{p} a_k R_{xx}[m-k] + \sigma_w^2 & m = 0 \\ R_{xx}[m-k] & m < 0 \end{cases} \tag{B.5.2}$$

In matrix form it becomes:

$$\begin{bmatrix} R_{xx}[0] & R_{xx}[-1] & \dots & R_{xx}[-p] \\ R_{xx}[1] & R_{xx}[0] & \cdots & R_{xx}[-p+1] \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}[p] & R_{xx}[p-1] & \cdots & R_{xx}[0] \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma_w^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{B.5.3}$$

Figure B.8 shows the block diagram of the AR generator, which produces the correlated variates by filtering white Gaussian noise sources.

The Yule-Walker equations in (B.5.2) can be solved efficiently by the Levinson-Durbin recursion, which makes use of the symmetry of the correlation matrix. The condition of the correlation matrix comes out as an important consideration in determining the accuracy of the solution. In the case of the Clarke's ACF function,

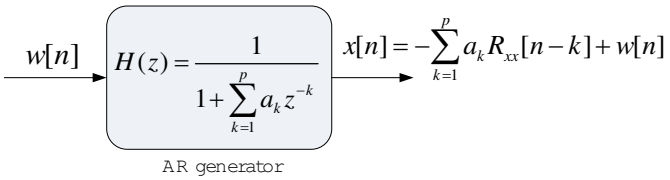


Figure B.8: AR generator.

the matrix is not well-conditioned. A simple heuristic approach that can be used to solve this problems is to increase the values along its principal diagonal by a very small positive amount ϵ.

B.6 Comparisons

In this section, several parameters are measured in order to compare the performance of the methods.

B.6.1 Comparisons based on the first and second moments

Table B.1 shows the measured mean and variance of the four methods, with the theoretical ones, which have been set to be 0 for the mean and 2 for the variance. The SoS generator is using 15 sinusoids and the AR is of order 100. The sampled mean and variance are measured over a realization of 2^{32} samples. As expected, all the methods are able to accurately approximated the desired statistics.

Table B.1: A comparison of the different Rayleigh channel generators: mean and variance

	$\overline{r}$	σ_r
MATLAB	1.2548	1.9874
SoS	1.2482	1.9659
IDFT	1.2485	1.9525
AR	1.2456	1.8652
Theoretical	1.2533	2

B.6.2 Comparisons based on the ACF

The ACF of the generators is shown in Figure B.9. Moreover, we have evaluated the power margin quality measures which are commonly employed as parameter of quality performance. In particular, the mean basis power margin and the maximum basis power margin are defined as [Young 2003]:

$$\begin{aligned} \mathcal{G}_{\text{mean}} &= \frac{1}{\sigma_{\text{g}}^2 L}\text{trace}(C_{\text{g}} C_{\widehat{\text{g}}}^{-1} C_{\text{g}}) \\ \mathcal{G}_{\text{max}} &= \frac{1}{\sigma_{\text{g}}^2}\text{max}(\text{diag}(C_{\text{g}} C_{\widehat{\text{g}}}^{-} 1 C_{\text{g}})) \end{aligned} \tag{B.6.1}$$

where $\sigma_{\text{g}}^2 = b_0$ is the variance of the theoretical distribution, L is the number of points used in the calculation of the metrics, C_{g} is the LxL covariance matrix of L neighboring samples of the theoretical fading process, and $C_{\widehat{\text{g}}}$ is the LxL covariance matrix of L neighboring samples of the fading model. The results are shown in Table B.2, for the partitions that approximate better Clarke's model: uniform and equiprobable. Perfect performance corresponds to 0 dB for both measures. An autocorrelation sequence length of 300 was considered with the same value of $f_D \cdot T_S$ fixed above. The measures of the different methods with 15 sinusoids in the SoS generator.

Both the figure and the power margin quality measures show the poor performance of the MATLAB and the AR method in approximating the ACF beyond the first lobes of Bessel function. It is also observed that the IDFT method is the best

in terms of ACF approximation.

Table B.2: Power Margin Quality Measure comparison

Method	$\mathcal{G}_{\text{mean}}$(dB)	$\mathcal{G}_{\text{max}}$(dB)
MATLAB channel	0.0383	0.0403
IDFT	0.0001	0.0001
SoS with 15 sinuoids	0.0016	0.0018
AR(100)	0.0219	0.0284

B.6.3 Comparisons based on execution time

To conclude the performance evaluation, the computational effort to generate samples using different methods is observed.

One advantage of both the direct FIR filtering and the SoS method is that samples can be generated as they are needed. In contrast, the IDFT method requires that all samples be generated using a single FFT operation. However, the reduced storage requirements of the former two methods come at the expense of overall computational effort and/or variate quality.

The time to generate 5e6 complex samples is shown in Table B.3. The best method in terms of computation time is the IDFT, which needs 10 times less computation time than the MATLAB method.

Table B.3: Computation Time to generate 5e6 complex samples

Method	Time (s)
MATLAB channel	5.64
SoS with 15 sinuoids	3.91
IDFT	0.40
AR	1.72

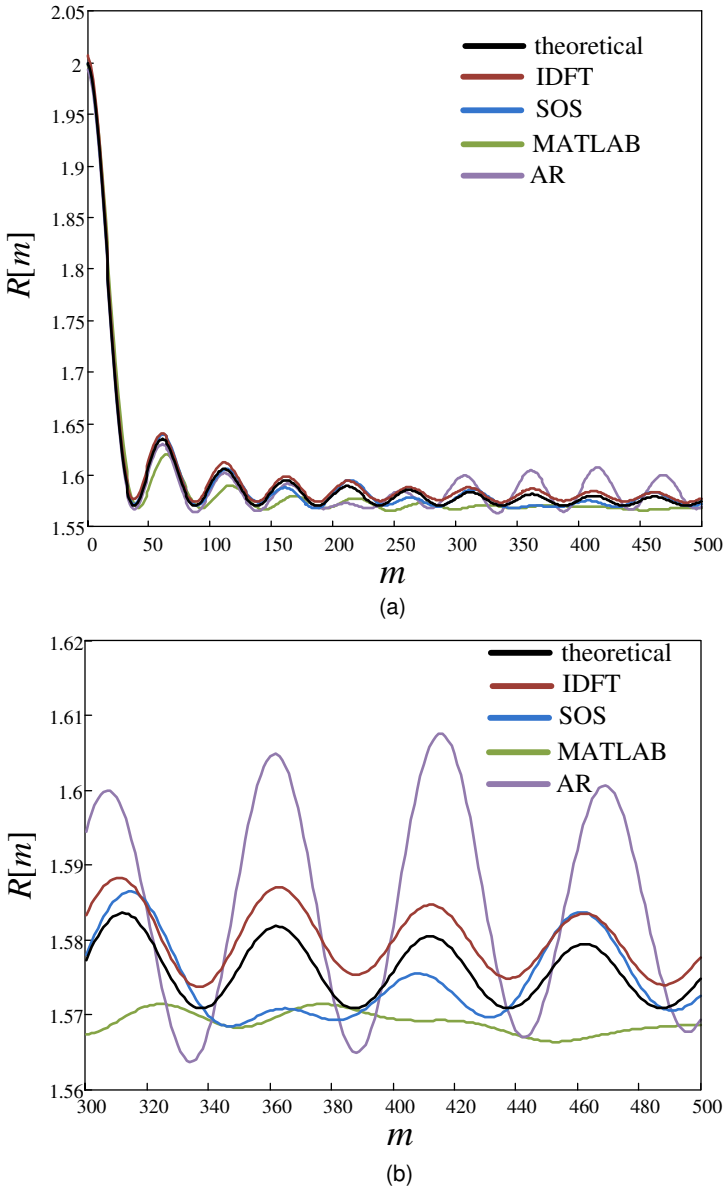


Figure B.9: (a) Comparison of the ACF for the different methods for Rayleigh fading generation (b) Zoom of (a)

Appendix C

Finite State Markov Chain

In late 1950s and early 1960s, Gilbert and Elliot at Bell Labs were modeling burst-noise telephone circuits with a very simple two-state channel model with memory. Despite its simplicity, the model made it possible to evaluate channel capacity and error rate performance through bursty wireline telephone circuits. However, it took another 30 years for the so-called Gilbert-Elliott channel (GEC) and its generalized Finite State Markov Chain (FSMC) to be applied in the design of second generation wireless communication systems. Since the mid 1990s, the GEC and FSMC models have been widely used for modeling wireless flat fading channels in a variety of applications, ranging from modeling channel error bursts to decoding at the receiver. FSMC models can capture, with suitable choices of model parameters, the essence of time-varying fading channels while avoiding complex mathematical formulations.

The use of finite state Markov chains (FSMC) for the simulation of the Rayleigh channel has been generalized in the last years [Shen 1995] [Zhang 1999] [Tan 2000] [Bergamo 2002] [Arauz 2004]. In this appendix, we detail the characteristics and parameters of a FSMC and present an overall evaluation of the model. The work presented in this Appendix has been partially published in [Ruiz 2009].

