Universidade de Aveiro
2013

Departamento de Electrónica,
Telecomunicações e Informática

Samah A. M. Ghanem

Analysis, Modeling, Design and Optimization of Future Communication Systems: From Theory to Practice

Universidade de Aveiro
2013

Departamento de Electrónica,
Telecomunicações e Informática

Samah A. M. Ghanem

Analysis, Modeling, Design and Optimization of Future Communication Systems: From Theory to Practice

Dissertação apresentada às Universidades de Minho, Aveiro e Porto para cumprimento dos requisitos necessàrios à obtenção do grau de Doutor no âmbito do doutoramento conjunto MAP-Tele, realizada sob a orientação científica do Doutor Paulo Jorge dos Santos Conçalves Ferreira, Professor Catedrático do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro, e do Doutor José Manuel Tavares Vieira Cabral, Professor Auxiliar do Departamento de Electrónica Industrial da Universidade do Minho.

Apoio financeiro da FCT bolsa de doutoramento: SFRH / BD / 68351 / 2010

To my family:
Who were always beside me
in the hardest times of my life

President: Dr. Armando da Costa Duarte
Full Professor at the University of Aveiro

Jury:

Dr. Izzat Darwazeh
Full Professor at University College London

Dr. José Manuel Bioucas Dias
Associate Professor at the Technical University of Lisbon

Dr. José Manuel Neto Vieira
Associate Professor at the University of Aveiro

Dr. Henrique Manuel de Castro Faria Salgado
Associate Professor at the University of Porto

Dr. Paulo Jorge dos Santos Gonçalves Ferreira
Full Professor at the University of Aveiro

Acknowledgement

First and foremost, I would like to thank Prof. Paulo J. S. G. Ferreira and Prof. Jose Cabral for holding the responsibility of my supervision and for their support that let me reach the final stage of my PhD successfully.
I am very grateful to Prof. Miguel R. D. Rodrigues and Prof. Daniel E. Lucani, who introduced me to core tools in the work and for fruitful discussions.
I am very grateful to Prof. Izzat Darwazeh (from University College London) who was always there for me when I require any advise and for his valuable feedback about my thesis writing.
I am very grateful to my friend and collaborator Munnujahan Ara who was always there for me and who introduced me to optimization theory.
Thanks to Gil Ramos who introduced me to matrix differentiation.
I am grateful to Prof. Atilio Gamerio and Prof. Adao Silva who introduced me to the cooperation concept.
I am thankful to the scientific committee of MAP-Tele program. In particular, I would like to express my special gratitude to Prof. Henrique Salgado and Prof. Anibal Ferreira for their help at the last year of my PhD.
I would like also to thank Prof. Frank Fitzek who invited me for three months within his research group at Aalborg University. Thanks also to all his team who were very friendly during my stay in Denmark.
Thanks to Alexandra Ferreira, who always shows professionality and support at all levels of administrative work during my stay at FCUP.
I would also like to thank the foundation of science and technology (FCT) in Portugal for providing the financial support during the course of my PhD.
Finally, I am very grateful to all my family for their patience and support. Without them, this work would never have come into existence.

Samah A. M. Ghanem
May 2013

Resumo

Esta tese descreve uma framework de trabalho assente no paradigma multi-camada para analisar, modelar, projectar e optimizar sistemas de comunicação. Nela se explora uma nova perspectiva acerca da camada física que nasce das relações entre a teoria de informação, estimação, métodos probabilísticos, teoria da comunicação e codificação. Esta framework conduz a métodos de projecto para a próxima geração de sistemas de comunicação de alto débito. Além disso, a tese explora várias técnicas de camada de acesso com base na relação entre atraso e débito para o projeto de redes sem fio tolerantes a atrasos. Alguns resultados fundamentais sobre a interação entre a teoria da informação e teoria da estimação conduzem a propostas de um paradigma alternativo para a análise, projecto e optimização de sistemas de comunicação. Com base em estudos sobre a relação entre a informação recíproca e MMSE, a abordagem descrita na tese permite ultrapassar, de forma inovadora, as dificuldades inerentes à optimização das taxas de transmissão de informação confiáveis em sistemas de comunicação, e permite a exploração da atribuição óptima de potência e estruturas óptimas de pre-codificação para diferentes modelos de canal: com fios, sem fios e ópticos. A tese aborda também o problema do atraso, numa tentativa de responder a questões levantadas pela enorme procura de débitos elevados em sistemas de comunicação. Isso é feito através da proposta de novos modelos para sistemas com codificação de rede (network coding) em camadas acima da sua camada física. Em particular, aborda-se a utilização de sistemas de codificação em rede para canais que variam no tempo e são sensíveis a atrasos. Isso foi demonstrado através da proposta de um novo modelo e esquema adaptativo, cujos algoritmos foram aplicados a sistemas sem fios com desvanecimento (fading) complexo, de que são exemplos os sistemas de comunicação via satélite. A tese aborda ainda o uso de sistemas de codificação de rede em cenários de transferência (handover) exigentes. Isso é feito através da proposta de novos modelos de transmissão WiFi IEEE 801.11 MAC, que são comparados com codificação de rede, e que se demonstram possibilitar transferência sem descontinuidades. Pode assim dizer-se que esta tese, através de trabalho de análise e de propostas suportadas por simulações, defende que na concepção de sistemas de comunicação se devem considerar estratégias de transmissão e codificação que sejam não só próximas da capacidade dos canais, mas também tolerantes a atrasos, e que tais estratégias têm de ser concebidas tendo em vista características do canal e a camada física.

Abstract

This thesis provides a novel framework based on multi-layer paradigms, to analyze, model, design, and optimize future communications systems. The thesis exploits a novel physical layer framework based on the interplay between information theory, estimation theory, probability methods, communication theory and coding techniques. This framework leads to proposals of design methods for next generation-very high data rate communication systems. In addition, this thesis exploits multiple access layer techniques based on the interplay between delay and throughput performance for the design of wireless delay tolerant networks. Building upon foundational work on the interplay between information theory and estimation theory, leading to proposals of an alternative paradigm for the analysis, design and optimization of communication systems. Based on new studies of the relation between the mutual information and the MMSE, this approach overcomes, in an innovative way, the inherent difficulties concerning the optimization of the reliable information transmission rates of key communication systems and allows the exploitation of the optimal power allocation and optimal precoding structures for different wireless and wired/optical channel models. The thesis also addresses the delay problem, to accommodate the huge demand of high data rates in communication systems. This is done through the proposal of new models and designs for systems that adopt network coding mechanisms in layers above its physical layer. In particular, this research work addresses the usage of network coding schemes for time varying channels that are delay sensitive. This was demonstrated via the proposal of a new model that has led to a novel adaptive scheme, which in turn is implemented in algorithms and applied to wireless systems with complex fading behavior, such as satellite systems. Furthermore, the thesis addresses the usage of network coding schemes in mobile and delay stringent handover scenarios. This is done through the proposal of novel models of the WiFi IEEE 802.11 MAC layer transmission techniques which are evaluated in comparison with network coding scheme, and demonstrated by solutions to novel formulations that allow seamless handover of the mobile station. Therefore, this thesis, through analytical work and design proposals backed by modeling and simulation studies establishes the fact that future telecommunication system designs should consider transmission and coding strategies that are not only capacity achieving, but also delay tolerant and that such design have to be done with care and awareness of the physical layer and the communication channel behavior.

Contents

List of Figures

List of Tables

Notation

The following notation is employed, boldface uppercase letters denote matrices, boldface lowercase letters denote vectors, and lowercase letters denote scalars.
The superscripts,

$\mathbf{I}$	The identity matrix
$\mathbf{X^{-1}}$	The inverse of matrix $\mathbf{X}$
$\mathbf{X^{T}}$	The transpose of matrix $\mathbf{X}$
$\mathbf{X}^{*}$	The conjugate of matrix $\mathbf{X}$
$\mathbf{X}^{\dagger}$	The conjugate transpose operations (Hermitian) of $\mathbf{X}$
$\|\mathbf{X}\|$	The determinant of a matrix
$diag(\mathbf{X})$	The diagonal elements of a matrix $\mathbf{X}$
$(\mathbf{X})_{ij}$	The (i,j)th elements of a matrix $\mathbf{X}$
$vec\,(\mathbf{X})$	The vector operation over a matrix $\mathbf{X}$
$\mathbf{X} \succ \mathbf{Z}$	$\mathbf{X} - \mathbf{Z}$ is positive definite
$\mathbf{X} \succeq \mathbf{Z}$	$\mathbf{X} - \mathbf{Z}$ is positive semidefinite
$\nabla_{\mathbf{X}}$	The gradient of a scalar function with respect to $\mathbf{X}$
$D_{\mathbf{X}}$	The differential of a scalar function with respect to $\mathbf{X}$
$\otimes$	The Kronecker product between two matrices
$\|\|.\|\|$	The norm
$\|\|.\|\|_2$	The Euclidean norm of a vector
$\|\|.\|\|_F$	The Forbenius norm of a matrix
$Tr\,\{.\}$	The trace of a matrix
$(.)^{\star}$	The optimum value
$(.)^{+}$	A positive value
C^{nxm}	The set of complex numbers of dimensions nxm
$log\,(.)$	The natural logarithm
$log_b\,(.)$	The logarithm in base b

$\mathbb{E}[.]$	The expectation operator
$\mathcal{R}(.)$	The real part
$a \in b$	a belongs to b
$a > b$	a is greater than b
$a < b$	a is less than b
$a \geq b$	a is greater than or equal b
$a \leq b$	a is less than or equal b
$\frac{\partial u}{\partial t}$	The derivative of u with respect to t
$a \to 0$	The limit of a approaches zero
$a \to \infty$	The limit of a approaches ∞
$a \longleftarrow c$	Replace the value of a by the new value of c
$\forall i$	For all values of i belonging to the defined set
$\sim$	Distributed according
$\mathcal{N}(\mu, \sigma^2)$	Gaussian vector distribution with mean μ and variance σ^2
$\mathcal{CN}(\mu, \Sigma)$	Complex Gaussian vector distribution with mean μ and covariance Σ
$log - \mathcal{N}(\mu, \sigma^2)$	log-normally distributed Gaussian vector with mean μ and variance σ^2

Abbreviations

ACK	Acknowledgement
ACK/NACK	Acknowledged/Not Acknowledged
AP	Access Point
AR	Auto-regressive
ARQ	Automatic Repeat reQuest
AWGN	Additive White Gaussian Noise
BC	Broadcast Channel
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
BS	Base Station
CANCO	Context-Aware Network Coding
CDMA	Code Division Multiple Access
CSI	Channel State Information
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CW	Contention Window
dB	Decibels
DCF	Distributed Coordination Function
DIFS	Distributed InterFrame Space
DL	Downlink
DoF	Degrees of Freedom
DSL	Digital Subscriber Line
DTN	Delay Tolerant Network
EDCF	Enhanced DCF
FDM	Frequency Division Multiplexing
FEC	Forward Error Correcting
FHMIP	Fast Handover Mobile IP
FiWi	Fiber-Wireless

GF	Galois Field
HARQ	Hybrid Automatic Repeat Request
HMIP	Hierarchical Mobile IP
IA	Interference Alignment
IC	Interference Channel
IP	Internet Protocol
KKT	Karush Kuhn Tucker
LAN	Local Area Network
LDPC	Low-Density Parity-Check
LP	Linear Program
LTE	Long Term Evolution
MAC	Multiple Access Channel
MCP	Multi-Cell Processing
MGF	Moment Generating Function
MIHF	Media Independent Handover Function
MIMO	Multi-Input-Multi-Output
MMSE	Minimum Mean Squared Error
MPDU	Message Packet Data Unit
MSDU	Message Service Data Unit
MRC	Maximal Ratio Combining
MRNC	MAC layer Random Network Coding
MSE	Mean Squared Error
MSNC	MAC Layer Systematic Network Coding
NC	Network Coding
NP-hard	Non-deterministic Polynomial-time hard
OFDM	Orthogonal Frequency Division Multiplexing
OSI	Open Systems Interconnection
PCF	Point Coordinated Function
PDP	Power Delay Profile
PSD	Power Spectral Density
PSK	Phase Shift Keying
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
RA	Reciprocal Alignment
RTS/CTS	Request to Send/Clear to Send
RTT	Round Trip Time

SIFS	Short InterFrame Space
SINR	Signal-to-Interference-Noise Ratio
SIP	Session Initiation Protocol
SISO	Single-Input-Single-Output
SNR	Signal-to-Noise Ratio
SR	Selective Repeat
SVD	Singular Value Decomposition
TCP	Transmission Control Protocol
TDD	Time Division Duplexing
TDM	Time Division Multiplexing
UL	Uplink
UT	User Terminal
UWB	Ultra Wide Band
WiFi	Wireless Fidelity
WiMax	Worldwide Interoperability for Microwave Access
WLAN	Wireless LAN

Chapter 1
Introduction

1.1 Motivation

Wireless communications have experienced an explosive growth with the demand for higher data rates to accommodate new services growing rapidly. A number of wireless systems have been gradually introduced including novel mobile and fixed wireless systems, e.g., second-generation and third-generation mobile systems as well as WiFi IEEE 802.11 and WiMax IEEE 802.16. This wireless communications revolution is being fueled by constant technological breakthroughs to enhance transmission performance in the wireless setting. The optimization of the reliable information transmission rate "the Holy Grail in communications" is still an open issue at large of paramount importance in communications systems research and practice, offering the prospect of communication at ever increasing information rates. On the one hand, a challenge towards approaching this goal is to design efficient transmission techniques that can increase the transmission rates. On the other hand, a tradeoff between the achievable rates and the delay in delay sensitive applications is another design challenge. The major difficulty relates to the absence of appropriate mathematical models, tools, and methodologies, due to the presence of complex features in communications systems.

Motivated by modular approaches like the OSI and TCP/IP that model the communication framework, we developed a novel framework for future communication systems with high data rate and delay tolerance, based on multi-layer paradigms. Traditionally, power control and precoding are viewed as features addressed in the physical layer, and all widely used medium access schemes are addressed in the link layer assuming the physical layer has already been established and provides an error-free link. The purpose of the physical-layer is to guarantee error-free transmission, however, state-of-the-art medium access schemes are insensitive to the characteristics

of the communications channels. Despite the fact that these modular approaches are widely accepted, a lot of enhancement layers can be added on top of existing layers and different schemes within the existing layers can be improved. Moreover, imposing some computational loads at the receiver side to the transmitter can be of particular relevance to allow an end to end performance gains. Therefore, we develop keen interest in introducing a framework for communication systems that, from the one hand, introduces physical layer transmission schemes by designing optimal power allocation and optimal precoding transmission strategies, and on the other hand, dealing with the delay problem on upper layers of the protocol stack. In particular, we develop interest on revisiting the linear precoding problem to maximize the data rates, and revisiting multiple access mechanisms on current wireless communication systems to propose channel-aware schemes, i.e., physical layer aware schemes on the upper layers which can improve current state of the art transmission technologies for delay tolerant networks. In addition to studying the impact of a network coding layer in improving the delay-throughput performance of the wireless network.

In this thesis, a study of a set of problems that has direct impact on the reliable achievable rates and can improve the delay performance has been established. Particularly, we address the design of linear precoders that can increase the achievable reliable data rates in a paradigm capitalizing on the interplay between information theory and estimation theory and using matrix linear algebraic techniques. In addition, we address network coding schemes which inherently include the diversity concept, through which the network coding part introduces a coding strategy that allows moderate and efficient use of the transmission resources with delay performance and diversity gains. Therefore, our second paradigm is based on developing transmission techniques that can improve the delay performance capitalizing on the interplay between the delay and throughput and using probabilistic modeling techniques.

