2008

Cross-Layer Optimization of Network Performance over MIMO Wireless Mobile Channels

Marco Luccini
Western University

Cross-Layer Optimization of Network Performance over MIMO Wireless Mobile Channels

(Spine title: Cross-Layer Optimization of Network Performance)

(Thesis format: Monograph)

by

Marco Luccini

Graduate Program
in
Engineering Science
Electrical and Computer Engineering

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Engineering Science

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

In the information theory, the channel capacity states the maximum amount of information which can be reliably transmitted over the communication channel. In the specific case of *multiple-input multiple-output* (MIMO) wireless systems, it is well recognized that the instantaneous capacity of MIMO systems is a random Gaussian process. Time variation of the capacity leads to the outages at instances when it falls below the transmission rate. The frequency of such events is known as outage probability.

The cross-layer approach proposed in this work focuses on the effects of MIMO capacity outages on the network performance, providing a joint optimization of the MIMO communication system. For a constant rate transmission, the outage probability sensibly affects the amount of information correctly received at destination. Theoretically, the limit of the ergodic capacity in MIMO time-variant channels can be achieved by adapting the transmission rate to the capacity variation. With an accurate channel state information, the capacity evolution can be predicted by a suitable autoregressive model based on the capacity time correlation. Taking into consideration the joint effects of channel outage at the physical layer and buffer overflow at the *medium access control* (MAC) layer, the optimal transmission strategy is derived analytically through the *Markov decision processes* (MDP) theory. The adaptive policy obtained by MDP is optimal and maximizes the amount of information correctly received at the destination MAC layer (throughput of the system). Analytical results demonstrate the significant improvements of the optimal variable rate strategy compared to a constant transmission rate strategy, in terms of both system throughput and probability of data loss.

Acknowledgments

And here we are, after almost two years in Canada it is the time for graduation and enter a new level of the game. These lines want to thank many many people, form professors to friends, which have helped me during all this time.

First of all, I am grateful to my supervisors, Dr. Serguei Primak and Dr. Abdallah Shami. Both of you trusted my potential and carefully leaded me to achieve the basic skills in academic research. These two years under your supervision represented for me an important growth, for which I will always thank you. I thanks Dr. Primak for introducing me to the research method, for always reminding me that it is only a good thing to have new problems to solve. I thank him for sharing with me his vaste knowledge, form communication theory to many others field. In the whole meaning, you represented for me a guide, not only academically speaking. I thanks Dr. Shami for his first research proposal and interest in bringing me into the master program. I thank you for the great patience you had in finding a suitable topic for my "cross-layer" design. Moreover, thanks for having always in mind the whole picture of my work, the final goals and pushing me to go always one step ahead. Finally, thanks for the having introduced me to the original libanese food, it was definitely part of my learning experience. I am really grateful to both of you, for your contribution in my academic and personal growth.

My recognition and appreciation go to the professors who reviewed this thesis work: Dr. Katchabaw, Dr. Rahman and Dr. Wang.

I would like to thanks Dr. Franco Berruti for his great effort in starting the double-degree agreement between the University of Western Ontario and Politecnico di Milano. Thanks to him I had the opportunity to come and complete my master program in this university. I want to thank you for all I learnt here and for the occasion it represented for me. I admire your enthusiasm in dealing with new projects and in making what you started better and better. Moreover, I would like to thank you and your wife Sandra for all the support and help you gave me during the time I spent in London.

Within the department of Electrical and Computer Engineering I met many special people. Special thanks goes to Sandra, Rachel and Jaquie. Special gratitude goes to Sharon for her enthusiasm and contagious way of dealing with everything.

My profound gratitude goes to all the friends I met in TEB 344, your help and friendship were fundamental for me.

I thank the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Communication Research Centre (CRC) for the financial support they provided during my graduate studies.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ACF	*Autocorrelation Function*
AFD	*Average Fade Duration*
AMC	*Adaptive Modulation and Coding*
AoA/AoD	*Angle of Arrival/Angle of Departure*
AWGN	*Additive White Gaussian Noise*
CDF	*Cumulative Distribution Function*
CDI	*Channel Distribution Information*
CSI	*Channel State Information*
CSIT	*Channel State Information at the Transmitter*
DAR	*Discrete Autoregressive*
DFT	*Discrete Fourier Transform*
DPSS	*Discrete Prolate Spheroidal Sequences*
FER	*Frame Error Rate*
FSMC	*Finite State Markov Chain*
IID	*Independent Identically Distributed*
LCR	*Level Crossing Rate*
MAC	*Medium Access Control* layer
MDP	*Markov Decision Process*
MDPSS	*Modulated Discrete Prolate Spheroidal Sequences*
MIMO	*Multiple-Input Multiple-Output*
MISO	*Multiple-Input Single-Output*
MSE	*Mean Square Error*
PDF	*Probability Density Function*
PER	*Packet Error Rate*
PHY	*Physical* layer
PSD	*Power Spectral Density*
PSS	*Prolate Spheroidal Sequences*
QoS	*Quality of Service*
QPSK	*Quadrature Phase Shift Keying*

SIMO	*Single-Input Multiple-Output*
SISO	*Single-Input Single-Output*
SNR	*Signal to Noise Ratio*
WCDMA	*Wideband Code Division Multiple Access*
WSS	*Wide Sense Stationary*

Chapter 1
Introduction

This chapter provides an overview of the thesis. Firstly, the definition of multiple antennas systems and their importance are discussed, along with the explanation of the main improvements introduced by these systems. Secondly, a brief overview on the meaning of cross layer design is presented, leading to the main contributions of the thesis. Finally, the structure of the thesis is briefly described.

1.1 Basics of MIMO systems

Multiple-input multiple-output (MIMO) are communication wireless systems in which the transmitter and receiver are equipped with N_T transmitting antennas and N_R receiving antennas. Special cases of MIMO are *single-input multiple-output* (SIMO), where $N_T = 1$ and $N_R > 1$, *multiple-input single-output* (MISO), where $N_T > 1$ and $N_R = 1$ and *single-input single-output* (SISO) in which $N_T = N_R = 1$. In the recent years, the attention around these systems has been growing rapidly because of their potential improvement in spectral efficiency, which is the amount of information bits that can be transmitted per second per Hertz [bits/s/Hz]. The advantages of MIMO communications can be divided in three main categories: *transmit diversity* using space-time coding for enhancing transmission robustness, *spatial multiplexing* for enhancing the data-transmission rate and *beamforming* for improving the received signal and reduce the interference from other users.

- *Transmit Diversity:* transmit diversity is a transmission technique in which similar data signals are transmitted from multiple antennas to improve the *signal to noise ratio* (SNR). This kind of gain can be quantified as $N_T N_R$ and can be achieved by using multiple antennas at the receiver (diversity reception) and/or by using multiple antennas in transmission (transmit diversity).

- *Spatial Multiplexing:* spatial multiplexing is a MIMO transmission technique in which independent and separately encoded data signals, called streams, are transmitted from multiple antennas. This technique requires multiple antennas at both the sides of the communication link. Ideally, the knowledge of the channel is required at the transmitting antenna to exploit the orthogonal transmission paths. However, if the transmitted streams arrive at the receiver with sufficiently different spatial signatures, the receiver can separate them leading to an increase in the channel capacity. Under spatial multiplexing mode, the achievable capacity gain is referred to the maximum multiplexing order $N_s = \min\{N_T, N_R\}$.

- *Beamforming:* beamforming is a signal-processing technique that is used to control the directionality of the radiation pattern of a multiple antennas system. In reception, beamforming can increase the receiver sensitivity in the direction of the desired signals and decrease the sensitivity to interference and noise. In transmission, beamforming can increase the radiated power in the intended direction. The change compared with an omnidirectional receiving/transmitting pattern is known as the directional gain. Those changes are performed by creating nulls and beams in the radiation pattern of the antennas [1].

In the 1990s, the term "smart antennas" introduced the idea of exploiting beamforming at the base station of a cellular network, where the antenna beam adaptively follows the movement of a mobile user. The result of beamforming is the increasing of the signal gain in a specific direction, both focusing the signal strength and excluding the interference from other directions. One limitation of beamforming is that in a urban environment, the directional signal can be scattered because of buildings and moving objects. By the late 1990s, this drawback was turned into an advantage when space-time coding and spatial multiplexing were developed. These methods exploit the multipath phenomena to increase the system throughput. Spatial multiplexing allows to increase the communication spectral efficiency, with a larger amount of information that can be transmitted over the channel. However, the transmission rate is not an absolute parameter. The key performance metrics associated with any communication system are the transmission rate and the *frame error rate* (FER). With respect to the channel SNR, a fundamental trade-off exists in any communication system between the transmission rate and FER. In the context of MIMO systems,

this trade-off is often referred to as the diversity - multiplexing trade-off [2, 3], with diversity signifying the FER reduction and multiplexing signifying an increase in the transmission rate. The diversity-multiplexing trade-off is central in MIMO communications theory.

MIMO techniques are mostly used in conjunction with *orthogonal frequency division multiplexing* (OFDM), showing a strong relationship and compatibility with the flat-fading channel models. Emerging examples of MIMO systems on OFDM structure are the wireless networks as defined by the IEEE 802.16e standard and IEEE 802.11n high-throughput standard. There are also efforts to standardize MIMO in *wideband code division multiple access* (WCDMA) systems, which are a form of third-fourth generation cellular networks.

1.2 Cross Layer Design

MIMO techniques have rapidly become the new frontier of wireless communications. The investigation of MIMO system is driven by the limited available bandwidth and the increasing demand for high data rate transmission systems. Substantial attention has been given to multiple antenna systems due to their potential high spectral efficiency in rich scattering wireless environments [4, 5]. While most of the efforts have focused on characterizing the MIMO channel at the physical layer, joint study of the MIMO physical layer and the network model, known as a cross-layer approach [6], is gaining momentum and significant interest.

Current ad hoc wireless network protocol design is largely based on a layered approach, as shown in Fig. 1.1. In this model each layer is designed and operated independently, with interfaces between layers that are static and independent of the the individual network constraints and applications. This paradigm has greatly simplified network design and led to robust scalable protocols in the Internet. However, research efforts on independent layers have mainly targeted isolated components of the overall network design, thereby ignoring important interdependencies. Inflexibility and sub-optimality of this paradigm result in poor performance for ad hoc wireless networks in general, especially when applications impose specific constraints (e.g. delay, energy, bandwidth) are present. To meet these requirements, a cross layer protocol design that supports adaptivity and optimization across multiple layers of the protocol stack is needed. Adaptation capabilities at each layer of the protocol

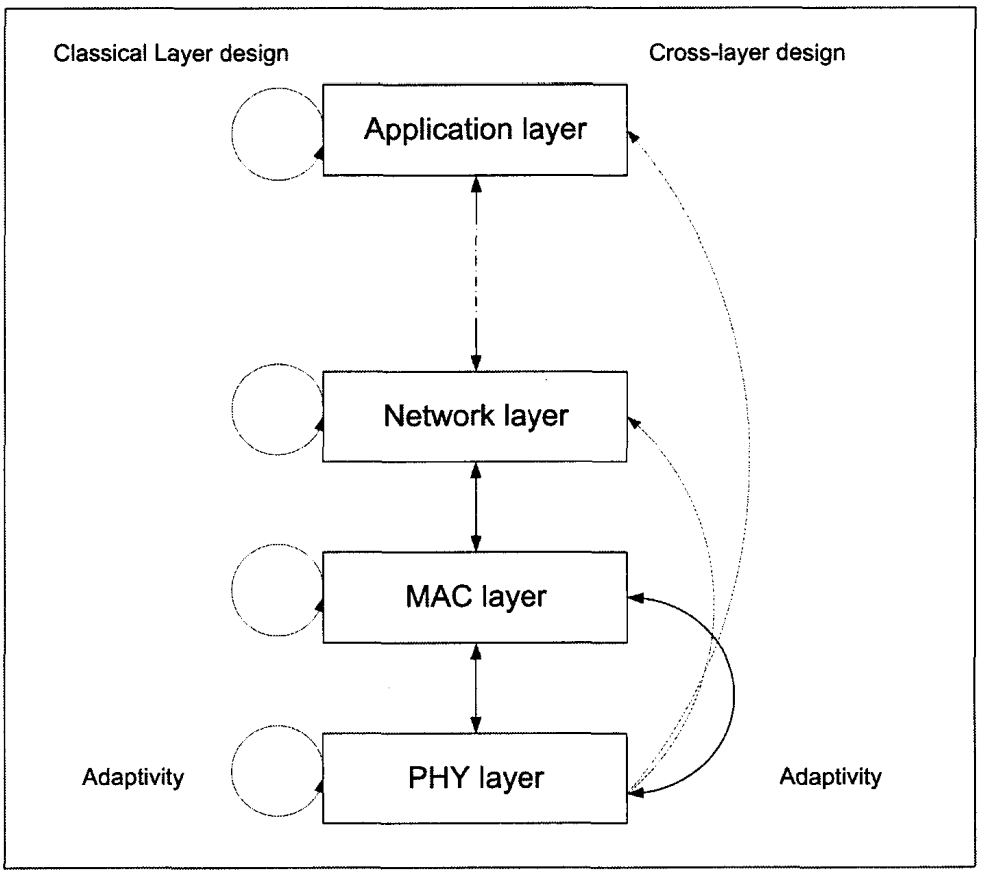

Figure 1.1: The OSI model for protocol stack design and operation. In classical layer design, adaptivity is independent of other layers. The interest in cross-layer design lies in the possibility of layer adaptivity in relation to other layers conditions.

stack should compensate for variations at that layer, based on the time-scale of these variations. Specifically, variations in link SNR are very fast, on the order of microseconds for vehicular mobile users. The network layer timescale is based on the frame duration of few milliseconds [7] and then is prone to the channel SNR variations. A a general compensation between the *physical* (PHY) and *medium access control* (MAC) becomes necessary in the perspective of an optimization analysis. Information about the time variation of single layers should be exchanged between the two layers in order to obtain a more general response. PHY-MAC cross-layer performance optimization for *single-input single-output* (SISO) wireless communications has been studied by many researchers [7, 8, 9, 10].

1.3 Thesis contributions

The cross layer approach proposed in this work focuses on the effects of MIMO capacity on network performance and analyze a joint optimization of the MIMO communication system. Specifically, the PHY layer is characterized by multiple antennas at both sides of the communication link, which results in a significant channel capacity increase. At the PHY layer, the amount of information sent over the MIMO channel per unit time/bandwidth is defined as the signalling rate of the transmitter. The MAC layer of the system is characterized by a single finite buffer, in which data arriving from the higher application layers are enqueued. The dependance between the PHY and MAC layers lies in the PHY signalling rate, which represents the number of bits (or packets) which are de-queued from the buffer and transmitted over the MIMO channel per unit time/bandwidth. This definition states the equivalence between the signalling rate at the PHY layer and the buffer service rate at the MAC layer.

According to the definition of Shannon's channel capacity, the signalling rate need to be lower than the channel capacity value to ensure reliable transmissions. Because of this definition, the channel capacity covers a primary role in the modeling of the communication system, when there are no constraints on energy and coding/decoding complexity at the transmitter/receiver side. In this conditions it is possible to focus on the MIMO channel capacity, which represents the maximum amount of information that is supported by the MIMO channel.

The main challenges of MAC-PHY cross-layer design for wireless communication are twofold. At the physical layer, on one side, the MIMO wireless mobile channel is characterized by a time-variant capacity which introduces the possibility of no reliable transmission (outage probability). If the signalling rate is not adapted to the channel capacity, outage events occur when the signalling rate is higher than the capacity value. At the MAC layer, on the other side, the limited buffer space is significantly vacated or exhausted from time to time, due to bursty arrivals from upper layers. For these reasons, both MAC and PHY layers introduce a probability of data loss, either because of buffer overflow or because of channel capacity outage. As a consequence, a joint consideration of the buffer and capacity states is necessary for a complete system understanding and efficient system design.

When transmission is performed with a constant signalling rate, the outage

probability is shown to be a significant component of the total data loss probability of the system. The amount of information reliably transmitted is sensibly affected by the outage events. In general, the capacity subject to outage is lower than the ergodic channel capacity, which represents the maximum amount of information supported by the channel [2]. The limit of the ergodic capacity can be achieved by adapting the transmission rate to the capacity evolution. From this perspective the role of the *channel state information* (CSI) becomes a primary issue not only for efficient decoding at the receiver, but also for exploiting adaptive rate strategies at the transmitter.

In order to fully exploit the MIMO spectral efficiency, an accurate knowledge of the CSI must be made available at the transmitter. Considering a mobile flat fading channel, the characteristics of the channel process suggest the use of a particular basis estimation method [11] to obtain the CSI. An appropriate channel estimation is obtained by the use of pilot symbols and a set of functions called *discrete prolate spheroidal sequences* (DPSS). The estimation accuracy shows that this basis set is appropriate for the representation of the mobile flat fading channel in isotropic scattering environments. More realistic environments subject to directional scattering are also investigated, showing that the accuracy of DPSS decreases significantly in those environments. A solution to this limitation is proposed with a new set of functions called *modulated prolate spheroidal sequences* (MDPSS).

With an accurate CSI at the transmitter side, the choice of a specific signalling rate has significant effects on the system throughput. Moreover, the optimal signalling rate must be chosen according to both the instantaneous channel capacity and the buffer occupancy, in order to minimize the probability of data loss at both layers. The process under analysis is defined by the pair of capacity and buffer state. If the knowledge of the capacity time correlation is available, it is possible to predict the capacity variation into a specific time interval. Specifically, a prediction method for the channel capacity evolution is proposed based on a *discrete autoregressive* model (DAR), which assigns a Markov nature to the capacity process. Consequently, a *finite state Markov chain* (FSMC) is developed to model the instantaneous channel capacity evolution. Considering the finite buffer at MAC layer, both capacity and buffer evolution in time can be modeled by two different FSMC. Introducing the concept of reward for each possible signalling rate, the optimal transmission strategy is analyzed through the *Markov decision processes* (MDP) theory. The result is a set of optimal signalling rates for each process state: the transmission rate is chosen

according to the buffer state and channel capacity to maximize the throughput of the system, namely the number of packets correctly received at the destination MAC layer.

1.4 Thesis structure

This work is organized as follows. In Chapter 2 the definition of channel capacity for a MIMO communication system is discussed. Particular attention is dedicated to the capacity gain of MIMO system, to the definition of ergodic capacity and outage capacity. Different definitions of capacity are proposed according to what kind of channel state information is available at the transmitter. Latest published results on the capacity time correlation are presented and, based on those results, a *discrete autoregressive* (DAR) model is proposed to predict the channel capacity evolution.

In Chapter 3 the effects of the MIMO capacity distribution on network performance are investigated. The case of a constant rate transmission is analyzed, showing the effects of the outage probability on the overall system performance. Some system design issues are addressed: the buffer size for a specific service rate, or the service rate for a given buffer size can be derived to fulfill the requirement on the total probability of data loss. Numerical solutions to those issues are obtained and analyzed.

Chapter 4 describes the powerful channel estimation method based on the use of pilot symbols and the set of DPSS functions. The accuracy of this basis expansion is discussed in detail, with numerical results showing the *mean square error* (MSE) of the estimation in both isotropic and non-isotropic environments.

In Chapter 5 the improvements of the variable rate over the constant rate transmission are presented. The optimal rate transmission strategy is investigated through the theory of MDP, which ensures the choice of the optimal signalling rate taking into account both the capacity state and the buffer occupancy. Numerical simulations show the improvement of variable rate transmissions over the constant rate transmission previously discussed. Benefits in terms of system throughput and total failure probability are evident.

Chapter 2
Capacity of MIMO Channel

Introduction

This chapter focuses on the definition of channel capacity in the particular case of MIMO wireless systems. The main results present in the literature are carefully described and supported by numerical simulations, to provide the understanding of the MIMO capacity process in wireless mobile channels.

For a MIMO system, the Shannon capacity limit dictates the maximum data rate that can be transmitted over the MIMO channel with asymptotically small error probability, assuming no constraints on the delay or the complexity of the encoder and decoder. Much of the attention on MIMO systems was due to the initial work by [4] predicting a remarkable capacity growth for wireless system with multiple antennas, when the channel exhibits rich scattering and its variations can be accurately tracked. For mobile wireless communication, the fading process due to the user mobility introduces a strong time variation of capacity. On one side, if the *channel state information* (CSI) is available to both the transmitter and the receiver, the channel ergodic capacity can be achieved by adaptive transmission strategies. On the other side, if no CSI is available at the transmitter, the best solution is a constant rate transmission based on the *channel distribution information* (CDI). In this case the channel ergodic capacity cannot be reached and capacity outage must be taken into consideration. The two cases lead to a different definition of channel capacity, which is the maximum information rate the channel can support.

Recent results on MIMO capacity are presented, indicating that the capacity gain obtained from multiple antennas heavily depends on the available channel information at either the transmitter or receiver, the channel *signal to noise ratio* (SNR) and the correlation between the channel gains on each antenna element. Temporal and spatial correlation are discussed, showing their effects on channel capacity. Moreover, the use of the general uncorrelated channel model is justified according to

the principle of maximum entropy, showing that spatially correlated channel can be described as a lower-dimensional uncorrelated channel. Finally, referring to [12], it is shown that the capacity time-correlation in fading environments is closely related to the channel time-correlation, meaning that the evolution in time of the two processes is comparable.

2.1 Fundamental capacity limits

In this section the definition of Shannon channel capacity and some fundamental assumptions beyond that model are discussed.

Channel capacity under average or peak power constraints has been subject of intense research for many years. The research was pioneered by Claude Shannon in 1948. He showed that capacity, defined to be the maximum rate at which reliable communication is possible, can be characterized in terms of the mutual information between the input and output of the channel. Shannon proved that the channel capacity is equal to the mutual information of the channel maximized over all possible input distribution. In case of an *additive white Gaussian noise* (AWGN) channel, the channel capacity is given by:

$$C = B\log_2(1+\gamma), \quad bits/s \tag{2.1}$$

where B is the channel bandwidth in Hz and γ is the received SNR within this bandwidth. The capacity of a channel dictates the maximum rate at which reliable communication can be performed, without any constraints on the transmitter and receiver complexity. Shannon showed that for any rate $r < C$, there exists a suitable channel code which achieve an arbitrarily small probability of block (or symbol) error. In general, the capacity achieving codes for wireless channels have asymptotically large block lengths. They can be used when there are no constraints in term of energy resources or information delay. This optimal coding scheme drives the probability of error to zero for any data rate below capacity, but the complexity of these schemes makes them hard to approximate with practical implementations. Moreover, Shannon showed that codes operating at $r > C$ cannot achieve an arbitrarily small error probability, in that case the error probability is bounded away from zero. Therefore, the channel capacity is truly the fundamental limit to communication.

Although it is theoretically possible to communicate at any rate below capacity, it might be a difficult task to design channel codes (or codes with a reasonable block length and encoding/decoding complexity) at rates close to capacity. In the specific case of MIMO systems, practical space-time coding and decoding techniques are shown to achieve near capacity limits in some scenarios [2]. Capacity investigations provide a theoretical information limit against which performance of different MIMO transmission and reception strategies can be compared.

A different capacity formulation arises when the energy resources are limited. Ad hoc wireless networks with limited energy must deal with capacity under energy constraints, as opposed to peak or average power constraint. With finite energy it is not possible to transmit any number of bits per unit energy with asymptotically small error probability. Capacity per unit energy is explored in [13, 14], which obtain the capacity of finite-energy channels in terms of bits. The capacity of a finite-energy channel in bits is an important concept, since it indicates that ad hoc wireless networks with finite energy nodes only have a finite number of bits that a given node can transmit before exhausting its energy. Allocating those bits to the different requirements of the network information transmission, exchange of routing information or channel estimation, becomes a challenging optimization problem that clearly requires cross-layer design.

The analysis of finite-energy channel constraints goes beyond the purposes of this work and represents an open issue in certain practical situations.

2.2 The MIMO channel capacity

Capacity for a communication channel is the maximum, asymptotic (in block length) error free transmission rate that can be achieved. For the specific case of *multiple-input multiple-output* (MIMO) channel, it is a complicated function of the channel conditions and transmit/receive processing constraints. In recent years, there has been great effort in obtaining capacity expressions for channel models that better reflect the channels characteristics underlying those systems. Recent results in this area include the capacity of channels with multiple antennas at both the transmitter and receiver [5]. Geometric scattering description leads to the definition of one-ring or two-ring model for MIMO wireless systems, which end in different expression for capacity [12]. The results described in the pioneering work [4] indicate that in rich

scattering propagation environment, the capacity of MIMO channels increases linearly with $min(N_t, N_r)$, where N_t, N_r are respectively the number of antennas at the transmitter and receiver side. Scattering and propagation conditions define the entries of the channel matrix $\mathbf{H}$, which characterizes the channel gains. In practice, a wireless channel is never simple: the received waves are a superposition of waves generated by reflections, scattering and the diffraction of the environment surrounding the transmitter and receiver. The superposition of all these effects cannot be characterized in a deterministic sense, for this reason the wireless channel must be described by statistical models. The use of *independent identically distributed* (i.i.d.) channel entries for the channel matrix is justified in [15] in accordance with the maximum entropy principle. The independence of the channel matrix components is the result of imposing only the condition of a fixed norm of the channel matrix. Assuming perfect channel knowledge at the receiver, the ergodic capacity of a $N_t \times N_r$ MIMO channel with input covariance matrix $Q = \mathrm{E}(\mathbf{x}\mathbf{x}^H)$ is [15]

$$\bar{C} = max_Q = \mathrm{E}\{C(\mathbf{Q})\} \tag{2.2}$$

with

$$C(\mathbf{Q}) = \log_2 \det\left(\mathbf{I}_{N_r} + \frac{\gamma}{N_t}\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathbf{H}}\right) \tag{2.3}$$

where γ is the SNR over the channel. The maximization is over the set of positive semidefinite Hermitian matrices $\mathbf{Q}$ satisfying the power constraint $tr(\mathbf{Q}) \leq P$ and the expectation is with respect to the random channel matrix $\mathbf{H}$. In the case of Gaussian i.i.d. $\mathbf{H}$ entries, ergodic capacity is achieved with $\mathbf{Q} = \mathbf{I}$ [4]. Moreover, the increasing of capacity for multiple antennas can be shown by a single value decomposition of the channel matrix $\mathbf{H}$. The channel matrix can be decomposed in multiple parallel independent sub-channels, whose number is equal to $g_m = \min(N_T, N_R)$. g_m is the multiplexing gain of the MIMO system and represents the asymptotic increasing of the channel capacity. Through an opportune data coding, the sub-channels can be exploited for independent data stream, leading to a linear increase of capacity according to g_m.

2.3 Time-varying channel

When the communication channel is time-varying the channel capacity has multiple definitions, depending on what is known about the channel state as well as the time scale of the underlying channel fading process. Considering a time scale based on the system frame duration T_f, the fading process is described by the block-fading model, in which the channel gains hold constant for the duration of the transmission frame T_f. From Eq. 2.3, it can be inferred that each channel realization defines a specific capacity level. When the instantaneous channel gains are known perfectly both at the transmitter and the receiver, the transmitter can adapt its transmission strategy (i.e. the signalling rate) relative to the instantaneous channel state. From now on, the information about the instantaneous channel gains will be referred as the *channel state information* (CSI). The Shannon capacity in this case is the maximum mutual information averaged over all the channel states. Ergodic capacity means that a reasonably long time sample of the capacity realizations has a distribution similar to the statistical distribution of the capacity process. Ergodic capacity is an appropriate capacity metric for channels that varies quickly in the time scale of interest: the channel is said to be ergodic in the time period of interest. With CSI at the transmitter (CSIT) ergodic capacity can be achieved using an adaptive transmission policy. This possibility will be explored in Chapter 5. If an adaptive rate is used, the signalling rate varies with the instantaneous capacity state and the ergodic capacity refers to the maximum possible long-term average of instantaneous rates.

An alternate capacity definition for time-varying channels is *outage capacity* [2]. Defining γ as the channel SNR and p as a percentage value, the outage capacity $C_{out,p}(\gamma)$ is defined as the transmission rate in bps/Hz that can be supported by $(100-p)\%$ of the fading realizations of the channel. This means also that the probability of correctly decoding at the receiver is equal to $(100-p)\%$ for a signalling rate $r = C_{out,p}(\gamma)$ and, consequently, the *frame error rate* (FER) is $p_e(\gamma, r) = p$. Given a constant signalling rate transmission, with perfect transmitter and receiver CSI, the signalling rate r does not depend on the channel variations except in outage states, in which no data is transmitted. As a result, the average rate associated with outage is typically smaller than the ergodic capacity. Outage capacity is an appropriate capacity metric in slowly varying channels, when the channel coherence time

exceeds the duration of the frame. If the channel state is not good enough to support the desired rate, an outage event is declared and no data is transmitted, since the transmitter knows that the channel is in outage. When only the receiver as perfect CSI, the transmission strategy must rely on the CDI instead of the instantaneous channel state. The transmitter must maintain a fixed-rate transmission strategy optimized with respect to its CDI. This means that knowing some statistic distribution of the fading process, the transmitter chooses a fixed rate, with no feedback from the receiver and no adaptive strategy. Without CSI available, the transmitter might send data at a rate that is not supported by all the channel states. Whenever the channel enters a deep fade the receiver declares an outage and the data are lost, since no reliable communications are possible. In this outage scenario, each transmission rate has an outage probability associated with it so capacity is parameterized by outage probability. This case will be considered and carefully described in Chapter 3.

From above, it is clear that the CSI and CDI play a major role in the definition of channel capacity under fading processes. The ergodic capacity in a fading environment represents the maximum amount of information that can be transmitted over the MIMO channel. It is achieved by knowing the CSI in transmission and adapting the signalling rate consequently. Knowing only the CDI in transmission results in a generally lower capacity utilization. Fig. 2.1 shows the difference between ergodic and outage capacity for an outage probability of $p = 0.05$.

In order to exploit the MIMO increasing in spectral efficiency, the analysis of the MIMO channel capacity distribution is required, as well as an appropriate method of obtaining the CSI that will be presented in Chapter 4.

2.4 Capacity in correlated fading

The impact of channel correlations on the capacity of MIMO channel is of interest because the channels encountered in practice invariably exhibit non-zero correlations in time and space. The correlation is the second order statistic of a random process, it describes the time interval (coherence interval) in which the process shows a dependance with the past samples. The coherence interval may also be referred as memory interval [2]. Temporal correlations are those that exist between the channel matrix realizations at different time instants. Spatial correlations are those that exist between the elements of the channel matrix for each realization.