In Section C.1 the different parameters of the chain are detailed. In Section C.2

first and second order statistics and other performance parameters are examined in order to shed more light on the advantages and limitations of the Markov modeling.

C.1 Parameters of the model

Several parameters influence the construction of the chain. The chain can be defined in terms of the complex channel response, the envelope of the received signal or the Signal to Noise Ratio. Moreover, the partition method has to be selected among various options.

Steady state probabilities

In a Finite State Markov Chain [Shen 1995] [Zhang 1999] [Tan 2000] [Bergamo 2002] [Arauz 2004], the range of the amplitude of the signal is divided into several consecutive regions. Region i is mapped into state i of the chain and is delimited by two thresholds, μ_i and μ_{i+1}. Thus, each state of the chain represents one fading region. The steady state probability for state i is just the probability that the received signal is between the thresholds of the region:

$$\pi_i = \int_{\mu_i}^{\mu_{i+1}} f(\mu) d\mu \tag{C.1.1}$$

where the density function $f(x)$ depends on the signal of interest: gaussian (in-phase and quadrature components), Rayleigh (envelope) or exponential (SNR), with the expressions detailed in Appendix A.

If the selected signal for the partition is the in-phase and quadrature components of the channel response, then the distribution is Gaussian and the symmetry of the pdf has to be taken into account. Thus, the first and last threshold are $\mu_1 = -\infty$

and $\mu_{n+1} = \infty$, and the remaining have to be chosen preserving the symmetry, so that the ones in the positive part of the pdf match up the ones in the negative axis with opposite sign. The steady state probabilities are:

$$\pi_i = \int_{\mu_i}^{\mu_{i+1}} f_G(x)dx = \int_{\mu_i}^{\mu_{i+1}} \frac{1}{2\pi b_0} e^{-\frac{x^2}{2b_0}} dx = \frac{1}{2}\left[\text{erf}\left(\frac{\mu_{i+1}}{\sqrt{2b_0}}\right) - \text{erf}\left(\frac{\mu_i}{\sqrt{2b_0}}\right)\right] \tag{C.1.2}$$

with the error function:

$$\text{erf}(y) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-t^2} dt \tag{C.1.3}$$

The envelope of the received signal follows a Rayleigh distribution. The thresholds are denoted δ_i. In that case, the first and last threshold are predetermined by the range of variation of a Rayleigh pdf, i.e. $\delta_1 = 0$ and $\delta_{n+1} = \infty$. The steady state probabilities are obtained:

$$\pi_i = \int_{\delta_i}^{\delta_{i+1}} f_R(r)dr = \int_{\delta_i}^{\delta_{i+1}} \frac{r}{b_0} e^{-\frac{r^2}{2b_0}} dr = e^{-\frac{\delta_i^2}{2b_0}} - e^{-\frac{\delta_{i+1}^2}{2b_0}} \tag{C.1.4}$$

The same applies when the desired parameter is the exponentially distributed received Signal to Noise Ratio, with thresholds Γ_i and state probabilities:

$$\pi_i = \int_{\Gamma_i}^{\Gamma_{i+1}} f_E(\gamma)d\gamma = \int_{\Gamma_i}^{\Gamma_{i+1}} \frac{1}{\bar{\gamma}} e^{-\frac{\gamma}{\bar{\gamma}}} d\gamma = e^{-\frac{\Gamma_i}{\bar{\gamma}}} - e^{-\frac{\Gamma_{i+1}}{\bar{\gamma}}} \tag{C.1.5}$$

Partitioning

Four different methods of partition have been examined [Ruiz 2009] [Arauz 2004]. They are described next.

- Uniform partition

  The methodology consists in defining first the desired thresholds and then deriving the state probabilities from them. In particular, the continuous pdf is divided by means of thresholds uniformly spaced, like shown in Figure C.1. The last threshold is not infinite but a finite number, so that the distribution

is truncated with that maximum finite value. Thresholds defined this way give rise to state probabilities that follow the shape of the pdf, doing a sampling of the density function.

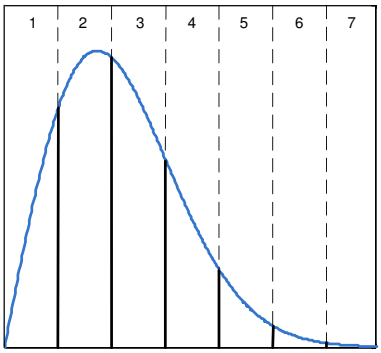


Figure C.1: Uniform partition of the Rayleigh distribution.

In the rest of methods the approach is the opposite: the state probabilities are fixed and the corresponding thresholds are obtained from them.

- Equal probabilities This is the very common approach for the partitioning of a FSMC. The thresholds are selected in such a way that the steady state probabilities of being in any state are equal:

$$\pi_1 = \pi_2 = ... = \pi_n = \frac{1}{n} \tag{C.1.6}$$

- Linearly increasing probabilities

In the linearly increasing probabilities partition, the goal is to increase linearly the probability of the states with higher amplitude. Thus, the state

probabilities are:

$$\begin{aligned}
&\text{if n even:}\\
&\pi_i = i\pi_1 \quad \text{with } \pi_1 = \frac{4}{n^2+2n} \text{ and } i = 1, 2, ..., \frac{n}{2}\\
&\text{if n odd:}\\
&\pi_i = i\pi_1 \quad \text{with } \pi_1 = \frac{4}{n^2+4n-2} \text{ and } i = 1, 2, ..., \frac{n+1}{2}
\end{aligned} \tag{C.1.7}$$

- Exponentially increasing probabilities

  This partition is similar to the partition above but with exponential increments in the probability:

$$\begin{aligned}
&\text{if n even:}\\
&\pi_{i+1} = 2\pi_i \quad \text{with } \pi_1 = \frac{1}{2^{\frac{n}{2}+1}-2} \text{ and } i = 1, 2, ..., \frac{n}{2}\\
&\text{if n odd:}\\
&\pi_{i+1} = 2\pi_i \quad \text{with } \pi_1 = \frac{1}{1+4\left(2^{\frac{n-1}{2}}-1\right)} \text{ and } i = 1, 2, ..., \frac{n+1}{2}
\end{aligned} \tag{C.1.8}$$

Output vector

Once the partition is defined, it is necessary a discrete value of the amplitude of the signal within the interval that represents the state of the chain. This value is the statistical mean between the two thresholds of the state. The set of these values is called output of the Markov chain and denoted $\{x_1, x_2, ..., x_n\}$.

In the case of the Gaussian distribution it yields:

$$x_i = \frac{\int_{\mu_i}^{\mu_{i+1}} x \cdot f_G(x)dx}{\pi_i} = \frac{1}{\pi_i}\left(\sqrt{\frac{b_0}{2\pi}}\left(e^{-\left(\frac{\mu_i^2}{2b_0}\right)} - e^{-\left(\frac{\mu_{i+1}^2}{2b_0}\right)}\right)\right) \tag{C.1.9}$$

The notation r_i is employed for the output vector of the envelope of the channel

and γ_i for the SNR. The obtained expressions are:

$$r_i = \frac{\int_{\delta_i}^{\delta_{i+1}} r \cdot f_R(r) dr}{\pi_i} = \frac{1}{\pi_i}\left[\delta_i e^{-\frac{\delta_i^2}{2b_0}} - \delta_{i+1} e^{-\frac{\delta_{i+1}^2}{2b_0}} - \sqrt{\frac{\pi b_0}{2}}\left(\text{erf}\left(\frac{\delta_i}{\sqrt{2b_0}}\right) + \text{erf}\left(\frac{\delta_{i+1}}{\sqrt{2b_0}}\right)\right)\right] \quad \text{(C.1.10)}$$

$$\gamma_i = \frac{\int_{\Gamma_i}^{\Gamma_{i+1}} \gamma \cdot f_E(\gamma) d\gamma}{\pi_i} = \frac{1}{\pi_i}\left[(\Gamma_i + \bar{\gamma}) e^{-\frac{\Gamma_i}{\bar{\gamma}}} - (\Gamma_{i+1} + \bar{\gamma}) e^{-\frac{\Gamma_{i+1}}{2b_0}}\right] \quad \text{(C.1.11)}$$

Transition probabilities

The Level Crossing Rate (LCR) is the average number of times per unit interval that a fading signal crosses a given signal level. The expressions for the in-phase and quadrature components of the channel gain (Gaussian), the envelope (Rayleigh) and the Signal to Noise Ratio (exponential) are presented in Table C.1:

Table C.1: Level Crossing Rate

$LCR_G(x) = \sqrt{2} f_D e^{-\frac{x^2}{2b_0}}$	$-\infty < x < \infty$
$LCR_R(r) = \sqrt{\frac{\pi}{b_0}} f_D r e^{-\frac{r^2}{2b_0}}$	$0 \leq r < \infty$
$LCR_E(\gamma) = \sqrt{\frac{2\pi\gamma}{\bar{\gamma}}} f_D e^{-\frac{\gamma}{\bar{\gamma}}}$	$0 \leq \gamma < \infty$

The transition probability from state i to state $i+1$ is defined as the probability that the chain is in state $i+1$ at instant k given that it was in state i at $k-1$. Mathematically, for the in-phase and quadrature components:

$$p_{i,i+1} = Pr[X_k = i+1 | X_{k-1} = i] = \frac{Pr[X_k = i, X_{k-1} = i]}{Pr[X_{K-1} = i]} = \frac{1}{\pi_i}\int_{\mu_i}^{\mu_{i+1}} \int_{\mu_{i-1}}^{\mu_i} f_{G-G}(x_1, x_2) dx_1 dx_2 \quad \text{(C.1.12)}$$

with $f_{G-G}(x_1, x_2)$ the joint probability function. The transition probability from state i to state $i-1$ is written in a similar way.