1.2 Thesis Organization

In this thesis, we consider transmission schemes that aim to increase the reliable data rates over MAC Gaussian channels, MIMO Gaussian channels, MAC Gaussian fading channels, and MAC Poisson channels. In particular, we investigate the optimal power allocation and optimal linear precoding for the design criterion defined which is the maximization of mutual information and with channel setups created via

cooperation. The physical layer information-theoretic estimation-theoretic approach stands as the first paradigm in solving the first problem in this thesis. However, we consider transmission schemes for efficiently decreasing the delay and so increasing the throughput as a counter part of data rates when dealing with layers above the physical layer. In particular, we investigate the usage of network coding mechanisms for practical systems like WiFi and satellite communications. The multiple access layer for applications with delay constraints requires probabilistic approaches to be integrated with communication theoretic and coding approaches. This stands as the second paradigm in solving the second problem in this thesis.

In Chapter 2, we give a brief overview of information-theoretic estimation-theoretic approaches at the physical-layer as well as prior work in the area of optimal precoding and power allocation and its relation to the work in this thesis.

In Chapter 3, we give a brief overview of network coding approaches at the multiple layers as well as prior work in the area of delay probabilistic models and its relation to the work in this thesis.

In Chapter 4, we derive closed-form expressions for the gradient of the mutual information with respect to arbitrary parameters of the two-user multiple-access channel (MAC). We investigate the linear precoding and power allocation policies that maximize the mutual information for the two-user MAC Gaussian channels with arbitrary input distributions, capitalizing on the relationship between mutual information and minimum mean-square error (MMSE). Our new derivation verifies the fundamental relations between information theory and estimation theory. We show that the optimal design of linear precoders may satisfy a fixed-point equation as a function of the channel and the input constellation under specific setups. We provide an interpretation for the interference with respect to the channel, power, and input estimates of the main user and the interferer.

In Chapter 5, we investigate the optimal power allocation and optimal precoding for a cluster of two base stations (BS) which cooperate to jointly maximize the achievable rate for two users connecting to each BS in a multi-cell processing framework. We provide a generalized fixed point equation of the optimal precoder in the asymptotic regimes of the low- and high-snr. We introduce a new iterative approach that leads to a closed-form expression for the optimal precoding matrix in the high-snr regime which is known to be an NP-hard problem. Two MCP distributed algorithms have been introduced: a power allocation algorithm for the uplink (UL), and a precoding algorithm for the downlink (DL).

In Chapter 6, we use the results of Chapter 4 and Chapter 5 to investigate the optimal power allocation and optimal precoding for a multi-cell-processing (MCP) framework with minimal cooperation between the base stations. We consider a cluster of two base stations (BSs) which maximize the achievable rate for two users connecting to each BS and only sharing their channel state information (CSI). We provide a generalized fixed point equation of the optimal precoder with respect to the estimated channel, power, and the MMSE. We provide the optimal power allocation with respect to the estimated channel and MMSE. The designs introduced are optimal for multiple access (MAC) Gaussian coherent time-varying fading channels with arbitrary inputs and can be specialized to MIMO channels by decoding interference. The impact of interference on the capacity is quantified by the gradient of the mutual information with respect to the power, channel, and error of the interferer. Two novel distributed MCP algorithms have been introduced: a power allocation algorithm and a precoding algorithm for the UL and DL with a two way channel estimation; to keep track over the channel variation over time in a block by block basis.

In Chapter 7, we derive closed-form expressions for the capacity of the SISO and the multiple access Poisson Channels under the assumption of constant shot noise. The optimal power allocation is also derived for the different models; results have been analyzed in the context of information theory and optical communications. We then provide a set of simulation results establishing a comparison between Gaussian channels and Poisson Channels. We shed also light on the existence of fundamental relations between the mutual information and the MMSE.

In Chapter 8, we investigate network coding over time varying channels. We consider a small scale fading channel over Ka-band. We propose a novel model that exploits the channel delay profile and the dependency between channel states via a first order auto-regressive model for satellite communications. We provide an approximation of the delay induced assuming finite number of time slots to transmit a given number of packets. We also propose a novel adaptive transmission scheme that compensates for the lost degrees of freedom by tracking the packet erasures over time. Our results show that network coding non-adaptive mechanism for time variant channels has 2 times throughput and delay performance gains for small size packets, and similar performance gains for large size packets, compared to network coding mechanisms with fixed channel erasures. In addition, its shown that network coding for time variant channels has similar performance to the selective repeat with ARQ, and better performance when packet error probability is high, while due to better utilization of channel resources SR performance is similar or moderately better at very small

erasures, i.e., at high SNR. However, our adaptive transmission scheme outperforms the network coding non-adaptive scheme and SR with more than 7 times in throughput and delay performance gains.

In Chapter 9, we propose new models for the DCF of the IEEE 802.11. We analyze the delay over the unicast and broadcast transmission for a network topology that includes two APs and one station. We provide a closed-form expression for the expected time to deliver the N packets for the DCF mechanism with unicast, uncoded and coded broadcast, with and without ACK. We have shown that coding across packets in an acknowledged broadcast scenario encounters less delays, higher reliability, and higher throughout than the uncoded broadcast or unicast cases. We propose a new network-coded protocol that utilizes network coding to broadcast coded packets to the station performing handover. This new proposed network-coded handover framework will immensely serve if implemented in the current standardized IEEE 802.11 systems. We build upon constraints that takes into consideration the distance of the station and the degrees of freedom it owns to decode the received packets before it switches the connection to the next AP. Therefore, we provide a framework under which the QoS over delay sensitive streaming applications can be radically improved.

Chapter 10 concludes this thesis, summarizing the main contributions and suggesting areas for future works.

1.3 Thesis Contributions

The main contributions of this thesis are:

- Derivation of a new fundamental relation between the mutual information and the MMSE for MAC Gaussian channels. The relation particularly introduces the closed-form of the gradient of the mutual information with respect to per user channel, per user precoder, and the MMSE of each user (see Chapter 4).
- Investigation and characterization, for the first time, of the optimal transmission power allocation strategies and optimal precoding strategies for MAC Gaussian channels with general inputs, arbitrary and Gaussian (see Chapter 4).
- Interpretation of the interference with respect to the covariance introduced from the interferer to the main user channel (see Chapter 4).

- Characterization of the optimal precoding matrix of BPSK inputs at the high-SNR which provides a novel solution of an NP-hard problem (see Chapter 5).

- Characterization of the optimal precoder for MIMO Gaussian channels in a fixed point equation with respect to the channel right singular vectors, power allocation, and a permutation structure of the error eigen vectors (see Chapter 5).

- Quantification of the data rate loss due to interference in MIMO setups with different arbitrary constellations (see Chapter 5).

- Formulation of a new minimal cooperation framework that models interference channels via two distinct MAC channels. The setup analyzes the impact of fading on the information rates, and when a two way pilot assisted, and auto-regressive channel estimation methods are implemented at the receiver and transmitter sides, respectively (see Chapter 6).

- Derivation of a new closed-form expression for the capacity of the k-users MAC Poisson channel under the assumption of constant shot noise and with average powers that are not necessarily equal (see Chapter 7).

- Characterization of the optimal power allocation for k-users MAC Poisson channel (see Chapter 7).

- Proposal of a novel new model for coded and uncoded packet transmission over time varying channel (see Chapter 8).

- Proposal of a novel adaptive transmission scheme which outperforms non-adaptive schemes like network coding and selective repeat ARQ (see Chapter 8).

- Proposal of novel new models of the DCF of the WiFi IEEE 802.11 for unicast with and without fragmentation, uncoded and coded broadcast with and without acknowledgment (see Chapter 9).

- Proposal of a novel mathematical formulation of a network-coded handover that optimally decides the optimal time when to switch to the new access point. This in conjunction with the techniques proposed underlies a proposal on adding a network coding layer on top of the MAC layer which can improve the current standardized IEEE 802.11 and boost its performance from a delay and throughput perspective (see Chapter 9).

These contributions have led to the following publications:

- **Conference papers:**

1. Samah A. M. Ghanem, Munnujahan Ara, Capacity and Optimal Power Allocation of Poisson Optical Communication Channels, International Conference on Communications Systems and Technologies, (ICCST), WCECS, San Francisco, USA, November, 2011. (Best paper award).

2. Samah A. M. Ghanem, MAC Gaussian channels with arbitrary inputs: optimal precoding and power allocation, IEEE International Conference on Wireless Communications and Signal Processing (WCSP), Huangshan, China, October, 2012.

3. Munnujahan Ara, Hugo Reboredo, Samah A. M. Ghanem, Miguel R. D. Rodrigues, A Zero-Sum Power Allocation Game in the Parallel Gaussian Wiretap Channel with an Unfriendly Jammer, IEEE International Conference on Communication Systems (ICCS), Singapore, November, 2012.

4. Samah A. M. Ghanem, Optimal Precoding and Power Allocation with Multi-Cell Processing, IEEE 77th Vehicular Technology Conference: VTC-Spring, Dresden, Germany, June, 2013.

5. Samah A. M. Ghanem, Network Coding Mechanisms for Ka-Band Satellite Time Varying Channels, IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), London, United Kingdom, September, 2013.

6. Samah A. M. Ghanem, Optimal Power Allocation and Optimal Precoding in Multi-Cell Processing with Minimal Cooperation, submitted.

7. Samah A. M. Ghanem, IEEE 802.11 Network-Coded Handover, submitted.

- **Journal papers:**

8. Samah A. M. Ghanem, Munnujahan Ara, The Poisson Optical Communication Channels: Capacity and Optimal Power Allocation, IAENG International Journal of Computer Science (IJCS), vol. 39, no.1, pp. 102-108, 2012.

- **Book chapters:**

9. Samah A. M. Ghanem, Munnujahan Ara, The MAC Poisson Channel: Capacity and Optimal Power Allocation, IAENG Transactions on Engineering Technologies, Lecture Notes in Electrical Engineering, ©Springer, vol. 170, pp. 45-60, 2013.

Chapter 2
Physical Layer Transmission Techniques

2.1 Introduction

During the past two decades, many communication techniques have been developed to achieve various goals such as higher data rate, more robust link quality, and more user capacity in more rigorous channel conditions; the well known transmission technologies CDMA, OFDM, MIMO, multi-user OFDM, and UWB systems as well as the powerful error correcting coding schemes, namely turbo codes [1] and LDPC codes [2]. The mobile and fixed wireless systems, e.g. second and third generation mobile systems as well as WiFi, WiMax, and LTE adopt different kinds of these transmission techniques and coding schemes in order to support a large variety of services. All these systems have their own uniqueness while they also induce other drawbacks that limit the system performance. The conventional way to overcome the drawbacks is to impose most of the computational effort in the receiver side and let the transmitter design much simpler than the receiver. The fact is that, however, by leveraging reasonable computational effort to the transmitter, the receiver design can be greatly simplified; thus achieving higher gains and performance [3]. New diversity techniques and optimal precoding followed the recognition of the possible benefits of transmit diversity [4], [5]. However, to achieve the maximum reliable data rate and/or the diversity gain, defined as the increase in the SINR due to some diversity scheme, afforded by the system hardware, appropriate precoding and modulation techniques are necessary. Two main approaches emerged through transmission technologies: The first approach uses appropriate mappings of the constellation in space and time such that, without CSI at the transmitter and with low complexity at the receiver, full

diversity gains become possible [4], [5], [6], [7]. The second approach addresses the optimization of the maximum reliable data rate in the case of flat fading, [8], [9], [10] and frequency-selective channels [11], [12], assuming full knowledge of the CSI at both the transmitter and receiver sides.

Optimal designs developed in the past, which were based on MIMO models such as [13], [14], [15], gained importance because of the new interest in joint transmit-receive diversity schemes. The optimization of pre- and post-filters was considered in [16] for a MIMO system distorted by additive noise only. Different design criteria have been introduced in the design of joint transmit-receiver linear processing (beamforming, also called precoding in transmitter and equalization at receiver) which are, the minimization of the MSE, the minimization of the BER, or the maximization of the SINR [17]. An optimal design for space-time linear precoders and decoders is introduced [17], [18]. Non-linear pre-equalization based on Tomlinson-Harashima-Miyakawa precoding is considered, together with a trellis signal shaping algorithm and zero-forcing adjustment [19]. New precoding methods have been developed through concepts of IA [20], [21] and RA, reversing the communication direction and using the receive combining vectors as precoding vectors, in the iterative distributed IA max-SINR algorithm [22].

Yet, the design of transmit precoding filters and receive equalization filters is an outstanding open problem that still abounds in the analysis, design and optimization of key elements of communication systems. In this context, significant progress has been made in the design of key communications system elements under the standard MSE or even the error probability criteria [17], [14], [18], and [23].

However, considerable less progress has been made in the design of communications system elements under the reliable information transmission rate criterion and interference alignment criterion. The difficulty relates to the fact that closed-form expressions of the mutual information mainly exist for communications systems driven by Gaussian inputs [24], [25], [26], [27], [28]. However, since practical complexity considerations pertaining to the transmission and reception of information dictate the use of discrete inputs (e.g. M-PSK or M-QAM, etc.) rather than the theoretically appealing Gaussian ones, one often has to resort to engineering experience and insight to optimize the system. One classical example relates to the optimization of the information transmission rate of DSL using the bit loading algorithm, [29], which delivers considerable gains though it is non-optimal.

Recently, Guo et al. have illuminated intimate connections between information the-

ory and estimation theory in a seminal paper, [30] (later generalized in follow-up papers [31], [32]. In particular, Guo et al. have shown that in the classical problem of information transmission through the conventional AWGN channel the derivative of the mutual information with respect to the SNR equals the non-linear MMSE, a relationship holding for scalar, vector, discrete-time and continuous-time channels regardless of the input statistics. The relevance of these recent connections derives from the fact that mutual information and MMSE are two canonical operational measures in information theory and estimation theory: mutual information measures the reliable information transmission rate between the input and the output of a system for a specific signaling scheme, while MMSE measures the minimum mean-squared error in estimating the input given the output.

The implications of a framework involving key quantities in information theory and estimation theory are countless both from the theoretical [33], [34] and the more practical perspective. Of particular relevance is the use of the simple connection between the derivative of the mutual information and the non-linear MMSE to address the open problems in the optimization of the reliable information transmission rate of communication systems in innovative manners.

In this context, Lozano et al. [35] have considered the maximization of the reliable information transmission rate for a bank of parallel independent Gaussian channels with arbitrary input distributions, subject to a total power constraint, a situation arising in many wired and wireless communication scenarios. In particular, by capitalizing on the mutual information-MMSE relationship, Lozano et al. have expressed the optimal power allocation policy in terms of the MMSE. They have also put forth a novel interpretation of the optimal power allocation policy: the so-called mercury/waterfilling interpretation represents an extension of the conventional waterfilling interpretation applicable to parallel independent Gaussian channels with Gaussian inputs [24].

More recently, Perez-Cruz et al. [36], [37] have considered the maximization of the reliable information transmission rate of multiple-input and multiple-output Gaussian channels with arbitrary inputs, subject to a total power constraint. This more general scenario captures many communication systems with practical relevance, such as multiple antenna systems, CDMA systems or OFDM systems, which currently form the basis of many wired and wireless standards. By capitalizing on the mutual information-MMSE relationship, Perez-Cruz et al. have shown that conventional communication techniques involving singular value decomposition and waterfilling [24], though long considered optimal, are highly suboptimal. They have also shown that the use of the optimal communication techniques offer the prospect of substantially higher reliable

information transmission rates in many relevant practical scenarios, namely in DSL systems [38] or magnetic recording systems [39].