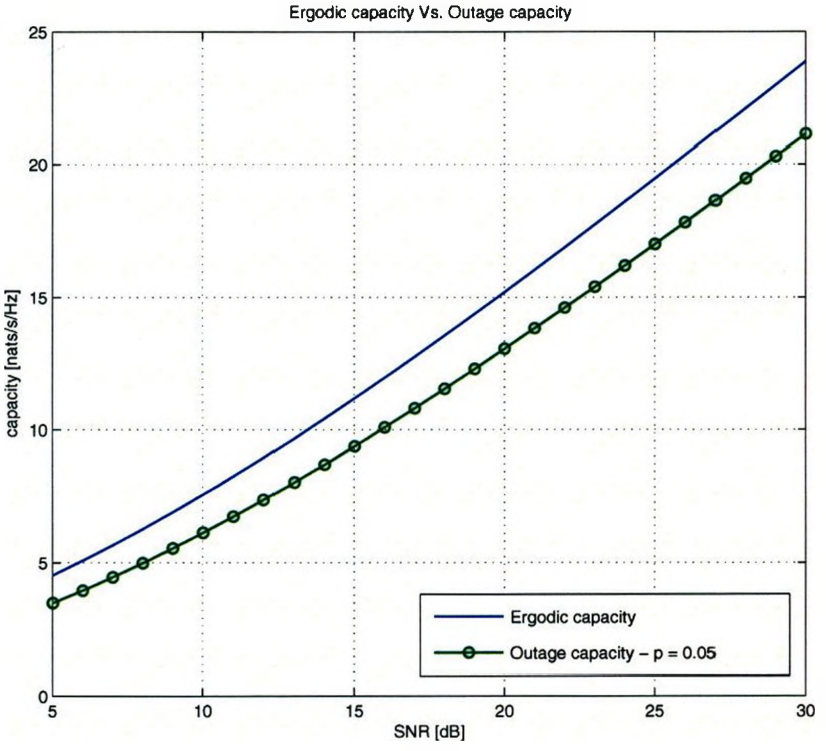

Figure 2.1: Difference between ergodic capacity and outage capacity. Ergodic capacity is the result of the analytical expression in Eq. 2.3, outage capacity is defined as the signalling rate that can be supported by the channel in $(100 - p)\%$ of the fading realizations of the channel. Outage capacity is always lower than ergodic capacity.

First, the impacts of temporal correlation are introduced. A general flat fading mobile wireless channel exhibits a temporal correlation dependent on the mobile velocity and the time scale underlying the system evolution. In a block fading model, the fading process is assumed to hold constant for the duration of the time frame T_f. The mobile velocity is characterized by a Doppler frequency f_d in Hz, the relation between the Doppler frequency and frame duration leads to the definition of the normalized Doppler frequency $f_m = f_d T_f$. If no CSI is available at the receiver, temporal correlation will increase the capacity. This is because the channel correlation allows some amount of channel estimation that is not possible in a memoryless channel. This means that with no channel estimation, a prediction can be possible if the channel correlation is known. This prediction possibility will be exploited in Chapter 5, to adapt the transmission rate to the capacity evolution on a frame by frame time scale. When the CSI is known at the receiver (CSIR), the temporal correlation does not affect ergodic capacity: at every time the receiver knows the channel state and any information on the past channel samples is useless.

A different analysis is required in understanding the effects of spatial correlation. In MIMO systems, spatial correlation is introduced by the system geometry, which includes the multiple antennas at the transmitter and receiver sides and scattering environment. Intuitively, spatial correlation between fades decreases as the density of scatterers in the vicinity increases or the antennas spacing increases. The most commonly used model for spatial correlation is the Kronecker product form

$$\mathbf{H} = \mathbf{R}_r^{1/2} \mathbf{H}_w \mathbf{R}_t^{1/2} \tag{2.4}$$

where $\mathbf{H}_w$ is an i.i.d. Rayleigh fading channel and $\mathbf{R}_t$,$\mathbf{R}_r$ are the transmit and receive correlation matrices respectively. This model has been shown to be reasonably tractable and accurate through field measurements [2]. To understand the impact of spatial correlation, different consideration must be carried for high and low SNR. For high SNR, substituting Eq. 2.4 in Eq. 2.3, capacity can be approximated as [16]

$$C \approx \log_2 \det \left(\mathbf{I}_{N_R} + \frac{\gamma}{N_T} \mathbf{H}_w \mathbf{H}_w^H \right) + \log_2 \left(\frac{|\mathbf{R}_T|}{N_T} \right) + \log_2 \left(\frac{|\mathbf{R}_R|}{N_T} \right), \tag{2.5}$$

from which can be inferred that spatial correlation decreases the capacity of uncorrelated channels. Intuitively, a strong spatial correlation reduces the multiplexing gain,

which is maximum in i.i.d. channels. For low SNR some geometrical considerations are needed. Consider a single cluster scenario, in which the received signal comes from a limited narrow area around a specific direction ϕ_R. The distribution of *angles of arrival* (AoA) and *angles of departures* (AoD) can be modeled as a uniform narrow spread Δ_R and Δ_T around some mean AoA ϕ_R and AoD ϕ_T respectively. In this case the spatial covariance function is approximately a sinc function, as it will be deeply analyzed in Chapter 4. As a consequence, the covariance matrices on both sides have a few approximately equal eigenvalues, with the rest very close to zero. In other words,

$$\mathbf{R} = \mathbf{U} \begin{bmatrix} \frac{1}{\beta}\mathbf{I}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{U}^H \tag{2.6}$$

where βN is the number of non-zero eigenvalues of the corresponding covariance matrix[1]:

$$\beta = \lfloor N\Delta |\sin\phi| \rfloor + 1; \tag{2.7}$$

$\mathbf{I}_\beta$ is the unity matrix of size $\beta N \times \beta N$ and the unitary matrix $\mathbf{U}$ is composed of modulated *discrete prolate spheroidal sequences* (DPSS) [17]. The DPSS will be deeply investigated in Chapter 4 as they represent an appropriate basis expansion for the wireless mobile channel fading. Particularly, it will be shown that the covariance matrix eigenvalues exhibit the property described above: a small number of them is clustered near 1 and the remaining number falls to zero very quickly. This analysis shows that a strong spatial correlated fading can be approximated by a lower-dimensional uncorrelated covariance matrix $N_{Teff} \times N_{Reff}$, whose dimensions are defined by the set of highest eigenvalues of the channel matrix. The same conclusion is obtained in [2], in which is underlined that the spatial correlation can make the channel matrix rank-deficient. This means that, in a low SNR case, only a subset of channels can be efficiently used with the limited available signal power.

A deep understanding of these considerations is beyond the purposes of this work, but they suggest that the study of channel matrices with i.i.d. entries can bring interesting results also in practical correlated fading environment.

[1] $N = N_T$ for transmit and $N = N_R$ for receive side.

2.5 MIMO capacity distribution

In the literature, many studies deal with the statistical properties of the MIMO wireless capacity. Some of those statistical properties are the *probability density function* (PDF), the *cumulative distribution function* (CDF), the *level crossing rate* (LCR) and the *average fade duration* (AFD) of the channel capacity. However, deriving exact analytical expressions for all those statistical properties appears a complicated task. Assuming that elements of the channel matrix are i.i.d. zero mean Gaussian process, MIMO capacity is approximately Gaussian for large number of transmitting and receiving antennas N_T,N_R [18]. The mean and the variance of the capacity process are given by [18]:

$$\frac{\bar{C}}{N_R} = \beta \ln\left[1+\frac{\gamma}{\beta}-\frac{1}{4}F\left(\frac{\gamma}{\beta},\beta\right)\right] + \ln\left[1+\gamma-\frac{1}{4}F\left(\frac{\gamma}{\beta},\beta\right)\right] - \frac{\beta}{4\gamma}F\left(\frac{\gamma}{\beta},\beta\right) \quad (2.8)$$

$$\sigma_C^2 = -\ln\left(1-\beta\left[\frac{1}{4\gamma}F\left(\frac{\gamma}{\beta},\beta\right)\right]^2\right) \quad (2.9)$$

where

$$F(x,z) = \left[\sqrt{x\left(1+\sqrt{z}\right)^2+1}-\sqrt{x\left(1-\sqrt{z}\right)^2+1}\right]^2 \quad (2.10)$$

For small values of SNR γ these equations could be simplified to produce

$$\frac{\bar{C}}{N_R} = \gamma - \frac{1}{2}\frac{\beta+1}{\beta}\gamma^2 \quad (2.11)$$

$$\sigma_C^2 = \gamma - \frac{1}{2}\frac{\beta+2}{\beta}\gamma^2 \quad (2.12)$$

while for large SNR γ the following approximations are valid for $\beta = 1$ [3]

$$\frac{\bar{C}}{N_R} \approx \ln\frac{\gamma}{e}+\frac{2}{\sqrt{\gamma}}-\frac{1}{\gamma} \quad (2.13)$$

$$\sigma_C^2 \approx \frac{1}{2}\left(\ln\frac{\gamma}{4}+\frac{2}{\sqrt{\gamma}}-\frac{1}{4\gamma}\right)-0.193; \quad (2.14)$$

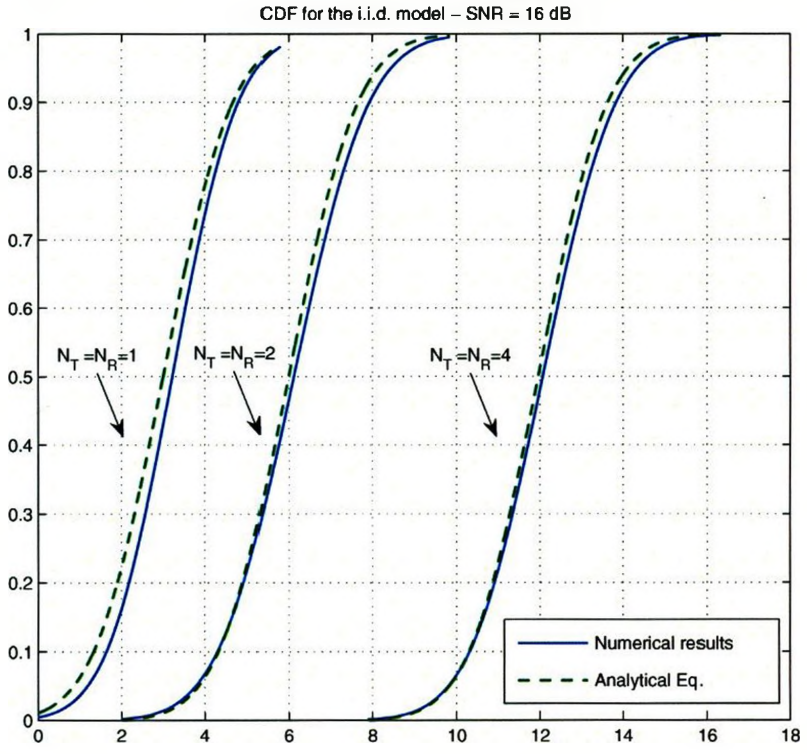

Figure 2.2: Numerical and theoretical CDF of channel capacity for different numbers of antennas. The theoretical distribution is obtained by the mean and variance in Eq.s 2.8,2.9. The numerical distribution is obtained by the values of Eq. 2.3.

Interestingly, authors of [5] considered the case of a reasonable number of antennas N_T, N_R, proving that the Gaussian distribution holds in the case of small numbers of antennas and i.i.d. channel coefficients. High accuracy in the capacity distribution is obtained even for 2×2 MIMO systems. The Gaussian approximation becomes more and more accurate for high number of N_T and N_R. Fig. 2.2 show the accuracy of the Gaussian distribution for different numbers of antennas. The SNR for the numerical simulations is 16 dB.

From this analysis it is clear that the MIMO channel capacity is a random process dependent on the system geometry and SNR. Fig. 2.3 and Fig. 2.4 show the numerical results obtained from Eq. 2.3 and the analytical forms in Eq.s 2.8, 2.9.

In the case of a time variant MIMO channel, capacity evolves in time according to the fading process [12]. Referring to the block fading model, the time variation of the channel is described by the normalized Doppler frequency $f_m = f_d T_f$, where f_d

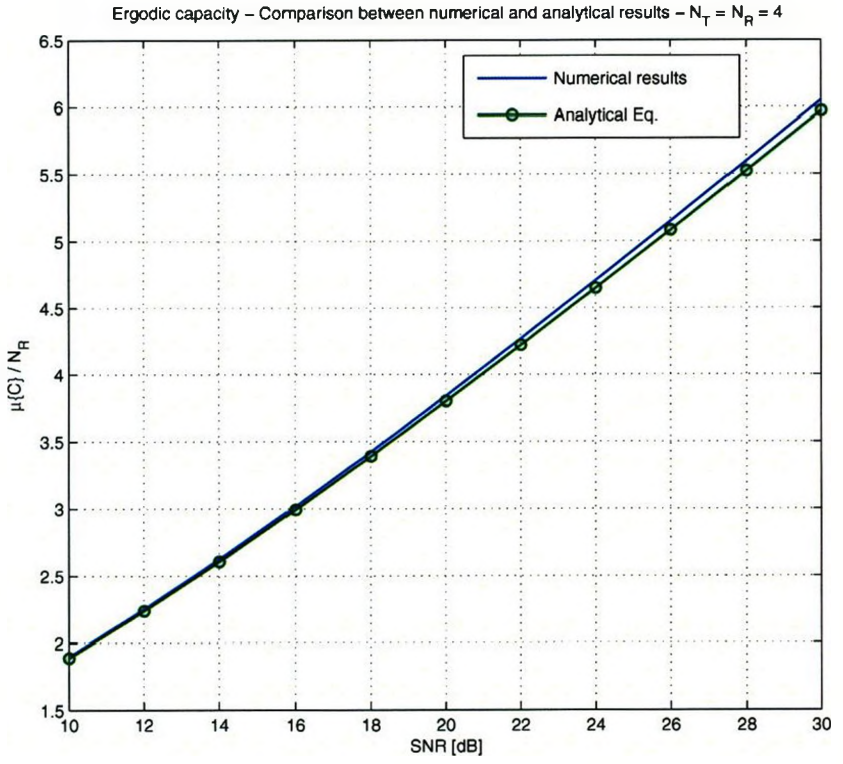

Figure 2.3: Mean capacity Vs. SNR. The numerical results from Eq. 2.3 are compared with the analytical Eq. 2.8. The analytical equation is shown to be very accurate.

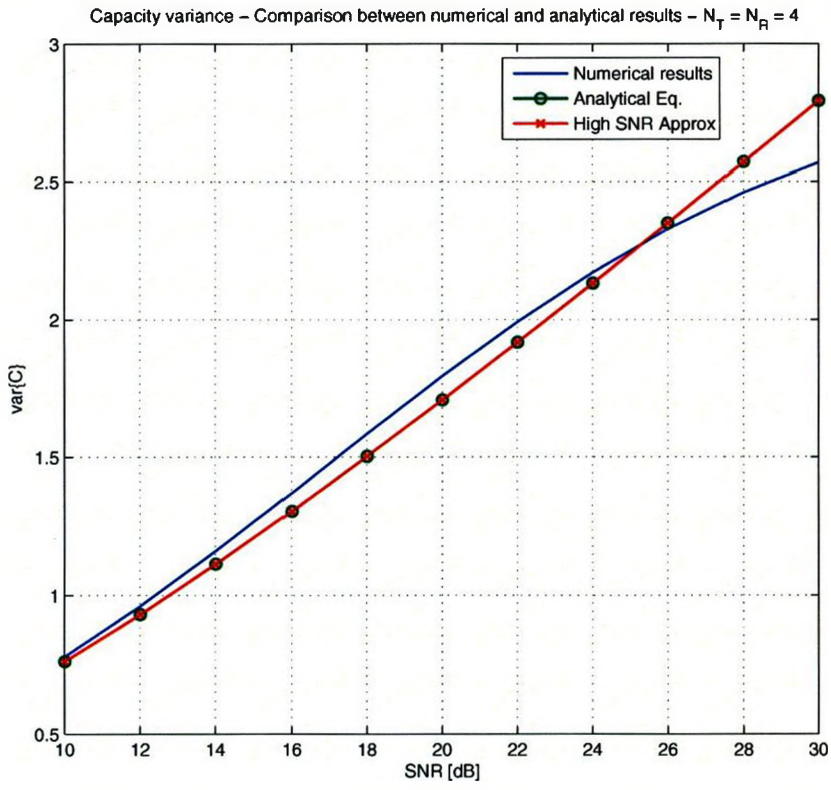

Figure 2.4: Variance of capacity Vs. SNR. The numerical results from Eq. 2.3 are compared with the analytical Eq. 2.9 and the approximation in Eq. 2.13. The two analytical equations are almost equal, the separation with the numerical results is due to the limited simulation length.

is the Doppler frequency in Hz and T_f is the system frame duration. The value of f_m does not affect the statistical channel description, once a proper observation interval is set to derive the statistical characteristics. If the process exhibits a long autocorrelation function, the need of independent samples to achieve accurate statistics leads to a long process observation time. The time variation of the instantaneous capacity could be described in terms of a *discrete autoregressive* (DAR) models, as presented in [19, 20]. In the simplest case of DAR-1 model, the capacity evolution equation is given by

$$\Delta\bar{C}(n+1) = \rho\Delta\bar{C}(n) + \sqrt{1-\rho^2}\xi(n+1) \tag{2.15}$$

where $\Delta\bar{C}(n+1) = \left(C(n+1) - \bar{C}\right)/\sigma_C$ is the normalized deviation of capacity $C(n+1)$ at the discrete time $n+1$ from its time average $\bar{C}$; $\xi(n+1)$ is a sequence of i.i.d. zero mean Gaussian variables with unit variance. The correlation coefficient ρ could be estimated from experimental data or its value could be deduced based on some theoretical development. Although a complete analytical description of the MIMO channel capacity is still under research, the study in [12] suggests that the instantaneous capacity evolves with the same statistical characteristics of the channel. Starting from the Gaussian *wide sense stationary* (WSS) uncorrelated scattering model used in Chapter 4, the autocorrelation sequence for ideally generated in-phase and quadrature Gaussian processes at discrete time is given by

$$R[n] = J_0(2\pi f_m|n|) \tag{2.16}$$

Once the discrete prediction interval nT_f is set, the value of ρ can be derived from Eq. 2.16.

The DAR-1 model assigns Markov nature to $C(n)$. A more sophisticated DAR-2 model can also be developed based as follows. Let $\rho_1 = \bar{C}(n)\bar{C}(n-1)$ and $\rho_2 = \bar{C}(n)\bar{C}(n-2)$ be correlation coefficients at lag 1 and 2 respectively. In this case the DAR-2 model is given by

$$\Delta\bar{C}(n) = \alpha_1\Delta\bar{C}(n-1) + \alpha_2\Delta\bar{C}(n-2) + \beta\xi(n) \tag{2.17}$$

where parameters α_1, α_2, and β are given by

$$\alpha_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2},\ \alpha_2 = \frac{\rho_2-\rho_1^2}{1-\rho_1^2},\ \beta^2 = \frac{(1-\rho_2)(1+\rho_2-2\rho_1^2}{1-\rho_1^2} \tag{2.18}$$

as the solution of the following (normal) matrix equation

$$\begin{bmatrix} \rho_1 & \rho_2 & 1 \\ 1 & \rho_1 & 0 \\ \rho_1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta^2 \end{bmatrix} = \begin{bmatrix} 1 \\ \rho_1 \\ \rho_2 \end{bmatrix} \tag{2.19}$$

The DAR model of capacity will be used in Chapter 5 as a prediction model to allow the adaptive rate transmission.

2.5.1 Outage probability

Capacity specifies how much information the channel can support, which turns into a maximum amount of information the system can reliably transmit. In MIMO system capacity is a Gaussian random process [3, 18]; the probability of outage describes the frequency at which capacity falls under a given value: in that case no reliable transmissions are possible. Let a certain maximum probability of outage P_{out} is required for the constant data rate transmission over the MIMO channel. The outage appears when the signalling information rate r exceeds the instantaneous capacity C_n at time n, therefore

$$P_{out} = \text{Prob}(C_n < r) = \int_{-\infty}^{r} p_C(x)dx = Q(\frac{\bar{C} - r}{\sigma_C}) \tag{2.20}$$

where $Q(x)$ is the well known Q-function [1] and $\bar{C}$ and σ_C can be calculated using Eq. 2.8 and Eq. 2.9. Solving Eq. 2.20 for r in terms of desired P_{out} allows one to set a proper information rate on the channel. If $P_{out} \ll 1$, which is usually a case, one can obtain a simple approximation

$$r \approx \max\left\{\bar{C} - \sqrt{-2\sigma_C^2 \ln 2P_{out}}, 0\right\} / \ln(2) \qquad [bits/s/Hz] \tag{2.21}$$

where r is the maximum signalling rate the channel can support given a desired probability of outage, $\bar{C}$ is the average channel capacity per antenna specified in Eq. 2.8, P_{out} is the desired outage probability and ln is natural logarithm.

2.6 Conclusion

The MIMO channel capacity is a random process not only because of user mobility but also because of the system geometry. Analytical expression for the mean and variance of capacity are available in the literature [18] and allow to defined analytically the concept of outage probability [2, 3]. The speed of variation of the capacity process can be described by its autocorrelation function, analytically described in [12]. With the knowledge of the capacity time correlation, a prediction method based on a AR model is proposed to describe the capacity process evolution. In the next chapter the effects of capacity variation and outage on the transmission performance are discussed.

Chapter 3
Effects of MIMO Capacity on Network Performance

Introduction

The analysis carried in this chapter considers a transmission system as shown in Fig. 3.1. A single finite queue is served with constant transmission rate defined by the physical parameters and the MIMO channel. The nature of the capacity process in the MIMO channel leads to the definition of probability of outage, presented in [3]. The possibility that the fading channel cannot support the transmission rate must be considered in a complete analysis of the transmission system, which takes into account the system queue. Referring to the system model in Fig. 3.1, data arrive at the server according to the Poisson model [21]. Data in the queue are served with a constant signalling rate r (service rate at the MAC layer) and sent over a MIMO channel characterized by a given average SNR. The use of constant signalling rate permits the application of known queuing theories such as M/D/1 and M/D/1/N, [21, 22].

From the analysis presented in Chapter 2, the outage probability in Eq. 2.20 defines the possibility of no reliable transmission over the MIMO channel. The probability of outage and the probability of buffer overflow define the possibility of loosing data in the system, therefore they must be combined to obtain the total probability of failure in the system. This analysis investigates the effect of the MIMO channel on the queuing system, taking into account the probability of outage to derive the system behaviour in terms of delay in the system and total probability of failure. A joint consideration of the error rate due to the queuing system and MIMO channel is provided, including the outage probability in the classical $M/D/1/N$ model. Numerical solutions to some system design issues are presented, with particular attention to the signalling rate that minimize the joint probability of failure of the system.

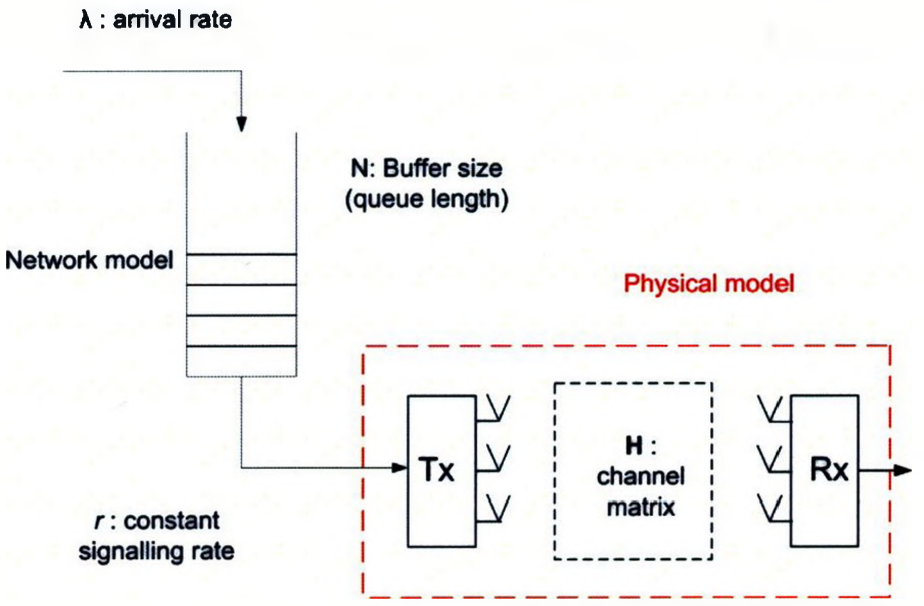

Figure 3.1: System Model

The chapter structure is the following. Firstly, the queuing behaviour for the M/D/1/N model is introduced referring to the main analytical results present in the literature. Next, the analysis of the MIMO channel capacity is included in the model. The joint consideration of outage probability and buffer overflow probability is deeply analyzed with a first set of results obtained by numerical simulations.

3.1 M/D/1/N: steady states probabilities

The system in Fig. 3.1 can be referred to the network theory of $M/D/1/N$ queuing systems [21]. The relevant performance characteristics of the $M/D/1/N$ queuing models are presented in this section. The closed-form analytical equations present in the literature are confirmed by numerical simulations in App. A, Sec. A.4.

The $M/D/1/N$ model is a finite capacity queuing system, with a queue length of $N-1$ packets. The maximum number of packets the system can hold is N: $N-1$ in queue and one under service. Packets arrive according to a Poisson process with parameter λ (arrival rate). Packets which upon arrival see a full system are rejected and do not further influence the system. The time needed to serve a packet is constant $T = 1/r$, where r is defined as service rate. The utilization factor $\rho = \lambda/r$ defines

the service capability of the system; a general assumption is that a stable system is described by $\rho < 1$. The average state probabilities $P_j(N)$ represents the probability of having j packets in the system given a queue size of N. Authors of [22] have defined a set of coefficients b_n

$$b_n = \sum_{k=0}^{n} \frac{(-1)^k}{k!}(n-k)^k e^{(n-k)\rho}\rho^k \tag{3.1}$$

which allows derivation of the probability distribution of the number of packets in the system [22]:

$$P_0(N) = \frac{1}{1+\rho b_{N-1}} \tag{3.2}$$

$$P_N(N) = 1 - \frac{b_{N-1}}{1+\rho b_{N-1}} \tag{3.3}$$

$$P_j(N) = \frac{b_j - b_{j-1}}{1+\rho b_{N-1}} \qquad j = 1, ..., N-1. \tag{3.4}$$

These equations represent the steady state results in terms of state probabilities: $P_j(N)$ represents the probability of having j packets in the system for a queue size of N packets. The probability $P_N(N)$ describes the probability of a full queue: when the system holds N packets a new arrival will be lost (overflow). In [23] it is proved that the probability $P_N(N)$ can be considered as a good description of the packet loss probability due to the buffer overflow. The mean number of packets in the $M/D/1/N$ queue is derived as [22]:

$$X_N = N - \frac{\sum_{k=0}^{N-1} b_k}{1+\rho b_{N-1}} \tag{3.5}$$

where X_N is the average number of packets in the queue. By Little's Theorem the average delay in the system T_N is given by:

$$T_N = \frac{X_N}{\lambda(1-P_N(N))}. \tag{3.6}$$

The average waiting time in the queue is defined as:

$$W_N = T_N - T = (N - 1 - \frac{\sum_{k=0}^{N-1} b_k - N}{\rho b_{N-1}})T. \tag{3.7}$$

3.2 Effects of MIMO capacity on network performance

3.2.1 System and queue delay for MIMO system

Given the MIMO wireless channel, the whole transmission system includes a single finite buffer of size $N-1$ bits in which data arrive with a certain arrival rate λ [bits/s]. The data is processed with a fixed signalling rate, which represents the amount of information the system can process and send on the channel per unit of time. Given a MIMO channel and a probability of outage defined as in [3], the maximum service rate can be derived from the channel capacity distribution. Data are sent on a MIMO channel with a specified number of transmitting and receiving antennas, N_T and N_R respectively, with a fixed signalling rate r defined by Eq. 2.21. For the simulations in this section the number of antennas is $N_T = N_R = 4$, which represent a practical system dimension [5]. The system described above can be modeled according to the theory of M/D/1/N queuing systems [21, 22]: the arrival process is modeled as a Poisson random process with parameter λ, the service rate is constant and equal to r, the buffer is one with finite size $N-1$. The M/D/1/N model allows to derive important information about the size of the queue, delay in the queue and delay in the system for data bit. Closed-form equations for the queue size, delay and state probability have been presented in Sec. 3.1. The theory of queuing systems states that in a stable system, the service rate r (amount of information the system transmits) must be higher than the incoming amount of information that need to be sent. In other words, the service rate r must be higher than the arrival rate λ, leading to a utilization factor $\rho = \lambda/r < 1$. If the service rate r is limited by the channel capacity, there will be a maximum arrival rate λ_c which will ensure the system stability. The critical arrival rate is defined as the limit arrival rate for which $\rho = \lambda_c/r = 1$. Therefore, given the channel capacity process, λ_c is the maximum arrival rate the system can support. Fig. 3.2 shows the relationship between the SNR in dB and the critical arrival rate λ_c in [bits/s] for different probabilities of outage. The maximum service rate r is provided by Eq. 2.21. Different probabilities of outage impose different restrictions to the achievable service rate and to maximum arrival rate that can be supported.

Equations presented in Sec. 3.1 allow to evaluate the average system delay for

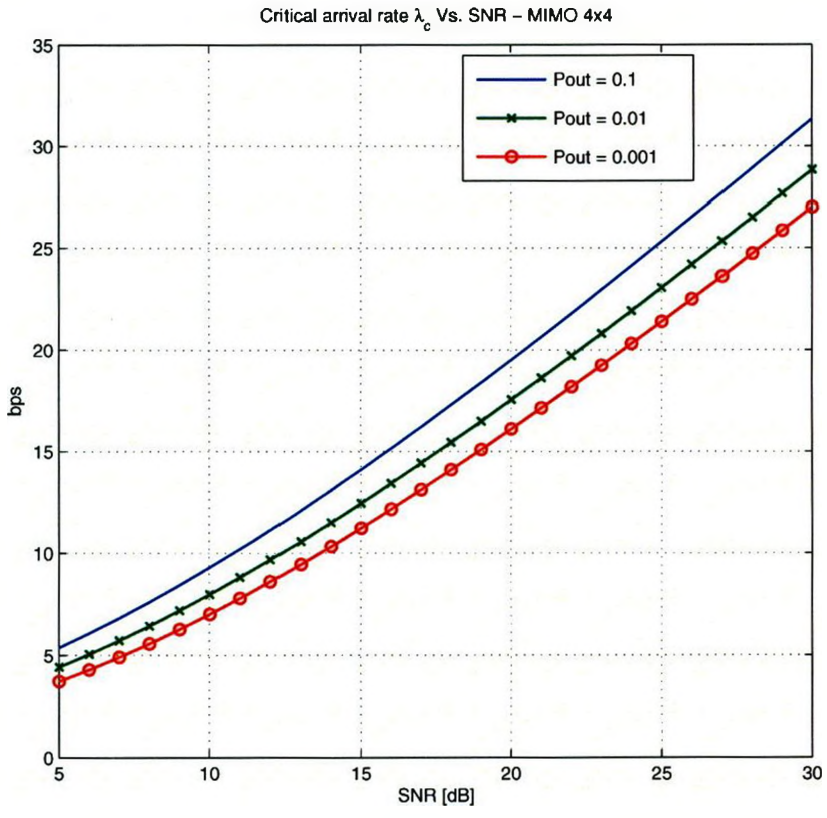

Figure 3.2: Critical arrival rate λ_C [bits/s] vs. SNR. The limit on λ_C is imposed by the maximum service rate r to satisfy a specific outage probability. This graph shows the results obtained by Eq. 2.21. An important result is that a finite set of arrival rates can be supported for a given average SNR, in order to guarantee reliable transmissions.