In a FSMC the transition probabilities can be approximated by means of the LCR. Thus, $p_{i,i+1}$ is approximated by the ratio of the level crossing rate at threshold u_{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 u_i and the average number of packets per second staying in state i. These approximations will be valid as far as $LCR << \pi_i/T_S$, i.e., as far as the fading process is sufficiently slow as to the transitions of the chain occur only between adjacent states ($p_{k,i} = 0, \;\; if \; |k-i| > 1$). In the case of the in-phase and quadrature components it yields:

$$p_{i,i+1} \approx \frac{N_G(\mu_{i+1}) \cdot T_S}{\pi_i} = \frac{1}{\pi_i} T_S \sqrt{2} f_D e^{-x_{i+1}^2/(2b_0)} \tag{C.1.13}$$

$$p_{i,i-1} \approx \frac{N_G(\mu_i) \cdot T_S}{\pi_i} = \frac{1}{\pi_i} T_S \sqrt{2} f_D e^{-x_i^2/(2b_0)} \tag{C.1.14}$$

When the envelope is considered, the transition probabilities are:

$$p_{i,i+1} \approx \frac{1}{\pi_i} T_S \sqrt{\frac{\pi}{b_0}} f_D \delta_{i+1} e^{-\delta_{i+1}^2/(2b_0)} \tag{C.1.15}$$

$$p_{i,i-1} \approx \frac{1}{\pi_i} T_S \sqrt{\frac{\pi}{b_0}} f_D \delta_i e^{-\delta_i^2/(2b_0)} \tag{C.1.16}$$

Finally, the expressions of the transition probabilities of the received SNR are as follows:

$$p_{i,i+1} \approx \frac{1}{\pi_i} T_S \sqrt{\frac{2\pi\Gamma_{i+1}}{\gamma_0}} f_D e^{-\Gamma_{i+1}/\gamma_0} \tag{C.1.17}$$

$$p_{i,i-1} \approx \frac{1}{\pi_i} T_S \sqrt{\frac{2\pi\Gamma_i}{\gamma_0}} f_D e^{-\Gamma_i/\gamma_0} \tag{C.1.18}$$

Autocorrelation function

The ACF of a FSMC is one of the limitations of the Markov modeling, as it decays exponentially, in contrast to the theoretical Bessel function of Clarke's model. The

analytical expression of the ACF is:

$$\begin{aligned} R[m] = E[f(X_0)f(X_m)] &= \sum_{i=1}^{n}\sum_{j=1}^{n} Pr[X_0 = j, X_m = i] \\ &= \sum_{j=1}^{n} f(j)\pi_j \sum_{i=1}^{n} f(i)p_{ij}^m \end{aligned} \tag{C.1.19}$$

where $f(\cdot)$ is the output of the chain in the corresponding state at instant m. In matrix form we can write:

$$R[m] = f^T \Phi \mathbf{P}^m f \tag{C.1.20}$$

with Φ a diagonal matrix whose elements are the state probabilities and $\mathbf{P}$ the transition matrix.

C.2 Evaluation

The FSMC has been evaluated with the goal of determining the influence of the choice of the parameters. We show here the results corresponding to the envelope of the signal. For the in-phase and quadrature components and the instantaneous SNR the conclusions are similar.

First of all, the estimated mean $\overline{r}$ and standard deviation $\overline{\sigma}_r$ of the fading envelope are shown in Table C.2, and compared to the theoretical, with a realization of 10^6 samples, $b_0 = 1$, 10 states and $f_D \cdot T_S = 10^{-2}$. All the partition methods estimate correctly the mean. For the standard deviation, the uniform partition is the one that obtains the best approximation for the specified length of the simulation.

As it was previously mentioned, the ACF is expected to be exponentially decreasing. We are interested in checking if it is possible to improve the grade of approximation to the theoretical bessel function by selecting properly the parameters of the chain (partition method and number of states).

Table C.2: Mean and standard deviation of the fading envelope

	$\overline{r}$	σ_r
Uniform	1.2548	1.9874
Equal probabilities	1.2482	1.9659
Linearly increasing probabilities	1.2485	1.9525
Exponentially increasing probabilities	1.2456	1.8652
Theoretical	1.2533	2

The ACF function of the different partition methods is compared in Figure C.2 (a), with the rest of parameters of the chain fixed ($f_D \cdot T_S = 10^{-2}$ and 10 states). The theoretical ACF from Clarke's model is plotted with solid line. In the figure, one can observe that the uniform and equal probabilities and linearly increasing probabilities obtain similar results, being the first two partitions slightly better, as they are closer to Bessel's function in the first samples of observation (the only ones that the FSMC can approximate properly). A zoom of the figure is shown in Figure C.2 (b) to corroborate it.

On the other hand, the influence of the number of states in the ACF function is shown in Figure C.3. In principle, defining more states would correspond to better approximations to Clarke's model. However, we may keep in mind the assumption of transitions only happening to adjacent states. If the number of states is increased too much, it will not be true anymore. The dependence with the correlation of the channel is closely related with the number of states, e.g. given a number of states, if the channel is too fast the assumption is not accomplished.

To finish with the study of the ACF, we evaluate the power margin quality measures as another parameter of quality performance. In particular, the mean basis power margin and the maximum basis power margin, defined as:

$$\begin{aligned} \mathcal{G}_{\text{mean}} &= \frac{1}{\sigma_{\text{g}}^2 L}\text{trace}(C_{\text{g}} C_{\widehat{\text{g}}}^{-1} C_{\text{g}}) \\ \mathcal{G}_{\text{max}} &= \frac{1}{\sigma_g^2}\max(\text{diag}(C_g C_{\widehat{g}}^{-} 1 C_g)) \end{aligned} \tag{C.2.1}$$

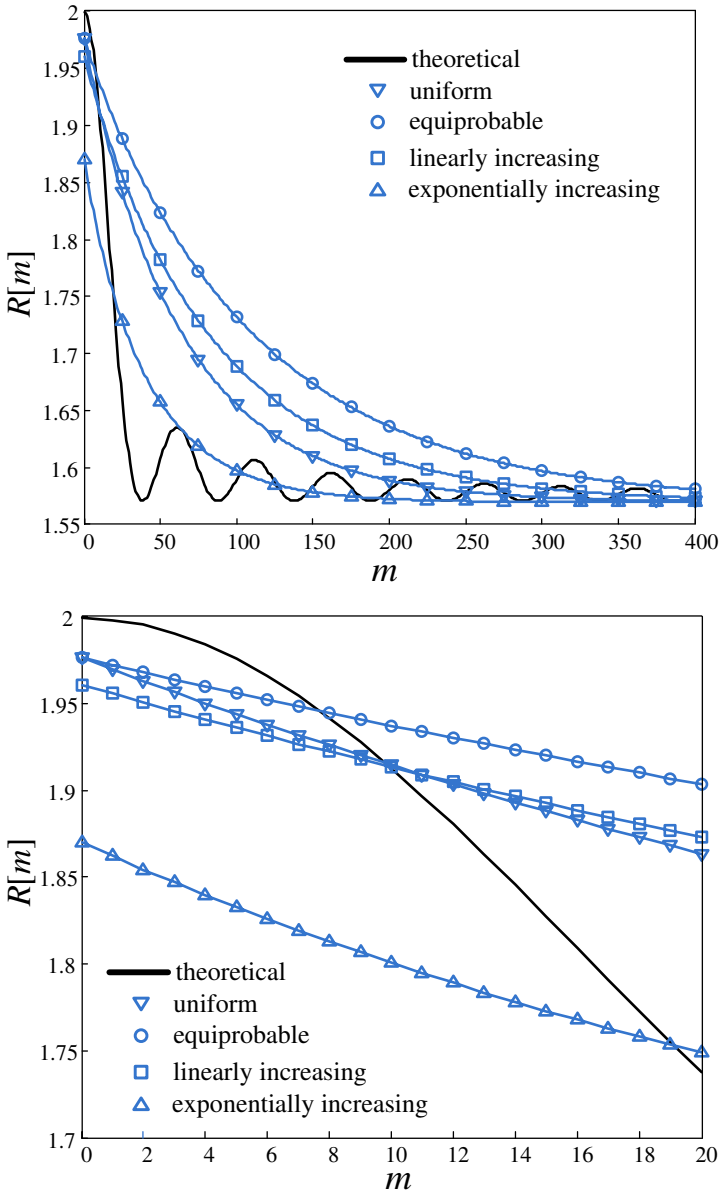


Figure C.2: (a) ACF of the envelope for different partition methods (b) Zoom of (a)