These recent results bear witness to the fact that the intersections between key quantities in information theory and estimation theory have important applications in the optimization of communications systems. Yet, the full implications of this framework in the analysis, design and optimization of communications systems has to be uncovered. In this Chapter, we present the fundamental framework of physical layer transmission techniques by the aspect of information-theoretic and estimation-theoretic approaches that will be used throughout the solution of the first paradigm. We review several works related to physical layer optimal power allocation and optimal precoding, the methodology to tackle the first problem in this thesis, and how they relate to our approach to the problem of optimal precoding and optimal power allocation for maximizing the reliable achievable data rates.

2.2 Parallel Gaussian Channels with Gaussian Inputs

Parallel independent Gaussian channels model a set of wireless and wireline domains. For example, multi-tone orthogonal transmission like OFDM constitutes a bank of independent parallel channels [40]. Multi-antenna communications with transmitter knowledge of the transfer coefficients between antennas is another example, [41], [11]. Block fading channels with time variant gains can be also seen as a bank of parallel channels, [42]. However, DSL is an example in the wireline world, see also [35]. Consider a bank of k-independent parallel non-interfering Gaussian channels, as shown in Figure 2.1. On the ith sub-channel, the input$-$output relationship is:

$$y_i = h_i x_i + n_i, \tag{2.1}$$

where the noise is n_i is a zero-mean unit-variance complex Gaussian random variable independent of the noise on the other channels. The complex scalar channel h_i is a deterministic nonzero gain. The normalized unit-power inputs s_i are complex Gaussian inputs such that, $x_i = \sqrt{p_i P} s_i$ are independent Gaussian random variables. We want an optimal power allocation policy that maximizes the mutual information of a bank of k independent parallel channels with Gaussian inputs is the solution of the following optimization problem:

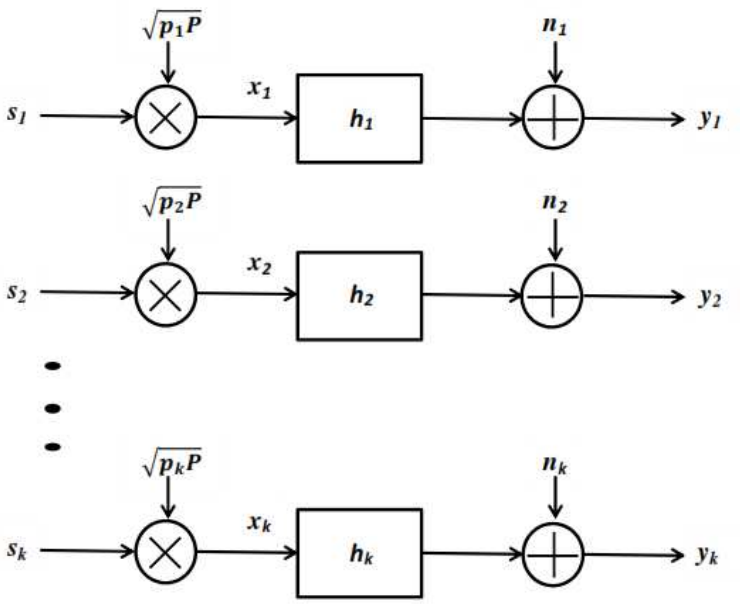

Figure 2.1: Bank of k independent parallel channels.

$$max \sum_{i=1}^{k} log\left(1 + p_i\gamma_i\right), \tag{2.2}$$

with the power allocation $p_i,\ \forall i = 1, .., k$ channels, is constrained by the total power P as follows:

$$\frac{1}{k}\sum_{i=1}^{k}\mathbb{E}[|\mathbf{x_i}|^{\mathbf{2}}] \leq P, \tag{2.3}$$

and $\gamma_i = P|h_i|^2$, where $p_i\gamma_i$ represents the signal to noise ratio on the ith channel.

This is a standard optimization problem and can be solved using Lagrange multipliers. Writing the Lagrangian for the optimization problem as:

$$\mathcal{L}\left(p_1, ..., p_k, \lambda\right) = -\sum_{i=1}^{k} log\left(1 + p_i\gamma_i\right) - \lambda\left(P - \frac{1}{k}\sum_{i=1}^{k} p_i\right), \tag{2.4}$$

Using the KKT conditions, we derive the optimal power allocation for the Gaussian inputs setup which is known as the waterfilling.

$$p_i^{\star} = \left(\frac{1}{\lambda} - \frac{1}{\gamma_i}\right)^{+}, \tag{2.5}$$

with the Lagrange multiplier λ is chosen such that $\sum_{i=1}^{k}\left(\frac{1}{\lambda}-\frac{1}{\gamma_i}\right)^{+}=P$. The optimal power allocation is such that, if the noise level $\frac{1}{\gamma_i}$ exceeds the base level $\frac{1}{\lambda}$, no power is allocated to the channel. While, if the noise level is less than the base level, more power is allocated to less noisy channels. The process through which the power is distributed among the various channels is identical to the way water distributes itself in a vessel, hence this process is referred to as waterfilling, [43].

2.3 Parallel Gaussian Channels with Arbitrary Inputs

The mutual information of independent parallel Gaussian channels is maximized, under an average power constraint, by independent Gaussian inputs whose power is allocated according to the waterfilling policy. In practice, discrete signaling constellations with limited peak to average ratios (M−PSK, M−QAM, etc.) are used in lieu of the ideal Gaussian signals. It is therefore important to study power allocation under these circumstances. Yet, no general solution has been found to date for the power allocation that maximizes the mutual information over parallel channels with non-Gaussian inputs. The obstacle in the way of this optimization problem is the lack of explicit expressions for the corresponding mutual information. Therefore, we can define the input output mutual information on the ith channels and try to maximize it subject to the average power constraint as follows:

$$max \;\; \frac{1}{k}\sum_{i=1}^{k} I_i(snr) = max \;\; \frac{1}{k}\sum_{i=1}^{k} I\left(s_i;\sqrt{snr}s_i+n_i\right) \tag{2.6}$$

Subject to:

$$\frac{1}{k}\sum_{i=1}^{k}\mathbb{E}[|\mathbf{x_i}|^{\mathbf{2}}]\leq P \tag{2.7}$$

Where the signal to noise ratio: $snr_i = p_i P|h_i|^2$. Capitalizing on the recently unveiled fundamental relationship between mutual information and MMSE [30], defined as:

$$\frac{d}{dsnr}I_i(snr) = mmse_i(snr). \tag{2.8}$$

The power allocation policy that maximizes the mutual information over parallel channels with arbitrary input distributions, referred to as mercury/waterfilling, generalizes the waterfilling solution and allows retaining some of its intuition [35], [44].

Denoting the inverse of the $mmse_i(\zeta)$ with respect to the composition of functions by $mmse_i^{-1}(\zeta)$, with domain equal to $[0,1]$, and $mmse_i^{-1}(1) = 0$, the optimal power allocation over all k channels can be explicitly expressed as:

$$p_i^\star = \frac{1}{\gamma_i} mmse_i^{-1}\left(min(1, \frac{\lambda}{\gamma_i})\right), \tag{2.9}$$

where λ is the unique solution to the non-linear equation:

$$\sum_{i=1\gamma_i>\lambda}^{k} \frac{1}{k\gamma_i} mmse_i^{-1}\left(\frac{\lambda}{\gamma_i}\right) = 1 \tag{2.10}$$

2.4 MIMO Gaussian Channels with Gaussian Inputs

MIMO systems can greatly increase the spectral efficiency also called the maximum reliable data rate, over a limited bandwidth since it has one additional dimension to carry information. The capacity gain of MIMO systems over the SISO system is a remedy to the fast increasing demands of higher data rates in wireless communications. Therefore, MIMO channels constitute a unified way of modeling a wide range of different physical communication channels, which can then be handled with a compact and elegant vector-matrix notation. Consider a linear superposition MIMO channel model with k transmit and k receive antennas, as shown in Figure 2.2. The received vector signal is as follows:

$$\mathbf{y} = \mathbf{HPx} + \mathbf{n}, \tag{2.11}$$

where $\mathbf{H}$ is an k x k complex channel matrix, h_{ij} represents the path gain from the ith receive antenna to the jth transmit antenna. $\mathbf{P}$ is a k x k precoding or power allocation matrix, $\mathbf{x}$ is a k x 1 transmitted vector signal, and $\mathbf{n}$ is a k x 1 zero mean complex Gaussian noise with independent real and imaginary part components. We assume the noise of each receiving antenna is independent, i.e., the covariance of $\mathbf{n}$ is $\mathbb{E}[\mathbf{nn}^\dagger] = \mathbf{I}$. For every $\mathbf{H}$, we can do a singular value decomposition, such that $\mathbf{H} = \mathbf{U\Lambda V}^\dagger$, $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices, and $\mathbf{\Lambda}$ is a diagonal matrix. Therefore, the MIMO channel model becomes:

$$\mathbf{y} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\dagger}\mathbf{P}\mathbf{x} + \mathbf{n}, \tag{2.12}$$

After defining $\tilde{\mathbf{y}} = \mathbf{U}^{\dagger}\mathbf{y}$, $\tilde{\mathbf{x}} = \mathbf{V}^{\dagger}\mathbf{P}\mathbf{x}$, and $\tilde{\mathbf{n}} = \mathbf{U}^{\dagger}\mathbf{n}$, we obtain an equivalent channel from the mutual information perspective with equivalent noise properties:

$$\tilde{\mathbf{y}} = \mathbf{\Lambda}\tilde{\mathbf{x}} + \tilde{\mathbf{n}}, \tag{2.13}$$

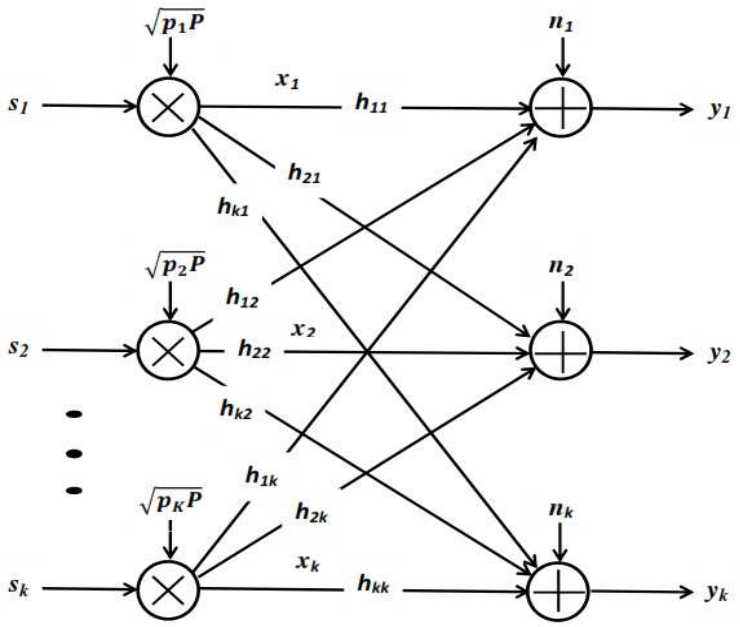

Figure 2.2: $k \times k$ MIMO channel.

Note that if each other received vector at the receiver antenna i is considered as interference, then we could consider a MIMO channel as a correlated parallel Gaussian channel. Note also that $\mathbf{\Lambda}$ is diagonal, thus we decompose the correlated parallel channel into independent parallel channels. So, to maximize the mutual information $I(\tilde{\mathbf{x}};\tilde{\mathbf{y}})$, we need to choose $\tilde{\mathbf{x}}$ to be independent, and each has independent Gaussian, zero-mean real and imaginary part. Then, allocating power to each virtual channel is done via waterfilling, [41]. Similar to the previous sections, we can define the input output mutual information on the channel and try to maximize it subject to the total power constraint as follows:

$$max \quad I(\mathbf{x};\mathbf{y}) \tag{2.14}$$

Subject to:

$$\mathbb{E}[\mathbf{x}\mathbf{x}^{\dagger}] \leq \mathbf{P} \tag{2.15}$$

The closed-form expression of the capacity of the MIMO channel with Gaussian inputs is given by:

$$I(\mathbf{x};\mathbf{y}) = log\ |\mathbf{I} + \mathbf{HPH}^{\dagger}| \tag{2.16}$$

Using eigenvalue decomposition for $\mathbf{HH}^{\dagger} = \mathbf{U}\widetilde{\mathbf{\Lambda}}\mathbf{U}^{\dagger}$, where $\widetilde{\mathbf{\Lambda}} = \ diag\,(\gamma_1, ..., \gamma_k)$, corresponds to the non-negative square roots of the eigenvalues of $\mathbf{HH}^{\dagger}$ on the main diagonal. Therefore,

$$I(\mathbf{x};\mathbf{y}) = log\ |\mathbf{I} + \widetilde{\mathbf{\Lambda}}^{\mathbf{1/2}}\mathbf{UPU}^{\dagger}\widetilde{\mathbf{\Lambda}}^{\mathbf{1/2}}|, \tag{2.17}$$

By Hadmard's inequality:

$$I(\mathbf{x};\mathbf{y}) \leq log \prod_i (1 + p_i\gamma_i), \tag{2.18}$$

with equality holds when $\mathbf{P}$ is diagonal. Therefore, the optimal precoding policy that maximizes the mutual information for MIMO channels with Gaussian inputs is a matrix that diagonalizes the channel. Thus, we get a classical independent parallel Gaussian channel. Therefore, capacity can be achieved by the following precoding technique: First, we form the data stream into a vector signal $\tilde{\mathbf{x}}$ with k substreams. Second, encode and modulate according to the waterfilling of the power to the different substreams. Third, rotate the signal $\tilde{\mathbf{x}}$ by multiplying $\mathbf{V}$ at the receiver. We rotate the received signal $\tilde{\mathbf{y}}$ by $\mathbf{U}^{\dagger}$ to get the equivalent parallel channel, and then we can decode the substreams. In the process of coding, we can either jointly or independently encode the substreams, [45].

2.5 MIMO Gaussian Channels with Arbitrary Inputs

In MIMO Gaussian channels as well as parallel channels already discussed, practical constraints dictate the use of discrete constellations, such as PSK and QAM, which depart significantly from the optimal Gaussian distributions. The channel model is still the same as the MIMO Gaussian channel discussed in the previous section, but the optimization of the mutual information will be based on sub-optimal arbitrary input constellations. Recall that in the Gaussian input world, mutual information is maximized by imposing a covariance structure on the Gaussian input vector that satisfies two conditions. First, transmitting along the right eigenvectors of the channel, the vector channel reduces to a set of parallel non-interfering sub-channels. Second,

the power allocation follows the waterfilling policy based on the signal-to-noise ratio in each non-interacting sub-channel. Consider the deterministic complex-valued vector channel:

$$\mathbf{y} = \mathbf{HPx} + \mathbf{n}, \tag{2.19}$$

$\mathbf{y}$ is the k x 1 received vector, $\mathbf{H}$ is the k x k channel matrix, $\mathbf{P}$ is the k x k precoding matrix, $\mathbf{x}$ is the k x 1 input (Gaussian or arbitrary), and $\mathbf{n}$ is the k x 1 noise vector. The optimization of a linear precoder and equalizer for MIMO channels has been typically addressed using an MMSE criterion, [14], [18], [46], with the optimal solution being the diagonalization of the channel matrix. In [17], a unifying approach using different criteria, such as the MMSE, the SINR and the BER, leads to identical results. To optimize a linear precoder to maximize the input-output mutual information for general (not necessarily diagonal) rectangular deterministic channel matrices; a necessary condition for the solution is expressed as a fixed point equation in terms of the relation between the MMSE and mutual information, [30], [31]. The optimization of the mutual information is carried out over all k x k precoding matrices $\mathbf{P}$ that do not increase the transmitted power. The precoding problem can be cast as a constrained non-linear optimization problem:

$$max \quad I(\mathbf{x}; \mathbf{y}) \tag{2.20}$$

Subject to:

$$Tr\{\mathbb{E}[\mathbf{Pxx}^{\dagger}\mathbf{P}^{\dagger}]\} = Tr\{\mathbf{PP}^{\dagger}\} \leq 1 \tag{2.21}$$

The optimum precoding matrix $\mathbf{P}^{\star}$ that solves the optimization problem is as follows:

$$\mathbf{P}^{\star} = \nu \mathbf{H}^{\dagger}\mathbf{HP}^{\star}\mathbf{E}, \tag{2.22}$$

with $\nu = \left\|\mathbf{H}^{\dagger}\mathbf{HP}^{\star}\mathbf{E}\right\|$, and $\mathbf{E}$ is the MMSE matrix. This MMSE matrix is estimated using any iterative method like Monte-Carlo. The solution gives a lower and upper bounds for the MMSE which leads to lower and upper bounds for the mutual information that relates to the minimum distance between the points in the arbitrary input constellations, the precoding matrix in the high-snr regime converges to the matrix that maximizes the minimum distance between constellation vectors, [37]. Note that in contrast to the situation for Gaussian inputs [24], the optimal linear precoder for arbitrary inputs does not diagonalize the channel and, in particular, it does not reduce to a diagonal matrix for parallel non-interfering channels. When the inputs are not Gaussian, information transmission rates are maximized when a full precoder is used instead of a power allocation strategy coupled to channel diagonalization. However,

using a full precoding matrix, the mutual information increases significantly for all but not low-snr regime, since at low-snr the receiver cannot distinguish complex Gaussian from quadrature symmetric arbitrary inputs [47].