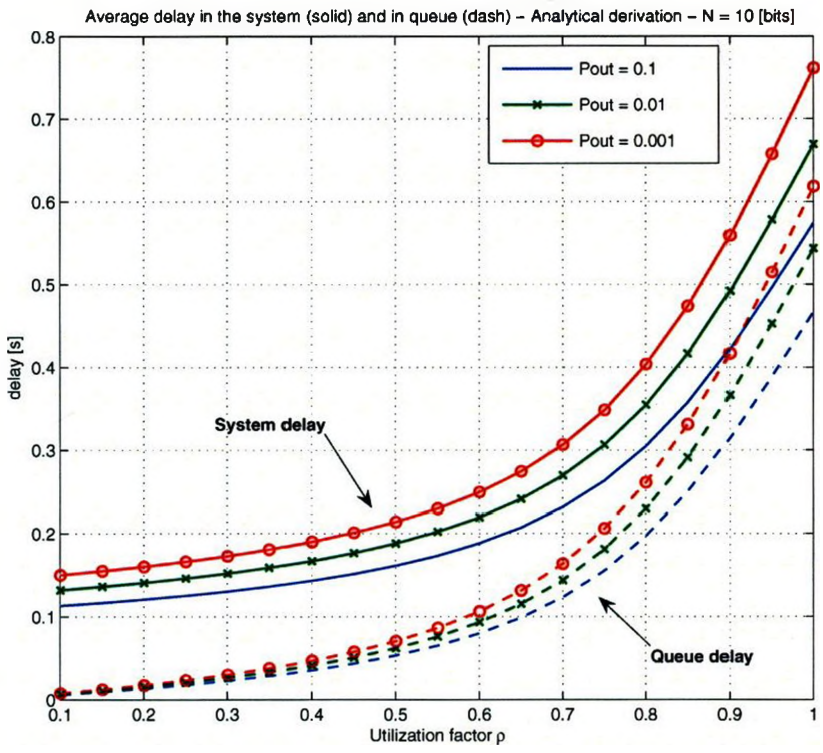

Figure 3.3: Average system delay (solid) and queue delay (dash) for different ρ and P_{out}: a smaller P_{out} leads to a smaller service rate which turns into an higher average system time for the whole set of utilization ρ. The average queue delay for small ρ is not sensible to the outage probability: the queue is almost empty therefore the service rate has a very little effect on the delay.

different outage probabilities. The system delay is defined as the sum of the average time spent waiting in the queue and the average time needed to process the data (service time) [21]. As before, the number of antennas is $N_T = N_R = 4$ with an SNR equal to 10 dB. The service rate r is computed to satisfy the outage probability P_{out} in three different cases: $P_{out} = 10^{-1}$, 10^{-2}, 10^{-3}. The system works at the service rate r for different arrival rates, according to the utilization factor $\rho = \lambda/r < 1$. Fig. 3.3 shows how different outage probabilities affect the average time per bit spent in the system and in queue. A strict requirement in terms of outage probability leads to a lower service rate, which turns into higher average system time. The time difference is appreciable for the whole set of utilization factor ρ. For small probabilities of outage the time spent in the system, which is the delay in transmission, is considerably higher: more reliable transmissions require a larger transmission delay. In case of

delay sensitive traffic, it is important to minimize the average transmission time and, therefore, the effects of outage must be taken into account. The same conclusion can be obtained by the study of the average time per bit spent in the queue (delay in queue), shown in Fig. 3.3 (dash line). For high utilization factor ρ the average time per bit spent in queue is sensible to the required probability of outage. When ρ is small, the queue is almost empty and the probability of outage does not affect the delay.

3.2.2 Total probability of failure for a MIMO system

In this section, the performance of the whole transmission system described in Fig.3.1 are studied in terms of the data loss probability. The model under analysis is a single queueing system with a finite buffer. The physical layer is characterized by a constant data rate transmission over a MIMO channel.

The two sources of data loss are the probability of outage P_{out} at the physical layer and the probability of buffer overflow $P_{overflow}$ at the MAC layer. In order to describe the whole system behaviour, the total probability of failure is defined as the sum of those two components: $P_{FAIL} = P_{out} + P_{overflow}$. Note that the sum of the two probabilities is the upper bound of the successful transmission probability, defined as

$$\nu = (1 - P_{out})(1 - P_{overflow}). \tag{3.8}$$

The buffer overflow and the outage events are independent, the first one is related to the arrival process and buffer size, the second one depends on the channel capacity random process. We assume that a certain maximum outage probability P_{out} needs to be met for the constant data rate MIMO transmission. Eq. 2.21 provides the system service rate r to satisfy the requirement on the outage probability. r also represents the critical arrival rate λ_C in [bits/s] that the system can support in the case of a finite buffer: if the arrival rate exceeds the service capability of the system, the buffer will rapidly become full and buffer overflow will be inevitable. Our interest is in defining the system probability of data loss $P_{overflow}$ due to the finite buffer. In the following example, we consider a 4×4 MIMO system, a fixed buffer size $N = 10$ [bits] and average SNR of 10 [dB]. If the desired probability of outage is 10^{-2}, the maximum service rate r will be 7.9 [bits/s/Hz] obtained by Eq. 2.21. Fig. 3.4 shows the probability of overflow for different buffer sizes N against the utilization factor

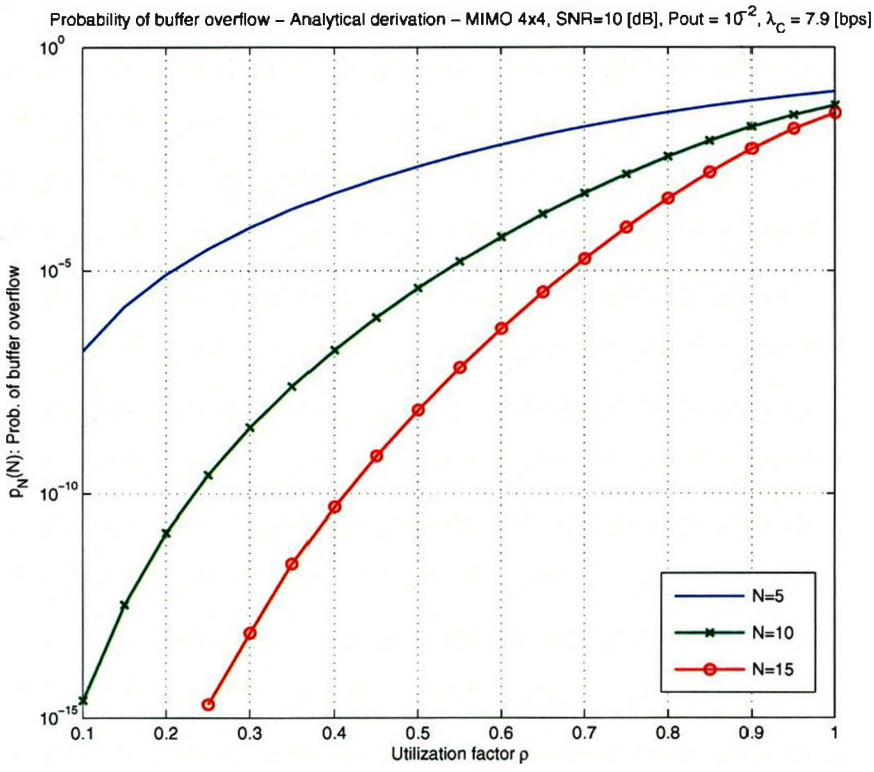

Figure 3.4: Probability of buffer overflow for different buffer sizes. SNR = 10 dB, $P_{out} = 10^{-2}$. $\rho = \lambda/r$ where $r = 7.9$ [bits/s/Hz] is the maximum service rate supported by the system. For a given arrival rate λ_c and desired outage, the buffer size has an evident influence on the overflow probability.

$\rho = \lambda/r$. The buffer sizes are specified according to the arrival process, the values are $N = 5, 10, 15$ bits.

According to the definition of total failure probability P_{FAIL} presented above, sum of the probabilities of outage and buffer overflow is considered to describe the whole system behaviour. Fig. 3.5 shows the total failure for the specified system. The outage probability is $P_{out} = 10^{-2}$ and the buffer overflow probability is specified in Fig. 3.4. For small utilization factors ρ, the fixed outage probability is dominant. Increasing the utilization factor ρ, the buffer overflow probability becomes relevant depending on the buffer sizes.

Next, a fixed buffer size of $N = 10$ [bits] is considered. Fig. 3.6 shows the probability buffer overflow for different outage requirements. The desired probability of outage has a direct effect on the service rate and plays an important role on the buffer overflow probability. Probability in Fig.3.6 is plotted versus the range of arrival

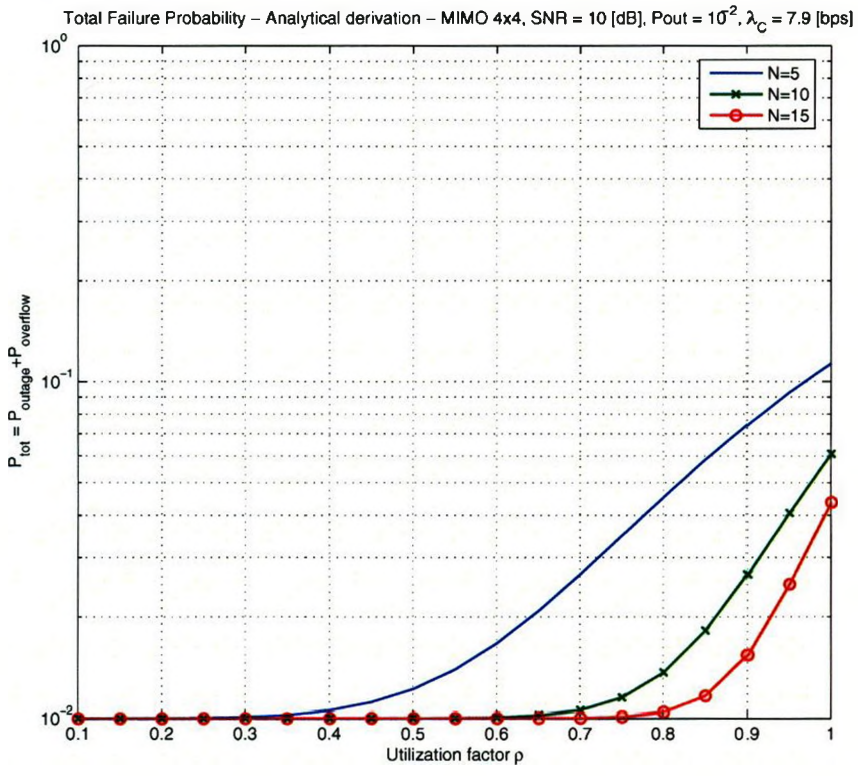

Figure 3.5: Total failure probability for different buffer sizes N [bits]. SNR = 10 dB, $P_{out} = 10^{-2}$. $\rho = \lambda/r$ where $r = 7.9$ [bits/s/Hz] is the maximum service rate supported by the system. For small utilization factors ρ, the fixed outage probability is dominant. Increasing the utilization factor ρ, the buffer overflow probability becomes relevant depending on the buffer sizes.

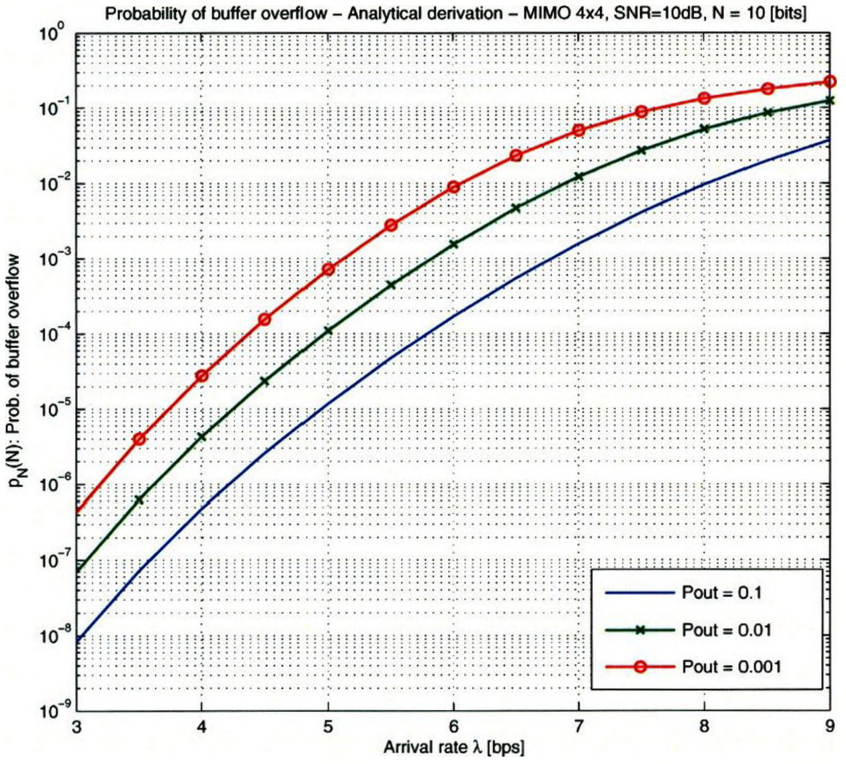

Figure 3.6: Probability of buffer overflow for different desired probabilities of outage. SNR = 10 dB, N = 10 [bits]. The arrival rate λ [bits/s] is increased until the critical value λ_c is met.

rates λ [bits/s] the system can support: values of λ grow until the critical arrival rate $\lambda_c = 9.3$ [bit/s] is met. For $\lambda > \lambda_C$ the outage probability requirement is no longer satisfied. Fig. 3.7 shows the total failure probability for a fixed buffer size N [bits] and different outage requirements. The maximum service rate r is defined by the outage probability, which also affects the maximum arrival rate λ_c the system can support.

From a design point of view, it is interesting to derive some system parameters to fulfill specific performance requirements. Let P_{th} be the required total failure probability in the system. Failure in the system happens either when the buffer overflows or outage occurs. One possible goal is to design the buffer size to achieve the desired total failure probability. To answer this question, a desired outage probability must be specified, in order to derive the system service rate r as a function of the MIMO channel parameters and SNR. The arrival rate in the system is then related

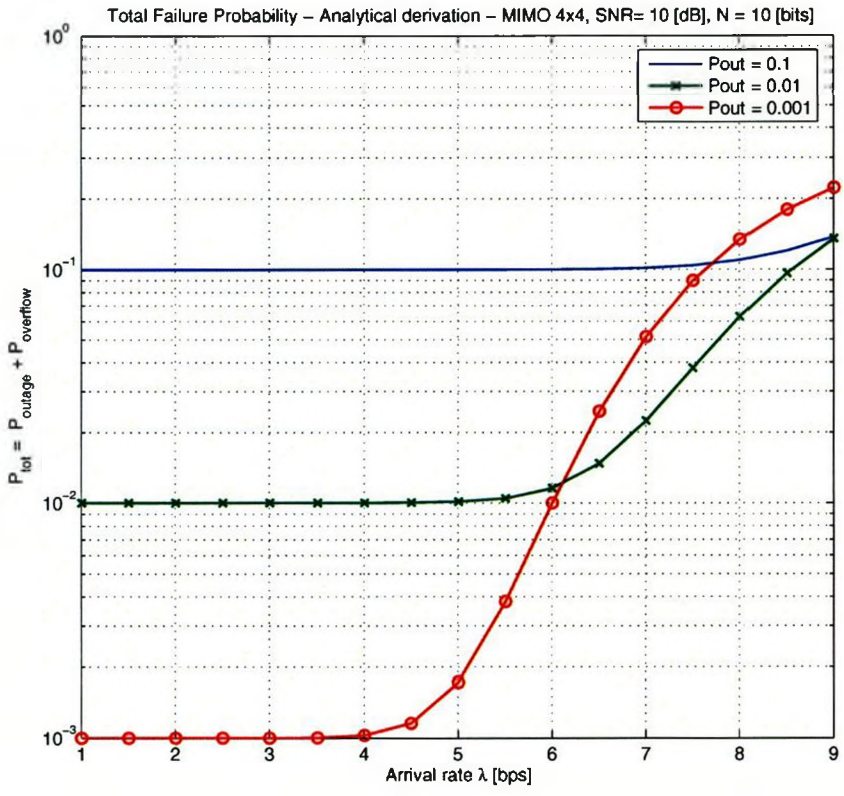

Figure 3.7: Total failure probability for different outage probabilities. The arrival rate λ [bits/s] is increased until the critical arrival rate λ_c is met. The different behaviour dependent on the required P_{out} is evident.

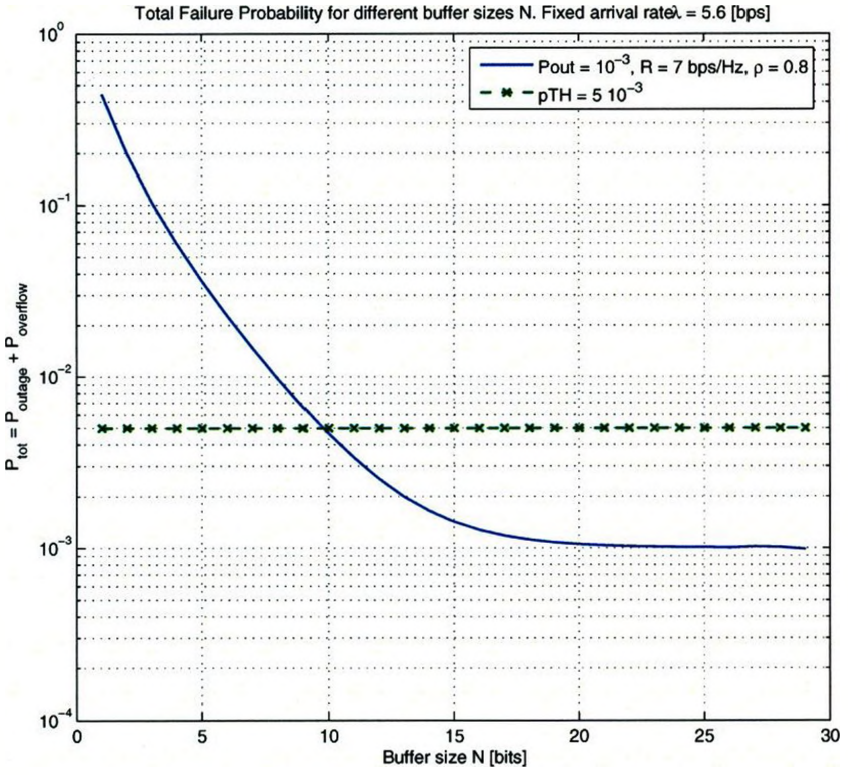

Figure 3.8: Total failure probability for different buffer sizes. For large buffers, the only contribution to the failure probability is given by outage. Given a desired threshold, a suitable buffer size can be chosen to fulfill that requirement. The desired outage imposes the system service rate r and, consequently, a maximum arrival rate. System parameters: SNR 10 dB, $P_{out} = 10^{-3}$, $\rho = 0.8$, $\lambda = 5.6$ [bits/s].

to r through the system utilization factor $0 < \rho < 1$. Fig. 3.8 shows the total failure probability for different buffer sizes N. Given a desired threshold, i.e. $P_{th} = 5 \cdot 10^{-3}$, the buffer size N can be derived numerically to fulfill the requirement. In Fig. 3.8 the desired outage is $P_{out} = 10^{-3}$. Given a MIMO channel with SNR 10 dB, the outage probability value P_{out} corresponds to a maximum signalling rate of 7 [bits/s/Hz]. The system utilization factor is set to $\rho = 0.8$ leading to the arrival rate of 5.6 [bits/s]. Both outage and buffer overflow probabilities depend on the service rate r. An optimal value of N can be computed numerically to fulfill the requirement on the total failure probability. In the case shown in Fig. 3.8, for buffer sizes greater than $N = 20$ the buffer overflow becomes negligible and the only component of P_{FAIL} is the capacity outage.

From another point of view, a possible goal would be to obtain the optimal system service rate according to the system buffer size N and a specific λ. The optimal service rate should minimize the total probability of failure in the system. Fig. 3.9 shows the behaviour of the total probability of failure versus the system utilization factor, for a fixed arrival rate and MIMO channel state. The channel SNR is 10 dB and the arrival rate is 5.6 [bits/s]. The total failure probability is the sum of the two components shown in figure 3.9, which have an opposite trend with respect to the utilization factor $\rho = \lambda/r$. For low ρ the primary contribution to the failure probability is given by outage, as the service rate r is not supported by the channel state. By increasing ρ (which means decreasing r) the outage becomes negligible while the probability of loosing data because of the buffer overflow becomes predominant (the service rate r is not enough to support the incoming data in the system). The main result is the presence of a trade-off point between the outage and buffer overflow probabilities. Figure 3.9 suggests the existence of an optimal service rate r_{opt} and corresponding utilization factor ρ which minimizes the total system failure probability for a specific arrival rate λ. Figure 3.9 shows the behaviour of the total failure probability (solid line) for different buffer size N. A lower total failure probability can be achieved by increasing the buffer size, which shifts the tradeoff point toward higher values of the utilization factor ρ. The reason is that the buffer size N affects only the probability of data loss due to overflow: the outage does not change since the channel state is the same. From a service rate point of view, the optimal r tends to be lower as the buffer size increases, following the asymptotic behaviour of outage probability: for N equal to infinity the only component of the

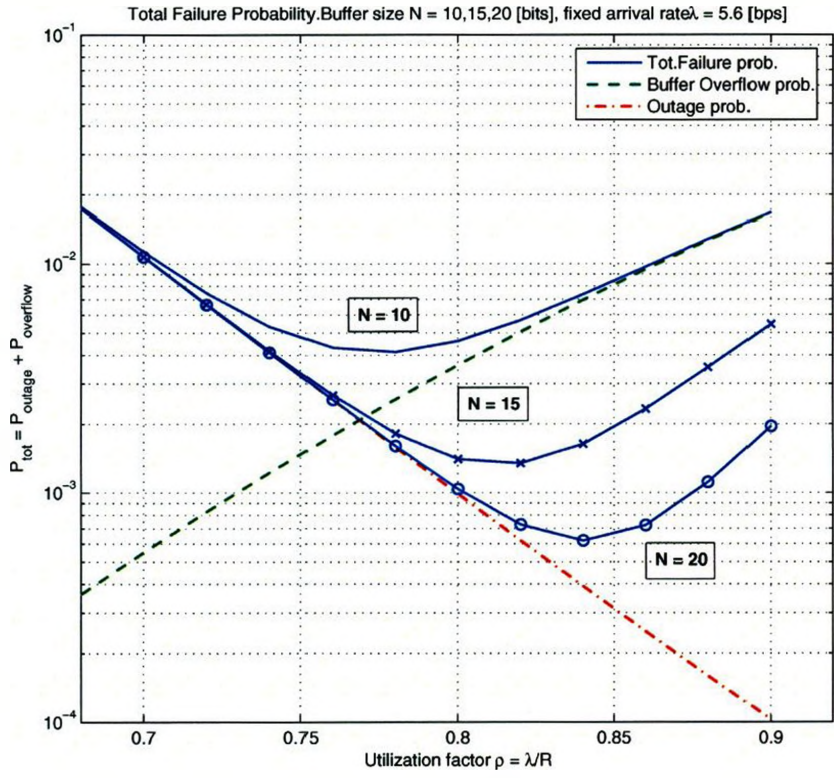

Figure 3.9: Total failure probability for different utilization factors. The arrival rate $\lambda = 5.6$ [bits/s] is fixed while the service rate is decreased monotonically maintaining the utilization factor $\rho = \lambda/r < 1$. The average SNR is 10 dB. For low ρ the outage is the dominant effect due to a high service rate that is not supported by the channel SNR. High ρ leads to high probability of buffer overflow since the service rate r is lower. The figure shows the behaviour of the total failure probability for different buffer size N. Notice the possible tradeoff in terms of ρ.

failure probability would be the outage.

3.3 Conclusion

The random nature of the MIMO channel capacity introduces a probability of outage which translates into a probability of no reliable transmission. This study showed that this outage has a significant effect on the network performance, particularly on the probability of data loss. For high service rates, the outage is the dominant effect, while buffer overflow is the dominant effect for smaller service rates. A good trade-off between outage and buffer overflow probabilities is therefore necessary. Optimal service rates and system buffer sizes have been discussed using numerical examples. According to the queue size and taking into account the MIMO PHY layer, an optimal constant signalling rate can be chosen to minimize the total failure probability of the system. The study of the queuing behaviour with variable transmission rates is the next step of the work.

Chapter 4
Estimation of the Channel State Information

Introduction

The capacity analysis developed in Chapter 2 relies on an accurate channel knowledge. Specifically, MIMO links promise very high data rates with low error probability when the wireless channel response in known at the receiver. The receiver is able to obtain the *channel state information* (CSI) with the use of pilot symbols sent by the transmitter. A precise evaluation of the channel capacity is a fundamental information for the transmitter, as discussed in Chapter 2. It is recognized that the error in the channel estimation can be modeled as a noise component in the system, which leads to an effective SNR lower than the actual one [24] and consequently reduces the channel capacity. Therefore, in this chapter a specific channel estimation technique is described, with particular attention to its estimation error.

In the specific case of time-variant channel due to user mobility, a promising estimation technique is based on the use of a set of functions known as *discrete prolate spheroïdal sequences* (DPSS). The use of DPSS basis set allows an accurate channel estimation even in the case of moderately fast fading channels, with the use of an affordable number of pilots. Generally speaking, an accurate representation of a moderately fast fading channel using bases functions is achievable when both channel and bases bands align. If a mismatch exists, usually a larger number of bases functions is needed to achieve the same accuracy.

In this chapter it is shown that the DPSS based estimation is a powerful technique in case of isotropic scattering environments. The *mean square error* (MSE) of the estimation is deeply discussed and compared with the Fourier basis expansion, to show the significative improvement. However, real world measurements reveal that the scattering angle distribution can deviate significantly from the uniform case.

In those cases a large number of DPSS functions is needed to achieve the desired accuracy, but numerical problems related to the nature of those basis set are unavoidable. A possible solution consists in the definition of a new set of modulated DPSS (MDPSS). Numerical simulations are used to show the better accuracy of this new basis set for directional scattering environments.

4.1 Model of time-variant wireless fading channel

A mobile environment, the Doppler shift in the frequency domain is a finite quantity that is proportional to the mobile velocity. For this reason, the bandlimited Rayleigh fading process, whose *power spectral density* (PSD) is zero past the maximum Doppler frequency, appears in many physical models of mobile radio channels. On one hand, because of its complexity, real data measurements of mobile radio channel are expensive and not easy to obtain; on the other hand the design and optimization of modern communication systems cannot be carried out without computer simulations which take into account the behaviour of the channel. The classical fading simulation application is to generate a single sequence of correlated Rayleigh variates in accordance with Clarke's *wide-sense stationary* (WSS) isotropic model [25].

4.1.1 Correlated fading model

The Gaussian WSS uncorrelated scattering fading model is a complex Gaussian process. The variability of the wireless channel over time is reflected in its *auto-correlation function* (ACF). In the assumption that the radio propagation path consists of a two dimensional isotropic scattering, the theoretical PSD associated with either in-phase or quadrature portion of the received fading signal shows a U shaped bandlimited form given by

$$S(f) = \begin{cases} \frac{1}{\pi f_d \sqrt{1-(\frac{f}{f_d})^2}}, & |f| \leq f_d \\ 0, & elsewhere \end{cases} \tag{4.1}$$

where f_d is the maximum Doppler frequency in Hertz [Hz], derived by the user velocity v and the carrier wavelength λ according to $f_d = v/\lambda$. The corresponding normalized (unit variance) continuous-time autocorrelation of the received signal under these

conditions is $R(\tau) = J_0(2\pi f_d \tau)$, in which $J_0(.)$ is the zeroth-order Bessel function of the first kind [1]. For the purposes of a discrete-time simulation, the autocorrelation sequence becomes

$$R[n] = J_0(2\pi f_m |n|) \tag{4.2}$$

where $f_m = f_d T$ is the normalized Doppler frequency. One method to simulate the channel gain is the sum-of-sinusoids approach also known as Jakes' model. Authors of [26] implemented a careful design to obtain the low-pass discrete fading process. In [26] it is shown that the statistical properties of the fading process approach those of Clarke's isotropic model as the number of sinusoids considered N_s approaches infinity. A good approximation of the ensemble statistics has been reported for $N_s > 8$.

One of the main assumptions of the WSS isotropic model is the uniform distribution of the *angle of arrival* (AoA) of multipath components at the mobile receiver. It has been demonstrated by experimental investigations that the scattering encountered in many environments is non-isotropic, condition that strongly affects the second order statistics and the power spectrum of the channel complex envelope. In directional fading scenarios, real and imaginary Gaussian sequences underlying the Rayleigh channel can exhibit cross-correlations. As a consequence, the PSD in non-isotropic environments can be bandlimited but not symmetric. In [27] it is shown that the received signal correlation and power spectra depend on the *probability density function* (PDF) of the AoA of the scattered wave. The Von Mises PDF is shown to be a versatile and powerful instrument to describe a directional scattering environment. Introduced in 1918 to study the deviations of measured atomic weights from integral values, the Von Mises PDF for the scattered component of χ in direction θ is

$$p_\chi(\theta) = \frac{exp[kcos(\theta - \phi)]}{2\pi I_0(k)} \tag{4.3}$$

where $I_0(.)$ is the zeroth-order modified Bessel function, $\phi \in [-\pi, \pi)$ accounts for the mean direction of AoA of scattered components and $k \geq 0$ controls the width of AoA of the scattered components. $k = 0$ corresponds to a uniform distribution of phase and large values of k correspond to a concentration of angles around a given direction ϕ in a Gaussian fashion with variance $\sigma^2 = 1/k$. The Von Mises PDF, also known as Tikhonov PDF, can be seen as a circular analog of a Gaussian distribution: if the unwrapped phase is distributed normally, its wrapped version is distributed according to the Von Mises PDF [28]. The normalized autocorrelation function $\rho\tau$ of the mobile

channel is associated to the distribution of AoA $p_\chi(\theta)$ as [29]

$$\rho(\tau) = \int_{\pi}^{\pi} exp(j2\pi f_d \tau \cos\theta) p(\theta) d\theta. \tag{4.4}$$

The discrete-time autocorrelation is derived analytically in [27] and proved to be

$$R[n] = R_{II}[n] + jR_{IQ}[n] = \frac{I_0\left(\sqrt{k^2 - z^2 + j2kz\cos(\phi)}\right)}{I_0(k)}; \tag{4.5}$$

where $z = 2\pi f_m |n|$, $R_{II}[n], R_{QQ}[n]$ denote the sampled autocorrelation of the real in-phase and quadrature Gaussian process respectively, and $R_{IQ}[n]$ denotes the cross-correlation function.