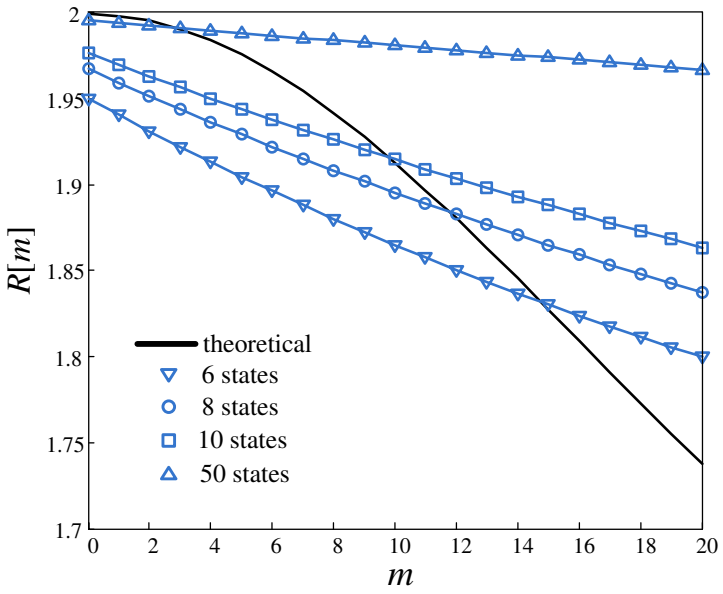


Figure C.3: ACF of the envelope for different number of states

where $\sigma_g^2 = b_0$ is the variance of the theoretical distribution, L is the number of points used in the calculation of the metrics, C_g is the LxL covariance matrix of L neighboring samples of the theoretical fading process, and $C_{\hat{g}}$ is the LxL covariance matrix of L neighboring samples of the fading model. The results are shown in Table C.3, for the partitions that approximate better Clarke's model: uniform and equiprobable. Moreover, two other methods for Rayleigh channel generation are plot: the simulator provided by MATLAB and the Sum-of-Sinusoids method proposed in [Zheng 2002] and simulated with 15 sinusoids. Perfect performance corresponds to 0 dB for both measures. An autocorrelation sequence length of 300 was considered with the same value of $f_D \cdot T_S$ fixed above. The results show the poor approximation of the FSMC to the theoretical ACF compared with the rest of models.

The main advantage of a Markov model for Rayleigh channels is the simplicity of the model, which leads to short times for generating long realizations. We show in Table C.4 the results of the time comparison for the same four methods. One

Table C.3: Power Margin Quality Measure comparison

Method	$\mathcal{G}_{\text{mean}}$(dB)	$\mathcal{G}_{\text{max}}$(dB)
FSMC with uniform partition and 20 states	1.1644	1.2227
FSMC with equiprobable states and 20 states	2.0427	2.1216
MATLAB channel	0.0503	0.0521
SoS with 15 sinuoids	0.0015	0.0022

can observe significant differences in terms of computation time for generating 5*e*6 samples of the complex channel response, being the FSMC 100 times better than the IDFT method and 1000 times better than the one provided by MATLAB.

Table C.4: Computation Time to generate 5e6 complex samples

Method	Time (s)
FSMC 20 states	0.0031
MATLAB channel	5.64
SoS with 15 sinuoids	3.91
IDFT	0.391

Appendix D

Summary (in Spanish)

D.1 Introducción

Garantizar calidad de servicio (Quality of Service, QoS) es un gran reto que se presenta en el diseño de las próximas generaciones de redes inalámbricas. En concreto, se estima que los servicios de tiempo real, que implican fuertes restricciones de retardo, sean cada vez más populares entre los usuarios de equipos móviles.

En esta tesis se analiza la relación entre los desvanecimientos del canal, el proceso de fuente de datos y parámetros de calidad de servicio (con especial atención al retardo) en un sistema inalámbrico.

Se propone la aplicación de la teoría del ancho de banda efectivo como base para el análisis. Se trata de una teoría ampliamente aplicada en redes cableadas y que ha sido adaptada recientemente para su uso en sistemas inalámbricos.

Modelo de sistema

En primer lugar, se aborda el estudio del sistema monousuario monocanal de la Figura D.1.

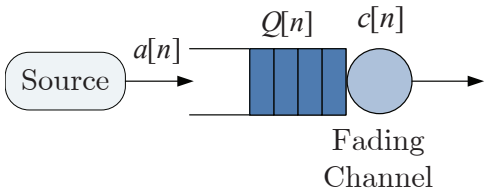


Figure D.1: Modelo de sistema

En este sistema, un flujo de información de velocidad instantánea $a[n]$ es transmitido a través de un canal. El servidor modela el proceso canal y puede transmitir a una velocidad variable $c[n]$, resultado de emplear técnicas de modulación adaptativas sobre un canal con desvanecimientos, de forma que el tamaño de la constelación se adapta a las condiciones instantáneas del canal (constelaciones más densas cuando el canal está en buenas condiciones). Así, la velocidad instantánea $c[n]$ depende de la respuesta del canal, que es en general correlada.

Se definen las funciones velocidad de fuente y de canal acumuladas como: $A[n] = \sum_{m=0}^{n-1} a[m]$ and $C[n] = \sum_{m=0}^{n-1} c[m]$.

Para cada uno de los procesos, fuente y canal, se define la función ancho de banda efectivo (EBF, Effective Bandwidth Function) como [Chang 1995a] [Wu 2003b]:

$$\begin{aligned} \alpha_A(\upsilon) &= \lim_{n\to\infty} \frac{1}{n\upsilon} \log \mathrm{E}\left[e^{\upsilon A[n]}\right] \text{ (fuente)} \\ \alpha_C(\upsilon) &= \lim_{n\to\infty} \frac{1}{n\upsilon} \log \mathrm{E}\left[e^{\upsilon C[n]}\right] \text{ (canal)} \end{aligned} \tag{D.1.1}$$

En un canal Rayleigh, intentar garantizar requisitos deterministas de QoS es inviable. Por ello resulta más aconsejable intentar garantizar requisitos de QoS estadísticos. En el caso del retardo, se ha definido una restricción probabilística de retardo dada por el par (D^t, ε), donde D^t es el retardo objetivo y ε es la probabilidad de exceder dicho retardo objetivo.

La teoría del ancho de banda efectivo nos da la siguiente expresión para la

probabilidad de violación ε [Wu 2003b]:

$$\varepsilon = Pr\{D(\infty) > D^t\} \approx \eta \cdot e^{-\theta \cdot \alpha_A(\theta) D^t} = \eta \cdot e^{-\theta \cdot \alpha_C(-\theta) D^t} \tag{D.1.2}$$

donde η es la probabilidad de que la cola no esté vacía y θ es el exponente de calidad de servicio, y es la solución a la ecuación:

$$\alpha_A(\upsilon) - \alpha_C(-\upsilon) \mid_{\upsilon=\theta} = 0 \tag{D.1.3}$$

La ecuación (D.1.2) sirve como base para el análisis del retardo en esta tesis. Como se puede observar, para estudiar la restricción de retardo (D^t, ε) es necesario evaluar las EBF del proceso canal y del proceso fuente.

Función ancho de banda efectivo

El significado de las funciones ancho de banda efectivo se ilustra en la Figura D.2. En el caso de la fuente, la EBF indica la mínima velocidad de canal necesaria para garantizar ciertos requisitos de calidad de servicio. Un valor alto de υ indica requisitos más estrictos de QoS (D^t o ε más pequeños). Así, la curva comienza en la velocidad media cuando $\upsilon = 0$. A partir de ahí, la curva crece hacia la velocidad de pico, que es el valor asintótico cuando $\upsilon \to \infty$. En el caso del proceso canal, la EBF expresa la máxima velocidad de fuente que un determinado proceso canal puede soportar manteniendo los requisitos de calidad de servicio. La curva $\alpha_C(-\upsilon)$ comienza en la capacidad de Shannon cuando $\upsilon = 0$, y decrece con υ hasta el valor máximo υ_{max}. Valores más allá de υ_{max} no pueden ser garantizados por el canal.

Si unimos las dos curvas, aparece un punto de trabajo que corresponde con la intersección de ambas. Este punto es el exponente de calidad de servicio θ que se obtiene con la ecuación (D.1.3) y representa el punto en el que tanto el proceso fuente como el proceso canal podrán garantizar los requisitos de QoS.