2.6 Methodology

In this section we will present the two main methodologies used to address the first paradigm in this thesis. First, we will present the methodology used to address the mathematical foundations to derive the solutions of the optimization problems. In particular, we will present the mutual information-MMSE (I-MMSE) identity used to find optimal precoding and optimal power allocation designs for general inputs, with arbitrary and Gaussian distributions. This methodology is of particular relevance to provide such solutions due to the fact that there are no explicit closed-form expressions of the mutual information for arbitrary constellations. Second, we will present the methodology used to simulate the mutual information and the MMSE of binary inputs, and to optimize such key quantities numerically. In particular, we will present Monte-Carlo method, a computational algorithm that is of particular relevance to address the simulation and optimization of such key quantities. Therefore, the second methodology has been used to simulate and quantify key quantities connecting information measures to estimation measures, particularly, the mutual information and the MMSE. In addition, it is used to optimize the mutual information or the MMSE, under predefined constraints, and to simulate optimal designs with respect to different variables in the communication system model.

2.6.1 The Relation between the Mutual Information and the MMSE

The conventional method to address the aforementioned optimization problems uses optimization theory, [48]. There are two key challenges associated with the optimization of the mutual information or data rates. One relates to the fact that the optimization problem is possibly non-convex. Consequently, the KKT conditions are only necessary rather than necessary and sufficient. The other challenge relates to the absence of closed-form mathematically tractable capacity expressions in a wide range of scenarios. Note that closed-form capacity expressions exist for the MIMO Gaussian

channel driven by Gaussian inputs. However, such expressions do not exist for MIMO channels driven by arbitrary inputs, e.g. M−PSK or M−QAM inputs. Consequently, the determination of the solution of these optimization problems will require the use of a novel methodology involving connections between information theory and estimation theory. Consider the classical AWGN channel:

$$\mathbf{y} = \sqrt{\mathbf{snr}}\,\mathbf{x} + \mathbf{n}, \tag{2.23}$$

where $\mathbf{y}$ denotes the k x 1 channel output vector, $\mathbf{x}$ denotes the k x 1 channel input vector, $\mathbf{n}$ denotes a k x 1 vector of Gaussian random variables with zero mean and unit variance and *snr* represents the signal-to-noise ratio. Guo et al. have illuminated intimate connections between information theory and estimation theory in a seminal paper, [30], later generalized in follow-up papers [31], [32] in this channel model. In particular, Guo et al. have shown the derivative of the mutual information with respect to the SNR equals the non-linear MMSE:

$$\frac{d}{dsnr} I_i(snr) = \frac{1}{2}\mathbb{E}[(\mathbf{x} - \mathbb{E}[\mathbf{x}|\mathbf{y}])^2] = \frac{1}{2} mmse_i(snr). \tag{2.24}$$

2.6.2 Monte-Carlo Method

Monte-Carlo method is one kind of computational algorithms that relies on repeated random sampling to obtain numerical results, i.e., by running simulations over a finite number of samples with all possible output permutations known in order to calculate the probabilities heuristically. Monte Carlo methods are mainly used in three distinct problems: optimization, numerical integration and generation of samples from a probability distribution. The lack of existence of explicit closed-form expressions of the mutual information and the MMSE for binary constellations makes Monte-Carlo method a well acknowledged means to estimate the mutual information and the MMSE of binary constellations (M-PSK, M-QAM, etc). The main steps of the Monte-Carlo method in calculating the channel conditional probability and the probability of the received vector at the channel output as well as the mutual information and the MMSE can be summarized into the following steps:

1. *Inputs:* channel $\mathbf{H}$, power $\mathbf{P}$.
2. *Step 1:* Generate a vector $\mathbf{x}$ of M symbols of a constellation, if BPSK, i.e., $M = 2$, then $\mathbf{x} = \{\mathbf{1}, -\mathbf{1}\}$, inputs power is normalized by the scaling factor $\sqrt{2}$.

3. *Step 2:* Generate the noise vector as a randomly generated complex Gaussian random variable $\mathbf{n}$ with zero mean and unit variance.

4. *Step 3:* Calculate $\mathbf{y} = \sqrt{\mathbf{snr}}\ \mathbf{HPx} + \mathbf{n}$.

5. *Step 4:* Find the channel conditional probability $\mathbf{p_{y|x}}(\mathbf{y}|\mathbf{x})$ corresponding to the noise distribution.

6. *Step 5:* Find the received vector probability $\mathbf{p_y}(\mathbf{y})$. This is done over the received vector and all possible permutations of the input constellation. If two user inputs with SISO or two sub-channels for a MIMO, then: $[\mathbf{x_1}, \mathbf{x_2}] \in \mathbf{x}' = [\{\mathbf{1}, \mathbf{-1}\}, \{\mathbf{-1}, \mathbf{-1}\}, \{\mathbf{1}, \mathbf{-1}\}, \{\mathbf{1}, \mathbf{1}\}]$.

7. *Step 6:* Estimate the mutual information averaged over all M symbols, and similarly estimate the MMSE as defined with respect to the probabilities and the inputs.

Note that, if the channel is random, like the scenario of Gaussian fading channels, an averaging step over a sufficiently large number of randomly generated channel realizations (with known channel probability distribution, like Rayleigh, Rician, etc) is required.

The first paradigm in this research work will capitalize on the relationship between the derivative of the mutual information and the non-linear MMSE to address the problem of the design and optimization of linear precoders that maximize the mutual information and so the reliable data rates. It is important to highlight that this relationship is a means to overcome the absence of explicit mutual information expressions in a range of scenarios, since it enables the computation of the mutual information in terms of a more tractable quantity, the non-linear MMSE. For example, it is possible to express the form of the optimal key communication system elements in terms of the non-linear MMSE, by feeding the derivative of the mutual information directly into the KKT optimality conditions. In addition, in a communications system operating in a regime of low-snr, it is possible to express the non-linear MMSE in closed-form using Taylor series and hence, by invoking the relationship, the mutual information can be expressed in closed-form too. In contrast, in a communication system operating in a regime of high-snr it is possible to tightly lower and upper bound the non-linear MMSE using genie and Euclidean based estimators, respectively, and hence, by invoking the relationship, tightly upper and lower bound the mutual information too. Since the

closed-form expressions/bounds are functions with respect to key system elements (e.g. precoders), one can readily derive further insight into the optimal structure of such communication elements, [36], [37], [35], [38], and [39].

2.7 Conclusion

In this Chapter, we present a brief overview of information-theoretic estimation-theoretic approaches at the physical-layer as well as prior work in the area of optimal precoding and optimal power allocation. In particular, we present main contributions on the optimal power allocation and optimal precoding designs for a set of channels driven by different types of inputs distribution. We focus on such designs, where the design criterion used is the maximization of the mutual information, for the parallel and MIMO Gaussian channels driven by inputs with Gaussian and arbitrary distributions. We finally present two main methodologies adopted in deriving the set of solutions of the first paradigm of this thesis, particularly, the I-MMSE identity and Monte-Carlo method. We established a direct connection between the set of methodologies and their relation to the work that will be addressed throughout the first part of this thesis.

In spite of the numerous theoretical contributions, the general problem of the design of physical-layer transmission schemes for maximizing the reliable data rates over Gaussian channels, multi-user Gaussian channels, multi-user fading channels, and channels with other distributions like Poisson channels is still widely open. However, another problem of long-standing interest is the study of communication scenarios, where cooperation exists, within a cluster of base stations. The objective of the first part of this PhD research is to fill this vacuum, attacking the problem from an information-theoretic estimation-theoretic approach.

Chapter 3
Network Coding and Delay

3.1 Introduction

Network coding has been successfully applied to a number of networking scenarios. These scenarios demonstrate many advantages of network coding over forwarding, including resource (e.g., bandwidth, energy, storage space) efficiency, computational efficiency, lower communication delay, and robustness to network dynamics. The idea of network coding was first proposed by Alswede et al. [49] to enhance the capacity of the noiseless wired network, therefore, feedback is not required. Although the noiseless assumption used in [49] is no longer valid in wireless communication. However, the additional diversity gain introduced by network coding in wireless networks is of particular relevance [50]. In fact, network coding is useful for wireless networks, since its advantages match very well with the broadcast nature and other inherent characteristics of the wireless medium.

For a multicast scenario, a network coding based solution has been shown to offer superior performance compared to state-of-the-art forwarding based solutions. The authors of [51] and [52] showed that linear codes over a network are sufficient to implement any feasible multicast connection, again considering a channel with no erasures. Much of the previous work in this area, including [53], [54] focuses on the use of network coding for multicast traffic. Even for unicast scenarios, network coding offers an attractive solution because its robustness can better deal with the dynamics in wireless networks. As shown in [55], network coding can outperform link-by-link and end-to-end ARQ and FEC strategies in terms of the number of transmissions in the network for every packet received at the destination. However, obtaining such efficiency requires careful design of the rate at which coded packets are transmitted on every path.

Network coding can be seen as a particular form of channel coding which makes it very interesting in decentralized streaming architectures. However, the application of network coding principles in multimedia streaming systems is not a trivial task as streaming applications generally impose strict timing and complexity constraints that limit the coding opportunities. Nonetheless, network coding is proved to be an essential ingredient to achieve the capacity of the network, [50]. This makes it of particular relevance to applications that require high data rates, and stringent delay considerations for example content distribution [56], distributed storage [57], or data dissemination [58].

Future communication systems that demand high data rates should be delay tolerant. There has been a proliferation of interest in applying network coding to DTNs to improve data transmission efficiency. However, at times of increasing traffic demands, the highly central nodes and parts of the network can get congested. Current approaches do not cope well with congestion because they are not adaptive and result in increased dropped-packet rates and delays. However, it has been shown that dropping coded packets can significantly degrade the delivery success rates [59]. Therefore, congestion aware forwarding with NC has been proposed to combine adaptive NC and adaptive forwarding in DTNs. In such framework, each node learns the status of its neighbors, and their ego-networks in order to detect coding opportunities, and codes as long as the recipients can decode [60].

The second paradigm in this research work will try to make the coding approach to be more delay tolerant. Network coding can be thought of as one kind of precoding techniques through which coding across the packets is done by selecting random linear coding coefficients from a Galois field. It is somehow a counter part to selecting precoding coefficients in a bit level or packet level. The first exploits the nature of the network architecture from the physical to the network layer. The later inherently exploits the space through its design. Therefore, the possible redundancy inherently existing in both paradigms puts part of the processing to the transmitter, and allows the ultimate receivers to essentially average spatial channel variations resulting from physical channel effects such as fading and shadowing or from interference.

Therefore, it was of particular relevance to complete the theme of this PhD research introducing the coding concept to delay. In particular, introducing transmission schemes that uses network coding as a special case of precoding, to meet delay requirements, constitutes the multi-layer paradigms to a novel framework for future communication systems.

3.2 Delay in Networks and Network Coding

Delay is a fundamental problem to address for future communication systems and networks. From a classical coding theory perspective, we need an arbitrary design for large codebook to achieve capacity. However, achieving capacity or increasing the data rates doesn't necessarily mean that we decrease the delay. Delay is addressed from the information theoretic-estimation theoretic paradigm through models that are limited to Poisson arrivals, [61], [62] which is more relevant to queuing theory, [63], [64].

In the networking community, delay has been addressed from an application perspective. For instance, [65], addresses the delay in the multiple access layer for systems where the process of accessing the channel is modeled via a Markov chain and the mean completion time has been analytically evaluated. The delay has been addressed from a network coding perspective where codes can be designed to meet delay requirements in lieu of only achieving capacity. For example, the work in [66] and [67] studied the delay performance gains and their scaling laws for network coding with and without channel side information, respectively. They provide a novel formulation where the elasticity of a traffic flow is based on the delay constraints associated with it. In [68] the delay performance of network coding for a tree-based multicast problem is studied. The authors define the expected number of transmissions per packet to be received and derive analytical expressions characterizing the performance of network coding in comparison with ARQ and FEC techniques. In [69] they studied network coding with feedback. The authors showed that the use of feedback can be employed for parameter adaptation to satisfy QoS requirements as well as for reliability purposes. They argue about the benefits of applying network coding to the feedback packets and examine the design of acknowledgment packets. The authors in [70] couple the benefits of network coding and ARQ by acknowledging DoF, defined as linearly independent combinations of the data packets, instead of original data packets to show that queue size in a node follows the DoF. In [71], the authors propose a network coding scheme for time division duplexing channels. The scheme uses a half duplex transmission with feedback through which the transmitter is able to transmit a predefined number of coded packets based on the DoF of the receiver. They showed that their scheme is optimal from a delay perspective and similar to selective repeat. However, in [72], they provide optimal and heuristic schemes to exploit feedback in the presence of asymmetric packet loss probabilities and propagation delay.

Our focus in the second part of this PhD research will be in attacking the delay problem via network coding schemes for delay challenging channels, particularly the

satellite channel and at the MAC layer of the WiFi IEEE 802.11. The next sections will first introduce state of the art work on channel aware strategies to study the network coding schemes for time varying channels. Second we will introduce a compact state of that art in the delay considerations at the MAC layer.

3.4 Network Coding in Wireless Channels

The study of network coding mechanisms in wireless networks need to take into consideration different aspects in the wireless medium like noise, interference, and fading. The diversity benefits of network coding that can mitigate the wireless fading was shown in [50]. In [73] the authors adopt a framework similar to the cooperation framework addressed in the precoding problem, particularly, their framework encourages strategically picked senders to interfere. Instead of forwarding packets, routers forward the interfering signals. The destination leverages network-level information to cancel the interference and recover the signal destined to it. The result is analog network coding because it mixes signals not bits. Moreover, from a decodability perspective, the authors in [74] proposed ZigZag decoding that is based on interference cancellation, and hence, requires a precise estimation of channel coefficients for each packet involved in a collision. In [75] an opportunistic network coding approach was introduced. In particular, the codewords are adapted according to the information received from the neighbors. In [76], the authors show that fixed network codes without CSI cannot achieve instantaneous min-cut. However, they proved that adaptive network codes with one bit global CSI have lower erasure probability than the codes without CSI.