Authors of [30] proposed an *autoregressive* (AR) modelling approach for the accurate generation of correlated Rayleigh process, in which the process covariance matrix is obtained by the Cholesky factorization. The exact generation of N Gaussian variables with the desired correlation can be achieved in principle by decomposing the desired $N \times N$ covariance matrix $\mathbf{R} = \mathbf{R}^{\frac{1}{2}}\mathbf{R}^{\frac{1}{2}H}$, where $\mathbf{R}^H$ indicates the Hermitian transpose of $\mathbf{R}$, then multiplying N independent variables by $\mathbf{R}^{\frac{1}{2}}$. Given the desired ACF sequence, the AR filter coefficients can be determined by solving the set of p Yule Walker equations. Those equations can be solved efficiently by the Levinson-Durbin recursion. Since $\mathbf{R}_{xx}$ is an autocorrelation function, it is positive semidefinite and can be shown to be singular only if the process is purely harmonic and consists of $p-1$ or fewer sinusoids. In all the other cases, the inverse $\mathbf{R}_{xx}^{-1}$ exists and the Yule-Walker equations are guaranteed to have a unique solution $\mathbf{a} = \mathbf{R}_{xx}^{-1}\mathbf{v}$. The generated AR(p) process has the ACF

$$\hat{R}_{xx}(k) = \begin{cases} R_{xx}(k), & 0 \le k \le p \\ -\sum_{m=1}^{p} a_m \hat{R}_{xx}[k-m], & k > p \end{cases} \tag{4.6}$$

The simulated process has the attractive property that its sampled ACF perfectly matches the desired sampled ACF up to lag p.

The following figures show the autocorrelation of the fading process obtained by the AR filtering described in [30]. The filter order is $p = 50$, the time autocorrelation is evaluated in case of isotropic scattering ($k = 0$) and directional scattering ($k = 5$).

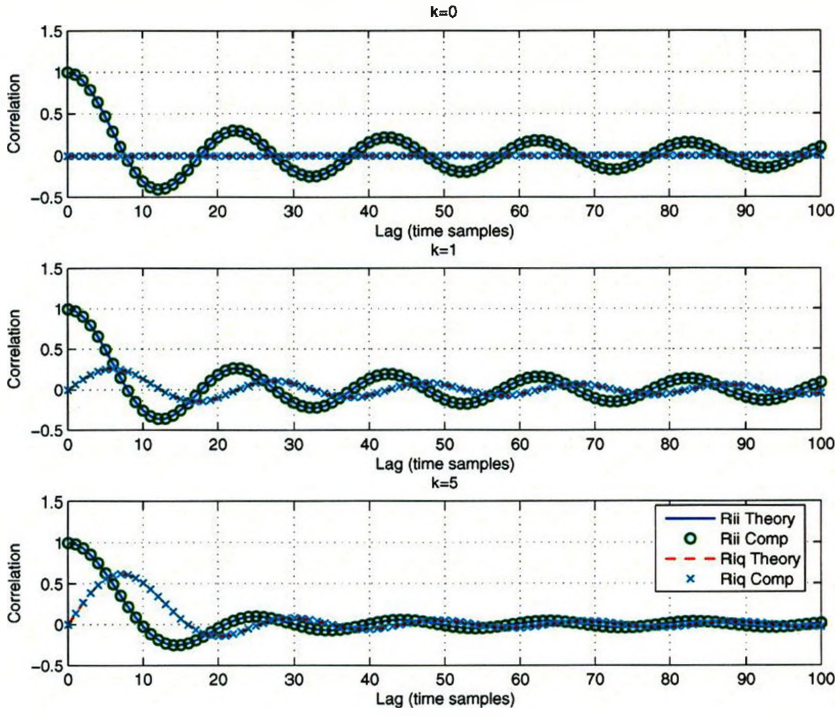

Figure 4.1: Channel in-phase correlation and I/Q cross-correlation in non-isotropic simulation examples for different k

The normalized Doppler frequency is f_m = for a mobile velocity v = 19.4 m/s. The carrier frequency is f_c = 2 GHz. Both those values are chosen for numerical comparison with the results in [30]. Fig. 4.1 shows the process correlation for isotropic and directional scattering. In the latter case the autocorrelation is characterized by a significative imaginary component. The theoretical correlation in Eq. 4.5 is plotted over the numerical results obtained by the AR filtering. Fig. 4.2 shows the channel correlation in isotropic environments. Fig. 4.2 compares the theoretical channel correlation, the AR filtering output from the theoretical correlation in Eq. 4.5 and the AR filtering output obtained by the numerical integral in Eq. 4.4. The results of the AR filtering show a high matching, proving the validity of the analytical closed form in Eq. 4.5. The matching with the theoretical correlation is accurate up to the filter order p, as clear from Eq. 4.6. In case of strong directional scattering, the matching between the real and imaginary components of Eq. 4.5 is shown in Fig. 4.3. A model which generates the correct second-order statistics of the fading channel is the first step to the problem of channel estimation in MIMO systems

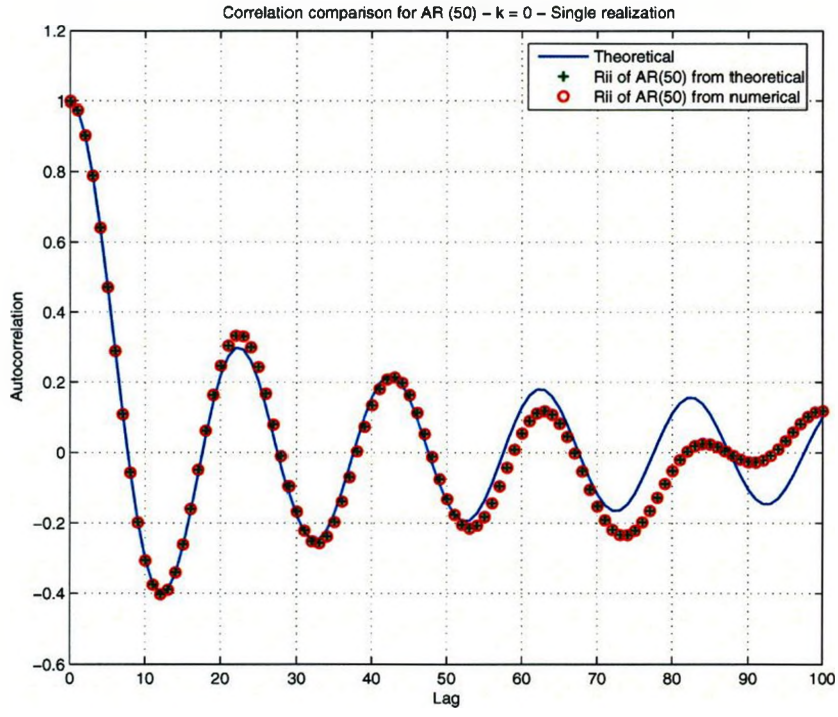

Figure 4.2: In-phase channel correlation R_{II} for isotropic scattering. Comparison among the theoretical Bessel correlation, the AR filtering output correlation obtained from Eq. 4.5 and Eq. 4.4

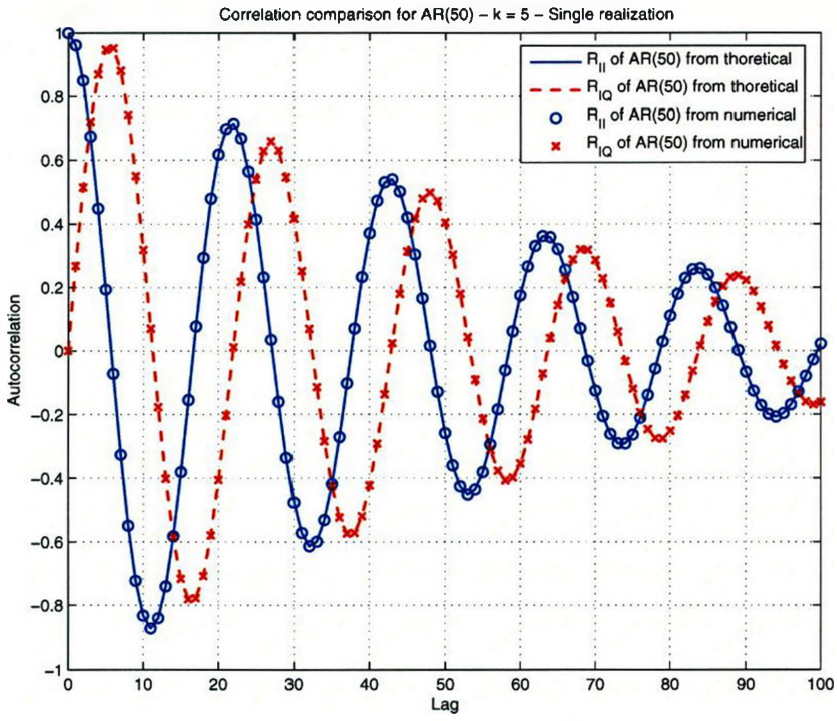

Figure 4.3: Channel in-phase correlation and I/Q cross-correlation resulting from the AR filtering procedure. The matching between numerical and theoretical curves prove the validity of the closed analytical form in Eq. 4.5.

4.2 Time-variant channel estimation with Discrete Prolate Spheroidal Sequences

4.2.1 Basis expansion

The theory of basis expansion states that a general process can be expanded as a sum of deterministic functions, eventually weighted by random coefficients if the interested process is random [31]. The functions used to expand the process are usually orthonormal on a specified domain, the coefficients are chosen to minimize a specific convergence function, i.e. the MSE between the process and its expansion. Given a process $x(t)$, the aim is to describe the process as a sum of specific functions $\{\phi_k\}_{k=1}^{\infty}$ weighted with appropriate coefficients:

$$x(t) = \sum_{k=1}^{\infty} \alpha_k \phi_k(t) \qquad 0 \leq t \leq T \tag{4.7}$$

where T is a specified time interval. The purpose is to understand if the process $x(t)$ can be described by a composition of known functions in a limited time window. α_k are the weights for each single function, defined according to the following relation:

$$\alpha_k = \int_0^T x(t)\phi_k^*(t)dt \tag{4.8}$$

which is a projection of the basis function ϕ_k over the process on the time interval T. In general, the *mean square* (MS) convergence of the sum in Eq. 4.7 is the objective function:

$$lim_{n\to\infty} \quad E\{|x(t) - \sum_k \alpha_k \phi_k|^2\} = 0, \qquad 0 \leq t \leq T. \tag{4.9}$$

Now the question is which set of functions should be chosen, so which basis is good to describe the process. The logical choice of the set ϕ_k leads to *uncorrelated* coefficients α_k, which means:

$$\begin{aligned} E\{\alpha_k\} &= m_k \\ E\{(\alpha_k - m_k)(\alpha_j - m_j)^*\} &= \lambda_k \delta_{kj} \end{aligned}$$

Assuming $m_k = 0$, $E\{|\alpha_k|^2\} = \lambda_k$ is the expected value of energy along $\phi_k(t)$ and $\lambda_k \geq 0$. The hypothesis of uncorrelated coefficients leads to the definition of a particular integral equation:

$$\lambda_k \delta_{kj} = E\{\alpha_k \alpha_j\} = E\{\int_0^T x(t)\phi_k(t)dt \int_0^T x(u)\phi_j(u)du\} = \int_0^T \phi_k(t) \int_0^T K_x(t,u)\phi_j(u)dudt, \tag{4.10}$$

for all the values k, j; then to:

$$\lambda_k \phi_k(t) = \int_0^T K_x(t,u)\phi_j(u)du \qquad 0 \leq t \leq T; \tag{4.11}$$

where $K_x(t,u)$ is the covariance matrix of the process $x(t)$. Eq. 4.11 is called Fredholm equation, λ are the eigenvalues of the covariance matrix K_x and ϕ are the eigenvectors. By definition, K_x is symmetric and non-negative. Eq. 4.11 has K real non-negative values of λ as solutions, obtained by solving an homogeneous linear integral equation. The only assumption about the process $x(t)$ is that its energy is limited, thus the following relation must be satisfied:

$$\int_0^T \int_0^T K_x(t,u)dtdu \leq \int_0^T E\{|x(t)|^2\}dt \leq \infty. \tag{4.12}$$

The integral in Eq. 4.11) has some properties, deeply described in [31]. The sum of the eigenvalues is the expected value of the energy of the process in the interval $(0,T)$. This property guarantee that is possible to find a set of functions ϕ_k that leads to uncorrelated coefficients. Moreover, since $K_x(t,u)$ is non-negative defined, it can be expanded in the series:

$$K_x(t,u) = \sum_{k=1}^{\infty} \lambda_k \phi_k(t)\phi_k(u) \qquad 0 \leq t,u \leq T; \tag{4.13}$$

with this expression, the expectation of the mean square error can be evaluated and the convergence of the expansion for $N \rightarrow \infty$ can be proved [31]:

$$lim_{N\rightarrow\infty} \quad E\{(x(t) - \sum_{k=1}^{N} \alpha_k \phi_k(t))^2\} = 0. \tag{4.14}$$

The procedure described is called Karhunen-Loeve expansion. It provides a second moment characterization of the process $x(t)$ in terms of uncorrelated coefficients. If $x(t)$ is a random process, the coefficients will be random variables and nothing change in the basis set, that will be deterministic in time. For the particular case of Gaussian random processes, the expansion coefficients are statistically independent Gaussian random variables.

4.2.2 Bandlimited processes and prolate spheroidal sequences

A particular case arises when the process under analysis has a limited spectrum in the frequency domain. Considering the simple model where the process has a uniform spectrum $S_x(\omega)$ defined as:

$$S_x(\omega) = \frac{\pi P}{W} \qquad |\omega| \leq W, \qquad 0 \quad elsewhere; \tag{4.15}$$

the corresponding covariance function is the Fourier transform of $S_x(\omega)$:

$$K_x(t,u) = P\frac{sin(W(t-u))}{W(t-u)}. \tag{4.16}$$

and the integral equation of interest becomes:

$$\lambda\phi(t) = \int_{-T/2}^{T/2} P\frac{sin(W(t-u))}{W(t-u)}\phi(u)du. \tag{4.17}$$

In [31], the procedure to solve Eq. 4.17 is to find the related differential equation and to examine its solution. The associated differential equation in the normalized

interval $-1 \leq t \leq 1$ is :

$$(1-t^2)\ddot{\phi}(t) - 2t\dot{\phi}(t) + (\lambda - c^2t^2)\phi(t) = 0 \tag{4.18}$$

where λ is the eigenvalue and c is called bandwidth parameter equal to $c = \frac{WT}{2} = \pi f_{max} T$, with f_{max} the maximum frequency of the spectrum in Hz. This equation has continuous solutions called angular prolate spheroidal functions, obtained for a discrete set of values $\lambda(c)$. These functions are tabulated in [32]. This set of function is characterized by a very compact set of eigenvalues: for values of k larger than $(2f_{max}T + 1)$ the values of λ_k rapidly approach zero. When a bandlimited process is observed for T seconds, there are only $(2f_{max}T + 1)$ significant eigenvalues, that means that the energy of the process is concentrated in this little set of coefficients. Moreover, this means that a good description of the process can be achieved with a little number of coefficients, which can be determined following the procedure exposed in the previous section. In conclusion, the *prolate spheroidal sequences* (PSS), also known as Slepian functions, are an appropriate basis set for the expansion of bandlimited processes. From this property arises the interest in using this functions for the estimation and prediction of a time-varying channel, whose spectrum is limited by the Doppler frequency value.

4.2.3 Discrete Prolate Spheroidal Sequences

The *discrete prolate spheroidal sequences* (DPSS) represent the discrete version of the PSS described in the previous section. Throughout this section and the rest of the chapter, it is assumed that only N discrete samples of the channel are available and that they were obtained with a sampling period T. Hence , the discrete frequency ν represents the continuous frequency f normalized by the sampling time: $\nu = fT$. Given N the k^{th} DPSS $v_k(n, N, W)$ for $k = 0, 1, ..., N-1$ is defined as the real solution of the system of equations [17]

$$\sum_{m=0}^{N-1} \frac{sin(2\pi W(n-m))}{\pi(n-m)} \cdot v_k(m, N, W) = \lambda_k(N, W) \cdot v_k(n, N, W) \tag{4.19}$$

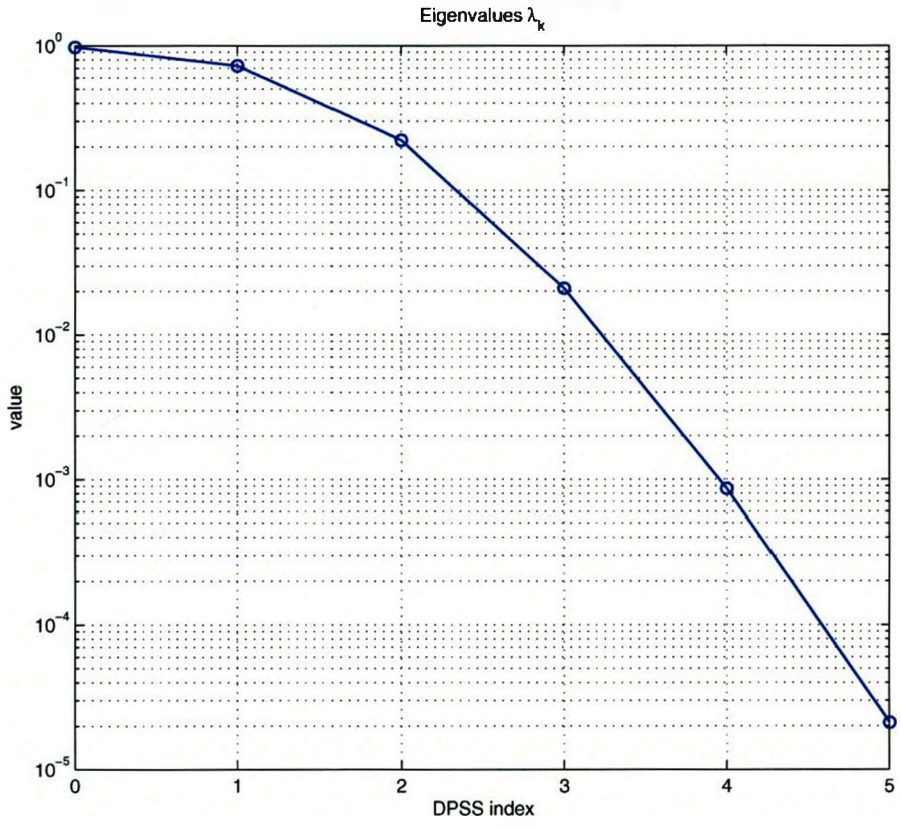

Figure 4.4: Eigenvalues for Discrete Prolate Functions. The prolate sequences are designed for a block length $N = 256$ and a normalized Doppler frequency of 3.8×10^{-3} Hz; the bandwidth parameter is $D = 2NW = 2$.

where $\lambda_k(N, W)$ are the ordered non-zero eigenvalues of Eq. 4.19

$$\lambda_0(N, W) > \lambda_1(N, W), ..., \lambda_{N-1}(N, W) > 0 \tag{4.20}$$

The first $2NW$ eigenvalues are very close to 1, while the rest rapidly decay to zero [17]. Fig, 4.4 shows the eigenvalues λ_k for the prolate sequences. The largest eigenvalues are concentrated in the first few sequences, meaning that the process energy is well described by the first few basis functions. The DPSS are doubly orthogonal: they are orthogonal on the infinite set $\{-\infty, ..., \infty\}$ and orthonormal on the finite set

$\{0, 1, ..., N-1\}$:

$$\sum_{-\infty}^{\infty} v_i(n, N, W) v_j(n, N, W) = \lambda_i \delta_{ij} \tag{4.21}$$

$$\sum_{n=0}^{N-1} v_i(n, N, W) v_j(n, N, W) = \delta_{ij} \tag{4.22}$$

where $i, j = 0, 1, ..., N-1$. This property enables parameter estimation that performs better than other methods. In the following simulations the prolate basis expansion is compared with the Fourier expansion to prove the previous statement.

4.2.4 Signal model for flat-fading time-variant channels

Following the same notation of [33], the model under analysis is the transmission of data symbols $d[n]$, where the index $[n]$ denotes the discrete time. The symbol rate is $R_s = 1/T_s$ and symbols are transmitted over a flat fading time-variant channel. The flat fading assumption is fulfilled for a symbol duration T_s much longer than the delay spread of the channel T_D; looking at the frequency domain this means that the channel bandwidth is larger than the symbol one, leading to the case of a non selective channel. With this assumptions, the received sequence $y[n]$ is obtained by multiplying the symbol sequence with the sampled time-variant channel $h[n] \triangleq h(nT_s, 0)$, with the addiction of complex white Gaussian noise $z[n] \backsim N(0, \sigma_z^2)$.

$$y[n] = h[n]y[n] + z[n], \qquad n \quad \epsilon\{0, ..., N-1\}. \tag{4.23}$$

The transmitted sequence is a block of N symbols. In each data block, J pilot symbols are sent in order to allow the channel estimation at the receiver side. Each block consists of $N - J$ data symbols $b[n]$ with J interleaved pilot symbols $p[n]$. Data symbols are chosen from a $QPSK$ alphabet, so they satisfy the following relation: $b[n] \in \{\pm 1, \pm j\}/\sqrt{2}$. The pilot positions are uniformly distributed in the block according to the following relation, which provides a regular pilot placement as function of the number of pilots itself:

$$P = \{\lfloor i\frac{N}{J} + \frac{N}{2J} \rfloor | 0 \leq i \leq J-1\}. \tag{4.24}$$

The pilot sequence is made by a sequence of orthogonal complex symbols, with normalized energy $E_s = 1$. Author of [34] suggest an optimal training sequence for the particular case of *MIMO* systems, building the pilot sequences from a *discrete Fourier Transform* (DFT) matrix. The pilot sequences used for the simulations are built according to this technique.

4.2.5 Time-variant flat fading channel estimation

According to figure 4.4, the eigenvalues λ_k are clustered near 1 for $k \leq 2NW$ and rapidly drop to zero for $k > 2NW$. Considering $D' = 2NW + 1$ the minimum signal dimension to have a good description of the process, the time variant channel can be described in terms of prolate functions as:

$$h[n] \approx \tilde{h}^{(s)} = \sum_{k=0}^{D^{(s)}-1} v_k^{(s)}[n]\gamma_k^{(s)}, \tag{4.25}$$

where $n \in \{0, ..., N-1\}$. The dimension of the basis expansion fulfills:

$$D' \leq D^{(s)} \leq N. \tag{4.26}$$

According to [33], by choosing $D^{(s)}$ it is possible to control the mean square error of the basis expansion defined as:

$$MSE_N = \frac{1}{N} \sum_{n=0}^{N-1} E\{|h[n] - \tilde{h}[n]|\}. \tag{4.27}$$

The channel estimation is based on the use of pilot symbols. The pilot pattern p allows to obtain the channel knowledge at the discrete time $n \in P$. The estimation is performed defining the basis vector

$$\mathbf{f}[n] = [v_0[n], ..., v_{D-1}[n]]^T \in \mathbb{R}^D, \tag{4.28}$$

which contains the instantaneous values of the basis functions and the autocorrelation matrix of the basis function

$$G = \sum_{l \in P} \mathbf{f}[l]\mathbf{f}[l]^H. \tag{4.29}$$

The vector of coefficients γ to weight the basis set in Eq. 4.25 can be obtained by the least square estimation, knowing the pilot pattern $p[n]$ and the received signal $y[n]$:

$$\hat{\gamma} = (\sum_{n \in P} \mathbf{f}[n]\mathbf{f}^H[n]|p[n]|^2)^{-1} \cdot \sum_{n \in P} y[n]p^*[n]\mathbf{f}^*[n] \tag{4.30}$$

where $\gamma = \{\hat{\gamma}_0, ..., \hat{\gamma}_{D-1}\}^T$. The pilot sequence is made by a set of DFT coefficients with $|p[n]| = \sqrt{E_s} = 1$, according to [34]. For a complete explanation of the least square estimation that leads to Eq. 4.30 see App. B.

The following numerical simulations show the performance of the prolate basis estimation, with particular attention on the variance of the basis estimation. The parameters used for the simulations are the following. The carrier frequency is $f_c = 2$ GHz, the symbol duration is $T_s = 20.5710^{-6}$ s, the speed of the user is $v_{max} = 27.8$ m/s corresponding to a maximum normalized Doppler frequency of $W = 0.0038$, the data block length is $N = 256$ symbols. With these system parameters, the approximate dimension of the signal space is $D' = 2NW + 1 = 3$. The following figure shows the comparison between the prolate basis expansion and the Fourier expansion. The Fourier basis is built according to [33]:

$$u_k^{(F)}[n] = \frac{1}{\sqrt{N}} e^{(j2\pi(i-(D-1)/2)n/N)}, \quad for\ i = 1, ..., D \tag{4.31}$$

The dimension D' is the minimum number of basis functions required to approximate the channel. Referring to figure 4.4, $D' = 3$ are the most high eigenvalues of the channel autocorrelation matrix. In [33] the authors propose the use of the minimum dimension of the signal space D'. Although, the energy not covered by the first D' bases is considerable, equal to the sum of all the eigenvalues from λ_4 to λ_∞. For this reason, in order to obtain a good estimation of the process, the basis dimension should be slightly larger than D'.

4.2.6 DPSS basis MSE analysis

By definition in [31, 33], the *mean square error* (MSE) is the composition of the square bias of the estimator and its variance:

$$MSE_N = bias_N^2 + var_N \tag{4.32}$$

where N is the time length of the process. Authors of [33] prove that DPSS perform better than the Fourier basis set. The Fourier expansion made with a truncated discrete Fourier transform (DFT) is not appropriate to describe a bandlimited process due to its infinite support. Truncating the spectrum created with the Fourier functions introduces spectral leakage which leads to significant phase and amplitude errors in the estimation. As shown in the following figures, the DPSS estimation offers better performances with the same grade of complexity of the Fourier expansion.

Following [33, 11], the squared bias can be described analytically as function of the basis error characteristic and the power spectral density of the channel $h[n]$:

$$bias^2[n] = \int_{-1/2}^{1/2} E[n, W] S_{hh}(W) dW, \tag{4.33}$$

defining the block squared bias as:

$$bias_N^2 = \frac{1}{N} \sum_{n=0}^{N-1} bias^2[n] \tag{4.34}$$

Fig. 4.5 shows the squared bias of the DPSS basis and the Fourier basis, as function of the discrete time n. The prolate basis offers a significative increase in the estimation accuracy compared to the Fourier basis set. The graph shows clearly that the $bias_N^2$ is higher at the beginning and at the end of the frame, where the basis functions try to extrapolate the channel behavior only from one pilot available. Both the DPSS and Fourier basis work with the same complexity, equal to a the problem dimensions $D = 5$.

The analytic expression for the estimator variance is [11]:

$$var_N \approx \sigma_z^2 \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{f}^H[n] \mathbf{G}^{-1} \mathbf{f}[n] \approx \sigma_z^2 \frac{N-J}{J}, \tag{4.35}$$

where $(N-J)/J$ is the ratio between the number of data symbols and pilot symbols. Thus, increasing the number of pilots the variance of the estimation should decrease, as the noise effect is limited to a shorter interval between two consecutive pilots. Moreover, a strong dependance on the noise power σ_z^2 is expected. Fig. 4.6 shows the basis MSE dependance on the SNR, for the DPSS and Fourier expansion. The

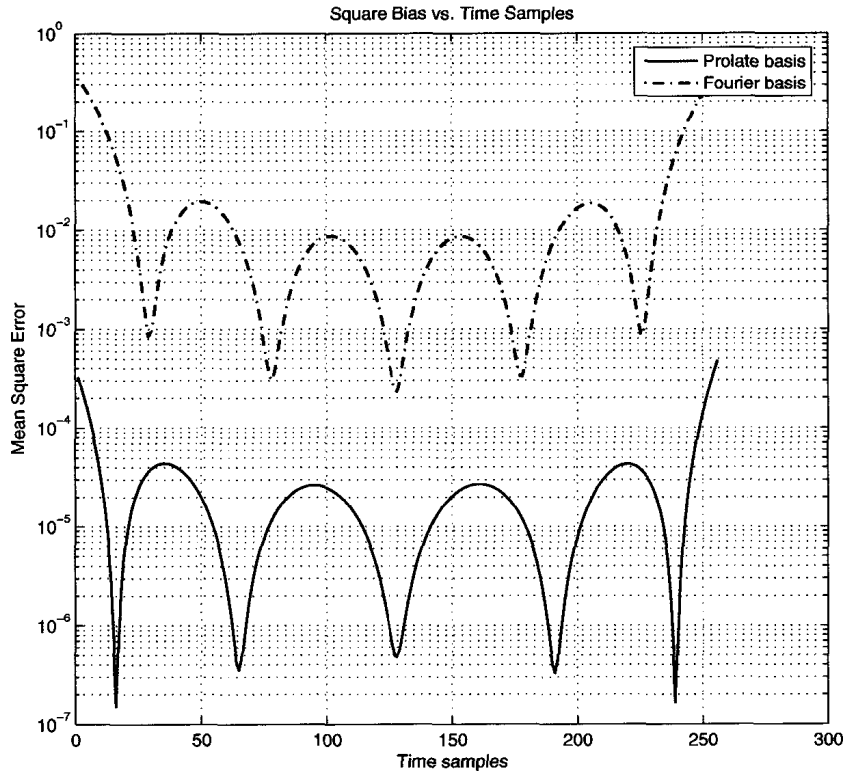

Figure 4.5: $Bias_M^2$ for the Fourier and Prolate expansions, both with dimension $D = 5$ and number of pilots $J = 10$ in a block of $N = 256$ symbols. In the whole block duration, the prolate $Bias_N^2$ is more than two magnitudes lower than the Fourier Bias

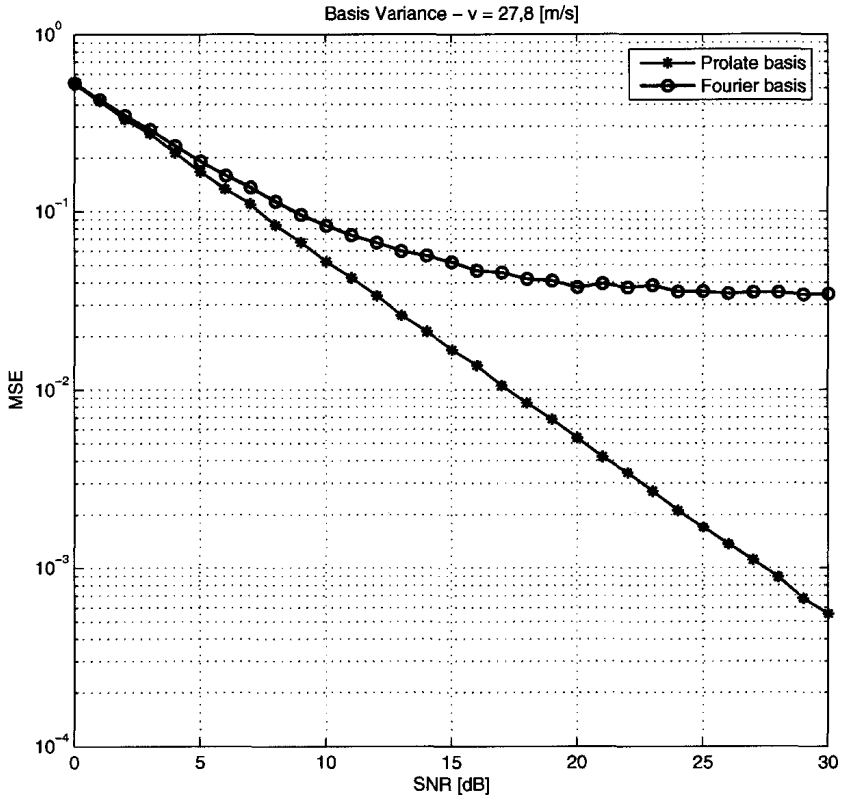

Figure 4.6: MSE_N of the Prolate and Fourier basis for different SNR,number of pilots $J = 10$ and normalized Doppler $W = 0.0038$. The $Bias_N^2$ of the Fourier basis affects significantly the estimation performances and it is responsible of the saturation of MSE_N at high SNR values. The prolate estimation is unbiased in the interest range

Fourier expansion is biased since at high SNR the MSE_N saturates at a specific value. On the other hand the prolate basis is practically unbiased and the MSE_N decreases uniformly increasing the SNR. The behavior of the MSE_N for the Fourier basis expansion can be explained according to Eq. 4.32: with high SNR the variance injected by the noise is very little and the $bias_N^2$ of the estimator becomes predominant as it does not depend on the SNR. The DPSS basis revealed to be an appropiate basis estimation set for wireless mobile flat fading channels. The performance analysis has been carried out in the case of isotropic scattering, which leads to a baseband limited channel spectrum.