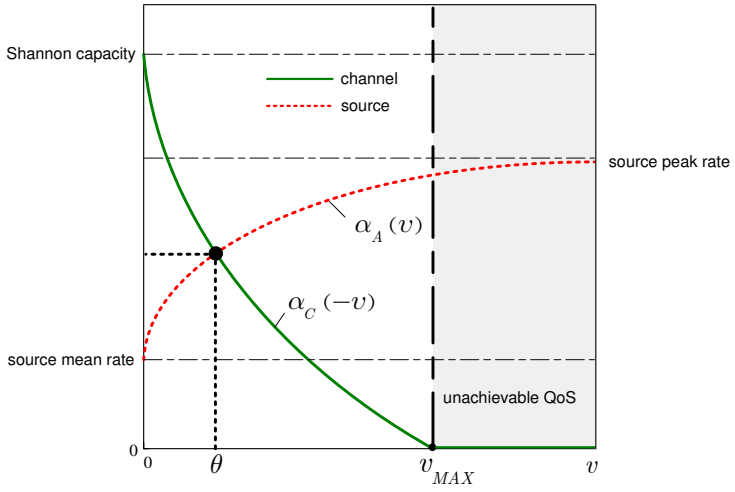


Figure D.2: Función ancho de banda efectivo.

D.2 Análisis del retardo en canales Rayleigh planos

EBF de canales Rayleigh planos

Para analizar el retardo en un sistema con canal plano, es necesario calcular la función ancho de banda efectivo del canal. En primer luegar, se calcula la EBF para canales incorrelados. En este caso, la velocidad de transmisión acumulada $C[n]$ es la suma de variables i.i.d. y se puede aplicar el teorema del límite central. Entonces, la EBF corresponde a la de una distribución Gaussiana:

$$\alpha_C(v) = m_c + \frac{v}{2}\sigma_c^2 \tag{D.2.1}$$

donde m_c y σ_c^2 son la media y la varianza de $c[n]$, que se calculan según la política de adaptación, continua o discreta. El resultado final depende de la BER objetivo que se quiera fijar y de la SNR media del canal.

Si el canal presenta correlación temporal, la estrategia a seguir es la siguiente. Se divide la velocidad total de transmisión en bloques de longitud k. Si la longitud de los bloques es suficientemente grande, entonces podemos considerar que los bloques

son independientes unos de otros y aplicar de nuevo el teorema del límite central. La clave está en escoger adecuadamente el valor de k. Cuanto mayor sea la correalción temporal, mayor es el valor de k necesario para que la aproximación sea válida. La aproximación gaussiana ha sido validada con el test de Lilliefors y con la comparación de los resultados con las simulaciones. Gracias a la aplicación del teorema del límite central la EBF del canal es de nuevo una función de la media y la varianza de $c[n]$, aunque su cálculo al incluir la correlación resulta mucho más engorroso. Además de la dependencia con la BER objetivo y con la SNR media, ahora la varianza depende también de la función de autocorrelación del canal.

En la Figura D.3 comparamos la EBF del canal obtenida en esta tesis con la solución aportada en [Wu 2003b], válida sólo para valores pequeños de SNR. Se representan dos valores de SNR media, 5 y 10dB, y el tamaño de bloque k es 10000. La ACF sigue una función Bessel con frecuencia Doppler normalizada con el periodo de símbolo $f_D \cdot T_S = 0.1$. Como cabía esperar, las curvas se separan conforme aumenta la SNR media.

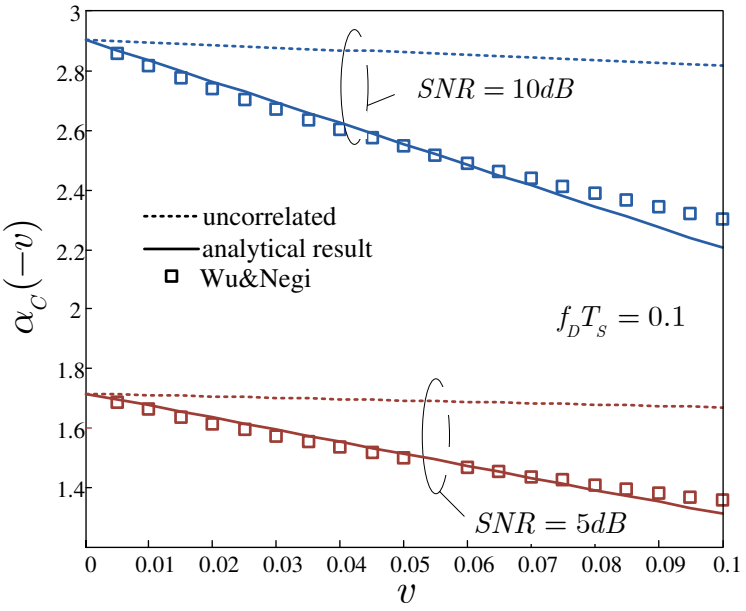


Figure D.3: EBF de un canal Rayleigh con ACF de Bessel.

Percentil del retardo

Una vez obtenida la EBF del canal, es posible evaluar el percentil del retardo como $1-\epsilon$.

Supongamos una fuente constante a una velocidad λ, la EBF de la fuente es una constante: $\alpha_A(v) = \lambda$. Para dar una expresión del percentil, hay que despejar el exponente de QoS. Si sustituimos la EBF de un canal plano, θ viene dado por:

$$\lambda - \alpha_C(-\theta) = 0 \Rightarrow \theta(\lambda) \triangleq \theta(m_c, \sigma_c^2, \lambda) = \frac{2(m_c - \lambda)}{\sigma_c^2} \tag{D.2.2}$$

Finalmente, se obtiene el percentil del retardo:

$$1 - \varepsilon = 1 - Pr\{D(\infty) > D^t\} = 1 - e^{-\frac{2(m_c-\lambda)}{\sigma_c^2} \cdot \lambda D^t} \tag{D.2.3}$$

Un ejemplo se muestra en la Figura D.4. Se pinta el percentil del retardo como una función de la velocidad de fuente λ, para un retardo objetivo dado. La ACF sigue una caída exponencial, con parámetro de correlación $\rho = 0.8$ y $\rho = 0.9$. Se pinta también el caso de canal incorrelado. La BER objetivo es 10^{-2} y D^t es 20 símbolos. El canal tiene SNR media 5, 10 y 15dB. Se puede comprobar cómo el percentil del retardo se acerca al 100% cuando la velocidad de fuente disminuye. A medida que aumenta λ, el percentil decrece hasta el 0%. Cabe destacar la drástica caída a cero en el caso de un canal incorrelado, de forma que pequeños cambios en la velocidad de fuente provocan enormes variaciones del percentil del retardo. Éste es el comportamiento esperado cuando no hay ni variabilidad en la fuente ni correlación en el canal. Cuando se incluye correlación en el canal u otro tipo de fuentes, la caída de la curva del percentil es mucho más suave.

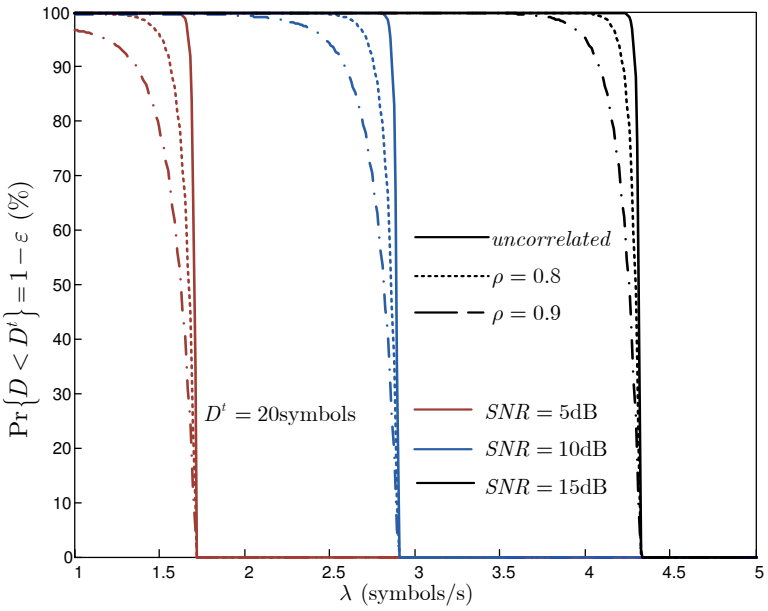


Figure D.4: Percentil del retardo para tráfico constante. ACF exponencial. $D^t = 20$ símbolos. $BER^t = 10^{-2}$.