Network codes adaptation strategies that are based on channel state information are limited to packet erasure channel model, that is, a two-state Markov model of Gilbert-Elliott channel, in which a packet is whether dropped with a certain probability or received without error, see [77], [78]. In [77], the authors developed a rate-controlled, multipath strategy using network coding. They showed that such strategy can provide throughput performance comparable to multipath flooding of the network while utilizing bandwidth nearly as efficiently as single-path routing. In [78], the authors studied the delay and energy performance under bursty erasures. They proved that channel-aware policies reduce delay by up to a factor of 3. Further, the authors in [71] explored the timing nature of coding across packets over time division duplexing channels. However, they consider time invariant channels, i.e., with fixed erasure

probability. Of particular relevance to consider scenarios of practical relevance where channel variation and fading evolves over time.

The first scenario we are going to address in the second paradigm of this PhD thesis is network coding for time varying channels. We focus on channels that are delay challenging, particularly, we focus on the Ka-band satellite channel.

3.3 Network Coding in Multiple Access Networks

Delay at the MAC layer has been addressed through different analytical models. For instance, in the WiFi IEEE 802.11, different models have been devised based on Bianchi's model of the DCF mechanism [65], and different proposed schemes to optimize the network capacity or to enhance the QoS have been conducted. The authors in [79] proposed an analytical model to study the throughput of a p-persistent IEEE 802.11 protocol that selects a backoff window size that balance collision and idle period costs. Other contributions built on top can be found in [80], [81], etc. Recently, EDCF that employs a radically different contention window size as compared to DCF of IEEE 802.11 was introduced into the IEEE 802.11e. A set of proposals to improve the QoS via introducing minor changes in the mechanisms like fixing the maximum CW to be equal to the first CW, or limiting the number of retrials for the sake of decreasing the collisions, increasing fairness, or to mitigate the hidden terminal problem can be found in [82] and [83]. In [84] and [85], they tried to improve the delay performance in the handover scenario focusing on providing mechanisms that minimize the most contributor to the delay, which is the probing delay.

Network coding constitutes a disruptive, yet simple concept that considers data packets as algebraic entities on which one can operate. In particular, NC allows intermediate nodes in the network to re-encode information flows, rather than just routing them. NC has been shown to achieve network capacity for multicast sessions, as given by the max-flow min-cut theorem [1] [86], which is not attainable using only routing, [49]. For the case of multicast flows, this bound is shown that it can be achieved using linear NC [51], where coded packets are generated through linear combinations of the original data packets.

[1] Max-Flow Min-Cut theorem states that in a flow network, the maximum amount of flow passing from the source to the sink is equal to the minimum capacity that when removed from the network, causes the situation that no flow can pass from the source to the sink.

The practicality of NC has been demonstrated in [87]. The authors propose COPE, a new architecture for wireless mesh networks. In addition to forwarding packets, routers mix (i.e., code) packets from different sources to increase the information content of each transmission. It has been also shown in [88] that random linear NC in lossy networks can achieve packet-level capacity for both single unicast and single multicast connections in wireline and wireless networks. In [89] the authors propose MORE, a MAC independent opportunistic protocol for wireless networks and provide experimental results emphasizing the throughput gains provided by network coding. In [90] the authors propose simple algorithms that allow the realization of the benefits of broadcasting using network coding. In particular, they provide theoretical analysis that shows that network coding improves performance by a constant factor in fixed networks. Then, they show that in networks where the topology dynamically changes, for example due to mobility, network coding can offer improvements of a factor of $log(n)$, where n is the number of nodes in the network.

Mobile nodes in some challenging network scenarios suffer from intermittent connectivity and frequent partitions. There has been several work on mobility protocols, such as mobile IP based approaches [91], fast mobility approaches like FMIP [92], HMIP [93], and new concepts for local mobility (PMIP), [94]. Recently, the IEEE 802.21 MIHF and its messages, can enable handover across heterogeneous wireless networks, [95]. Their proposal aims to assist proactive authentications in order to reduce the delay due to medium access, authentication, and key establishment with the target network. In [96], the authors propose CANCO scheme for DTNs that achieves higher delivery ratio than the existing network coding schemes in networks where nodes move according to non-homogeneous mobility models. They also show that their scheme encounters higher data efficiency than a MORE-like scheme which suffer from the stop-and-wait feature. In [97] the authors propose MRNC for WiMax IEEE 802.16. They evaluated network coding in three cases: single-hop transmissions, handovers, and multi-hop transmissions. It has been shown that MRNC achieves a higher transmission efficiency than HARQ as it avoids the problem of ACK/NAK packet overhead and the additional redundancy resulting from their loss. Moreover, [98] defines the conditions for the optimality of coding for delay sensitive traffic, then shows that the MSNC scheme is the optimum NC scheme for delay sensitive applications in WiMAX unicast.

The second scenario we are going to address in the second paradigm of this PhD thesis is network coding for the MAC layer of the WiFi IEEE 802.11. In this scenario a station is performing handover from one AP to another.

3.4 Methodology

In this section we present the methodology that describes the mathematical foundations we used to model the scenarios in the second paradigm in order to evaluate the delay. We will first give a simple introduction about Markov chains and provide basic definitions. Then we will present absorption Markov chains [99], the type of Markov chains we used to measure the mean completion time. We describe a Markov chain as follows: We have a set of states $S = \{s_1, s_2, ..., s_r\}$. The process starts in one of these states and moves successively from one state to another. Each move is called a step. If the chain is currently in state s_i, then it moves to state s_j at the next step with a probability denoted by p_{ij}, and this probability does not depend upon which states the chain was in before the current state. The probabilities p_{ij} are called transition probabilities. The process can remain in the state it is in, and this occurs with probability p_{ii}. An initial probability distribution, defined on S, specifies the starting state. The transition matrix $\mathbf{P}$ is a matrix whose elements are the transition probabilities between all states to all states. Therefore, if the number of states are r, the transition matrix will be of size r x r.

Definition 1: Let $\mathbf{P}$ be the transition matrix of a Markov chain. The ijth entry p_{ij}^n of the matrix $\mathbf{P}^n$ gives the probability that the Markov chain, starting in state s_i, will be in state s_j after n steps.

Definition 2: A state s_j of a Markov chain is called absorbing if $p_{jj} = 1$, that is, it is impossible to leave it once reached. A Markov chain is absorbing if it has at least one absorbing state, and if from every state it is possible to go to an absorbing state (not necessarily in one step).

Definition 3: In an absorbing Markov chain, a state which is not absorbing is called transient.

In an absorbing Markov chain, the probability that the process will be absorbed is 1. From each non-absorbing state s_i it is possible to reach an absorbing state. Let n_i be the minimum number of steps required to reach an absorbing state, starting from s_i. Let p_i be the probability that, starting from s_i, the process will not reach an absorbing state in n_i steps. Then $p_i < 1$. Let m be the largest of the n_i and let p be

the largest of p_i. The probability of not being absorbed in n steps is less than or equal to p, in $2n$ steps less than or equal to p^2, etc. Since $p < 1$ these probabilities tend to 0. Since the probability of not being absorbed in n steps is decreasing monotonically, these probabilities also tend to 0, i.e., $p_i \to 1$.

Definition 4: For an absorbing Markov chain with transition matrix $\mathbf{P}$, the matrix $\mathbf{N}$ is called the fundamental matrix for $\mathbf{P}$. The n_{ij}th entry of $\mathbf{N}$ gives the expected number of times that the process is in the transient state s_j if it started in the transient state s_i.

Definition 5: Let $\mathbb{E}[T_i]$ be the expected number of steps before the chain is absorbed, given that the chain starts in state s_i, and let T be the column vector whose ith entry is T_i. Then:

$$\mathbb{E}[T] = \mathbf{N}c, \tag{3.1}$$

where c is a column vector all of whose entries are 1. Therefore, the expected time to absorption starting from state s_i is given by:

$$\mathbb{E}[T_i] = 1 + \sum_{i=1}^{r} p_{ij}\mathbb{E}[T_j], \tag{3.2}$$

with $\mathbb{E}[T_k] = 0$ if k is the absorption state.

The second paradigm in this research work will capitalize on the Markov modeling of the scenarios of interest. In particular, we use the absorption Markov chains to address the problem of developing and optimizing transmission schemes that introduce network coding in order to minimize the expected delay associated with the scenarios considered. It is important to highlight that the mean time to absorption provides means to have closed-forms of the delay encountered mapping the physical system into states of transmission and reception and under practical assumptions of finite number of resources. This provides a novel framework for modeling the system under study as well as optimizing the transmission schemes or integrating the solutions into current existing technologies.

3.5 Conclusion

In this Chapter, we give a brief overview of network coding approaches at the multiple layers as well as prior work in the area of delay probabilistic models. In particular, we present main contributions that addressed the delay problem in networks, with classical design and with the introduction of network coding. We present state of the art work on network coding for wireless channels and the diversity gains obtained via network coding, similar to the first paradigm which addresses precoding as a transmit diversity technique. In addition, we show the directions of the research towards physical layer aware designs, and their tendency to integrate such designs to upper layers. Therefore, we focus on prior works on delay-throughput enhancements on multiple access networks, with classical design and with the introduction of network coding. Finally, we present the main methodology adopted to devise analytical models of the scenarios under study. We established a direct connection between such modeling methodology, and their relation to the delay problem addressed throughout the second part of this thesis.

In spite of the numerous theoretical contributions and practical setups, the general problem of designing transmission schemes in conjunction with network coding for minimizing the delay over wireless channels is still widely open. Moreover, in delay challenging environments, where understanding the existing setups from a modeling perspective, evaluating the current systems, and proposing new enhancement layers is of long-standing interest. The objective of the second part of this PhD research is to fill this vacuum, attacking the delay problem with a network coding-probabilistic approach. In particular, we focus the last part of this PhD research work in solving the delay-throughput problem in two challenging scenarios. The first is the delay performance in a time varying satellite environment, where network coding scheme plays a significant role in improving their performance. The second is the delay performance of a station performing handover in a WiFi network under mobility.

Chapter 4
MAC Gaussian Channels with Arbitrary Inputs: Optimal Precoding and Power Allocation

4.1 Introduction

Linear precoding for mutually interfering channels corrupted with independent Gaussian noise can be maximized under a total power constraint [100]. In the multiple user setting, the mutual information is maximized by imposing a covariance structure on the Gaussian input vector that satisfies two conditions: First, transmitting along the right eigenvectors of the channel. Second, the power allocation follows the waterfilling based on the SNR in each parallel non-interacting sub-channel [43], [37]. The MAC channel stands as a special case of interference channels. With arbitrary inputs, the mutual information of the MAC channel is minimized due to the fact that at a certain point the inputs may lie in the null space of the channel matrix, and this causes the mutual information to decay at such point. Therefore, the capacity-achieving strategies can be employed either at the transmitter side and/or the receiver side: by orthogonalizing the inputs using TDM/FDM techniques, or via using interference mitigation techniques, such as, dealing with interference as noise, rate splitting, successive interference cancellation, decoding interference, as well as receive diversity techniques [101], [102], [103], [104], [105], [106]. Precoding is another technique that can be used at the transmitter side to maximize the mutual information by exploiting the transmit diversity. On the one hand, precoding plays the role of power allocation diagonalizing the channel while, on the other hand, rotating the channel eigenvectors; thus shaping the constellation for better detection. Since practical considerations enforce the use of discrete constellations, such as PSK and QAM, which depart

significantly from the optimal Gaussian distributions, it is of great importance to revisit the linear precoding problem under the constraint that each input follows a specified discrete constellation. In [37], [31], and [107], the optimal power allocation for parallel non-interacting channels with arbitrary inputs, and the optimal precoding for MIMO channels with different setups were obtained exploiting the relation between the mutual information and the MMSE [31], [30]. The results in [100] demonstrate that capacity-achieving strategies for Gaussian inputs may be suboptimal for discrete inputs. In [37] a general solution for the optimum precoder has been derived for the high and low-snr regimes capitalizing on first order expansions for the MMSE and relating it to mutual information. The mercury/waterfilling solution for power allocation maximizes the mutual information compensating for the non-Gaussianess for the discrete constellations [35]. However, linear precoding techniques introduce rescaling among the channel inputs that achieves cleaner detection of the received signals and thus higher information rates. In this Chapter, we propose a linear precoder structure that aims to maximize the input output mutual information for the two-user MAC setting with channel state information known at both receiver and transmitter sides.

In this Chapter, (i) we derive new closed-form expressions for the gradient of the mutual information with respect to arbitrary parameters of the two-user MAC. (ii) we investigate the linear precoding and power allocation policies that maximize the mutual information for the two-user MAC Gaussian channels with arbitrary input distributions, capitalizing on the relationship between mutual information and MMSE. (i) verifies for the first time the MAC channel applicability of the fundamental relations between information theory and estimation theory. And (ii) verifies that the optimal design of linear precoders may satisfy a fixed-point equation as a function of the channel and the input constellation under specific setups. We provide a novel interpretation for the interference with respect to the channel, power, and input estimates of the main user and the interferer.

4.2 Problem Formulation

Consider the deterministic complex-valued vector channel,

$$\mathbf{y} = \sqrt{\mathbf{snr}}\ \mathbf{H_1 P_1 x_1} + \sqrt{\mathbf{snr}}\ \mathbf{H_2 P_2 x_2} + \mathbf{n}, \tag{4.1}$$

where the n_r x 1 dimensional vector $\mathbf{y}$ and the n_t x 1 dimensional vectors $\mathbf{x_1}$, $\mathbf{x_2}$ rep-

resent, respectively, the received vector and the independent zero-mean unit-variance transmitted information vectors from each user input to the MAC channel. The distributions of both inputs are not fixed, not necessarily Gaussian nor identical. The n_r x n_t complex-valued matrices $\mathbf{H_1}$, $\mathbf{H_2}$ correspond to the deterministic channel gains for both input channels (known to both encoder and decoder) and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is the n_r x 1 dimensional complex Gaussian noise with independent zero-mean unit-variance components. The optimization of the mutual information is carried out over all n_t x n_t precoding matrices $\mathbf{P_1}$, $\mathbf{P_2}$ that do not increase the transmitted power. The precoding problem can be cast as a constrained non-linear optimization problem as follows:

$$max \quad I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y}) \tag{4.2}$$

Subject to:

$$Tr\left\{\mathbb{E}[\mathbf{P_1 x_1 x_1^\dagger P_1}^\dagger]\right\} = Tr\left\{\mathbf{P_1 P_1}^\dagger\right\} \leq Q_1 \tag{4.3}$$

$$Tr\left\{\mathbb{E}[\mathbf{P_2 x_2 x_2^\dagger P_2}^\dagger]\right\} = Tr\left\{\mathbf{P_2 P_2}^\dagger\right\} \leq Q_2 \tag{4.4}$$

To solve the optimization problem in (4.2) subject to (4.3) and (4.4), we need to know the relation between the mutual information and the non-linear MMSE in the context of MAC channels. The following theorem puts forth a new fundamental relation between the mutual information and the non-linear MMSE for MAC channels. In particular, the gradient of the mutual information with respect to arbitrary parameters has been derived. First we will present the gradient of the mutual information, with respect to the channel in Theorem 1, then with respect to the precoder (or power allocation) in Theorem 2.