4.3 Modulated Discrete Prolate Spheroidal Sequences

The channel spectrum presented in Sec. 4.1 resembles a uniform spectrum for most of the frequencies, therefore its Karhunen-Loeve basis expansion [31] is close to the one defined by a sinc type covariance function. For this reason, the representation in terms of *discrete prolate spheroidal sequences* (DPSS) suggested in [33] is shown to be very accurate for isotropic environments. When the *angle of arrival* (AoA) distribution deviates significantly from the uniform case in $[-\pi\ \pi]$, the received signal is well represented by a sum of sinusoids arriving from a narrow band of angles, corresponding to individual clusters [35]. If the DPSS are used for channel estimation, then usually accurate and sparse representations are obtained when both the DPSS and channel under investigation occupy the same frequency band [33]. Some problems may arise when the channel band is centered around some frequency $|\nu_0| > 0$ and the occupied bandwidth is smaller than $2W$, as depicted in Fig. 4.7.

In such situations, a larger number of DPSS is required to approximate the channel with the same accuracy, despite the fact that such narrowband channel is more predictable than a wider band channel. In order to find a suitable basis, the *modulated discrete prolate spheroidal sequences* (MDPSS) are defined as:

$$M_k(n, N, W; \omega_m) = exp(j\omega_m n) v_k(n, N, M); \tag{4.36}$$

where $\omega_m = 2\pi\nu_m$ is the modulating frequency. It is easy to show that MDPSS are bandlimited to the frequency band $[-W + \nu_m : W + \nu_m]$, obey Eq. 4.19 and

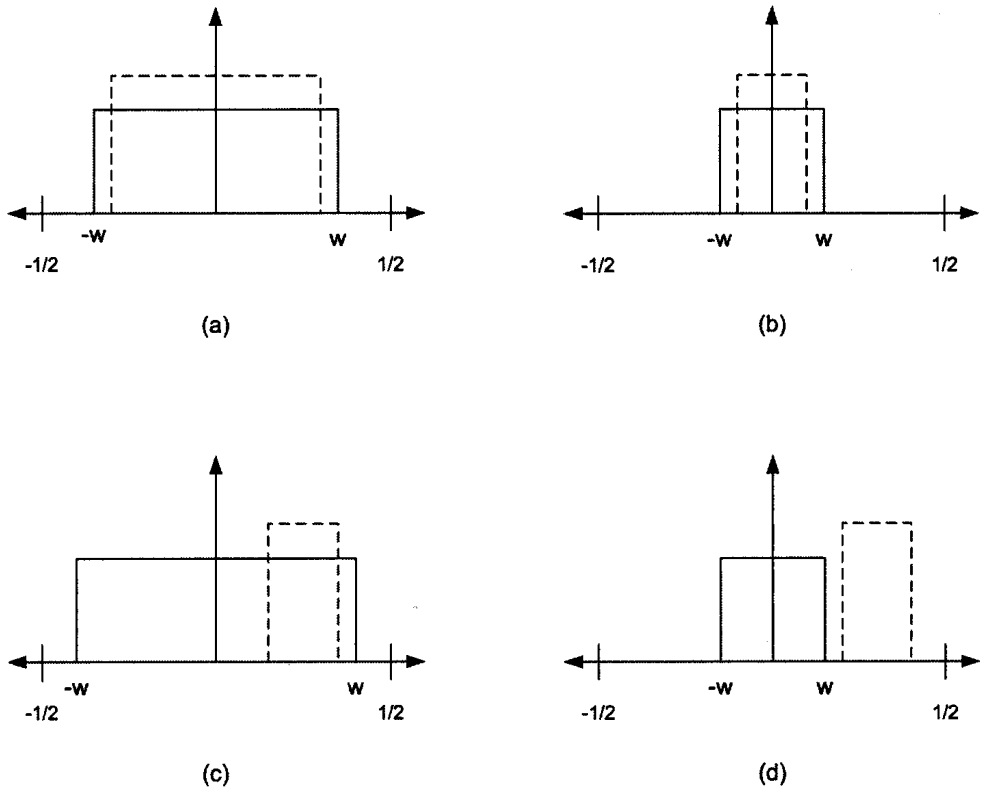

Figure 4.7: Comparison of the bandwidth for a DPSS (solid line) and a channel (dashed line): (a) both have the same bandwidth; (b) both have narrow bandwidth; (c) DPSS has a wide bandwidth, while the channel bandwidth is narrow and centered around $\nu_0 > 0$; (d) both have narrow bandwidth, but centered at different frequencies.

are doubly orthogonal. The next question is how to properly choose the modulating frequency ν_m. In the simplest case in which the spectrum of channel $S(\nu)$ is confined in a known band $[\nu_1; \nu_2]$, i.e.

$$S(\nu) = \begin{cases} \gg 0, & \forall \nu \in [\nu_1; \nu_2] and |\nu_1| < |\nu_2| \\ \approx 0, & elsewhere \end{cases} \tag{4.37}$$

the modulating frequency ν_m and the bandwidth of the DPSS basis is defined by

$$\nu_m = \frac{\nu_1 + \nu_2}{2} \tag{4.38}$$

$$W = \left| \frac{\nu_2 - \nu_1}{2} \right| \tag{4.39}$$

as long as both satisfy

$$|\nu_m| + W < \frac{1}{2}. \tag{4.40}$$

In practical applications, the exact frequency band is known only with a certain degree of accuracy. Especially in mobile communications, the channel time evolution suggests that only a relative band defined by the mobile velocity and carrier frequency is known. To improve the estimation robustness, the concept of frames has been introduced in [36].

The performance of the MDPSS estimator is compared with the DPSS basis expansion in the next figures. The mobile channel has been modeled following the AR approach in Sec. 4.1, with correlation properties described by Eq. 4.5. The parameters of the simulated system are the same as in Sec. 4.2. The carrier frequency is $f_c = 2$ GHz, the symbol duration is $T_s = 20.5710^{-6}$ s, the speed of the user is $v_{max} = 27.8$ m/s corresponding to a maximum normalized Doppler frequency of $W = 0.0038$, the data block length is $N = 256$ symbols and the number of pilots is $J = 10$. The number of DPSS used in estimation is given by $D' = \lceil 2NW \rceil + 1$. Fig. 4.8 shows the MSE of the estimation in a noise-free case, i.e. the $bias_N^2$ component of Eq. 4.32. The angle profile of the considered channel has a central AoA $\phi_0 = 5$ degrees and spread $\Delta = 20$ degrees. Fig. 4.8 shows the average over 1000 realizations. The MSE for both MDPSS and DPSS estimators have the highest values at the edge of the data block. However, the MSE for MDPSS estimator is several orders of magnitude lower than the value of the DPSS basis. Fig. 4.9 shows the MSE of the estimation for different angle spread. The SNR is 20 dB and the central AoA is 45 degrees. The angle

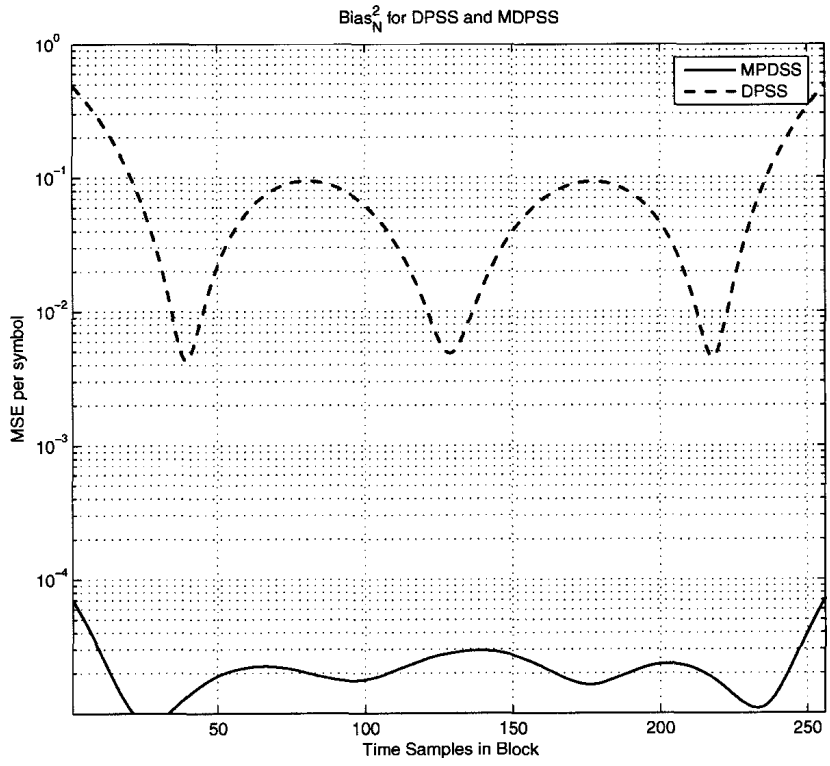

Figure 4.8: $bias_n^2$ per symbol for MDPSS (solid) and DPSS (dashed) mobile channel estimator for the noise-free case.

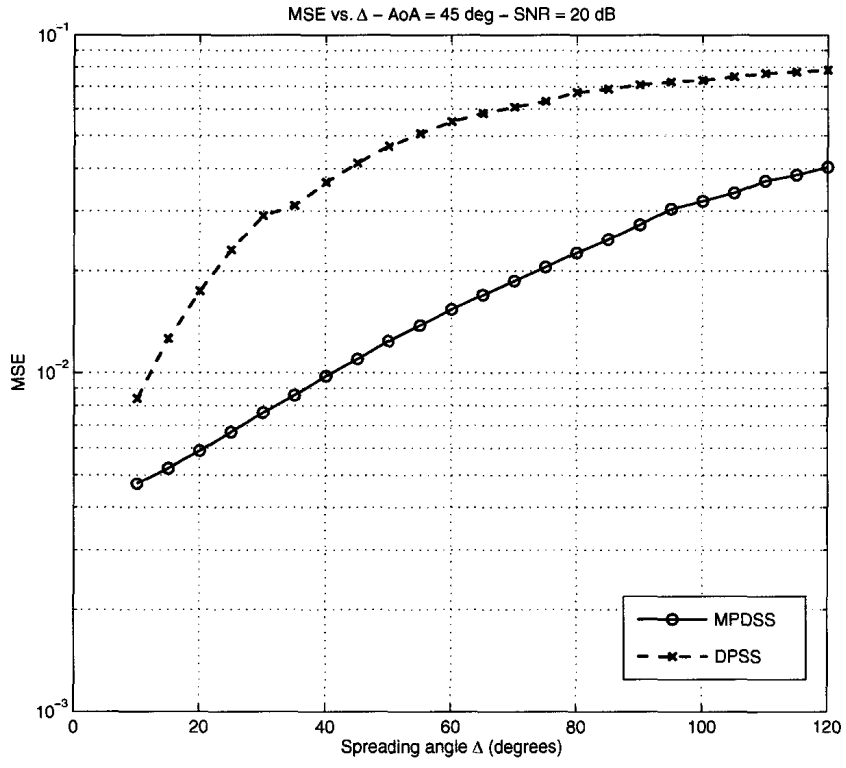

Figure 4.9: Dependence of the MSE on the angular spread Δ, AoA 45 degrees and SNR 20 dB. The MDPSS show a significant increase in the estimation accuracy using the same number of functions of the DPSS method.

spread Δ describes the width of the incoming scattering component, for high spread the scattering gets closer to the isotropic case. The MSE of the modulated basis set MDPSS is significantly lower, especially for limited spreading angles ($20 < \Delta < 80$). The difference between the two basis sets decreases as the scattering becomes similar to the isotropic case. For very small Δ, the scattering can be assumed to be composed by few sinusoidal components and the performance of the two basis sets in that case is almost equivalent.

In conclusion, the MDPSS show an higher estimation accuracy in the realistic case of directional scattering environments, with the same numerical complexity (same number of functions for the channel expansion). The complexity of this solution lies in the need of estimating the mobile velocity to perform an efficient modulation. However, this task does not seem to be prohibitive at least in the case of average vehicular speeds.

4.4 Conclusion

The characteristics of a flat fading time variant wireless channel can be well described by an AR model once the process autocorrelation is available. Simulation results showed that the fading process obtained by the AR filtering is very accurate, with a high matching between the desired and generated second order statistic of the process.

The bandlimited characteristic of the fading spectrum suggests that it can be well expanded by a set of deterministic functions with a similar limited bandwidth. Using a set of pilot symbols spread in the data sequence, the DPSS basis expansion allows to model a time-variant channel for the duration of the data block. The DPSS expansion offers major performance gains compared to other basis expansion for the same numerical complexity (number of unknowns D). The analysis of the MSE of the prolate basis expansion shows that this particular expansion is not biased: the $bias^2$ is several magnitudes less than the Fourier basis one in the whole data block duration.

This optimal accuracy is lost when the scattering environment is not isotropic. A directional scattering leads to a mismatch between the channel and DPSS bandwidth. The MDPSS basis set is proposed for estimation of fast fading channels to preserve the sparsity of the representation and enhance the estimation accuracy. The members of the basis set are obtained by modulation and bandwidth variation of the

original DPSS functions in order to reflect various scattering scenarios. The results obtained by numerical simulations showed that the MDPSS method provides more accurate estimation than the DPSS scheme.

Chapter 5
Variable Rate Transmission

Introduction

Chapter 4 showed that an accurate *channel state information* (CSI) can be obtained at the receiver through the use of pilot symbols. Moreover, if the CSI is available not only at the receiver but also at the transmitter, a high multiplexing gain can be achieved in multi-antennas transmissions [5], increasing the spectral efficiency. Specifically, with the knowledge of the CSI, the theoretical limit denoted by the channel ergodic capacity can be achieved in transmission by adapting the signalling rate to the capacity evolution. As stated in Chapter 3, the signalling rate is defined as a percentage of the channel capacity. The knowledge of the channel state at the transmitter can be obtained by a feedback channel [37], as depicted in Fig. 5.1. The channel capacity seems to be the more appropriate feedback: feeding the predicted channel capacity to the transmitter would allow to choose the right modulation, coding and power to suit the wireless channel, with a reasonable amount of feedback information. In this scenario it is assumed that delay in the feedback loop is negligible. However, this approximation is not always realistic: if the delay loop was considerable, it would become necessary to study the changes in the channel capacity during the feedback loop interval, so that the transmitter can take them into account [1].

A possible solution is to predict the capacity variation in a specific time interval through a prediction model. Particularly, the investigation on the MIMO capacity correlation in [12] allows to describe its time evolution through the *autoregressive* model (AR) described in Chapter 2. The capacity evolution can be predicted frame by frame and the actual CSI can be made available at the transmitter on a larger time scale, reducing the complexity of the feedback system.

---

[1]This investigation goes beyond the purposes of this work and will be left for a future work

The AR-1 model discussed in Chapter 2 assigns a Markov nature to the instantaneous capacity process. Dividing the continuous capacity process into a finite number of discrete states, a *finite state Markov chain* model (FSMC) for the instantaneous MIMO capacity is obtained. The principle of FSMC is to discretize a continuous process into a finite number of states, over which the process itself can be qualified separately. The FSMC is a well accepted block fading channel model for slow-varying flat fading channels, where the channel is assumed to stay in the same state within one block period. In wireless communication, the markovian assumption is widely used to model the channel fading and the SNR variation [7, 38, 39]. The FSMC model and *adaptive modulation and coding* (AMC) strategies are commonly accepted as the fundamental techniques for developing effective cross-layer protocols and algorithms.

Assuming that the instantaneous channel capacity distribution is Gaussian [18], in this chapter the FSMC model is derived analytically for the MIMO channel capacity. The purpose of this study is to investigate the optimal approach to the signalling rate control problem for a MIMO wireless channel, improving the performance of the constant rate transmission described in Chapter 3. Considering the whole transmission system, higher layer data packets are enqueued into a finite size buffer space before being released onto the time-variant wireless channel. The choice of the optimal signalling rate must take into account the buffer state: over the PHY layer, the buffer at the MAC layer is itself a FSMC dependent on the arrival process. The joint consideration of the capacity state and buffer state leads to a two-dimensional optimization process, where the optimal signalling rate should be chosen according to the state of the instantaneous capacity and how many packets are present in the buffer. The choice of the signalling rate at every time frame must maximize the performance at each state: the probability of successful transmission (throughput) at that specific signalling rate. With the outage definition provided in Chapter 3, the objective of this analysis is to minimize the average joint packet loss rate due to both outage and buffer overflow. In this case there are two design objectives (outage and buffer overflow) that jointly define the optimization target: minimizing the end-to-end packet loss. The optimal transmission policy design is investigated analytically with the application of *dynamic programming* (DP) and *Markov decision process* (MDP) theories.

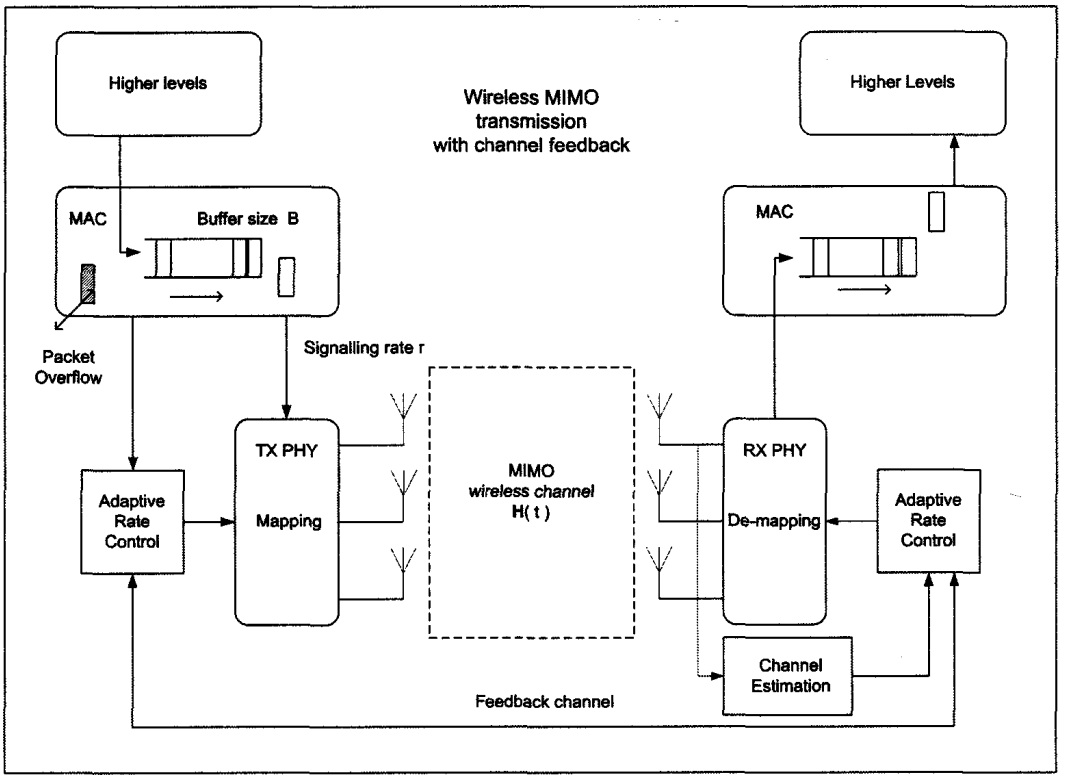

Figure 5.1: Communication system scheme with feedback channel

5.1 Contributions

This study addresses the cross-layer optimization issues by jointly considering adaptive transmission scheduling and wireless channel capacity evolution to achieve better performance gain. The major contributions of this work are summarized as follows:

1. this chapter proposes an analytical definition of FSMC for the MIMO channel capacity;

2. this chapter develops an analytical model of the proposed algorithm for evaluating *Quality of Service* (QoS) performance metrics at the MAC layer, such as the system throughput and probability of packet loss;

3. combined with previous research, here is proposed an efficient method for cross-layer performance optimization.

5.2 System Model

Fig. 5.1 illustrates the general structure of the system model used for this study. At the transmitter side, packets arriving from upper layers are enqueued into a finite MAC buffer space. Backlogged packets are de-queued, on first-come-first-serve basis, and further processed by the digital modulator before being transmitted. Undergoing distortion and attenuation by the wireless fading channel, received symbols are sequentially passed through demodulator and decoder, with uncorrectable codewords being dropped, to retrieve MAC layer packets. At the PHY layer, AMC is applied to achieve adaptive multirate transmission over the time-variant MIMO channel.

The physical layer is characterized by a MIMO wireless channel subject to fading and noise: the instantaneous SNR is a random process described by the Nakagami's law [1]. The instantaneous channel capacity can be well approximated by a Gaussian process [5, 18]. Recalling the analysis presented in Chapter 3, Sec. 2.1, the channel capacity can be interpreted as the maximum amount of information that can be reliably transmitted on the MIMO channel, disregarding the code and modulation scheme at the PHY layer of the system. Considering delay tolerant traffic, the instantaneous capacity represents the limit of information that can be transmitted by the system. The channel spectral efficiency (bits/s/Hz) is related to the signalling rate (packets/s) by the bandwidth required by the system [1] and the packet size. The system works with time frames of fixed duration T_f. Data packets of size N_b bits are collected in a finite buffer of B packets.

5.3 Theory of Discrete Finite Markov Chain

The two most important parameters which characterize a finite state Markov chain are the transition probability matrix P and the state probability vector $\boldsymbol{\pi}$. In this section the theoretical definition of those two parameters is presented. Transition and state probabilities are evaluated in Sec. 5.6 to model the MIMO channel capacity as a finite state Markov chain .

For the purpose of this study, our attention is focused on discrete time processes $x(n)$ with the time index $n = 0, 1, ...$ If the process x can assume a finite set of values (or states) $x = s_i$ with $0 \leq i \leq K-1$, it is said to be Markov chain if, whenever it is

in state i, there is a fixed probability P_{ij} that it will next be in state j, regardless of the process history before the state i:

$$\begin{aligned} p_{ij} &= Pr\{x(n) = j | x(n-1) = i, x(n-2) = i(n-2), ..., x(0) = i(0)\} \\ &= Pr\{x(n) = j | x(n-1) = i\}. \end{aligned} \tag{5.1}$$

The values p_{ij} for $0 \leq i, j \leq K-1$ are called transition probabilities between the two states i, j

A finite Markov chain is a Markov process such that the transition probabilities $p_{ij}(n)$ do not depend on the time index n [40]. In network theory, this property leads to the definition of *memoryless* model, in which the current behaviour does not depend on the history of the process and the current time index. A probability distribution $\{p_j | j \geq 0\}$ is said to be a *stationary distribution* for the Markov chain if

$$p_j = \sum_{i=0}^{K-1} p_i p_{i,j}, \quad j = 0, 1, ..., K-1 \tag{5.2}$$

which, in a continuous domain is equal to the marginal distribution

$$p(j) = \int p(i) p(i, j) di. \tag{5.3}$$

For an irreducible and aperiodic Markov chain can be shown that

$$p_j = \lim_{n \to \infty} Pr\{x(n) = j | x(0) = i\}, \quad i = 0, 1, ..., K-1. \tag{5.4}$$

When it exists, the stationary distribution is unique according to the following Theorem [41]:

> in an irreducible, aperiodic Markov chain there are two possibilities for the scalars $p_j = \lim_{n \to \infty} Pr\{x(n) = j | x(0) = i\}$
>
> - $p_j = 0$ for all $j \geq 0$ in which case the chain has no stationary distribution.
> - $p_j > 0$ for all $j \geq 0$ in which case $\{p_j | j \geq 0\}$ is the unique stationary distribution of the chain.

The state probabilities p_j for $j = 0, 1, ..., K-1$ are arranged in the row vector $\boldsymbol{\pi}$. All the transitions among the finite states can be arranged into the $K \times K$ transition

matrix $\mathbf{P}$, with its elements $p_{i,j}$ defined as in Eq. 5.1. Referring to the independence of the transition probabilities from the time index n, the n-steps transition matrix is proved to be the n^{th} power of the single step transition matrix $\mathbf{P}$ [40]. A state transition probability matrix has the property that the sum of the elements on each row is equal to 1:

$$\sum_{j=0}^{K-1} p_{i,j} = 1, \forall i \in \{0, 1, ..., K-1\}. \tag{5.5}$$

Eq. 5.2 can be extended to all the states in the vector equation

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}, \quad \textstyle\sum_{i \in S} \pi_i = 1. \tag{5.6}$$

Eq. 5.6 shows that $\boldsymbol{\pi}$ is left eigenvector of the matrix of transition probabilities $\mathbf{P}$ corresponding to the eigen-value 1. This condition will be used later on to confirm the transition probability matrix obtained by different simulation models.

5.4 Theory of Markov Decision Processes

In the single buffer communication system under analysis, a reasonable strategy for variable rate transmission would adapt the signalling rate to the state of the capacity process and the buffer occupancy. The optimal signalling rate would minimize the failure in transmission due both to the buffer overflow and outage events, as shown by the analysis presented in Chapter 3. The rate adaptation is a sequential process, in which the signalling rate is optimized frame by frame according to a set of reference metrics. In case of Markov processes, a way to solve this optimization problem is found in the theory of dynamic programming and Markov decision processes [41]. The purpose of this section is to explain the main steps to solve the optimization process presented.

5.4.1 Expected Total Earning function

Given a finite Markov chain x, the knowledge of the transition probabilities, state probabilities and rewards is assumed. The reward can be any metric of interest (money, profit, space) and must be associated with each state and transition. The idea beyond the optimization procedure is that the total expected reward must be

maximized after a finite number of iterations or in the perspective of an indefinite working horizon. The *expected total earning function* $v_i(n)$ is defined as the expected total earning starting from the state i in the next n steps in time. Knowing the transition probability $p_{i,j}$ and the transition reward $r_{i,j}$ associated to the transition from state i to state j, the expected total earning can be written in the iterative form [41]:

$$v_i(n) = \sum_{j=0}^{K-1} p_{i,j}[r_{i,j} + v_j(n-1)], \quad i = 0, ..., K-1, \quad n = 1, 2, ... \tag{5.7}$$

The quantity q_i

$$q_i = \sum_{j=0}^{K-1} p_{i,j} r_{i,j} \tag{5.8}$$

is called the expected immediate reward for the state i (or one-step reward). Therefore, the total expected reward in the next n steps is:

$$v_i(n) = q_i + \sum_{j=0}^{K-1} p_{i,j} v_j(n-1), \quad i = 0, ..., K-1, \quad n = 1, 2, ... \tag{5.9}$$

The extension to all the possible starting states leads to the vector form equation:

$$\mathbf{v}(n) = \mathbf{q} + \mathbf{P}\mathbf{v}(n-1) \tag{5.10}$$

in which $\mathbf{v}(n)$ is a column vector with K components $v_i(n)$. The vector $\mathbf{v}(n)$ is the total expected reward in n steps.