Capacidad con una restricción probabilística de retardo

La capacidad de un canal indica la máxima velocidad de datos que puede ser transmitida sobre él. En comunicaciones inalámbricas, se han definido varias capacidades sujetas a distintas restricciones y modelos. La capacidad limitada por retardo [Biglieri 1998] es la máxima velocidad fija que puede garantizar un retardo determinístico. En el caso de canal Rayleigh, esta capacidad es cero. Por ello, proponemos definir una nueva capacidad en términos de retardo estadístico. Así, la **Capacidad con una restricción probabilística de retardo** $\mathrm{C}_{D^t,\varepsilon}$ es la máxima velocidad de fuente soportada por el canal bajo una BER objetivo y mantiendo el retardo por debajo de D^t excepto para una probabilidad ε.

Para calcular $C_{D^t,\varepsilon}$, la restricción de retardo se despeja de (D.1.2) suponiendo que el sistema trabaja en condiciones de carga alta ($\eta \to 1$):

$$-\frac{\log(\varepsilon)}{D^t} = \theta \cdot \alpha_A(\theta) \tag{D.2.4}$$

En el caso de una fuente constante, si se substituye el exponente de QoS (D.2.2) en (D.2.4), el valor de λ despejado es la **Capacidad con una restricción probabilística de retardo** $\mathrm{C}_{D^t,\varepsilon}$ y puede escribirse:

$$\mathrm{C}_{D^t,\varepsilon} = \frac{m_c}{2} + \frac{1}{2}\sqrt{m_c^2 - 2\sigma_c^2 \frac{(-\log \varepsilon)}{D^t}} \tag{D.2.5}$$

Se pueden comprobar dos límites en (D.2.5). Si el retardo objetivo aumenta o la probabilidad de violación aumenta, entonces el requisito de QoS se relaja y $\mathrm{C}_{D^t,\varepsilon}$ se aproxima a m_c, que es la capacidad ergódica del canal. En el otro extremo, si D^t o ε disminuyen, el canal puede soportar una velocidad de tráfico menor para poder seguir garantizando la restricción de retardo. Lógicamente, $\mathrm{C}_{D^t,\varepsilon}$ es siempre menor que la capacidad ergódica. La influencia de la BER objetivo está incluída en la media y la varianza del canal. Cuanto menor sea la BER objetivo, menor será la capacidad. La varianza del canal incluye también la influencia de la correlación del canal. Así, cuando la correlación temporal aumenta (σ_c^2 aumenta), la capacidad disminuye.

La Figura D.5 presenta la evaluación de $\mathrm{C}_{D^t,\varepsilon}$ para un canal incorrelado y políticas de velocidad continua y discreta. La SNR media es 10dB y 20dB, y la probabilidad de exceder el retardo objetivo es ε=0.1 (percentil 90%) y ε=0.05 (percentile 95%). $\mathrm{C}_{D^t,\varepsilon}$ se pinta como función de D^t. La BER objetivo es $BER^t = 10^{-3}$.

El caso de un canal correlado en el tiempo se muestra en la Figura D.6 para una política continua. La SNR media es 5 y 10dB. $\varepsilon = 0.10$ y la correlación sigue una caída exponencial con parámetro ρ fijado a 0.95 y 0.99. El resultado del canal incorrelado se representa con triángulos. Se puede observar que la correlación temporal degrada las prestaciones en términos de retardo. Así, cuando la correlación aumenta, valores mayores de ρ, la capacidad disminuye.

Por último, se pueden considerar fuentes variables. Por ejemplo, supongamos un

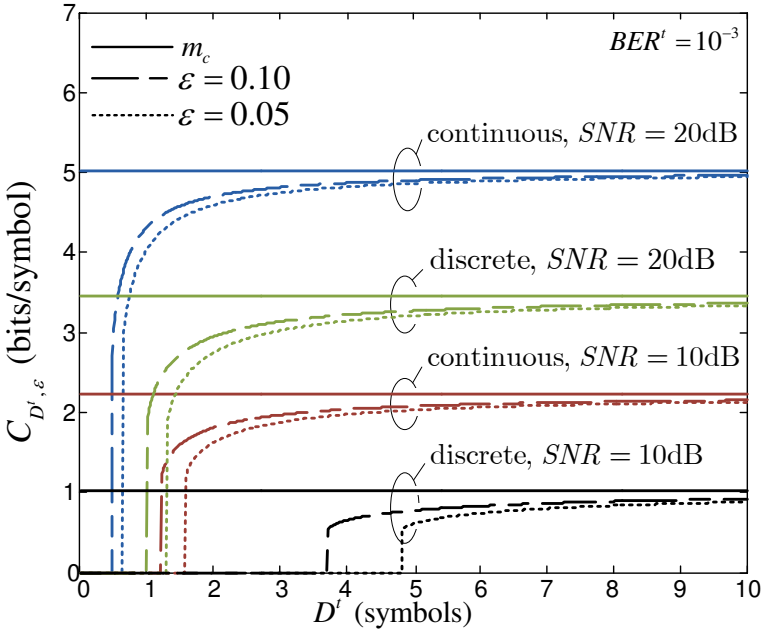


Figure D.5: Capacidad con restricción de retardo. Canal incorrelado.

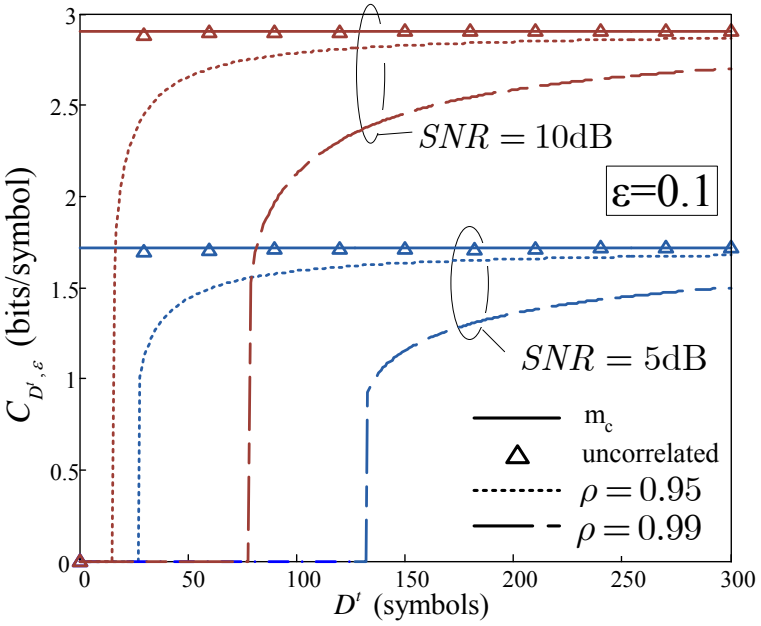


Figure D.6: Capacidad con restricción de retardo. Correlación temporal.

modelo ON-OFF que simule una fuente de voz. Su EBF es [Kelly 1996]:

$$\alpha_A(u) = \frac{1}{2u}\left[h \cdot u - \mu - \lambda + \sqrt{(h \cdot u - \mu + \lambda)^2 + 4 \cdot \lambda \cdot \mu}\right] \tag{D.2.6}$$

donde los parámetros λ y μ son las probabilidades de transición de ON a OFF y viceversa y h la tasa en el estado ON.

En este caso, la máxima velocidad media de fuente sujeta a la restricción (D^t, ε) es:

$$m_A(D^t, \varepsilon) = -\frac{1}{2} \cdot \frac{\lambda}{\lambda + \mu} \cdot \frac{(\sigma_c^2)^2 s^3 - 4\sigma_c^2 m_c s^2 + 4m_c^2 s - 2\sigma_c^2(\lambda + \mu)s + 4m_c(\lambda + \mu)}{\sigma_c^2 s^2 - 2m_c s - 2\lambda} \tag{D.2.7}$$

con s:

$$s = \frac{-m_c D^t - \sqrt{(m_c D^t)^2 - 2\sigma_c^2 D^t \log \varepsilon}}{\sigma_c^2 D^t} \tag{D.2.8}$$

En la Figura D.7 se ilustra la influencia de los parámetros de la fuente en el retardo. Las probabilidades de transición ON-OFF son iguales y se han fijado tres valores diferentes (0.1, 0.01 y 0.001). La ACF es Bessel con parámetro $f_D \cdot T_S = 5 \cdot 10^{-2}$ y $1 \cdot 10^{-2}$. Periodos ON-OFF más cortos son beneficiosos para el retardo, de forma que se obtienen valores más altos de capacidad. Además, se puede observar que las curvas con distintos valores de correlación del canal se cruzan, mostrando el efecto conjunto que tienen el proceso fuente y el proceso canal en el retardo. Así, para cumplir ciertos requisitos estadísticos de retardo, se puede optar por fuentes con transiciones ON-OFF más cortas o por canales con menor correlación.