Theorem 1. *The relation between the gradient of the mutual information with respect to the channel and the non-linear MMSE for a two user-MAC channel with arbitrary inputs* (4.1) *satisfies:*

$$\nabla_{\mathbf{H_1}} \mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y}) = \mathbf{H_1 P_1 E_1 P_1}^\dagger - \mathbf{H_2 P_2} \mathbb{E}[\widehat{\mathbf{x_2}}\widehat{\mathbf{x_1}}^\dagger]\mathbf{P_1}^\dagger \tag{4.5}$$

$$\nabla_{\mathbf{H_2}} \mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y}) = \mathbf{H_2 P_2 E_2 P_2^\dagger} - \mathbf{H_1 P_1} \mathbb{E}[\widehat{\mathbf{x_1}}\widehat{\mathbf{x_2}}^\dagger]\mathbf{P_2}^\dagger \tag{4.6}$$

Proof. See Appendix A. □

Theorem 2. *The relation between the gradient of the mutual information with respect to the precoding matrix and the non-linear MMSE for a two user-MAC channel with arbitrary inputs* (4.1) *satisfies:*

$$\nabla_{\mathbf{P_1}} \mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y}) = \mathbf{H_1^\dagger H_1 P_1 E_1} - \mathbf{H_1^\dagger H_2 P_2} \mathbb{E}[\widehat{\mathbf{x_2}}\widehat{\mathbf{x_1}}^\dagger] \tag{4.7}$$

$$\nabla_{\mathbf{P_2}} \mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y}) = \mathbf{H_2^\dagger H_2 P_2 E_2} - \mathbf{H_2^\dagger H_1 P_1} \mathbb{E}[\widehat{\mathbf{x_1}}\widehat{\mathbf{x_2}}^\dagger] \tag{4.8}$$

Proof. See Appendix B. □

The per-user MMSE is given respectively as follows:

$$\mathbf{E_1} = \mathbb{E}[(\mathbf{x_1} - \widehat{\mathbf{x_1}})(\mathbf{x_1} - \widehat{\mathbf{x_1}})^\dagger] \tag{4.9}$$

$$\mathbf{E_2} = \mathbb{E}[(\mathbf{x_2} - \widehat{\mathbf{x_2}})(\mathbf{x_2} - \widehat{\mathbf{x_2}})^\dagger]. \tag{4.10}$$

The input estimates of each user input is given respectively as follows:

$$\widehat{\mathbf{x_1}} = \mathbb{E}[\mathbf{x_1}|\mathbf{y}] = \sum_{\mathbf{x_1},\mathbf{x_2}} \frac{\mathbf{x_1}\mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\mathbf{p}_{\mathbf{x_1}}(\mathbf{x_1})\mathbf{p}_{\mathbf{x_2}}(\mathbf{x_2})}{\mathbf{p}_{\mathbf{y}}(\mathbf{y})} \tag{4.11}$$

$$\widehat{\mathbf{x_2}} = \mathbb{E}[\mathbf{x_2}|\mathbf{y}] = \sum_{\mathbf{x_1},\mathbf{x_2}} \frac{\mathbf{x_2}\mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\mathbf{p}_{\mathbf{x_1}}(\mathbf{x_1})\mathbf{p}_{\mathbf{x_2}}(\mathbf{x_2})}{\mathbf{p}_{\mathbf{y}}(\mathbf{y})}. \tag{4.12}$$

The conditional probability distribution of the Gaussian noise is defined as:

$$\mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) = \frac{\mathbf{1}}{\pi^{\mathbf{n_r}}}\mathbf{e}^{-\left\|\mathbf{y}-\sqrt{\mathbf{snr}}\mathbf{H_1}\mathbf{P_1}\mathbf{x_1}-\sqrt{\mathbf{snr}}\mathbf{H_2}\mathbf{P_2}\mathbf{x_2}\right\|^2} \tag{4.13}$$

The probability density function for the received vector $\mathbf{y}$ is defined as:

$$\mathbf{p}_{\mathbf{y}}(\mathbf{y}) = \sum_{\mathbf{x_1},\mathbf{x_2}} \mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\mathbf{p}_{\mathbf{x_1}}(\mathbf{x_1})\mathbf{p}_{\mathbf{x_2}}(\mathbf{x_2}). \tag{4.14}$$

Henceforth, the system MMSE with respect to the SNR is given by:

$$mmse(snr) = \mathbb{E}\left[\left\|\mathbf{H_1}\mathbf{P_1}(\mathbf{x_1} - \mathbb{E}[\mathbf{x_1}|\mathbf{y}])\right\|^2\right] + \mathbb{E}\left[\left\|\mathbf{H_2}\mathbf{P_2}(\mathbf{x_2} - \mathbb{E}[\mathbf{x_2}|\mathbf{y}])\right\|^2\right] \tag{4.15}$$

$$= Tr\left\{\mathbf{H_1}\mathbf{P_1}\mathbf{E_1}(\mathbf{H_1}\mathbf{P_1})^\dagger\right\} + Tr\left\{\mathbf{H_2}\mathbf{P_2}\mathbf{E_2}(\mathbf{H_2}\mathbf{P_2})^\dagger\right\} \tag{4.16}$$

Note that we can derive this relation with respect to any arbitrary parameter following similar steps of the proof. Note also that the derived relations in (4.5), (4.6), (4.7), and (4.8) reduce to the relation between the gradient of the mutual information and the non-linear MMSE derived for the linear vector Gaussian channels [31] if the multiplication of the estimates of the inputs is zero. In other words, if and only if the users inputs were orthogonalized or interference cancellation is performed. However, the derived relation between the gradient of the mutual information and the MMSE of the two-user MAC setting in Theorem 1 will lead to a new formulation of the input non-linear estimates, that is discussed in the following Theorem.

Theorem 3. *The estimates of the inputs* $\mathbf{x_1}$ *and* $\mathbf{x_2}$ *of the two-user MAC channel with arbitrary inputs given the output* $\mathbf{y}$ *can be expressed as:*

$$\mathbf{H_1 P_1}\mathbb{E}[\mathbf{x_1}|\mathbf{y}] + \mathbf{H_2 P_2}\mathbb{E}[\mathbf{x_2}|\mathbf{y}] = \mathbf{y} + \frac{\nabla_{\mathbf{y}}\mathbf{p}_{\mathbf{y}}(\mathbf{y})}{\mathbf{p}_{\mathbf{y}}(\mathbf{y})} \tag{4.17}$$

Proof. See Appendix C. □

Equation (4.17) can be solved numerically to obtain the estimates of the inputs. However, we can extract a special case of Theorem 3 in the following lemma.

Lemma 1. *If the channel is a linear vector Gaussian channel, Theorem 3 tends to the following formulation:*

$$\mathbb{E}[\mathbf{x}|\mathbf{y}] = \left(\mathbf{y} + \frac{\nabla_{\mathbf{y}}\mathbf{p}_{\mathbf{y}}(\mathbf{y})}{\mathbf{p}_{\mathbf{y}}(\mathbf{y})}\right)(\mathbf{HP})^{-1} \tag{4.18}$$

The non-linear estimates given in (4.11) and (4.12) give a statistical intuition to the problem. However, in practical setups, such estimates can be found via the linear MMSE. In particular, the input estimates can be found by deriving the optimal Wiener receive filters solving a minimization optimization problem of the MSE. The following theorem provides the linear estimates.

Theorem 4. *The linear estimates of the inputs* $\mathbf{x_1}$*,* $\mathbf{x_2}$ *of the two-user MAC channel given the output* $\mathbf{y}$ *can be expressed as:*

$$\widehat{\mathbf{x_1}} = \mathbf{P_1^\dagger H_1^\dagger}(\mathbf{I} + \mathbf{P_1^\dagger H_1^\dagger P_1 H_1} + \mathbf{P_2^\dagger H_2^\dagger P_2 H_2})^{-1}\mathbf{y} \tag{4.19}$$

$$\widehat{\mathbf{x_2}} = \mathbf{P_2^\dagger H_2^\dagger}(\mathbf{I} + \mathbf{P_1^\dagger H_1^\dagger P_1 H_1} + \mathbf{P_2^\dagger H_2^\dagger P_2 H_2})^{-1}\mathbf{y} \tag{4.20}$$

Proof. See Appendix D. □

Worth to note that the linear estimation process of the inputs corresponds to multiplying the user Wiener MMSE receive filters to the received vector. Note that each receive filter contains the precoding (or power allocation), the user channel, in addition to a structure of the linear MMSE, that is, the inverse of the covariance structure of the received vector.

4.3 Optimal Precoding with Arbitrary Inputs

Capitalizing on the relation between the gradient of the mutual information with respect to the precoding matrix and the MMSE, we can solve the optimization problem in (4.2) subject to (4.3) and (4.4). The following theorem provides the optimal precoding of the two-user MAC channel with arbitrary inputs which can be also specialized to Gaussian inputs as will be seen later.

Theorem 5. *Let the distribution of the arbitrary inputs* $\mathbf{x_1}$, $\mathbf{x_2}$ *to the channel in* (4.1) *be* $\mathbf{p_{x_1}}(\mathbf{x_1})$, $\mathbf{p_{x_2}}(\mathbf{x_2})$, *respectively. The optimal precoding matrices satisfy the following:*

$$\mathbf{P}_1{}^{\star} = \nu^{-1}\mathbf{H}_1^{\dagger}\mathbf{H_1}\mathbf{P_1}{}^{\star}\mathbf{E_1} - \nu^{-1}\mathbf{H}_1^{\dagger}\mathbf{H_2}\mathbf{P_2}{}^{\star}\mathbb{E}[\widehat{\mathbf{x_2}}\widehat{\mathbf{x_1}}^{\dagger}] \tag{4.21}$$

$$\mathbf{P}_2{}^{\star} = \nu^{-1}\mathbf{H}_2^{\dagger}\mathbf{H_2}\mathbf{P_2}{}^{\star}\mathbf{E_2} - \nu^{-1}\mathbf{H}_2^{\dagger}\mathbf{H_1}\mathbf{P_1}{}^{\star}\mathbb{E}[\widehat{\mathbf{x_1}}\widehat{\mathbf{x_2}}^{\dagger}] \tag{4.22}$$

Where ν are the received *snr* normalized by the Lagrange multipliers of the optimization problem in (4.2).

Proof. The proof of Theorem 5 relies on the KKT conditions [48] and the relation between the gradient of the mutual information with respect to the precoding matrices and the MMSE, see Appendix E. □

There are unique $\mathbf{P_1}^{\star}$, $\mathbf{P_2}^{\star}$ that satisfy the KKT conditions when the problem is strictly concave, corresponding to the global maximum. Using Monte-Carlo method, we can compute the MMSE matrices (4.9) and (4.10) for discrete constellations. To obtain the solution of (4.21) and (4.22), we can use an iterative approach, similar to [100], [37], and [107]. It is very important to notice that the optimal precoders of (4.21) and (4.22) satisfies a fixed point equation under two specific setups: *First*, for the multiple-channels-per user in the two-user-MAC if and only if the second term in the gradient of the mutual information with respect to the precoding matrix is zero, this occurs when both inputs are orthogonal, or in this case, we can think about it as if the second user multiple-inputs have no influence to the first user multiple inputs, therefore, each user precoding is over his mutually interfering channels only. If each user is using OFDM signaling, i.e., transmitting over parallel independent Gaussian channels, the problem breaks into two separate problems, and the precoding solution for each is a fixed point equation. *Second*, for the SISO-two-user-MAC case with each user is mutually interfering with the other, in this case the relation between the gradient of the mutual information with respect to the precoding matrix can

be manipulated as in [31], thus, the precoder structure is a special case of the one introduced in [37]. We define in the following theorem a general form of the optimal precoder's structure for the two-user MAC channel in (4.21) and (4.22) under the two specific setups where the precoder would satisfy a fixed point equation, therefore, the solution applies to the low-snr and high-snr regimes. However, it also reduces to the optimal power allocation solution when the precoding is precluded as will be shown later. Therefore, the precoder should admit a structure that performs matching of the strongest source modes to the weakest noise modes, and this alignment enforces permutation process to appear in the diagonal power allocation setup.

Theorem 6. *The non-unique first-order optimal precoders that maximize the mutual information for a two-user MAC subject to per-user power constraint can be written as follows:*

$$\mathbf{P_1}^{\star} = \mathbf{U_1 D_1 R_1^{\dagger}} \tag{4.23}$$

$$\mathbf{P_2}^{\star} = \mathbf{U_2 D_2 R_2^{\dagger}} \tag{4.24}$$

With $\mathbf{U_1}$, $\mathbf{U_2}$ *are unitary matrices,* $\mathbf{D_1}$, $\mathbf{D_2}$ *are diagonal matrices, and* $\mathbf{R_1}$, $\mathbf{R_2}$ *are rotation matrices.*

Proof. The theorem follows the relation between the gradient of the mutual information with respect to the precoding matrix and the decomposition of the components of the channel, precoder, and MMSE matrices, and is a direct consequence to [Theorem 1, 5], see Appendix F. □

The optimal precoders are as follows, $\mathbf{P_1} = \mathbf{U_1 D_1 R_1^{\dagger}}$ and $\mathbf{P_2} = \mathbf{U_2 D_2 R_2^{\dagger}}$ with $\mathbf{U_1} = \mathbf{V_{H_1}}$, $\mathbf{U_2} = \mathbf{V_{H_2}}$, correspond to the per user channel right singular vectors, respectively. $\mathbf{D_1} = \mathbf{diag}(\sqrt{\mathbf{p_{1,1}}}, \ldots, \sqrt{\mathbf{p_{1,n_t}}})$ and $\mathbf{D_2} = \mathbf{diag}(\sqrt{\mathbf{p_{2,1}}}, \ldots, \sqrt{\mathbf{p_{2,n_t}}})$ are power allocation matrices, such that $\mathbf{p_{l,m}}$ is admitted by user l over sub/channel m; this in fact corresponds to the mercury/waterfilling. $\mathbf{R_1} = \mathbf{\Pi U_{E_1}}$ and $\mathbf{R_2} = \mathbf{\Pi U_{E_2}}$ are rotation matrices that contain in their structure the eigenvectors of the user MMSE matrix which can be permuted and/or projected with Π based on the correlation of the inputs and their estimates. Therefore, the rotation matrix insures firstly, allocation of power into the strongest channel singular vectors, and secondly, diagonalizes the MMSE matrix to insure un-correlating the error or in other words independence between inputs, enforcing the setup defined in (4.21) and (4.22) for the two special setups discussed previously about each user inputs.