Let consider the Markov process with rewards by means of the z-transform. This analysis is useful to underline the behaviour of $\mathbf{v}$ for a large number of steps n. The z-transform of the total-value vector $\mathbf{v}(n)$ will be called $\mathbf{v}(z)$ where $\mathbf{v}(z) = \sum_{n=0}^{\infty} \mathbf{v}(n) z^n$. The z-transform of Eq. 5.10 leads to:

$$\mathbf{v}(z) = \frac{z}{1-z}(\mathbf{I} - z\mathbf{P})^{-1}\mathbf{q} + (\mathbf{I} - z\mathbf{P})^{-1}\mathbf{v}(0), \tag{5.11}$$

that requires the inverse of the matrix $(\mathbf{I} - z\mathbf{P})$, which also appeared in the solution of the state probabilities from Eq. 5.6. This means that the presence of rewards does not affect the probability structure of the process. As in Eq. 5.6, $\pi_i(n)$ represents the state probability for state i at time index n. Considering the vector of all the possible

states, the state probabilities at step $n+1$ is given by:

$$\boldsymbol{\pi}(n+1) = \boldsymbol{\pi}(n)\mathbf{P}, \tag{5.12}$$

from which, taking the z-transform

$$\begin{aligned} z^{-1}[\boldsymbol{\pi}(z) - \boldsymbol{\pi}(0)] &= \boldsymbol{\pi}(z)\mathbf{P}, & (5.13)\\ \boldsymbol{\pi}(0) &= \boldsymbol{\pi}(z)(\mathbf{I} - z\mathbf{P}). & (5.14) \end{aligned}$$

Recalling that

$$\boldsymbol{\pi}(n) = \boldsymbol{\pi}(0)\mathbf{H}(n) = \boldsymbol{\pi}(0)\mathbf{P}^n, \tag{5.15}$$

the following equality is obtained:

$$\mathbf{H}(n) = iztf\{(\mathbf{I} - z\mathbf{P})^{-1}\} = \mathbf{P}^n. \tag{5.16}$$

$\mathbf{H}(n)$ is the inverse z-transform of $(\mathbf{I} - z\mathbf{P})^{-1}$, the elements $H_{i,j}(n)$ are the probabilities that the system will be in state j at time n given it was in state i at time 0. In [41] is proved that the matrix $(\mathbf{I} - z\mathbf{P})^{-1}$ can always be expressed in the form

$$(\mathbf{I} - z\mathbf{P})^{-1} = \frac{1}{1-z}\mathbf{S} + F(z) \tag{5.17}$$

which shows two main components. The matrix $\mathbf{S}$ is a stochastic matrix [41] and its identical rows are the limiting-state probability vector of the process, independent of n. The term $\frac{1}{1-z}$ shows that the determinant of $(\mathbf{I} - z\mathbf{P})$ always vanish for $z = 1$ and then one eigen-value of the matrix is always $\lambda_i = 1$, as in Eq.5.6. The other term in Eq. 5.17 is $F(z)$ and it represents the transient behaviour of the process. Taking the inverse z-transform of $F(z)$, the result appears as a probability matrix weighted by a geometric coefficient α^n with $|\alpha| < 1$. The transient matrix represents the decreasing geometric sequences of probability components typical of Markov processes. In conclusion the transient probability matrix $\mathbf{H}(n)$ appears always in the form:

$$\mathbf{H}(n) = \mathbf{S} + (\alpha^n, n\alpha^n, n^2\alpha^n, ...)\mathbf{T}(n). \tag{5.18}$$

For an ergodic process, $|\alpha < 1|$ and rows of the matrix $\mathbf{T}(n)$ sums to zero. In [41], $\mathbf{T}(n)$ is called differential matrix.

The asymptotic behaviour of the total expected earning, which is the study of the earning of the process in a long time duration, can be studied according to the results shown above. Starting from Eq. 5.17, the expected total earning in Eq. 5.11 can be rewritten as

$$\mathbf{v}(z) = \frac{1}{1-z}\mathbf{Sq} + \frac{1}{1-z}F(z)\mathbf{q} + \frac{1}{1-z}\mathbf{Sv}(0) + F\mathbf{v}(0). \tag{5.19}$$

The asymptotic form for large n is shown in [41] to be:

$$\mathbf{v}(n) = n\mathbf{Sq} + F(1)\mathbf{q} + \mathbf{Sv}(0), \tag{5.20}$$

where $\mathbf{Sq}$ is the column vector whose elements g_i are the sum of the immediate rewards weighted by the limiting state probabilities if the system started in state i:

$$g_i = \sum_{j=1}^{K-1} s_{i,j}q_j. \tag{5.21}$$

In the case of a completely ergodic process, all the rows of the stochastic matrix $\mathbf{S}$ are equal to the limiting state probabilities $\boldsymbol{\pi}$ of the process. Then all the starting states i have the same gain g

$$g = \sum_{i=1}^{K-1} \pi_i q_i. \tag{5.22}$$

The remaining term in Eq. 5.20 is the sum $F(1)\mathbf{q} + \mathbf{Sv}(0)$, which represents the intercepts at $n = 0$ of the asymptotes of $\mathbf{v}(n)$, leading to the simple expression valid for large n:

$$\mathbf{v}(n) = n\mathbf{g} + \mathbf{v}. \tag{5.23}$$

The Markov processes with rewards have been analyzed with special attention to the asymptotic behaviour of the expected total earning function, which is important to proceed in the study of sequential decision problems.

5.4.2 Alternatives and policies

Till now, the time-discrete process x is described by a finite Markov chain with rewards associated with each transition. If x is ergodic, the evolution of the process is driven by a stationary probability distribution for the states and transitions, with

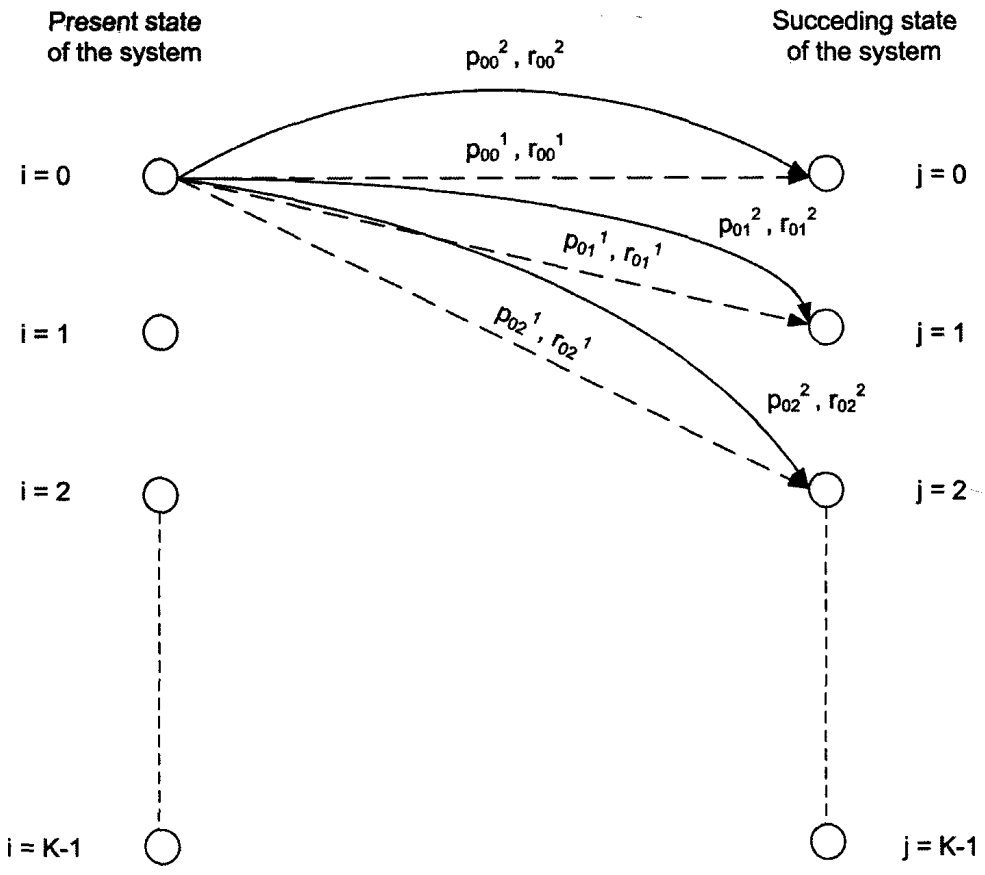

Figure 5.2: Diagram of state transitions for different actions. Each action taken modifies the process structure in terms of transition probabilities and rewards.

vector $\boldsymbol{\pi}$ and transition matrix $\mathbf{P}$. If different actions can be taken at any state, the transition matrix and the reward structure will change according to the chosen action. At each state, a specific action has effects on the probability distribution of transitions and on the rewards, since each action reasonably introduces a specific cost and result. The structure of the process with different actions is shown in Fig. 5.2. The number of alternative actions at each state must be finite, even if it is possible that the number of alternatives is different among states. A policy μ is obtained when an action a_i in the set of the possible $a_i \in A\{a_0, a_1, ..., a_M\}$ is specified for all the states $i = 1, ..., K-1$ [41]. Here A is a finite set of M possible actions for the specific state i. The optimal policy μ_o is the one that maximizes the total expected return for each state i in a working horizon of n transitions. In [41], two methods are presented to obtain an optimal policy, respectively in the case of a finite or indefinite time horizon n. For the communication system considered for this study, it is not possible to predict a finite working horizon. The process can evolve for a long time

and the decision policy must maximize the reward in a long-duration perspective. For this reason here is presented the policy-iteration method [41] for the solution of sequential decision processes, which defines the optimal policy μ_0 for a long-duration process.

5.4.3 The Policy-Iteration Method for the Solution of Sequential Decision Processes

Consider a completely ergodic K-states Markov process described by a transition probability matrix $\mathbf{P}$ and a reward matrix $\mathbf{R}$. The process is allowed to evolve for a very long time and the parameter of interest is the total expected earning of the process. Since the process is completely ergodic, the limiting steady state probabilities $\pi_{i,j}$ are independent of the staring state, the gain of the process is defined as

$$g = \sum_{i=0}^{K-1} \pi_i q_i \tag{5.24}$$

where the quantity q_i is the immediate reward in state i defined in Eq. 5.8. It is important to note that every ergodic Markov process with reward will have a gain defined as in Eq. 5.24. The same equation can be arranged in a vector form considering all the states $i = 0, ..., K-1$:

$$g = \boldsymbol{\pi} \times \mathbf{q}^T \tag{5.25}$$

The analysis of the process total earning refers to the quantity $v_i(n)$ expressed in Eq. 5.9. Since the process is allowed to evolve for a long time, it is reasonable to consider its asymptotic behaviour. The system of Eq. 5.9 and Eq. 5.23 leads to

$$g + v_i = q_i + \sum_{j=0}^{K-1} p_{i,j} v_j. \tag{5.26}$$

The linear system in Eq. 5.26 is composed by K equations for the $K+1$ unknown variables. The solution proposed in [41] sets the expected total earning of one state to zero (i.e. $v_{K-1} = 0$) so that the vector $\mathbf{v}$ contains the relative earning values respect to $v_{K-1} = 0$. In this way the number of unknown variable is K for K equations. As

explained in [41], shifting from the absolute expected earning to the relative expected earning does not influence the evaluation of different policies, mainly because the total expected earning is already a relative value to the earning at the end of the process $\mathbf{v}(0)$, as clear from Eq. 5.20.

The policy iteration method works on the relative values v_i, which hold the key for finding better and better policies. Suppose that for a given state i a set of actions $a \in \{A\}$ is available. The question is which action must be taken to maximize the total expected earning of the process. If an optimal policy has been used up to stage n, the best action for the state i at stage $n+1$ is found by maximizing the test quantity

$$q_i^a + \sum_{j=0}^{K-1} p_{i,j}^a v_j \tag{5.27}$$

for each state i, the action a which maximizes the test quantity can be found using the relatives values of the old policy. From here the iteration cycle is composed by two states.

1. **Value determination operation** This operation uses $p_{i,j}$ and q_i for a given policy to solve the system

$$g + v_i = q_i + \textstyle\sum_{j=0}^{K-1} p_{i,j} v_j, \quad i = 1, 2, ..., K-1 \tag{5.28}$$

For all the relatives values v_i and the gain g. As stated before, $v_{K-1} = 0$ by assumption. The value determination operation yields v_i and g corresponding to a given choice of $p_{i,j}$ and q_i, which are specified by a current policy μ.

2. **Policy improvement routine** for every state $i = 0, 1, ..., K-1$ this operation find the best alternative a' solving

$$\max_{a \in \{A\}} \{q_i^a + \sum_{j=0}^{K-1} p_{i,j}^a v_j\} \tag{5.29}$$

and using the relatives values v_i of the previous policy. When the maximum is found, the following parameters are set

$$d_i = a', \quad q_i = q_i^{a'}, \quad p_{i,j} = p_{i,j}^{a'} \tag{5.30}$$

This routine yields the probability $p_{i,j}$ and immediate reward q_i which increase the gain for a given set of v_i. It returns a policy starting from the values v_i of the previous one.

The iteration procedure can be entered in either states, respectively by setting a initial policy μ or a set of initial values v_j. Choosing to enter in the policy improvement routine and setting $v_j = 0$ for $j = 0, 1, ..., K-1$, the routine will find the best alternatives for all the states i, defining a first policy to enter the value determination state. In [41] is proved that each succeeding policy found in the iteration cycle has higher gain that the previous one. If a policy returns the same gain of the previous one, the previous stays as optimal to avoid useless updates. The optimal policy is reached when two successive iterations are identical and the optimization process stops. The final result is the optimal policy μ_0, which univocally defines a Markov process with the values of transition probability $p_{i,j}$ and reward $r_{i,j}$ for each possible state.

5.5 Finite State Markov model for radio communication channels

The literature reveals many attempts to model the fading envelope of a time variant communication channel and the resulting error flow using finite Markov chains. In this section the meaning and importance of FSMC as a model of radio communication channels is discussed, in order to introduce the problem of the FSMC model for the MIMO channel capacity.

The study of the FSMC to model a communication channel emerges from the early work of Gilbert [42] and Elliott [43]. They studied a two-states Markov channel known as the Gilbert-Elliott channel. Their model is composed of a good state G and a bad or burst state B, the transition probabilities are made in order to simulate the burst error conditions. More complex Markov models followed this first one, in the attempt of modelling more accurately the communication error statistical behaviour [44]. When the channel quality varies significantly, the two-state Gilbert-Elliott model is not adequate. A straightforward solution is to increase the possible states of the model to a finite set $S = \{s_0, s_1, ..., s_{K-1}\}$; if the process under analysis is ergodic, the resulting Markov process $\{S(n)\}$, for $n = 1, 2, ...$ will be a constant Markov process

with the properties described in Sec. 5.3. Looking at wireless communications in fading environments, many works such as [38, 39, 45] suggested the idea of creating a FSMC for the channel state, obtained by partitioning the SNR at the receiver side in K possible states. The transition probabilities between two channel states depend on the statistical description of the random fading process. Authors of [45] proposed a well known analytical model to calculate the transition probabilities by looking at the fading process and SNR statistics. The core of this model is the evaluation of the Level Crossing Rate (LCR) for the SNR process. Given a random process x, the LCR of a specific value $\bar{x}$ in a time interval Δt is defined as:

$$\int_{t}^{t+\Delta t} Pr\{x > \bar{x}\}dt \tag{5.31}$$

and it is dependent on the duration of the observation interval Δt. The LCR is defined as the number of crossing per second of a given threshold. Defining $\Delta x = \dot{x}\Delta t$, a possible way to obtain the LCR is to derive the statistical distribution of the first derivative $\dot{x}$, which represents the velocity of the process, and then evaluate the probability of x being around the threshold $\bar{x}$ with all the possible velocities:

$$LCR(\dot{x}) = \int_{0}^{\infty} \dot{x} Pr\{\bar{x}, \dot{x}\}d\dot{x} \tag{5.32}$$

where $Pr\{\bar{x}, \dot{x}\}$ is the joint density of the process x and its first derivative $\dot{x}$. In the case of Gaussian process this integral can be solved in a closed form which involves the second derivative of the correlation coefficient $\rho_{xx}(\Delta t)$ of the process x under analysis. Knowing the statistics of the received SNR, the probability inside the integral can be expressed in a closed form. The study of the LCR allows to approximate the transition probabilities to the adjacent channel states in the simple form :

$$p_{i,i+1} \approx \frac{LCR_{i+1}}{R_t^{(i)}}, \quad i = 0, 1, ..., K-2 \tag{5.33}$$

where LCR_{i+1} is the number of up-crossing rate per second and $R_t^{(i)} = R_t \times p_i$ is the symbol transmission rate weighted by the state probability p_i. The LCR is derived

analytically from the received SNR distribution [45]

$$LCR(a) = \sqrt{\frac{2\pi a}{\bar{\gamma}}} f_d \exp\left\{-\frac{a}{\bar{\gamma}}\right\}, \tag{5.34}$$

where $\bar{\gamma} = E\{A\}$ denotes the mean value of the received SNR and $f_d = v/\lambda$ is the Doppler frequency of the mobile user normalized to the carrier wavelength λ. Eq. 5.33 is referred as the Wang-Moayeri model for adjacent states transition probabilities. They can be determined as:

$$\begin{aligned} P_{k,k+1} &= \frac{N_{k+1}T_f}{P_k}, \qquad k = 0, ..., K-1 \\ P_{k,k-1} &= \frac{N_k T_f}{P_k}, \qquad k = 1, ..., K \end{aligned} \tag{5.35}$$

in which N_k is the cross rate for state k, either upward or downward. By the assumption in the model [45], the probability of remaining in the same state k is defined as:

$$P_{k,k} = \begin{cases} 1 - P_{k,k+1} - P_{k,k-1}, & if\ 0 < k < K \\ 1 - P_{0,1}, & if\ k = 0 \\ 1 - P_{K,K-1}, & if\ k = K \end{cases} \tag{5.36}$$

The Wang-Moayeri method for the evaluation of transition probabilities of a random process has been used in [7] to generate a FSMC of the channel state based on the received SNR. In [7], the modulation and coding at the transmitter sides are adapted to the variation of the *signal to noise ratio* (SNR) of the communication link, which is modeled as a FSMC. The proposed CPM maintains a certain level of average *packet error rate* (PER) over the time-variant channel when corresponding AMC mode is applied for each channel state. When the target PER is not fixed a priori, the overall packet loss at PHY and MAC layers can be minimized through cross-layer analysis. The use of LCR for MIMO capacity first appeared in [19], in which the estimation of LCR is a good parameter to obtain the transition probabilities.

The framework just discussed provides a simple way to obtain the transition probabilities only between adjacent states. Consequently it can be inferred that this model works well for slow varying processes, which evolve only to adjacent states in the time observation interval Δt. For high normalized Doppler frequency $f_m = f_d T_f$,

the process could jump to far states and the crossing probability between the adjacent states may turn in very low values. This is confirmed by simulation results for low and high varying processes, which will be discussed in the next section. In order to overcome this limitation a different approach to the problem must be taken.

The main disadvantage of most proposed models is the fact that, while the elements of the transition matrices can be analytically calculated, it is especially difficult to obtain analytical expressions for eigenvalues and eigenvectors of the matrix of transitional probabilities. The model proposed in [46] allows to determine the transition probability matrix for a K state Markov chain with the knowledge of the process correlation interval d, which is the second largest eigenvalue of the generated process [28]. The exact knowledge of eigenvalues greatly reduces the complexity and accumulation of numerical error. Let $\mathbf{P} = [p_{ij}]$ be a matrix of transitional probabilities of the *DAR-1* K-states Markov chain. It is shown in [46] that $\mathbf{P}$ is defined as

$$\mathbf{P} = \mathbf{Q} + d \times (\mathbf{I} - \mathbf{Q}), \quad 0 \leq d < 1 \tag{5.37}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{Q}$ is composed of the steady state probabilities as shown:

$$\mathbf{Q} = \begin{bmatrix} p_0 & p_0 & \cdots & p_0 \\ p_1 & p_1 & \cdots & p_1 \\ . & . & \cdots & . \\ p_{K-1} & p_{K-1} & \cdots & p_{K-1} \end{bmatrix}, \tag{5.38}$$

where $\{p_k\}_{k=0,\ldots,K-1}$ are the stationary probabilities. The resulting Markov process ι at discrete time $\kappa = 0, 1, \ldots$ has an exponential autocorrelation function

$$R_{\iota\iota}(\kappa) = R_{\iota\iota}(0) d^{\kappa}. \tag{5.39}$$

which match the desired autocorrelation of an AR-1 generated process. The value d corresponds to the autocorrelation of the process at time $\kappa = 1, 2, \ldots$. In the case of a mobile wireless communication system, the channel autocorrelation between two frames of duration T_f is $\rho = J_0(2\pi f_d T_f)$ [1], where f_d is the Doppler frequency of the mobile user and T_f is the time frame duration. In [46], the product $f_d T_f$ is referred as fading bandwidth.

The method proposed in [46] provides the transition probabilities among all the possible states K and not only between adjacent states. For this reason, especially

for significant fading bandwidth $f_d T_f$, the transition probabilities obtained by this model are to be preferred to the ones based on LCR in Eq.s 5.35 and 5.36.

5.6 FSMC for MIMO channel capacity

As in [3, 15], instantaneous capacity is well described by a random Gaussian process $C = N(\hat{C}, \sigma_C^2)$. Referring to the analysis in Sec. 2.5, the AR-1 evolution model of the channel capacity process in Eq. 2.15 assigns a Markov nature to the process, in the specific case of the first order. In this model, the information on the next state is gained only from the current state, assuming that information corresponding to previous states is negligible. Following the idea proposed in Sec. 5.5, a Finite State Markov Chain can be obtained by partitioning the instantaneous capacity process into a finite numbers of intervals or states $S = \{0, 1, ..., K-1\}$. We consider a finite set of capacity states $S = \{s_0, s_1, ..., s_{K-1}\}$ with corresponding thresholds $\{c_k\}_{k=0}^{K}$. Capacity is said to be in state s_k, $k = 0, 1, ..., K-1$ if the value C of the process is in the interval $[c_k, c_{k+1})$:

$$c_k \leq C < c_{k+1}. \tag{5.40}$$

As discussed in [46], partitioning must be performed such that the highest state probability is assigned to the state s_n^*, which contains the average value of the capacity process, by selecting boundaries $[c_k^*, c_{k+1}^*)$ such that:

$$\bar{C} = \frac{c_k^* + c_{k+1}^*}{2}. \tag{5.41}$$

The states surrounding s_k^* on the left and right of the process *pdf* must have decreasing probabilities with respect to s_k^*. If the partitioned process is ergodic, the Markov process S_k is a stationary process, with the property that transition probabilities in Eq. 5.1 are time invariant as discussed in Sec. 5.3.

The evolution of the capacity process is related to a fixed time scale defined by the frame duration T_f of the system. According to the general block fading model, the channel capacity is assumed to remain constant within one block period, with block length equal to T_f. The instantaneous capacity evolution model has been presented in Eq. 2.15. Given the value of capacity at time index nT_f with $n = 0, 1, ...$, the capacity process at time $n+1$ is a Gaussian random variable related to the original

process by the correlation coefficient $\rho(T_f)$ (from now on the parameter T_f will be omitted). Specifically, the Gaussian process at one step prediction is characterized by the following first two moments:

$$\begin{aligned} \bar{C}_{n+1} &= (1-\rho)\bar{C}_n + \rho C(n) \\ \sigma^2_{n+1} &= (1-|\rho|^2)\sigma^2_n; \end{aligned} \tag{5.42}$$

For a FSMC model the two probabilities of main interest are the steady state probability, which describes the asymptotic probability of being in a given state s_k and the transition probability, which drive the transitions among different states.

- **Steady State Probabilities**: the probability p_k^o that the instantaneous capacity is in state s_k is defined as

$$\pi_k = \int_{c_k}^{c_{k+1}} p(C)dC \tag{5.43}$$

for $k = 0, 1, ..., K-1$. For Gaussian random variables, the steady state probability is easily expressed by $Q(c_k) - Q(c_{k+1})$, where $Q(x)$ represents the well-known Q-function [1]. The steady state probabilities can be arranged in the following vector form:

$$\boldsymbol{\pi} = \lim_{n\to\infty} Pr(s(n) = s_k), \tag{5.44}$$

for time index $n = (0, 1, 2...)$ and for $k = (0, 1, ..., K-1)$.

- **Transition Probabilities**: the probability of transitions between two states is a conditional probability which can be obtained by the joint probability density function of the state distribution, according to the Bayes' rule:

$$p(b|a) = \frac{p(b,a)}{p(a)} \tag{5.45}$$

The two variables under attention are the capacity values C at time n and $n+1$. According to Eq. 2.15, the capacity evolution $C(n)$ is still a Gaussian variable. For the particular case of two Gaussian random variables, the joint probability

is given by:

$$Pr\{x_1, x_2\} = \int_{c_i}^{c_{i+1}} \int_{c_j}^{c_{j+1}} \left(\frac{1}{\sqrt{2\sigma_1^2\sigma_2^2(1-\rho^2)}} \right) \exp\left[-\left(\frac{1}{2(1-\rho^2)} \right) \left(\frac{\delta x_1^2}{\sigma_1^2} + \frac{2\rho\delta x_1 \delta x_2}{\sigma_1\sigma_2} + \frac{\delta x_2^2}{\sigma_2^2} \right) \right] dx_1 dx_2 \quad (5.46)$$

where δx is defined as $x - \mu_x$ and ρ is the correlation coefficient between x_1 and x_2 at the sampling time T_f. The transition probability is then the joint probability density function weighted by the marginal density of one of the two variables. Transition probabilities are arranged in a matrix $\mathbf{P}$ with property specified in Eq. 5.5.

The steady state probabilities and the transition probabilities should satisfy the equilibrium conditions proper of any Markov chain described in Sec. 5.3. The steady state and transition probabilities can be evaluated numerically once a set of thresholds $\{c_k\}_{k=0}^{K}$ has been defined. The steady state probabilities of the capacity process are evaluated by Eq. 5.43 as the probability of being a specified state:

$$Pr\{C \in s_k\} = \int_{c_k}^{c_{k+1}} p_C(x)dx \quad (5.47)$$

where $p_C(x)$ is the Gaussian probability density function of the instantaneous capacity. The transition probabilities can be arranged as follows [20]:

$$Y_{k,k+1} = X(k)T(k, k+1) = Pr(C(n) \in s_k, C(n+1) \in s_{k+1}) \quad (5.48)$$

where $k = 0, 1, ...K-1$ is the state index and $n = 1, 2, ...$ represents the discrete time evolution. Eq. 5.48 is a joint density function of two Gaussian random variables. Knowing the thresholds $\{c_k\}_{k=0}^{K}$ of the capacity states and referring to the model in Eq. 2.15 the following result is obtained:

$$Y_{k,k+1} = Pr\left(c_k < C(n) < c_{k+1}, \frac{c_{k+1} - \rho C(n)}{\sigma\sqrt{1-\rho^2}} < \xi(n+1) < \frac{c_{k+2} - \rho C(n)}{\sigma\sqrt{1-\rho^2}} \right); \quad (5.49)$$

recalling that $\xi(n+1)$ is a Gaussian random variable with zero mean and unitary variance, the transition probability between the adjacent states (s_k, s_{k+1}) is given by:

$$\begin{aligned} Y_{k,k+1} &= \int\int Pr(C(n) \in s_k, C(n+1) \in s_{k+1}) dC(n) dC(n+1) \\ &= \int_{c_k}^{c_{k+1}} f(C(n+1)) \left[F(\frac{c_{k+2} - \rho C(n)}{\sigma_C \sqrt{1-\rho^2}}) - F(\frac{c_{k+1} - \rho C(n)}{\sigma_C \sqrt{1-\rho^2}}) \right] dC(n+1) \end{aligned} \tag{5.50}$$

where σ_C is the standard deviation of the capacity Gaussian process $C(n)$, ρ is the correlation coefficient for the capacity evolution at time T_f and $F(x)$ denotes the *cumulative distribution function* (CDF) of the Gaussian random variable $C(n)$.

This approach has been compared to the other frameworks proposed in the literature and briefly introduced in Sec. 5.5. Referring to the Wang-Moayeri model in Eq. 5.33, the evaluation of the LCR is subject to the second derivative of the process correlation coefficient $\rho(\tau)|_{\tau=0}$, which depends on the process under analysis. In the proposed AR-1 model, the correlation coefficient is exponential (Eq. 5.39) and the second derivative cannot be calculated for $\tau = 0$. From here the need of a different approach to the LCR evaluation problem. Given a threshold c_k, one upcrossing event is defined as the joint probability

$$\begin{aligned} Pr\left(\tilde{C}(n+1) > c_k, \tilde{C}(n) < c_k\right) &= Pr(\xi(n+1) > \frac{c_k - \rho\tilde{C}(n)}{\sqrt{1-\rho^2}}, \tilde{C}(n) < c_k) \\ &= \int_{\alpha}^{\infty} f_\xi(x) \int_{-\infty}^{c_k} f_C(y) dy dx \\ &= \int_{\alpha}^{\infty} f_\xi(x) F_C(c_k) dx \end{aligned} \tag{5.51}$$

where $\alpha = \frac{c_k - \rho\tilde{C}(n)}{\sqrt{1-\rho^2}}$. Following the same calculation, the downward crossing of the same threshold is given by

$$Pr(\tilde{C}(n+1) < c_k, \tilde{C}(n) > c_k) = F_\xi(\alpha) - \int_{-\infty}^{\alpha} f_\xi(x) F_C(c_k) dx. \tag{5.52}$$

These quantities can be numerically obtained once the thresholds have been decided and the correlation parameter ρ is obtained. The adjacent states transition probabilities can be determined by the Wang-Moayeri model [7]:

$$\begin{aligned} P_{k,k+1} &= \frac{N_{k+1}T_f}{Prk}, \qquad k = 0, ..., K-1 \\ P_{k,k-1} &= \frac{N_k T_f}{Prk}, \qquad k = 1, ..., K \end{aligned} \tag{5.53}$$

in which N_k is the cross rate for state k, either upward or downward. By the assumption in the model [45], the probability of remaining in the same state k is defined as:

$$P_{k,k} = \begin{cases} 1 - P_{k,k+1} - P_{k,k-1}, & if\ 0 < k < K \\ 1 - P_{0,1}, & if\ k = 0 \\ 1 - P_{K,K-1}, & if\ k = K \end{cases} \tag{5.54}$$

As discussed in Sec. 5.5, the LCR based method limits the transitions only to the adjacent states. This limitation is realistic only for a slow varying process, for which the transition to far states in a prediction interval T_f is very small. The main result of Wang-Moayeri model for high values of f_m is a high probability of remaining in the same state, clearly shown by Eq. 5.36. In order to have a good representation of the variability of the process for all the possible values of f_m, the transition probability matrix is evaluated by the eigenvalue framework presented in [46].

One way to test the transition probabilities obtained by the different methods is to verify if they satisfy the Markov property of state and transition probabilities described in Eq. 5.6. Numerical results showed that the method based on the joint probability density function in Eq. 5.46 is easily subject to numerical error due to the precision of the integral evaluation, which can be only solved numerically. As expected, the LCR method is accurate only for low varying processes with low normalized Doppler f_m, showing for high f_m an unrealistic high probability remaining in the same capacity state. The eigenvalue framework proposed in [46] has shown the best behaviour, matching with high accuracy the Markov property in Eq. 5.6.