D.3 Análisis del retardo en canales Rayleigh selectivos en frecuencia

A continuación, se considera el sistema de la Figura D.8, en el que los F servidores en paralelo representan las F subportadoras de un sistema OFDM. La diferencia

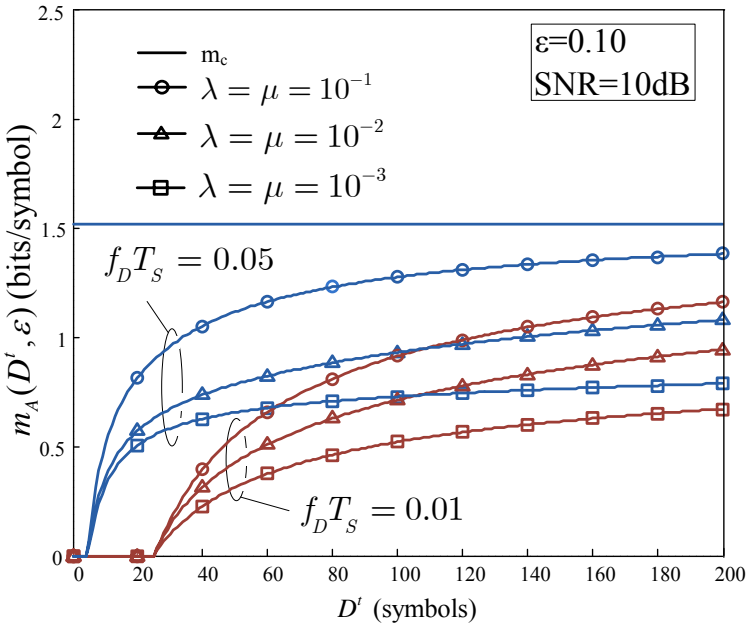


Figure D.7: Capacidad con restricción de retardo para tráfico de voz. Influencia de los parámetros de la fuente.

respecto al sistema monocanal es que ahora la portadora f puede transmitir a una velocidad $c^f[n]$, de forma que la velocidad total del canal es $c[n] = \sum_{f=1}^{F} c^f[n]$.

De forma análoga al canal plano, se calcula la EBF del nuevo canal selectivo. La estrategia es la misma: dividir la velocidad de canal acumulada en bloques de tamaño k, suficientemente grandes como para despreciar la correlación entre bloques. La aplicación del teorema del límite central da como resultado una expresión de la EBF en función de la media y la varianza del canal. La principal diferencia radica en que ahora el cálculo de la varianza considera no sólo la correlación temporal dentro del bloque sino también la frecuencial.

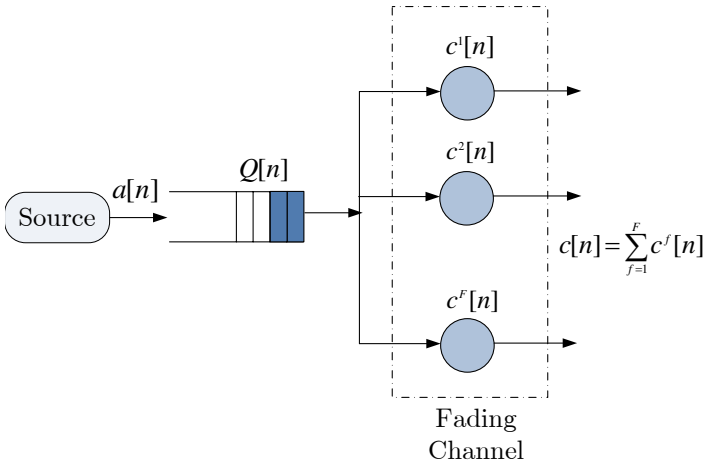


Figure D.8: Modelo de colas para el sistema OFDM.

Capacidad con una restricción probabilística de retardo

Una vez calculada la EBF del nuevo canal, es fácil evaluar la capacidad con una restricción probabilística de retardo en el caso de canales selectivos en frecuencia. La capacidad se representa en la Figura D.9. Se puede observar la influencia de la correlación frecuencial en las prestaciones de retardo. Se han pintado dos canales: un canal vehicular de ancho de banda de coherencia 430kHz (menos correlación frecuencial) y un canal pedestre cuyo ancho de banda de coherencia es de 3.4MHz (más correlación frecuencial). Hay 16 portadoras y la SNR media es 10dB y 15dB. Se puede observar que para valores menores del ancho de banda de coherencia se obtienen valores más altos de capacidad.

D.4 Análisis del retardo en sistemas multiusuario

Finalmente, se considera un sistema multiusuario en el que cada usuario tiene asociada una cola en la que sus datos son almacenados antes de ser transmitidos al

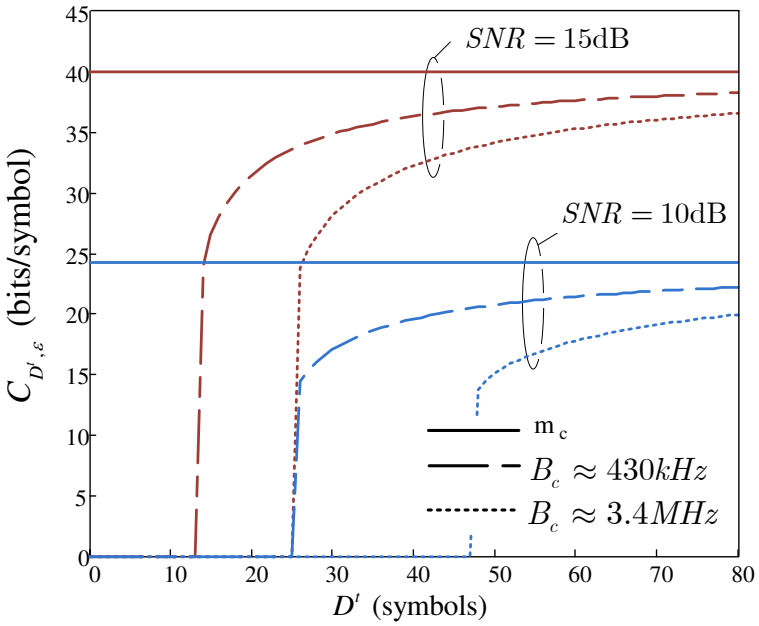


Figure D.9: Capacidad con restricción de retardo para una fuente constante y 16 portadoras. Influencia de la correlación frecuencial.

canal. Un nuevo elemento aparece en este escenario, el algoritmo de multiplexación, que va a tener una influencia importante en el retardo. Se han evaluado tres algoritmos de multiplexación representativos [Entrambasaguas 2007]: Round Robin (RR), Best Channel (BC) y Proportional Fair (PF). El modelo de sistema es ahora el representado en la Figura D.10

Los F servidores modelan las F portadoras del sistema OFDMA. El planificador asigna recursos a los usurios. Cada usuario tiene su propia restricción de retardo pero, por simplicidad, se mantiene la notación (D^t, ε).

La velocidad instantánea de transmisión que ve el usuario u es:

$$c_u[n] = \sum_{f=1}^{F} c_u^f[n],$$
$$c_u^f[n] = \begin{cases} c^f[n] \text{ si la portadora } f \text{ es asignada al usuario u} \\ 0 \text{ en otro caso} \end{cases} \quad \text{(D.4.1)}$$

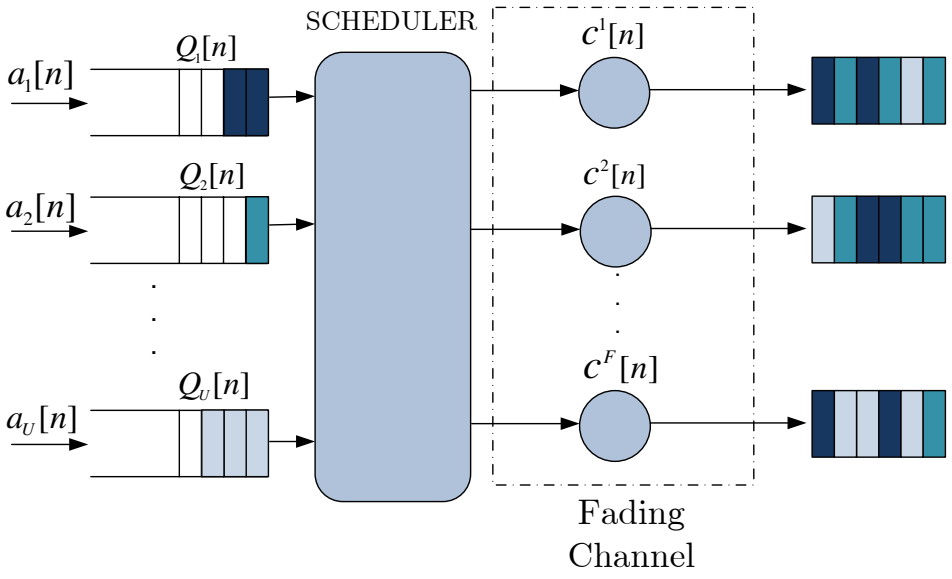


Figure D.10: Modelo de sistema OFDMA

La EBF del canal depende ahora del algoritmo de multiplexación y de la correlación temporal y frecuencial del canal. La aproximación Gaussiana se aplica de forma similar a como se hizo en el caso monousuario. Entonces, la máxima velocidad alcanzable por el usuario u-ésimo bajo una restricción en el retardo y una BER objetivo es:

$$\mathrm{R}^{u}{}_{D^t,\varepsilon} = \frac{m_{k_u}}{2kF} + \frac{1}{2}\sqrt{\frac{m_{k_u}^2}{(kF)^2} - 2\frac{\sigma_{k_u}^2}{kF}\frac{(-\log\varepsilon)}{D^t}} \tag{D.4.2}$$

donde la gran dificultad radica en el cálculo de los estadísticos m_k y σ_k^2, que son específicos de cada algoritmo de multiplexación.