4.4 The Low-SNR Regime

We now consider the optimal precoding policy for the two-user MAC Gaussian channel with arbitrary input distributions in the regime of low-snr. Consider a zero-mean uncorrelated complex inputs, with $\mathbb{E}[\mathbf{x_1}\mathbf{x_1}^\dagger] = \mathbb{E}[\mathbf{x_2}\mathbf{x_2}^\dagger] = \mathbf{I}$, and $\mathbb{E}[\mathbf{x_1}\mathbf{x_1}^\mathbf{T}] = \mathbb{E}[\mathbf{x_2}\mathbf{x_2}^\mathbf{T}] = \mathbf{0}$. We consider the low-snr expansion to the MMSE of equation (4.15). Note that it can be easily deduced that the Taylor expansion of the non−linear MMSE in (4.15) will lead to the first order Taylor expansion of the linear MMSE for the Gaussian inputs setup. Thus, the low-snr expansion of the MMSE matrix can be expressed as:

$$\mathbf{E} = \mathbf{I} - (\mathbf{H_1P_1})^\dagger \mathbf{H_1P_1}.snr - (\mathbf{H_2P_2})^\dagger \mathbf{H_2P_2}.snr + \mathcal{O}(snr^2), \tag{4.25}$$

with $\mathbf{E} = \mathbf{E_1} + \mathbf{E_2}$. Consequently,

$$mmse(snr) = Tr\left\{\mathbf{H_1P_1E_1}(\mathbf{H_1P_1})^\dagger\right\} + Tr\left\{\mathbf{H_2P_2E_2}(\mathbf{H_2P_2})^\dagger\right\} \tag{4.26}$$

$$= Tr\left\{\mathbf{H_1P_1}(\mathbf{H_1P_1})^\dagger\right\} + Tr\left\{\mathbf{H_2P_2}(\mathbf{H_2P_2})^\dagger\right\}$$
$$- Tr\left\{(\mathbf{H_1P_1}(\mathbf{H_1P_1})^\dagger)^\mathbf{2}\right\}.snr - Tr\left\{(\mathbf{H_2P_2}(\mathbf{H_2P_2})^\dagger)^\mathbf{2}\right\}.snr + \mathcal{O}(snr^2). \tag{4.27}$$

Capitalizing on the relationship between mutual information and MMSE [31], [30], the low-snr Taylor expansion of the mutual information is given by:

$$I(snr) = Tr\left\{\mathbf{H_1P_1}(\mathbf{H_1P_1})^\dagger\right\}.snr + Tr\left\{\mathbf{H_2P_2}(\mathbf{H_2P_2})^\dagger\right\}.snr$$
$$- Tr\left\{(\mathbf{H_1P_1}(\mathbf{H_1P_1})^\dagger)^\mathbf{2}\right\}.\frac{snr^2}{2} - Tr\left\{(\mathbf{H_2P_2}(\mathbf{H_2P_2})^\dagger)^\mathbf{2}\right\}.\frac{snr^2}{2} + \mathcal{O}(snr^3). \tag{4.28}$$

The wideband slope − which indicates how fast the capacity is achieved in terms of required bandwidth − is inversely proportional to the second order terms of the mutual information in the low-snr Taylor expansion (4.28). Therefore, this term is a key low-power performance measure since the bandwidth required to sustain a given rate with a given low power, i.e., minimal energy per bit, is inversely proportional to this term [108]. According to (4.28), for first-order optimality, subject to (4.3) and (4.4), the form of the optimal precoder follows from the low-snr expansions the form of optimal precoder for the complex Gaussian inputs settings. To prove this claim, lets re-define our optimization problem as follows:

$$max \quad Tr\left\{\mathbf{H_1P_1}(\mathbf{H_1P_1})^\dagger\right\}.snr + Tr\left\{\mathbf{H_2P_2}(\mathbf{H_2P_2})^\dagger\right\}.snr \tag{4.29}$$

Subject to:

$$Tr\left\{\mathbf{P_1}\mathbf{P_1}^{\dagger}\right\} \leq 1 \tag{4.30}$$

$$Tr\left\{\mathbf{P_2}\mathbf{P_2}^{\dagger}\right\} \leq 1 \tag{4.31}$$

Lets do an eigen value decomposition such that:

$$\mathbf{H_1}\mathbf{H_1^{\dagger}} = \mathbf{U}_{\mathbf{H_1}\mathbf{H_1^{\dagger}}}\mathbf{\Lambda}_{\mathbf{H_1}\mathbf{H_1^{\dagger}}}\mathbf{U}^{\dagger}_{\mathbf{H_1}\mathbf{H_1^{\dagger}}}, \tag{4.32}$$

$$\mathbf{H_2}\mathbf{H_2^{\dagger}} = \mathbf{U}_{\mathbf{H_2}\mathbf{H_2^{\dagger}}}\mathbf{\Lambda}_{\mathbf{H_2}\mathbf{H_2^{\dagger}}}\mathbf{U}^{\dagger}_{\mathbf{H_2}\mathbf{H_2^{\dagger}}} \tag{4.33}$$

Let:

$$\tilde{\mathbf{P_1}} = \mathbf{U}^{\dagger}_{\mathbf{H_1}\mathbf{H_1^{\dagger}}}\mathbf{P_1} \tag{4.34}$$

$$\tilde{\mathbf{P_2}} = \mathbf{U}^{\dagger}_{\mathbf{H_2}\mathbf{H_2^{\dagger}}}\mathbf{P_2} \tag{4.35}$$

Let $\mathbf{Z_1} \succeq \mathbf{0}$ and $\mathbf{Z_2} \succeq \mathbf{0}$ are such that:

$$\mathbf{Z_1} = \tilde{\mathbf{P}}_1\tilde{\mathbf{P}}_1^{\dagger} \tag{4.36}$$

$$\mathbf{Z_2} = \tilde{\mathbf{P}}_2\tilde{\mathbf{P}}_2^{\dagger} \tag{4.37}$$

Substitute (4.32) to (4.37) into (4.29), (4.30), and (4.31), the optimization problem can be re-written as:

$$max \quad Tr\left\{\mathbf{Z_1}\mathbf{\Lambda}_{\mathbf{H_1}\mathbf{H_1^{\dagger}}}\right\}.snr + Tr\left\{\mathbf{Z_2}\mathbf{\Lambda}_{\mathbf{H_2}\mathbf{H_2^{\dagger}}}\right\}.snr \tag{4.38}$$

Subject to:

$$Tr\left\{\mathbf{Z_1}\right\} \leq 1 \tag{4.39}$$

$$Tr\left\{\mathbf{Z_2}\right\} \leq 1 \tag{4.40}$$

Solving the KKT conditions for the Lagrangian of the objective function (4.38) subject to (4.39) and (4.40) leads to:

$$\mathbf{Z_1} = \lambda^{-1}\mathbf{\Lambda}_{\mathbf{H_1}\mathbf{H_1^{\dagger}}}.snr, \tag{4.41}$$

$$\mathbf{Z_2} = \lambda^{-1}\mathbf{\Lambda}_{\mathbf{H_2}\mathbf{H_2^{\dagger}}}.snr \tag{4.42}$$

The results in (4.41) and (4.42) prove that the optimal precoders in the low−snr perform mainly two operations: Firstly, it aligns the transmit directions with the eigenvectors of each sub-channel. Secondly, it performs power allocation over the sub-channels. Where λ is the Lagrange multiplier of the optimization problem.

4.5 Optimal Power Allocation with Arbitrary Inputs

To derive the optimal power allocation for MAC Gaussian channels driven by arbitrary inputs, we need to redefine the same objective function subject to the per user power constraints. The power allocation is a special case of the precoding which inherently includes a power allocation operation. However, the first reshapes the constellation by rotating or redefining the minimum distance between the points, while the second redistributes the power over the channels to guarantee a better utilization of the channel under the defined design criterion. Therefore, the power allocation problem can be cast as a constrained non-linear optimization problem as follows:

$$max \quad I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y}) \tag{4.43}$$

Subject to:

$$(\mathbf{P_1})_{\mathbf{ij}} = \sqrt{\mathbf{p_{1j}}}, \quad (\mathbf{P_2})_{\mathbf{ij}} = \sqrt{\mathbf{p_{2j}}} \tag{4.44}$$

$$\sum_{j=1}^{n_t} \mathbf{p_{1j}} \leq 1, \ \mathbf{p_{1j}} \ \geq \ 0 \ \ \forall j = 1, ..., n_t \tag{4.45}$$

$$\sum_{j=1}^{n_t} \mathbf{p_{2j}} \leq 1, \ \mathbf{p_{2j}} \ \geq \ 0 \ \ \forall j = 1, ..., n_t \tag{4.46}$$

where $\mathbf{p_1} = \mathbf{diag}(\sqrt{\mathbf{p_{1,1}}}, \ ... \ , \sqrt{\mathbf{p_{1,n_t}}})$, and $\mathbf{p_2} = \mathbf{diag}(\sqrt{\mathbf{p_{2,1}}}, \ ... \ , \sqrt{\mathbf{p_{2,n_t}}})$. Capitalizing on the relation between the gradient of the mutual information with respect to the power allocation matrix and the MMSE, we provide the optimal power allocation in the following theorem.

Theorem 7. *The optimal power allocation that solves* (4.43) *subject to* (4.44) *to* (4.46) *satisfies:*

$$\mathbf{p}_1{}^{\star} = \gamma_1^{-1} \left(\mathbf{P_1}^{\star} \mathbf{H}_1^{\dagger} \mathbf{H_1} \mathbf{P_1}^{\star} \mathbf{E_1} - \mathbf{P_1}^{\star} \mathbf{H}_1^{\dagger} \mathbf{H_2} \mathbf{P_2}^{\star} \mathbb{E}[\widehat{\mathbf{x_2}} \widehat{\mathbf{x_1}}^{\dagger}] \right)_{1j} \tag{4.47}$$

$$\mathbf{p}_2{}^{\star} = \gamma_2^{-1} \left(\mathbf{P_2}^{\star} \mathbf{H}_2^{\dagger} \mathbf{H_2} \mathbf{P_2}^{\star} \mathbf{E_2} - \mathbf{P_2}^{\star} \mathbf{H}_2^{\dagger} \mathbf{H_1} \mathbf{P_1}^{\star} \mathbb{E}[\widehat{\mathbf{x_1}} \widehat{\mathbf{x_2}}^{\dagger}] \right)_{2j} \tag{4.48}$$

Where γ_1 and γ_2 are the Lagrange multipliers normalized by the received *snr*, of the optimization problem in (4.43).

Proof. The proof of Theorem 7 relies on the KKT conditions [48] and the relation between the gradient of the mutual information with respect to the precoding matrices and the MMSE, see Appendix G. □

For the special case where we have a SISO per user in the two-user MAC, (4.47) and (4.48) can be simplified to the following:

$$\mathbf{p}_1{}^{\star} = \gamma_1^{-1}\mathbf{p_{1j}}{}^{\star}|\mathbf{h_{1j}}|^2\mathbb{E}[(\mathbf{x_1}-\widehat{\mathbf{x_1}})(\mathbf{x_1}-\widehat{\mathbf{x_1}})^{\dagger}] - \gamma_{\mathbf{1}}^{-\mathbf{1}}\mathbf{p_{2j}}{}^{\star}|\mathbf{h_{1j}}{}^{*}\mathbf{h_{2j}}|\mathbb{E}[(\mathbf{x_1}-\widehat{\mathbf{x_1}})(\mathbf{x_2}-\widehat{\mathbf{x_2}})^{\dagger}] \tag{4.49}$$

$$\mathbf{p}_2{}^{\star} = \gamma_2^{-1}\mathbf{p_{2j}}{}^{\star}|\mathbf{h_{2j}}|^2\mathbb{E}[(\mathbf{x_2}-\widehat{\mathbf{x_2}})(\mathbf{x_2}-\widehat{\mathbf{x_2}})^{\dagger}] - \gamma_{\mathbf{2}}^{-\mathbf{1}}\mathbf{p_{1j}}{}^{\star}|\mathbf{h_{2j}}{}^{*}\mathbf{h_{1j}}|\mathbb{E}[(\mathbf{x_2}-\widehat{\mathbf{x_2}})(\mathbf{x_1}-\widehat{\mathbf{x_1}})^{\dagger}] \tag{4.50}$$

In (4.49) and (4.50), due to the fact that the MSE is less than or equal to 1, and consider the special case where the second covariance term is zero, i.e., no user is interfering with the other, we can express the optimal power allocation with the single user mercury/waterfilling form [35] as follows:

$$\begin{cases} \mathbf{p_1^{\star}} = \frac{1}{snr|h_{1j}|^2} mmse_1^{-1}\left(\frac{\gamma_1}{|h_{1j}|^2}\right) & \boldsymbol{\gamma_1} < |\mathbf{h_{1j}}|^{\mathbf{2}} \\ \mathbf{p_1^{\star}} = 0, & \boldsymbol{\gamma_1} \geq |\mathbf{h_{1j}}|^{\mathbf{2}} \end{cases} \tag{4.51}$$

$$\begin{cases} \mathbf{p_2^{\star}} = \frac{1}{snr|h_{2j}|^2} mmse_2^{-1}\left(\frac{\gamma_2}{|h_{2j}|^2}\right) & \boldsymbol{\gamma_2} < |\mathbf{h_{2j}}|^{\mathbf{2}} \\ \mathbf{p_2^{\star}} = 0, & \boldsymbol{\gamma_2} \geq |\mathbf{h_{2j}}|^{\mathbf{2}} \end{cases} \tag{4.52}$$

There is a unique set $\mathbf{p_1^{\star}}$, $\mathbf{p_2^{\star}}$ that satisfy the KKT conditions when the problem is strictly concave, corresponding to the global maximum.

4.6 Numerical Results

In this section we will introduce a set of illustrative results of the two-user MAC with arbitrary inputs. In particular, we will present the simplified case where the two user channels are SISO. We used Monte-Carlo method to generate the results. Figure 4.1 illustrates the decay in the mutual information at the 45^o when the two BPSK inputs cancel each other as they lie in the null space of the channel; called the Voronoi region, with the channels are such that $h_1 = 1$, and $h_2 = 1$. If the channel gains or the total powers are different, the decay will be shifted into another line in the coordinate axis. However, we can easily check that if we induce orthogonality, for example if $h_1 = 1$ and $h_2 = 1j$, the mutual information will be the maximum achievable one without this decay region. In Figure 4.2 we show the total MMSE for the two-user MAC. It is interesting to notice that the MMSE behavior corresponds to swapping the mutual information curve in Figure 4.1. This can let us visualize the relation between the

gradient of the mutual information and the MMSE [31], [30]. For the case of Gaussian inputs the relation still holds, but the decay encountered in the arbitrary input setup will not exist. This can be easily verified theoretically and by simulation. Notice that the rate loss in the first user mutual information is due to the interference incurred by the second user. This can be understood via the negative second term in the gradient of the mutual information, similarly, for the other user.

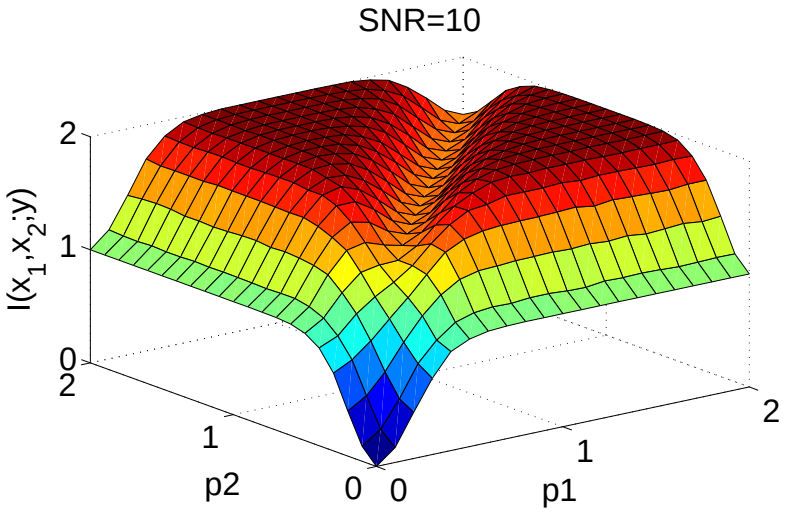

Figure 4.1: The two-user MAC mutual information with BPSK inputs.

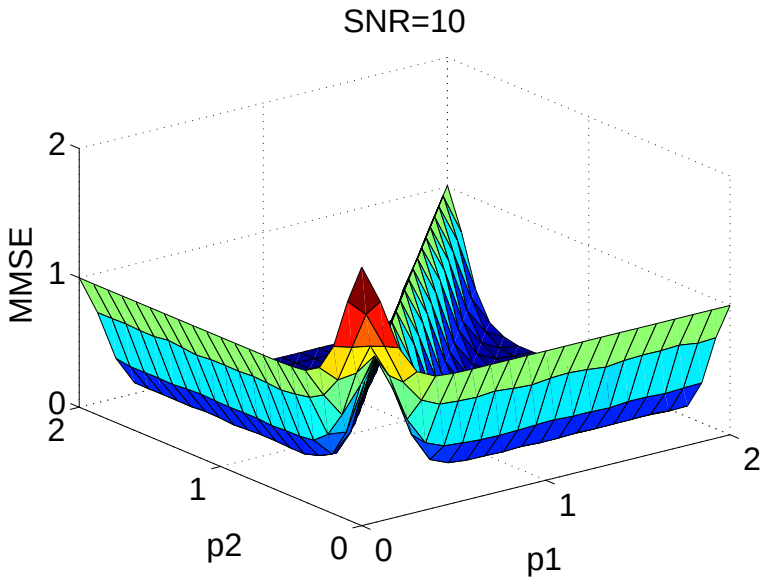

Figure 4.2: The two-user MAC MMSE with BPSK inputs.