In conclusion, once the transition probabilities are computed, the FSMC for the channel capacity is modeled as a $(K+1) \times (K+1)$ transition probability matrix,

with the form:

$$\begin{bmatrix} P_{0,0} & P_{0,1} & \cdots & 0 \\ P_{1,0} & P_{1,1} & \cdots & \vdots \\ 0 & \ddots & \ddots & 0 \\ \vdots & P_{K-1,K-2} & P_{K-1,K-1} & P_{K-1,K} \\ 0 & \cdots & P_{K,K-1} & P_{K,K} \end{bmatrix} \tag{5.55}$$

5.7 Two-dimensional cross layer optimization

The knowledge of the evolution of the instantaneous capacity process leads to describe the evolution of the maximum amount of information that can be sent over the channel. In particular, outage occurs when the signalling rate at a specific time instant is above the instantaneous channel capacity. Given the duration frame T_f, the signalling rate can be adapted following the capacity evolution at discrete steps T_f. Once a set of possible signalling rates is available at the PHY layer, the question of which rate must be chosen according to the capacity state becomes of primary importance. Authors of [8] underline that the optimal choice of the signalling rate must take into consideration not only the channel state evolution but also the system buffer state. This leads to a deep cross-layer analysis, whose purpose is the use of the optimal signalling rate given the conditions of the buffer and communication channel. This kind of problem is referred in the literature as a two-dimensional optimization problem, which can be investigated by the application of *dynamic programming* (DP) and *Markov decision process* (MDP) theories for the optimal transmission policy design. To avoid a deep analysis of the modulation and code scheme at the physical layer, this work proceeds with the assumption that the information level described by the channel capacity can be achieved by a proper code and modulation scheme. The optimization target turns into the MAC layer throughput for the optimal transmission policy design.

Consider the FSMC model of the capacity evolution described in Sec. 5.6. The frame duration T_f represents the time interval in which the system transmits with a fixed information rate, established at the beginning of the time frame. Once the channel capacity mean value and variance have been estimated according to Eq. 2.8 and Eq. 2.9, the AR prediction model discussed in Sec. 2.5 is used to obtain the next values of instantaneous capacity. Consider the knowledge of the instantaneous

capacity $C(n)$ at time n. $C(n+1)$ is the predicted instantaneous capacity at the next time step according to the specific AR-1 evolution model in Eq. 2.15. The probability distribution of the predicted instantaneous capacity is given by $p_C(C(n+1)|C(n))$: it is conditioned to the previous value $C(n)$ and it is still Gaussian. The mean value and variance of prediction are obtained by taking the expectation of the AR model, resulting in Eq. 5.42. The transmitter station predicts the instantaneous state of the channel capacity $C(n+1)$ based on the knowledge of the previous state $C(n)$ and establish a possible information rate $r(n+1) \leq C(n+1)$. The corresponding probability of outage is given by [3]:

$$P_{out} = Prob(C(n+1) < r(n+1)) = \int_{-\infty}^{R} p_C(x)dx = Q\left(\frac{\bar{C}(n+1) - r(n+1)}{\sigma_{n+1}}\right) \tag{5.56}$$

For any given rate $r(n+1)$ the corresponding outage probability can be calculated with the use of the mean and variance of the AR-1 predicted capacity. As from Eq. 5.42, the moments of the distribution depend on the correlation coefficient ρ of the capacity process. For a long prediction interval the correlation coefficient decays to zero resulting into the mean and variance of prediction equal to the mean and variance of the original process: no additional information is available for the prediction. If a specific outage is required in the transmission, the corresponding rate can be derived according to [3]:

$$r \approx max\left\{\bar{C}_{n+1} - \sqrt{-2\sigma_{n+1}^2 \ln(2P_{out})}\right\} \tag{5.57}$$

where P_{out} is the desired outage probability and $\bar{C}_{n+1}$,σ_{n+1}^2 are the mean and variance of the AR-1 prediction.

From a MAC layer point of view, deciding for a specific rate $r(n+1)$ has two effects: on one side a value of $\varphi(n+1)$ packets will be transmitted in the next T_f, allowing to free a precise space in the buffer; on the other side a price will be paid in terms of probability of loosing packets due to outage. The FSMC of the instantaneous capacity defines the transition probability matrix of the process. The capacity transition in the next time step is driven by the values of the transition matrix, with each arriving state associated with a probability value. For each one of the capacity states we define a proper set of permitted information rates $R =$

$\{r_k\}_{k=0}^{K-1}$, where

$$r_k \leq c_k, \quad r_k \in R, \quad 0 \leq k \leq K-1; \tag{5.58}$$

For each one of the possible rates, the outage probability would be the probability that capacity at the predicted state is under the rate value. Knowing the moments of the Gaussian distribution of the predicted capacity, the outage probability for a specific rate r_k is given by Eq. 5.56. When outage occurs no reliable transmission are possible since the channel do not support and carry the information that is being sent. If the signalling rates r_k are expressed in packets/s, the *packet error rate* (PER) is given by $r(n+1) \times P_{out}$ while a successful data transmission is obtained for $r(n+1) \times (1-P_{out})$.

5.7.1 Policy domain performance optimization

The set of available signalling rates constitutes the action space $A(s)$ for each capacity state $s \in S$. A policy is defined when one specific action $a \in A$ has been defined for each state of the model. Bringing into the model the system buffer, a convenient signalling rate can be chosen according to both the capacity state and the queue buffer state. Suppose that the system has a single buffer of B packets. The buffer state defines how many packets are waiting in the buffer and it is itself a FSMC dependent on the arrival process. The higher-level arrival process is described by a Poisson model [21]:

$$Pr\{A(\tau) = m\} = \frac{e^{-\lambda\tau}(\lambda\tau)^m}{m!}, \quad m = 0, 1, ... \tag{5.59}$$

where $A(\tau) = m$ denotes the event of m arrivals in the time interval τ. λ is the average arrival rate in packets/s. The capacity transitions and the arrival process are considered two independent processes due to their different nature. Considering both the capacity and buffer states leads to a two-dimensional problem, in which the system is completely characterized by the state pair $s(k, q)$, where k is the capacity state index and q is the MAC layer queue length.

In order to refer to the Markov Decision Processes theory, the state transition probability matrix and the reward matrix must be specified, as discussed in 5.3. Referring to [7, 8] the following matrices are defined.

1. State Transition Probability matrix **T**: is a three dimensional matrix which orders the possible actual states, the possible succeeding states and the possible

actions (or alternatives) which can be chosen in each state. Each "slice" ζ_a corresponds to the set of all possible states $(k,q)(k',q')$ for $k = 0,1,...,K-1$, $q = 0,1,...,B$.

$$\zeta_a = \begin{bmatrix} p^a_{(0,0);(0,0)} & \cdots & p^a_{(0,0);(1,0)} & \cdots & p^a_{(0,0);(K-1,B-1)} \\ p^a_{(0,1);(0,0)} & \cdots & \cdots & \cdots & \vdots \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p^a_{(K-1,B-1);(0,0)} & \cdots & p^a_{(K-1,B-1);(1,0)} & \cdots & p^a_{(K-1,B-1);(K-1,B-1)} \end{bmatrix} \tag{5.60}$$

The transition probability $p^a_{(k,q),(k',q')}$ is function of the parameters (a,k,q,k',q') and represents the probability of passing from the state (k,q) to the new state configuration (k',q') in terms of capacity and buffer size. The value 'a' specifies the action to take at time T_f, in this case the choice of the signalling rate for the current block. The optimal choice of 'a' is the optimization problem which requires the knowledge of the rewards for each action. The value of $p^a_{(k,q),(k',q')}$ in $\mathbf{T}$ is defined for any possible action in the current $a \in \mathcal{A}(s_{(k,q)})$, as:

- if $(q - min(q, \varphi^a_{max}) \leq q' < B)$,then

$$p^a_{(k,q),(k',q')} = p\{A(T_f) = q' - [q - min(q, \varphi^a_{max})]\} \times p_{k;k'} \tag{5.61}$$

- if $(q - min(q, \varphi^a_{max}) \leq q' = B)$,then

$$p^a_{(k,q),(k',q')} = 1 - \sum_{q'=0}^{B-1} p\{A(T_f) = q' - [q - min(q, \varphi^a_{max})]\} \times p_{k;k'}; \tag{5.62}$$

where φ^a_{max} is the maximum number of packets that can be served in the time interval T_f when a specific action a is taken. $A(T_f)$ denotes the expected number of packets that will arrive in the next interval T_f according to the arrival process. The probability of arrival are modeled according to the Poisson arrival process, the capacity transition probabilities are derived for every state transition according to the FSMC of channel capacity.

2. State Transition Reward Matrix $\mathbf{R}$: each element specifies the reward associated with a state transition $(k,q),(k',q')$ for a given action 'a'. The definition

of "reward" refers to a specific performance metric. The decision of working with rewards should bring to the same conclusions obtained by minimizing the process costs. Costs and rewards involve different process analysis: a smart solution is to choose the metric which simplifies the process analysis and then minimize/maximize it according to the costs/rewards description. In this study the value of "reward" is defined as the number of packets correctly received at the destination MAC layer. With this definition, elements of $\mathbf{R}$ are function of the set $\{a, k, q, k', q'\}$ and defined as:

$$r^{a}_{(k,q);(k'q')} = \begin{cases} min(q, \varphi^{a}_{max}) \times \left[1 - P^{a}_{out}(k)\right], & p^{a}_{(k,q);(k',q')} \neq 0 \\ 0, & otherwise \end{cases} \quad (5.63)$$

where $P^{a}_{out}(k)$ denotes the probability of outage on the channel, which is the probability of no reliable transmission as seen in Sec. 2.5.1.

The optimization problem works on the variable a, which denotes the possible actions to take at every decision epoch T_f. In an optimal model, at each step T_f the transmitter predicts the instantaneous capacity and obtains a set of possible signalling rates associated with outage probability. The action 'a' is the decision of a specific rate to use in the next time interval T_f. A decision on the signalling rate will have consequences on the buffer queue (how many packets the system will serve) and on the outage probability in transmission. When the optimal action is decided for every possible state, a policy μ is obtained. From the matrices $\mathbf{T}$ and $\mathbf{R}$, the optimal policy for signalling rate selection can be solved by the policy iteration method discussed in Sec. 5.4. The resultant policy μ_0 is a vector $\mathbf{d}$, which specifies the best action 'a' to take at each state. Each policy μ specifies a unique Markov process with rewards, with specific state probability $p^{a}_{(k,q),(k',q')}$ and reward $r^{a}_{(k,q),(k'q')}$ for each process state.

5.7.2 Numerical simulations

In this section the improvement in the system performance brought by the two-dimensional optimization is analyzed. The possibility of adapting the signalling rate to the channel capacity state and the buffer state should reduce the total probability of failure in the system, defined as the sum of the probability of buffer overflow and

capacity outage. The starting point for this analysis are the final results obtained in Chapter 3. Fig. 3.9 showed the possible tradeoff between the buffer overflow and capacity outage probability, for a MIMO system with constant signalling rate derived from the outage capacity. The study proposed in this chapter suggests that, if the signalling rate can be adapted to the buffer and capacity states, the ergodic channel capacity can be achieved by an optimal adaptive signalling rate strategy. The optimal strategy is found analytically by the MDP analysis. The optimal solution is characterized by the maximum expected earning or highest process gain defined in Eq. 5.25. In the model specified in the previous section, the reward structure of the problem is defined as the number of packets which are correctly received on the other side of the communication link, as specified in Eq. 5.63. For this reason and recalling Eq. 5.8, the process gain g can be related directly to the throughput η of the system through the time frame duration:

$$\eta = g/T_f \tag{5.64}$$

A second parameter of interest is the packet loss rate at the source. Disregarding the origin of the loss, the total packet loss rate is the complementary part of the total packet arrived at the source and the packet correctly received. For this reason, the source packet loss rate is defined as

$$\varepsilon = 1 - \frac{g}{\lambda T_f} \tag{5.65}$$

The gain of the constant rate transmission must be obtained in order to compare it with the optimal adaptive rate solution. Using the same definition of reward of Eq. 5.63, the reward matrix for the constant rate transmission is a matrix of equal elements since the signalling rate is constant for all the states. The signalling rate is obtained by the numerical analysis presented in Chapter 3, Sec. 3.2.2, as the rate which minimize the joint effects of outage and buffer overflow. Once the rate is specified, the corresponding outage probability is used for the process reward. The following figures show the difference of process gains and data loss probability for the constant rate transmission and the adaptive rate transmission obtained by the MDP analysis. The simulations parameters are presented in the following table.

In the case of constant signalling rate, the rate is chosen as the optimal solution

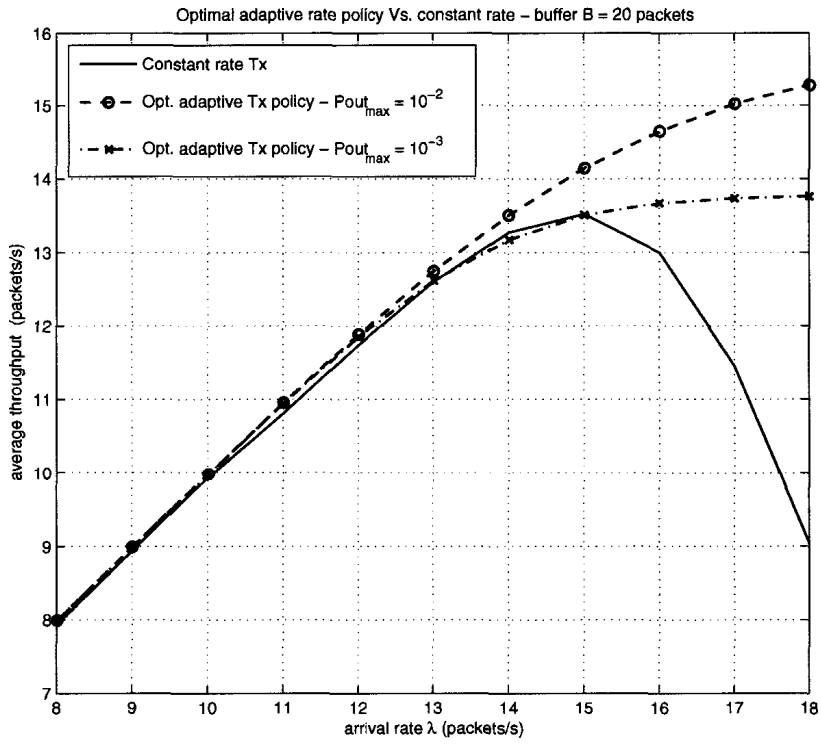

Figure 5.3: Gain of constant rate transmission and adaptive rate transmission for different arrival rates λ. The gain is the system throughput per time frame T_f. The buffer of the system is $B = 20$ packets.

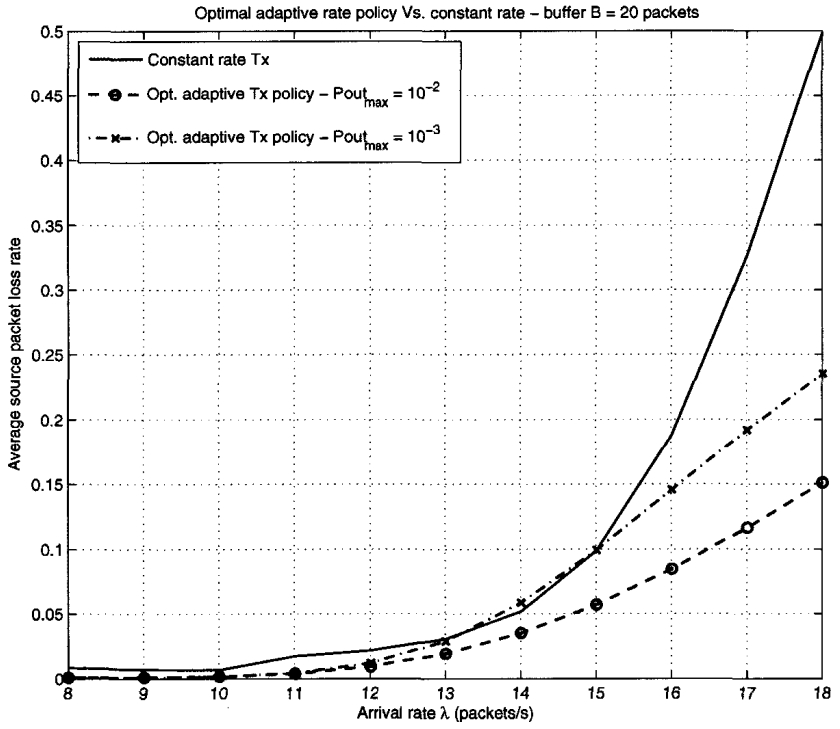

Figure 5.4: Total packet loss rate of constant rate transmission and adaptive rate transmission for different arrival rates λ. The buffer of the system is $B = 20$ packets.

Description	Parameter	Value
Signal to Noise Ratio	SNR	$10dB$
Band	W	$1.25\ MHz$
Tx,Rx Antennas	N_T, N_R	4
Bits per packet	N_b	1080
Carrier frequency	f_c	$2\ GHz$
User speed	v	$10m/s$
Frame duration	T_f	$0.002s$
Buffer size (packets)	B	20

Table 5.1: System Parameters

obtained by solving the analytical equations of buffer overflow and outage presented in Chapter 3, Sec. 3.2.2. Increasing the arrival rate λ, the tradeoff solution of Fig. 3.9 shifts toward high values of outage, due to the increase in the required signalling rate. The optimal policy obtained by the MDP analysis adapts the signalling rate to the state of buffer and capacity at every time frame T_f, forcing the system to work at high values of signalling rate only when the buffer is full. Moreover, a target outage probability P_{out} will lead to the choice of a suitable set of signalling rate to match the requirement.

Fig. 5.3 shows the process gain of constant rate transmission and adaptive rate transmission for different arrival rates λ. The gain represents the system throughput per time frameT_f. The buffer of the system is $B = 20$ packets. For the adaptive rate scheme, the results in terms of packets correctly received at the destination are significantly better than the constant rate solution. Particularly, the constant signalling rate is subject to a strong decrease in gain as the arrival rate is allowed to increase. The reason of this behaviour is the high outage that becomes predominant for high signalling rates. Adapting the rate according to the optimal policy μ_0, the heavy effect of outage is avoided for high arrival rates: the gain curves of adaptive rate show an asymptotic stability toward the maximum allowed signalling rate. A more strict requirement in terms of outage imposes a lower signalling rate which turns into a lower gain: the number of packet per time frame correctly received is lower and an increase in the buffer overflow probability is expected.

The study of the total packet loss rate in Fig. 5.4 confirms the better performance of rate adaptation through the MDP. From Fig. 5.4 it is also clear that the outage probability for high arrival rates becomes the main component of the packet

loss rate, since the system is forced to work at high signalling rates to minimize the buffer overflow. The packet loss rate for adaptive rate transmission lies under the constant rate solution for almost all the values of arrival rate λ. Especially for high rates, the constant rate solution imposes a high signalling rate to decrease the buffer overflow but this choice turns into a high outage probability, which appears to be predominant in the total packet loss rate. The adaptive rate transmission, limiting the allowed signalling rates to match a specific outage requirement, works with a lower rate but results in overall better performance: results show a controlled increase of the packet loss rate. Again in this case, a more strict requirement on the target outage probability results into an increased packet loss probability: with a lower signalling rate the buffer overflow probability is higher for a given arrival rate.

5.8 Complexity Issues

The analysis of the complexity for the proposed algorithm should take into consideration the set of possible signalling rates and how often rates are adapted to the capacity and buffer states. From a PHY layer point of view, a fixed set of possible signalling rates would be available according to the specific modulation technique of the transmission system. The optimization algorithm proposed can easily include specific signalling rates and derive the outage probability for each one of them according to the capacity distribution. The resulting decision policy would lead to a sub-optimal process gain but would perform better than the constant signalling rate solution. Another complexity issue is how often a CSI estimation can be available at the transmitter to match the actual channel capacity state. The results shown in this chapter are based on the possibility of tracking the capacity process at every time frame T_f through the prediction model in Eq. 2.15. According to the normalized Doppler f_m of the process, the prediction can be exploited for more than one frame. In the worst case of very high user mobility, at the time frame T_f the correlation of the capacity evolution could be zero and the prediction for the next frame would be useless. In those cases the only solution is the use of a constant rate transmission based on the capacity CDI.

5.9 Conclusion

In this chapter the possibility of adapting the signalling rate to the actual channel capacity and the buffer state has been presented. If it is possible to have a feedback information about the channel capacity, the transmitter can adapt the signalling rate to the actual capacity state to minimize the outage. Moreover, this analysis is not optimal [8] without considering the buffer state, i.e. how many packet are in the buffer and if the buffer can experience overflow. The joint consideration of the capacity evolution and the buffer state leads to a two-dimensional optimization problem. The possibility of modelling the MIMO channel capacity as a FSMC allows to refer to the theory of MDP to tackle this optimization problem. In the specific case, the solution must be the optimal transmission signalling rate for every possible state of capacity and buffer. In the proposed model, the channel capacity is allowed to evolve in a finite set of states and the transition probabilities among states are investigated. Referring to widely accepted model like the Wang Moayeri [45] for the FSMC of fading channel, an extension became necessary to allow the transition not only to adjacent states but also to far states. The framework proposed in [46] has revealed to be useful for this task. Considering the optimal case in which the transmitter can choose a proper transmission rate to match the capacity state, the MDP theory returns the optimal policy μ_0 which must be followed to maximize the process gain, i.e. the throughput of the system, defined as the number of packet correctly received at the destination MAC layer. Assuming a specific frame duration T_f, the signalling rate is adapted at every frame according to the capacity transition behaviour and buffer state. The choice of the time scale for adapting the rate must take into consideration the evolution of the process, which can be described by its second order statistic (autocorrelation parameter ρ). Once the optimization problem is solved, the optimal policy μ_0 drives the choice of the most suitable rate according to the joint state of buffer and capacity. The performance of the optimal policy have been compared with the constant rate transmission proposed at the end of Chapter 3, showing a significant increase in the system performance, in terms of both throughout per frame and packet loss rate.

Chapter 6
Conclusion

Capacity for MIMO wireless channels is currently an open research subject, investigated in many different scenarios. Great effort has been spent to describe analytically the MIMO channel capacity under different conditions and system geometries [15]. The main reason for this attention is the considerable capacity increase of these system, predicted in the pioneering work [4], supported later on by [5, 15]. Capacity, also described in terms of spectral efficiency, represents the theoretical maximum amount information which can be transmitted with asymptotically small probability of error. The capacity gains of MIMO systems is the main reason of the attention about this promising technology [3]. Important achievements accounts the description of MIMO channel capacity correlation and behaviour in time. Especially, it was shown that instantaneous capacity can be modeled as a Gaussian random process, for which the first two statistical moments are analytically described [18].

That great effort in characterizing the channel features cannot be exploited without a serious cross layer approach to the transmission problem [6]. The joint study of the MIMO physical layer and the network model has recently gained momentum and significant interest. This research work focused on the impacts of MIMO channel capacity on network performance, both for constant and adaptive rate transmission.

Firstly, it was shown that the random nature of the MIMO channel capacity introduces a probability of outage which translates into a probability of no reliable transmission. The capacity outage has a significant effect on the network performance, particularly on the probability of data loss. Considering a finite buffer system and a constant rate transmission, the signalling rate represents the amount of data that can be taken from the buffer and transmitted over the MIMO channel. For high signalling rates, the outage is the dominant effect in the data loss probability, while buffer overflow is the dominant effect for smaller signalling rates. A good trade-off between outage and buffer overflow probabilities is therefore necessary. Optimal signalling rates and system buffer sizes have been discussed using numerical examples.

According to the queue size and taking into account the MIMO PHY layer, it was shown that an optimal constant signalling rate can be chosen to minimize the total failure probability of the system.

The maximum limit of information that the channel can support is represented by the ergodic capacity. Ergodic capacity can be achieved when the transmission rate is adapted to the random evolution of channel capacity [2]. The target of achieving the ergodic channel capacity poses two main tasks: the evolution of the channel must be accurately tracked and this information must be made available at the transmitter. An accurate *channel state information* (CSI) can be obtained by training pilot symbols [37]. An appropriate estimation technique was chosen according to the specific characteristics of a flat-fading time-variant wireless channel, which were described with particular attention to the channel second-order statistic. The bandlimited characteristic of the fading spectrum suggests that it can be well expanded by a set of deterministic functions with a similar limited bandwidth. An appropriate basis expansion is found in *discrete prolate spheroidal sequences* (DPSS), firstly introduced in [33]. It was shown that DPSS are very accurate and efficient method of channel estimation with a reasonable number of pilots and isotropic scattering model. Realistic measurements shown instead that the scattering encountered real scenarios is non-isotropic. The case of non-isotropic scattering environment was considered and a modulated version of the DPSS basis set was introduced to improve the estimation performance of the original basis set. The members of the new basis set are obtained by modulation and bandwidth variation of the original DPSS functions in order to reflect various scattering scenarios. The results obtained by numerical simulations showed that the modulated DPSS method provides more accurate estimation than the DPSS scheme.

Finally, the possibility of adapting the signalling rate to the actual channel capacity and the buffer state has been presented. The knowledge of the correlation of the capacity process [12] allows to model its evolution by an autoregressive prediction model. Based on the prediction model, the transmitter can adapt its signalling strategy to the capacity evolution. Moreover, this analysis is not optimal without considering how many packet are in the buffer and if the buffer can experience overflow [8]. The joint consideration of the capacity evolution and the buffer state leads to a two-dimensional optimization problem. The possibility of modeling the MIMO channel capacity as a *finite state Markov chain* (FSMC) allows to refer to the theory

of *Markov decision processes* (MDP) to tackle this optimization problem. Considering the optimal case in which the transmitter can choose a proper transmission rate to match the capacity state, the MDP theory returns the optimal policy μ_0 which must be followed to maximize the process gain, defined as the number of packets correctly received at the destination MAC layer (the system throughput). Once the optimization problem is solved, the optimal policy μ_0 drives the choice of the most suitable rate according to the joint state of buffer and capacity. The performance of the optimal variable rate strategy has been compared with the constant rate transmission subject to outage. Numerical results showed a significant increase in the system performance, in terms of both throughput per frame and packet loss rate.

Future directions include the analysis of energy/delay constraints on the proposed work. Capacity when energy is limited assumes a completely different meaning since the number of bits (or amount of information) which each user can send is limited. This case becomes relevant in ad hoc networks, in which energy is seen as an important resource to be maximized. Also, a possible extension is the analysis of more realistic conditions, such as a fixed set of signalling rate at the transmitter. A fixed set of rates imposed by the modulation constellation would lead to a lower achievable spectral efficiency since rates cannot match exactly the channel capacity values. Although this represents a limitation, the optimality of the MDP algorithm is still guaranteed by the analytical framework: given a set of possible signalling rate, the rate decision will be optimal in those system specifications.

References

[1] J.Proakis, *Digital Communications*, 4th ed. New York: Mc Graw Hill, 2001.

[2] E.Biglieri, R.Calderbank, A.Constantinides, A.Goldsmith, A.Paulraj, and H.V.Poor, *MIMO Wireless Communications*. Cambridge, UK: Cambridge University Press, 2007.

[3] S.Loyka and G.Levin, "Diversity-Multiplexing Tradeoff via Asymptotic Analysis of Large MIMO Systems," *IEEE ISIT 2007*, June 2007, Nice, France.

[4] I.E.Telatar, "Capacity on Multi-Antenna Gaussian Channels," *European Trans. on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[5] B.Hochwald, T.Marzetta, and V.Tarokh, "Multiple-Antenna Channel Hardening and its Implications for Rate Feedback and Scheduling," *IEEE Trans. on Information Theory*, vol. 50, no. 9, pp. 1893 –1909, September 2004.

[6] V.K.N.Lau and Y.Kwok, *Channel Adaptive Technoligies and Cross-Layer Design for Wireless Systems*. Wiley Interscience, 2006.

[7] Q.Liu, Z.Shengli, and G.B.Giannakis, "Queuing with adaptive modulation and coding over wireless links: cross-layer analysis and design," *IEEE Trans. on Wireless Communications*, vol. 4, no. 3, pp. 1142–1153, May 2005.

[8] X.Bai and A.Shami, "A new perspective of cross-layer optimization for wireless communication over fading channel," *Proc. of 16th International Conference on Computer Communications and Networks*, pp. 1–6, 13-16 Aug. 2007.

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

[10] X.Lin, N.B.Shroff, and R.Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1452–1463, Aug. 2006.

[11] M. Niedzwiecki, *Identification of Time-varying Processes*. New York: John Wiley & Sons, Ltd, 2000.

[12] B.O.Hogstad, "Mulitple-input mulitple-output fading channel models and their capacity," Ph.D. dissertation, Aalborg University, Department of Communication Technology, Faculty of Engineering and Science, Aalborg University, Aalborg, Denmark, 2007.

[13] S.Verdú, "On channel capacity per unit cost," *IEEE Trans. on Information Theory*, pp. 1019–30, Sept. 1990.

[14] H.Mandyam and A.J.Goldsmith, "Capacity of finite energy channels," *Proc. Allerton Conf. Commun. Cntl. Comp.*, Oct. 2001.

[15] M.Debbah and R.R.Muller, "MIMO Channel Modeling and the Principle of Maximum Entropy," *IEEE Trans. on Information Theory*, vol. 51, no. 5, pp. 1667 –1690, May 2005.

[16] A.Paulraj, R.Nabar, and D.Gore, *Introduction to Space-Time Wireless Communications.* Cambridge, UK: Cambridge University Press, 2003.

[17] D.Slepian, "Prolate spheroidal wave functions, fourier analysis, and uncertainty - V: The discrete case," *Bell System Technology Journal*, vol. 57, pp. 1371–1430, May/Jun. 1978.

[18] A.Tulino and S.Verdú', "Random Matrix Theory and Wireless Communication," *Foundations and Trends in Communication and Information Theory*, vol. 1, pp. 1–182, 2004.

[19] A.Giorgetti, M.Chiani, M.Shafi, and P.Smith, "Level Crossing Rates and MIMO Capacity Fades: Impacts of Spatial/Temporal Channel Correlation," *IEEE Proc. ICC 2003*, 2003.

[20] S.Vaihunthan, S.Haykin, and M.Sellathurai, "MIMO Channel Capacity Modeling using Markov Models," *IEEE VTC 2005*, 2005.

[21] D. Bertsekas and R. Gallager, *Data Network*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, Inc., 1987.

[22] O.Brun and J.Garcia, "Analytical Solution of Finite Capacity M/D/1 Queues," *Journal of Applied Probability*, vol. 37, pp. 1092 –1098, 2000.

[23] J.M.Pitts and J.A.Schormans, *Introduction to ATM Design and Performance.* New York: John Wiley & Sons, Ltd, 1996.

[24] B.Hassibi and M.Hockwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. on Information Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[25] R.H.Clarke, "A statistical theory of mobile-radio reception," *Bell System Technology Journal*, vol. 47, pp. 957–1000, Jul./Aug. 1968.

[26] Y.R.Zheng and C.Xiao, "Simulation model with correct statistical properties for rayleigh fading channels," *IEEE Trans. on Communications*, vol. 51, no. 6, 2003.

[27] A.Abdi, J.A.Barger, and M.Kaveh, "A parametric model for the distribution of the angle of arrival and the associated correlation function and power spectrum at the mobile station," *IEEE Trans. on Vehicular Technology*, vol. 51, no. 3, pp. 425–434, May 2002.