Las Figuras D.11, D.12 y D.13 muestran la evaluación de D.4.2 para las tres disciplinas analizadas: RR, BC and PF. Se ha considerado un canal pedestre de ancho de banda de coherencia $3.4MHz$. La ACF sigue una función de Bessel con parámetro $f_D T_S = 0.1$ y $f_D T_S = 0.01$. Hay 5 usuarios con SNR media 5, 6, 7, 8 y 9 dB, y 4 portadoras. Se pinta la máxima velocidad de transmisión $\mathrm{R}^{u}{}_{D^t,\varepsilon}$ para el mejor (9dB), el medio (7dB) y el peor (5dB) usuario.

Las velocidades individuales $R^{u}{}_{D^t,\varepsilon}$ se representan como función del retardo objetivo. La probabilidad de violación se fija a 0.1 para todos los usuarios. Se puede observar que los usuarios con mejores condiciones de canal obtienen velocidades mayores y pueden exigir requisitos de QoS más estrictos.

Las diferencias entre usuarios son mucho más acusadas en BC si lo comparamos con RR. Así, el usuario con mejores condiciones de canal se ve claramente beneficiado si se cambia a disciplina BC a costa del usuario con peor canal. Esto afecta no sólo al comportamiento asintótico de las curvas (acercándose a la media de la velocidad del canal), sino también a los retardos objetivo mínimos que cada usuario podría demandar. En el caso de PF, se observa cómo se reducen las diferencias entre usuarios a costa de reducir ligeramente la tasa media. Destaca en el algoritmo PF la igualación de los usuarios en términos de retardo objetivo

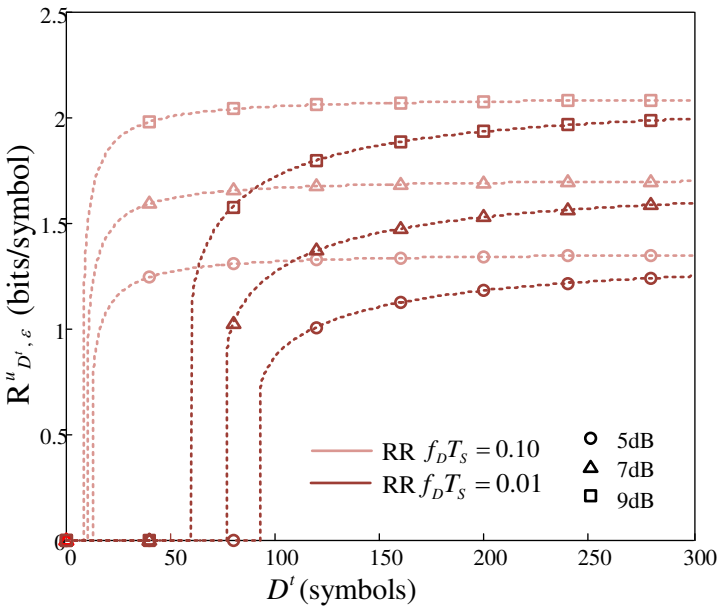


Figure D.11: Máxima velocidad de usuario admisible en un sistema OFDMA con disciplina Round Robin.

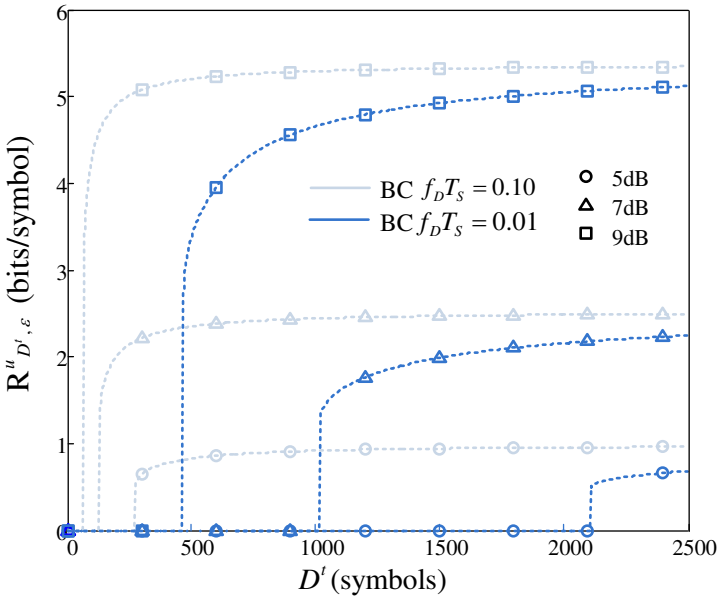


Figure D.12: Máxima velocidad de usuario admisible en un sistema OFDMA con disciplina Best Channel.

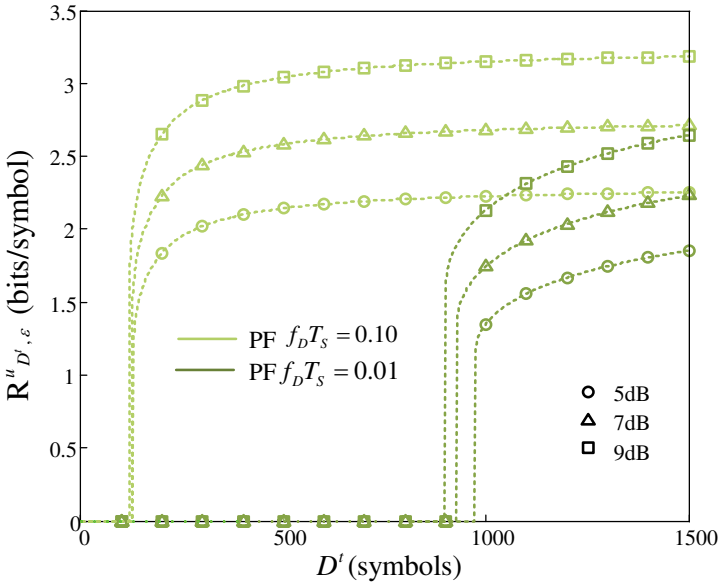


Figure D.13: Máxima velocidad de usuario admisible en un sistema OFDMA con disciplina Proportional Fair.

D.5 Conclusiones

En esta tesis se ha abordado el estudio de garantías de QoS estadísticas en un sistema inalámbrico. En concreto, se ha analizado la relación entre los desvanecimientos del canal Rayleigh, la fuente de información y algunos parámetros de QoS (en especial el retardo). Para ello, se ha propuesto la aplicación de la teoría del ancho de banda efectivo. Se han calculado las EBF de los procesos fuente y canal, en primer lugar para un canal Rayleigh plano. Con ellas, se ha analizado el percentil del retardo. Además, se ha propuesto la definición de una nueva capacidad en un sistema inalámbrico: la capacidad con una restricción probabilística de retardo, que es la máxima velocidad de fuente admisible por un canal manteniendo una BER objetivo y una restriccción en el retardo dada por (D^t, ε). La evaluación de la capacidad demuestra que la correlación temporal del canal o de la fuente son perjudiciales para el comportamiento del retardo.

A continuación, el análisis se ha repetido para el caso de un sistema OFDM con un canal Rayleigh selectivo en frecuencia. Para ello ha sido necesario calcular la EBF de este nuevo canal. Se ha comprobado que también la correlación frecuencial del canal reduce las prestaciones del sistema en términos de retardo.

Por último, se ha planteado un sistema multiusuario y se ha repetido el análisis del retardo. En este caso, se calcula la máxima velocidad admisible para cada usuario sujeta a una BER objetivo, su propia restricción probabilística de retardo y el algoritmo de multiplexación empleado. Se han analizado tres algoritmos: Round Robin, Best Channel y Proportional Fair. Los resultados obtenidos permiten analizar el balance entre justicia y prestaciones de cada uno de los algoritmos.

En resumen, en esta tesis se ha comprobado la alta sensibilidad del retardo a la variabilidad del tráfico, a la correlación temporal o frecuencial del canal y al algoritmo de multiplexación. Los resultados analíticos presentados se han validado comparándolos con simulaciones. Además, el procedimiento propuesto es genérico

y podría extenderse a otras disciplinas, modelos de tráfico y modelos de canal. En cualquier caso, las funciones de ancho de banda efectivo de una fuente y un canal no siempre se pueden evaluar analíticamente. Para esos casos, se aporta una estrategia semianalítica.