[28] S.Primak, V.Kontrovitch, and V.Lyandres, *Stochastic Methods & their Applications to Communications: Stochastic Differential Equations Approach.* London: John Wiley & Sons, Inc., 2004.

[29] W.C.Jakes, *Microwave Mobile Communications.* New York: Wiley, 1974.

[30] K.E.Baddour and N.C.Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Trans. on Wireless Communications*, vol. 4, no. 4, pp. 1650–1662, July 2005.

[31] H. Trees, *Detection, Estimation and Modulation Theory.* John Wiley & Sons, Inc., 1967, vol. 1.

[32] C.Flammer, *Spheroidal Wave Functions.* Stanford, CA: Stanford University Press, 1957.

[33] T.Zemen and C.F.Mecklenbrauker, "Time-Variant Channel Estimation Using Discrete Prolate Spheroidal Sequences," *IEEE Trans. on Signal Processing*, vol. 53, no. 9, 2005.

[34] M.Biguesh and A.B.Gershman, "MIMO Channel Estimation: Optimal Training and Tradeoff Between Estimation Techniques," *IEEE Trans. on Signal Processing*, 2004.

[35] K.E.Baddour, C.C.Squires, and T.J.Willink, "Mobile channel prediction with application to transmitter antenna selection for alamouti systems," *Vehicular Technology Conference, 2006. VTC-2006 Fall. 2006 IEEE 64th*, pp. 1–6, Sept. 2006.

[36] E.Sejdic, M.Luccini, S.Primak, K.Baddour, and T.Willink, "Channel estimation using dpss based frames," *ICASSP 2008, Las Vegas*, Apr. 2008.

[37] C.Xiaodong and G.B.Giannakis, "Adaptive psam accounting for channel estimation and prediction errors," *IEEE Trans. on Wireless Communications*, vol. 4, no. 1, pp. 246–256, Jan. 2005.

[38] M.Zorzi, R.R.Rao, and L.B.Milstein, "A markov model for block errors on fading channels," *Seventh IEEE Inter. Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 3, no. 3, pp. 1074–1078, 15-18 Oct 1996.

[39] C.C.Tan and N.C.Beaulieu, "On first-order markov modeling for the rayleigh fading channel," *IEEE Trans. on Communications*, vol. 48, no. 12, pp. 2032–2040, Dec 2000.

[40] J. G. Kemeny and J. Snell, *Finite Markov Chains.* D. Van Nostrand Company Inc., 1959.

[41] R.A.Howard, *Dynamic Programming and Markov Processes.* Cambridge, MA: The M.I.T Press, 1960.

[42] E.Gilbert, "Capacity of a burst noise channel," *Bell System Technology Journal*, vol. 39, pp. 1253–1265, Sept. 1960.

[43] E.O.Elliott, "Estimates of error rates for codes on a burst-noise channel," *Bell System Technology Journal*, vol. 42, pp. 1977–1997, Sept. 1963.

[44] L.N.Kanal and A.R.K.Sastry, "Models for channels with memory and their applications to error control," *Proceedings of the IEEE*, vol. 66, no. 7, pp. 724–744, July 1978.

[45] H. S.Wang and N.Moayeri, "Finite-state markov channel-a useful model for radio communication channels," *IEEE Trans. on Vehicular Technology*, vol. 44, no. 1, pp. 163–171, Feb 1995.

[46] V. Subotic, "Error performance evaluation of wireless communication systems in the presence of non-rayleigh channel fading," Ph.D. dissertation, The University of Western Ontario, Department of Electrical and Computer Engineering, Faculty of Engineering, the University of Western Ontario, London, Ontario, Canada, 2005.

[47] A.M.Law and W.D.Kelton, *Simulation, Modeling and Analysis*, 3rd ed. New York: Mc Graw Hill, 1997.

Appendix A
Simulation of single server queuing system

A.1 Discrete time simulations

In this section a description of the algorithm used to simulate a single server queuing system is presented. The theoretical framework goes under the name of *discrete events simulation* and is deeply described in [47]. Given a single queue system, at time $t = 0$ the simulation starts with no packet in the system and with the server in idle state. The first arrival will happen at time A_1.The simulation stops when the n^{th} packet is served. The simulation time is not fixed but is dependent on the arrival and service processes, it can be considered a random variable. The algorithm follows the model of a discrete time simulation, in which the arrival and service processes must be specified: inter-arrival times follow a Poisson process with parameter λ and the service times for each packet can be either exponentially distributed (Poisson process with parameter μ) or deterministic [21]. In the last case the service time is equal to $T = 1/\mu$ for all the packets. The purpose of this simulation process is to provide 4 main statistical results, which are described below.

- Average delay in queue: for each packet the time interval between its arrival in the system and the actual starting of the service is recorded. The average delay in queue is given by the sum of all those time intervals divided by the number of packets served. Must be underlined that this information is subject to random variables different for each simulation (inter-arrival times and service duration), the expectation over a certain number of simulation is necessary to obtain a meaningful result.

- Average number of packets in queue: the average queue length is defined as

$$q(n) = \sum_{i=0}^{\infty} i p_i, \tag{A.1}$$

it is the product between the number of packets i and the proportion of time the queue spends at length i. Numerically the estimated average queue length is

$$\hat{q}(n) = \frac{\sum_i iT_i}{T(n)}, \tag{A.2}$$

where $T_i/T(n)$ is the observed proportion of time during the simulation in which there were i customers in the queue. The same expression can be written for the continuous time with the following result:

$$\hat{q}(n) = \frac{\int_0^{T(n)} Q(t)dt}{T(n)}. \tag{A.3}$$

- Probability of packet loss (buffer overflow) $P_N(N)$: it is the percentage of time the system experiences an overflow status over the total duration of the simulation. The time the system experiences overflow status is related to the event of buffer full and the arrival of a new packet in the system, which is rejected and lost. The probability of overflow is expressed by the discrete sum

$$p_{over} = \frac{T_{over}}{T(n)} \tag{A.4}$$

where $T(n)$ is the duration of the simulation and T_{over} is the time spent with the system full rejecting new arrivals.

A.2 The M/M/1 Queue: derivation of basic equations

The M/M/1 queuing model is based on two main assumptions. Firstly the packets arrival pattern from different sources can be approximated by negative exponential inter-arrival times. This is the same as saying that the arrivals are described by a Poisson process with a given parameter λ (arrival rate). The second assumption is that the service time associated with the packets is described by a negative exponential distribution with parameter μ (service rate). A good question is how large should be the buffer of the system. The M/M/1 model assumes infinite buffer space, but it is possible to derive some results by considering the number or packets in the system,

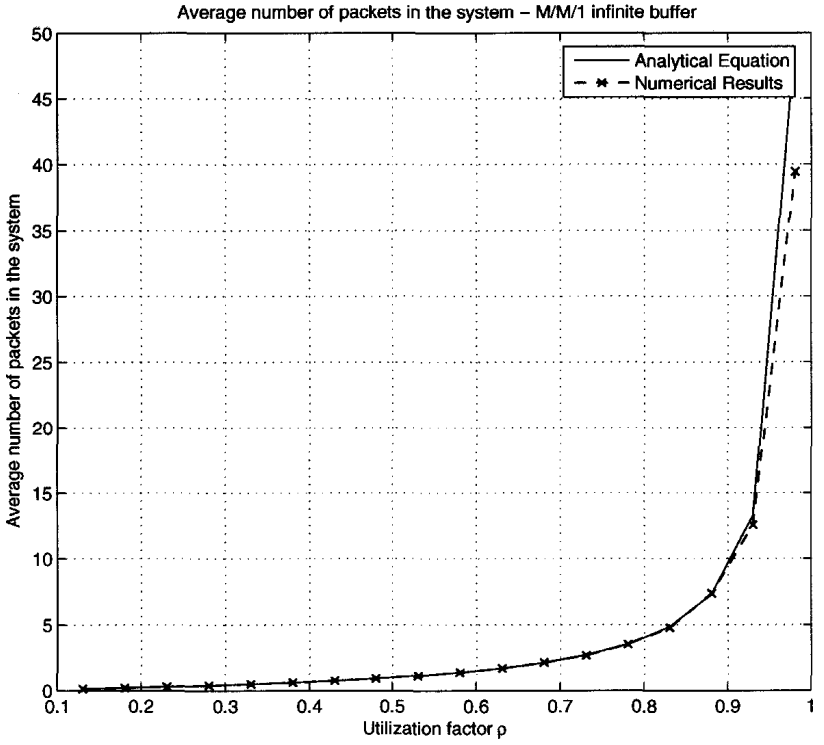

Figure A.1: Average number of packets in the system (in queue and under service) for the M/M/1 model. Comparison between the theoretical curve and the numerical simulation results. The discrepancy for high utilization ρ is due to the strong asymptotic behaviour of the theoretical curve, the numerical results could be closer to theory for a very high number of packets sent.

which for a M/M/1 system is given by [21]:

$$\overline{N}_{sys} = \frac{\rho}{1-\rho}, \tag{A.5}$$

where ρ is defined as the system utilization factor: it is the ratio between the arrival rate λ and the service rate μ. Looking at Figure A.1, which shows the average number of packets in the system against the utilization factor ρ, the key characteristic is the "knee" around $0.8 - 0.9$ ρ, which suggests that it is best to operate the system below the 0.8 utilization factor to avoid large queues building up. In order to have some ideas about the buffer size, the next step is to look at the probability distribution

of the number of packets in the system, which is given by [23]:

$$Pr\{n = x\} = (1 - \rho)\rho^x; \tag{A.6}$$

where n is the number of packets in the system. Eq. A.6 describes the probability of having a specific number of packets in the system. If the system has a finite buffer which can hold $N - 1$ packets, Eq. A.6 can be helpful to derive the probability of having a queue full, thus having N packets in the system (one under service and N-1 waiting in queue). This assumes that an infinite buffer model is a good model for a finite buffer, and that $Pr\{n = x\}$ is a reasonable approximation of the probability of loosing packets from a finite queue of size x. Authors of [23] worked with this assumption, deriving from the M/M/1 model the probability of exceeding a system size x, which is given by [23]:

$$Pr\{n > x\} = \rho^{x+1}, \tag{A.7}$$

If the system has a finite queue of length N packets, the $M/M/1$ model looses the assumption of infinite buffer and turns into the $M/M/1/N$ case. Equation A.7 is obtained by the sum of the probability of having more than x packets in the system. To confirm the reliability of the M/M/1 model for finite queue length, Figure A.2 shows the probability of packet loss for different utilization factor ρ. The solid curve is the plot of Equation A.6, the dotted curve is the numerical result of a finite queue system, where the packets are lost if the queue is full. The dimension of the queue (or buffer) is set to 10 packets. In the simulation, the probability of packet loss is obtained by checking for how much time the system experiences an overflow status compared to the total simulation time. The results show a good matching between the curves, the oscillation at small ρ is due to the finite number of packets sent in the system: for small utilization factor the probability of overflow is very little and difficult to be identified exactly with a small number of simulations. The results in Figure A.2 are confirmed by [23], which obtained similar curves for the estimated probability of packets loss from M/M/1 model and the results for a finite queue system M/M/1/N.

The results in Fig.A.2 are achieved averaging 10 trials with different Poisson arrivals and 10^6 packets sent in each trial.

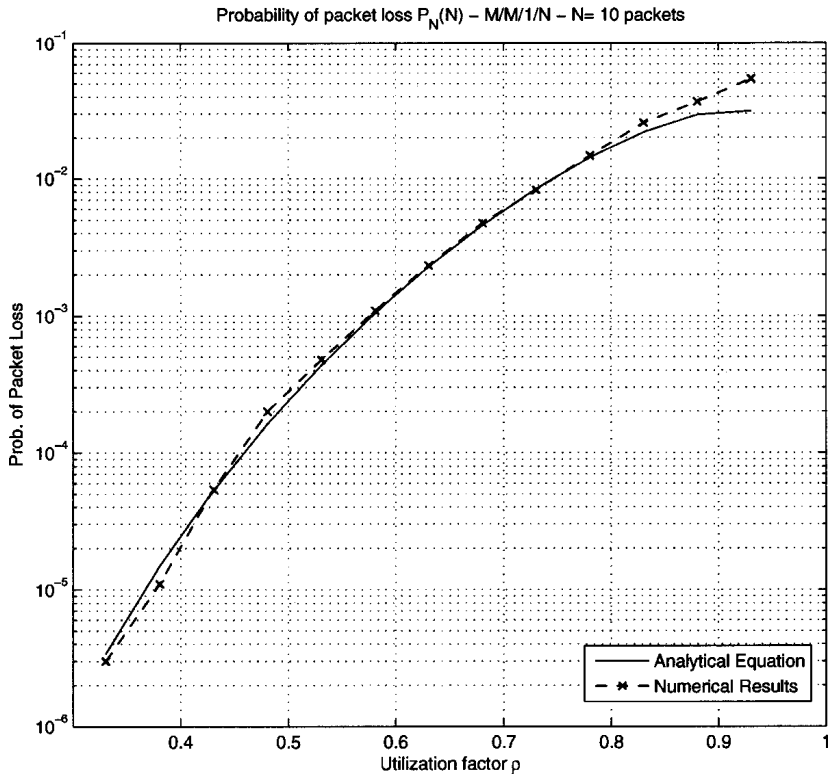

Figure A.2: Probability of packet loss due to the overflow of a finite buffer. Comparison between the estimated probability from a M/M/1 model and the actual results of numerical simulations. The queue size (buffer) is 10 packets.

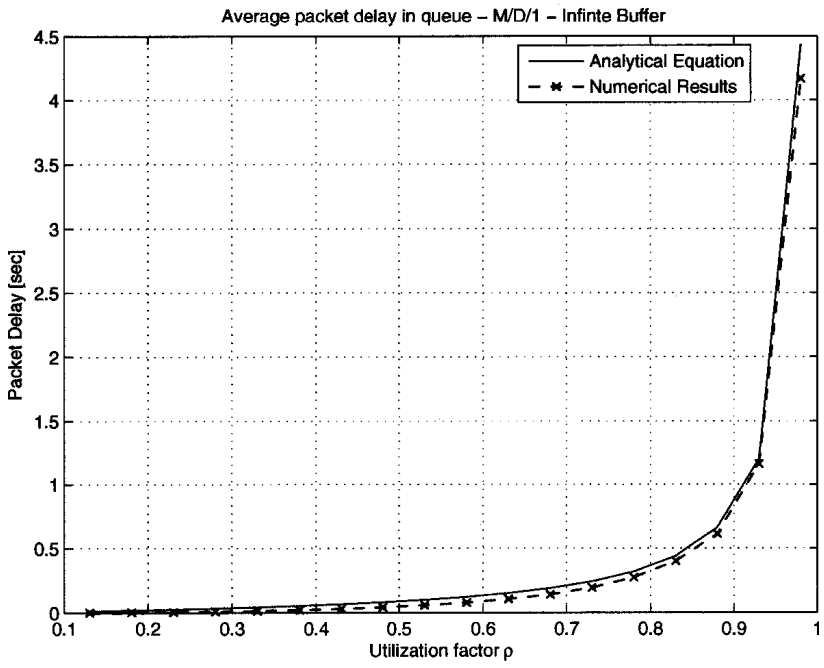

Figure A.3: Average delay in queue for M/D/1 system. For comparison purposes, the buffer size is assumed infinite.

A.3 M/D/1 queue model

In this section the main results and extensions for an $M/D/1$ model are presented. In the case of $M/D/1$ queuing model, the arrival assumptions are the same as for $M/M/1$, while the service time in this case is deterministic. It is the case of a system in which the service time for different users is constant: constant time slots are available for each user in the system. In many communication systems, the physical layer is based on a fixed frame duration, from this point of view the behaviour of the system can be compared to a M/D/1 model. The deterministic service is described by the service rate μ and the service time $T = 1/\mu$. First of all, a validation of the numerical simulations according to the theory of the M/D/1 model is needed. Figure A.3 shows the average delay per packet in queue under the assumption of infinite buffer. The analytical curve is derived using the Pollaczek - Khinchin equation [21], adapted for the particular case of deterministic service. Defining W the average waiting time in the queue, it can be expressed as function of the arrival rate λ, the utilization factor $\rho = \lambda/\mu$ and the second moment of the service time, which for a deterministic process

is $\overline{X}^2 = 1/\mu^2$ [21]. The result, know as Pollaczek - Khinchin formula is:

$$W = \frac{\lambda \overline{X^2}}{2(1-\rho)} \tag{A.8}$$

From the average waiting time in queue, the average number of packets in the queue can be derived by the Little's theorem [21]. For an infinite buffer, the average number of packets in queue N_q is given by the product between the arrival rate and the average waiting time: $N_q = \lambda W$. The average number of packets in the system N_{sys} is related to the average number in queue through the utilization factor ρ [21]:

$$Nsys = \rho + \frac{\lambda^2 \overline{X^2}}{2(1-\rho)} \tag{A.9}$$

where N_{sys} is the average number of packets in the system, $\overline{X^2}$ the second moment of service time, λ and ρ respectively the arrival rate and utilization factor. The simulation results for the $M/D/1$ model with infinite buffer matches very well the analytical equation derived above.

Considering now a finite buffer size N, we are interested in deriving the packet loss probability studying the system size distribution, following the same procedure presented for the $M/M/1$ case. The first step will be to present the analytical solution for the probability of having a certain number of packets in the system, from which the probability of packets loss can be derived. The model we will refer to is called $M/D/1/N$, underlining the finite queue length of N packets. When the system holds N packets, the queue is full and an eventual new arrival will be lost and will not influence the system state. The main result we we would like to derive is the probability $P_N(N)$ of having N packets in the system, which corresponds to the probability of having the queue full. As for the case of $M/M/1/N$, that probability can be interpreted as the probability of packet loss in the system due to the buffer overflow. The analytical solution for the $M/D/1/N$ system is presented in the next section.

A.4 M/D/1/N: Analytical solution

A.4.1 Derivation of steady state probabilities

The $M/D/1/N$ model is a finite capacity queuing system, with a queue length of $N-1$ packets. The maximum number of packets the system can hold is N: $N-1$ in queue, 1 under service. Packets arrive according to a Poisson process with parameter λ packets/s. λ is called arrival rate in the system. By model assumption, packets which upon arrival see a full buffer are rejected and do not further influence the system. An overflow event happens when a packet is rejected because of full buffer. In this model the service rate μ in packets/s is constant. Since the packet length in bits is fixed, the time needed to serve a packet is constant $T = 1/\mu$. The utilization factor $\rho = \lambda/\mu$ defines the service capability of the system, a general assumption is that a stable system is described by $\rho < 1$. Our interest is in a computational scheme for the average state probabilities $P_j(N)$, which represents the probability of having j packets in the system given a queue size of N packets. According to [22] a description of the transition probabilities can be computed by the z-transform of the transition matrix for the M/D/1/N system. The most important result are a set of coefficients b_n, defined as [22]:

$$b_n = \sum_{k=0}^{n} \frac{(-1)^k}{k!}(n-k)^k e^{(n-k)\rho}\rho^k \tag{A.10}$$

The probability distribution of the number of customer in the system is then derived as [22]:

$$P_0(N) = \frac{1}{1+\rho b_{N-1}} \tag{A.11}$$

$$P_N(N) = 1 - \frac{b_{N-1}}{1+\rho b_{N-1}} \tag{A.12}$$

$$P_j(N) = \frac{b_j - b_{j-1}}{1+\rho b_{N-1}} \qquad j = 1, ..., N-1; \tag{A.13}$$

The previous equations show the steady state results in terms of state probabilities: $P_j(N)$ represents the probability of having j packets in the system for a queue size of N packets. The probability $P_N(N)$ describes the probability of a queue full: when the system holds N packets a new arrival will be lost and overflow will occur. In [23] is proved that the probability $P_N(N)$ can be considered as a good description of the

packet loss probability due to the buffer overflow. This consideration has been used to derive the probability of overflow for a M/M/1/N system in Sec. A.2. The mean number of packets in the M/D/1/N queue is derived as [22]:

$$X_N = N - \frac{\sum_{k=0}^{N-1} b_k}{1 + \rho b_{N-1}} \tag{A.14}$$

where X_N is the average number of packets in the queue. The relation between X_N and the average delay in the system T_N is given by the Little's Theorem:

$$X_N = \lambda(1 - P_N(N))T_N; \tag{A.15}$$

in which the probability of having the queue full becomes relevant for high values of the utilization factor ρ. Finally, the average waiting time in queue is defined as the difference between the average system time T_N and the constant service time T, with the following closed equation [22]:

$$W_N = T_N - T = (N - 1 - \frac{\sum_{k=0}^{N-1} b_k - N}{\rho b_{N-1}})T. \tag{A.16}$$

W_N is the average packet delay in queue for the $M/D/1/N$ system.

A.4.2 Results

Figure A.4 shows the analytical probability $P_N(N)$ of a full queue of N packets compared with the probability of packet loss derived by numerical simulations. Despite of the little distance between the analytical queue probability distribution and the numerical results, the two curves shows the same trend. The reason for that discrepancy is the little number of independent simulations averaged. In the simulation process, the probability of packet loss is computed by measuring the time the system spends in the state N compared to the total simulation time, which means for how much time the system is full compared with the total simulation time. As the curves shows the same behaviour, the analytical $P_N(N)$ can be considered as a good upper bound for the packet loss probability. This means that the analytical results of [22] can be used to derive the system behaviour for different service rates. Figure A.5 shows the average number of packets in the system for the closed form presented in [22] and the

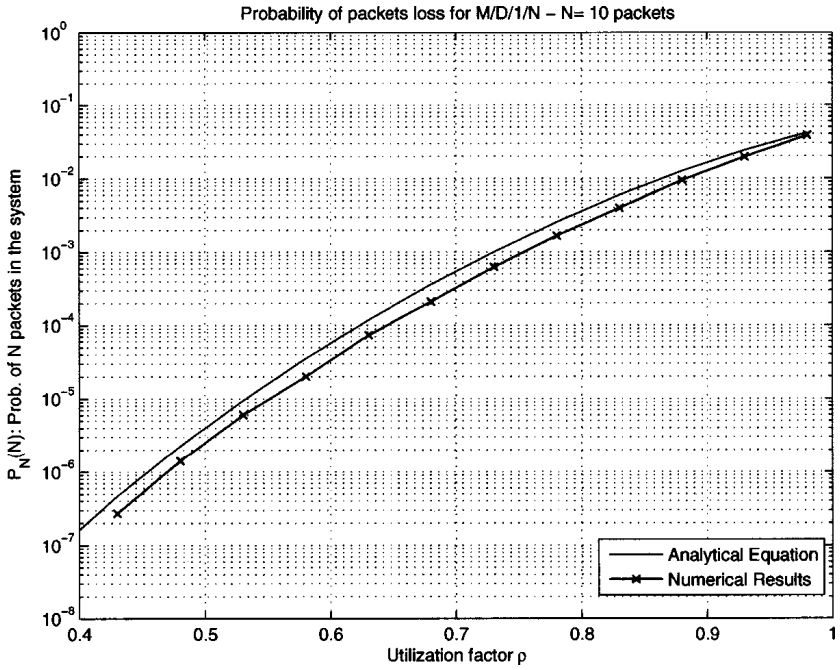

Figure A.4: Comparison between the analytical queue size distribution probability $P_N(N)$ and the numerical probability of packet loss. $P_N(N)$ is the probability of having the system queue full, the packet loss probability is computed by the number of packets lost due to the queue full.

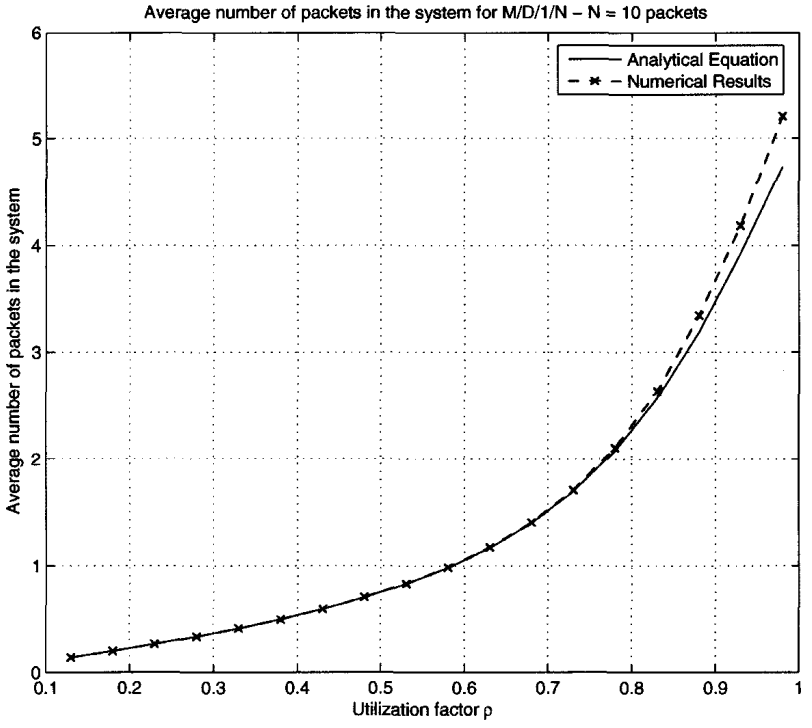

Figure A.5: Average number of customers in the system: comparison between the closed form presented in [22] and the numerical results. The match between the two curves is high, validating the closed equations presented.

numerical results of simulations. For small ρ the number of customers in the system can be less than 1 according to the arrival Poisson process: there might be a time interval in which the system is empty. In [22] the average number of customers in the system is computed by solving the following equation:

$$X_N = \sum_{k=0}^{N} kP_k(N) \tag{A.17}$$

which is a normal average operation for all the possible number of packets in the system. The final result is [22]:

$$X_N = N - \frac{\sum_{k=0}^{N-1} b_k}{1 + \rho bN - 1}, \tag{A.18}$$

where X_N is the average number of packets in the system and N is the maximum number of packets the system can hold. As clear from the figure, the behaviour of Equation A.18 follows the numerical results of the average number of packets in the system. This is an important conclusion since it means that the model presented in [22] is valid, with the possibility of using the closed equations presented as a good behaviour model for the M/D/1/N system. Given the average number of customers in the system, the average amount of time spent in the system and in queue can be derived by applying the Little's theorem. For a buffer size of $N - 1$ packets, the application of the Little's theorem yields:

$$X_N = \lambda(1 - P_N(N))T_N \tag{A.19}$$

where X_N is the average number of packets in the system, $P_N(N)$ is the probability of having the queue full and T_N is the average time spent in the system per packet. The term $P_N(N)$ becomes relevant for high utilization factor ρ, when the probability of having the buffer full is significant. From a simple inversion, the average time spent in the system per packet T_N is given by:

$$T_N = \frac{X_N}{\lambda(1 - P_N(N))}; \tag{A.20}$$

the time spent in queue per packet W_N can be derived by the previous equation by taking into consideration the constant service time T:

$$W_N = T_N - T \tag{A.21}$$

where $T = \rho/\lambda = 1/\mu$.

Appendix B
Estimation by Pilot Symbols for a General Transmission Model

In this section we describe the *Least Square* estimation for the series coefficients in the context of channel basis expansion. The following algorithm is an extension to the *MIMO* case of the one described in ([11], Ch. VI) for *SISO* systems. Let consider a frame of N samples in a general $N_R \times N_T$ MIMO channel. P symbols of the frame are used as pilots at the moments $1 \leq t_1 < t_2 < ... < t_{Np} \leq N$, $T = \{t_j\}$. If the transfer function of the channel at time $t \in T$ is $\mathbf{H}(t)$ and the pilot signal sent is $\mathbf{p}(t)$, then the received signal is given by:

$$\mathbf{r}(t) = \mathbf{H}(t)\mathbf{p}(t) + \mathbf{z}(t), \tag{B.1}$$

with the gaussian noise $\mathbf{z}(t)$. Using the vectorial notation, eq. B.1 can be rewritten as

$$\mathbf{r}(t) = [\mathbf{I}_{N_R} \otimes \mathbf{p}(t)]\mathbf{h}(t) + \mathbf{z}(t), \tag{B.2}$$

where $\mathbf{h}(t) = vec(\mathbf{H})$. Using the basis expansion notation, the channel vector can be expanded in the form $\mathbf{h}(t) = \mathbf{C}\mathbf{e}(t)$, where $\mathbf{e}(t)$ is the vector of basis function with size $D \times 1$ and C is the matrix of coefficients with size $N_R N_T \times D$. Then Eq. B.2 can be expanded in

$$\mathbf{r}(t) = [\mathbf{I}_{N_R} \otimes \mathbf{p}(t)]\mathbf{C}\mathbf{e}(t) + \mathbf{z}(t), \tag{B.3}$$

and, using vectorization, in

$$\mathbf{r}(t) = [\mathbf{I}_{N_R} \otimes \mathbf{p}(t) \otimes \mathbf{e}(t)]\mathbf{c} + \mathbf{z}(t), \tag{B.4}$$

with $\mathbf{c} = vec(\mathbf{C})$. Performing the *Least Square* estimation of vector $\mathbf{c}$ in the general system $\mathbf{r}(t) = \mathbf{Ac} + \mathbf{z}$, with $\mathbf{A} = [\mathbf{I}_{N_R} \otimes \mathbf{p}(t) \otimes \mathbf{e}(t)]$, we obtain:

$$\hat{\mathbf{c}} = (\sum_{t \in T} [\mathbf{I}_{N_R} \otimes \mathbf{p}(t) \otimes \mathbf{e}(t)][\mathbf{I}_{N_R} \otimes \mathbf{p}(t) \otimes \mathbf{e}(t)]^H)^{-1} (\sum_{t \in T} [\mathbf{I}_{N_R} \otimes \mathbf{p}(t) \otimes \mathbf{e}(t)]^H \mathbf{r}(t)) \quad \text{(B.5)}$$

(for all the passages refer to [11], Ch. VI). Using the *Kronecker* products, eq. B.5 can be simplified the following equation:

$$\hat{\mathbf{c}} = \mathbf{I}_{N_R} \otimes (\sum_{t \in T} [\mathbf{p}(t)\mathbf{p}(t)^H \otimes \mathbf{e}(t)\mathbf{e}(t)^H]^{-1} (\mathbf{I}_{N_R} \otimes \sum_{t \in T} [\mathbf{p}^H(t) \otimes \mathbf{e}^H(t)]\mathbf{r}(t)). \quad \text{(B.6)}$$

This final equation, in the case of a *SISO* system, can be simplified assuming $N_R, N_T = 1$. Finally, assuming also that the pilot sequences are orthonormal, the product $\mathbf{p}(t)\mathbf{p}(t)^H$ becomes $|p(t)|^2 = 1$, obtaining Eq. 4.30 in chapter 4.