This term is a function of the main channel, the interferer channel, the main and interferer powers, and the input estimates. Therefore, the interference from user 1 to user 2 can be interpreted as:

$$\mathbf{H_1}^{\dagger}\mathbf{H_2}\mathbf{P_2}^{\star}\mathbb{E}[\widehat{\mathbf{x_2}}\widehat{\mathbf{x_1}}^{\dagger}], \tag{4.53}$$

where we can re-write equation (4.53) for the SISO input case in terms of the covariance of the interferer as follows:

$$\frac{1}{snr\ h_1^{\dagger}h_2}cov(snr\ h_1^{\dagger}h_2p_2^{\star}), \tag{4.54}$$

And the interference from user 2 to user 1 can be interpreted as:

$$\mathbf{H_2^{\dagger}}\mathbf{H_1}\mathbf{P_1}^{\star}\mathbb{E}[\widehat{\mathbf{x_1}}\widehat{\mathbf{x_2}}^{\dagger}], \tag{4.55}$$

where we can re-write equation (4.55) for the SISO input case in terms of the covariance of the interferer as follows:

$$\frac{1}{snr\ h_2^{\dagger}h_1}cov(snr\ h_2^{\dagger}h_1p_1^{\star}). \tag{4.56}$$

Figure 4.3 and Figure 4.4 illustrate the MMSE per user in the two-user MAC channel. MMSE1 and MMSE2 correspond to the mean-squared error of the first and second user respectively. It is worth to note that the sum of the two MMSEs leads to the total MMSE in Figure 4.2. Figure 4.5 illustrates the covariance of one interferer into the main user. We can see the negative term that leads to the loss incurred in the achieved mutual information.

Finally, Figure 4.6 presents the optimal power allocation for the two-user MAC. When the total power for the first user is larger than that for the second user, the optimal power allocation for the first will be to allocate its total power, as far as the second user doesn't extremely interfere.

However, for the second user, the optimal power allocation will start with an allocation of the total power, then iteratively optimize it by decreasing it. Therefore, precoding is of great importance to such cases where we can align the transmit directions - or caused interference - and maximize the information rates.

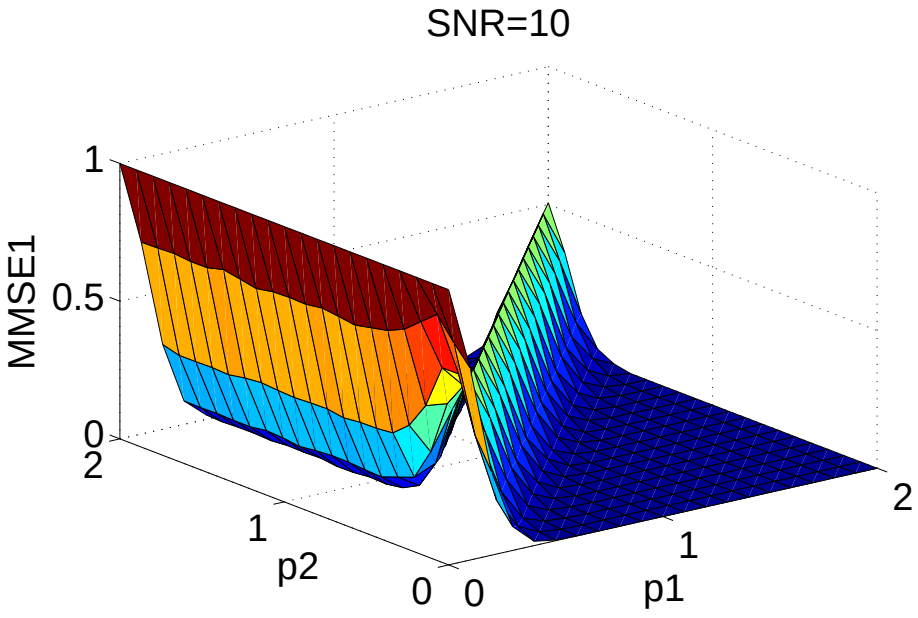

Figure 4.3: MMSE1 for the first user of the two-user MAC with BPSK input.

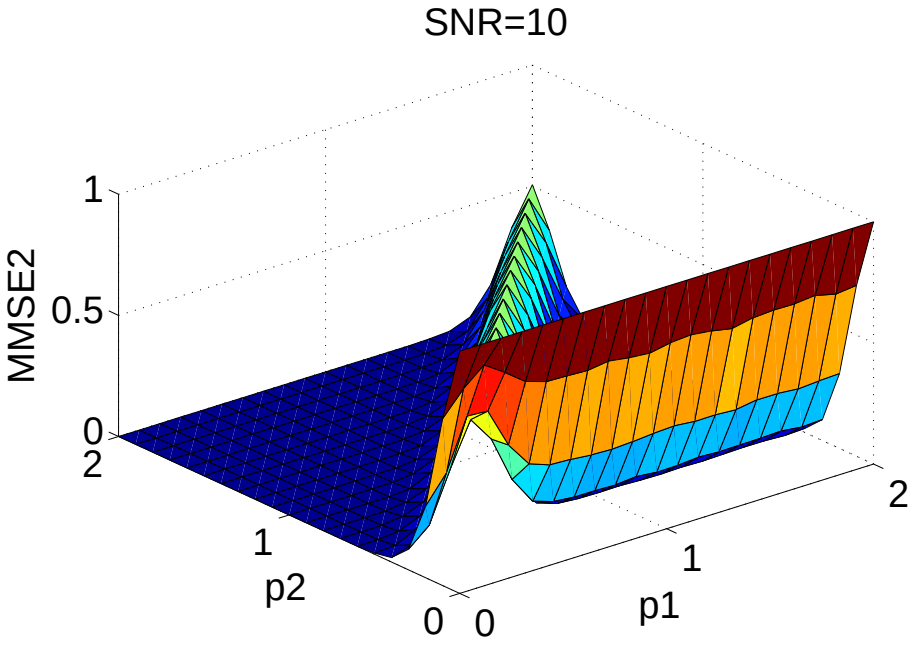

Figure 4.4: MMSE2 for the second user of the two-user MAC with BPSK input.

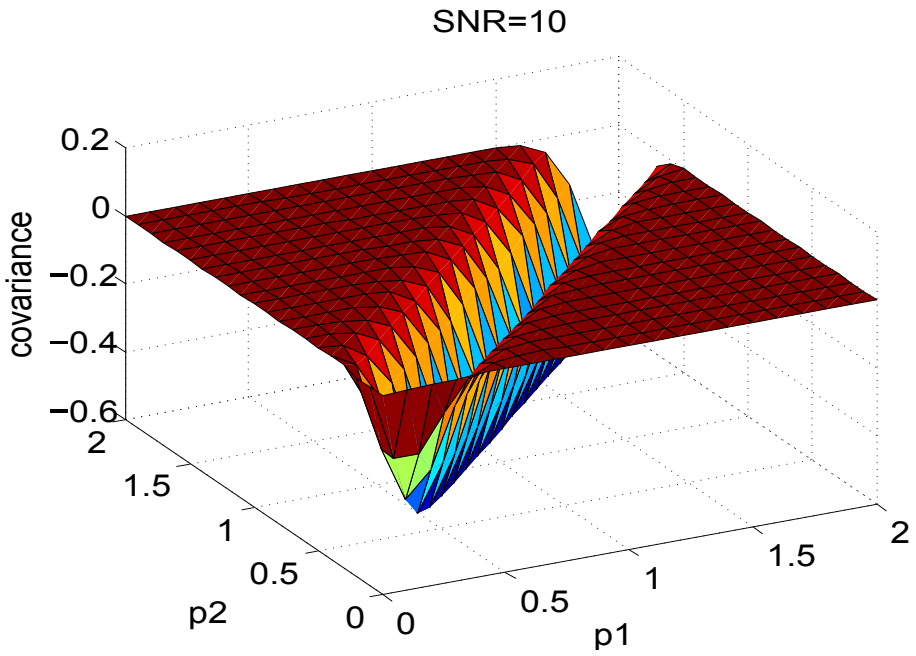

Figure 4.5: Covariance for the first user of the two-user MAC with BPSK input.

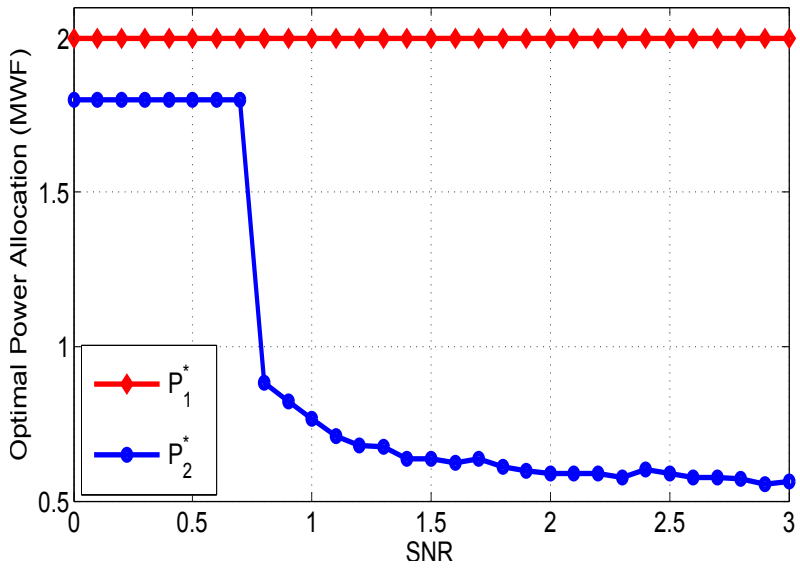

Figure 4.6: Optimal power allocation of the two user MAC with BPSK inputs.

4.7 Conclusion

In this Chapter, we studied the two-user MAC with arbitrary inputs. We derived the relation between the gradient of the mutual information and the MMSE for any input setup. We build upon derivations for the optimal precoding and optimal power allocation. We specialize our results to different cases of the input users, where for the simple case when each user transmits over a SISO channel, the relations can be mapped to the ones for the linear vector Gaussian channels with mutually interfering inputs [100], [31] and then, the optimal precoder can follow a fixed point equation, [37]. A more simplification to the setup where both inputs don't interfere with each other leads to the mercury/waterfilling interpretation for the single user case with binary inputs [35]. This casts further insights to have an interpretation of the interference as discussed. As well, it casts further research questions in quantifying the losses incurred in the mutual information due to interference with and without precoding, and in conjunction with error correcting codes. Of particular relevance are the implications of our derived new fundamental relation in the study of the fundamental limits of cooperation, if a multi-cell processing framework is to be implemented between the BSs in a network. In particular, the interpretation of the interference in the studied two-user MAC, which is basically extracted from the gradient formula, can be extended into proofs for the k-user MAC setups where the impact of interference incurs higher losses into the information rates. Therefore, this have direct impact on the optimal designs of power allocation and optimal precoding. In fact, a global CSI and/or data sharing will not be possible, and this enforces a framework where the MCP should be implemented in clusters of known size. Furthermore, it is straightforward to think about setups where the cooperation and a clustered knowledge, i.e., a MCP with different levels of cooperation, can provide robust techniques to cancel or decode the interference at the transmitter/receiver sides and so improving the network spectral efficiency. In addition, its of particular relevance to understand the significance of our derivations on setups that include higher number of users accessing the network. For instance, the application of such formulas into scenarios where main users and cognitive users co-exist. The main users can design there optimal power allocation and optimal precoding subject to, the known interferer(s), i.e., the cognitive user interference, whether, they will be transmitting over SISO channels, or using OFDM modulation.

Appendix A: Proof of Theorem 1

The conditional probability density for the two-user MAC can be written as follows:

$$p_{y|x_1,x_2}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) = \frac{\mathbf{1}}{\pi^{\mathbf{n_r}}}\mathbf{e}^{-\|\mathbf{y}-\mathbf{H_1P_1x_1}-\mathbf{H_2P_2x_2}\|^{\mathbf{2}}} \tag{4.57}$$

Thus, the corresponding mutual information is:

$$\mathbf{I}(\mathbf{x_1},\mathbf{x_2};\mathbf{y}) = \mathbb{E}\left[\mathbf{log}\left(\frac{\mathbf{p_{y|x_1,x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})}{\mathbf{p_y}(\mathbf{y})}\right)\right] \tag{4.58}$$

$$\mathbf{I}(\mathbf{x_1},\mathbf{x_2};\mathbf{y}) = -\mathbf{n_r log}(\pi\mathbf{e}) - \mathbb{E}\left[\mathbf{log}\left(\mathbf{p_y}(\mathbf{y})\right)\right] \tag{4.59}$$

$$\mathbf{I}(\mathbf{x_1},\mathbf{x_2};\mathbf{y}) = -\mathbf{n_r log}(\pi\mathbf{e}) - \int \mathbf{p_y}(\mathbf{y})\mathbf{log}\left(\mathbf{p_y}(\mathbf{y})\right)\mathbf{dy} \tag{4.60}$$

Then, the gradient of the mutual information with respect to the channel of user 1 on the two-user MAC is as follows:

$$\frac{\partial \mathbf{I}(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{\partial \mathbf{H_1^\dagger}} = -\frac{\partial}{\partial \mathbf{H_1^\dagger}}\int \mathbf{p_y}(\mathbf{y})\mathbf{log}\left(\mathbf{p_y}(\mathbf{y})\right)\mathbf{dy} \tag{4.61}$$

$$= -\int\left(\mathbf{p_y}(\mathbf{y})\frac{\mathbf{1}}{\mathbf{p_y}(\mathbf{y})} + \mathbf{log}\left(\mathbf{p_y}(\mathbf{y})\right)\right)\frac{\partial \mathbf{p_y}(\mathbf{y})}{\partial \mathbf{H_1^\dagger}}\mathbf{dy} \tag{4.62}$$

$$= -\int\left(1 + \mathbf{log}\left(\mathbf{p_y}(\mathbf{y})\right)\right)\frac{\partial \mathbf{p_y}(\mathbf{y})}{\partial \mathbf{H_1^\dagger}}\mathbf{dy} \tag{4.63}$$

Where the probability density function of the received vector $\mathbf{y}$ is given by:

$$\mathbf{p_y}(\mathbf{y}) = \sum_{\mathbf{x_1},\mathbf{x_2}} \mathbf{p_{y|x_1,x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\mathbf{p_{x_1,x_2}}(\mathbf{x_1},\mathbf{x_2}) \tag{4.64}$$

$$= \mathbb{E}_{x_1,x_2}\left[\mathbf{p_{y|x_1,x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\right] \tag{4.65}$$

The derivative of the conditional output can be written as:

$$\frac{\partial \mathbf{p_{y|x_1,x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})}{\partial \mathbf{H_1^\dagger}} = \\ -\mathbf{p_{y|x_1,x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\frac{\partial}{\partial \mathbf{H_1^\dagger}}\left(\mathbf{y}-\mathbf{H_1P_1x_1}-\mathbf{H_2P_2x_2}\right)^\dagger\left(\mathbf{y}-\mathbf{H_1P_1x_1}-\mathbf{H_2P_2x_2}\right) \tag{4.66}$$

$$= \mathbf{p_{y|x_1,x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\left(\mathbf{y}-\mathbf{H_1P_1x_1}-\mathbf{H_2P_2x_2}\right)\mathbf{x_1^\dagger P_1^\dagger} \tag{4.67}$$

$$= -\nabla_{\mathbf{y}} \mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) \mathbf{x}_1^\dagger \mathbf{P}_1^\dagger \tag{4.68}$$

Therefore, we have:

$$\mathbb{E}_{x_1,x_2}\left[\nabla_{\mathbf{H}_1^\dagger} \mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\right] = \mathbb{E}_{x_1,x_2}\left[-\nabla_{\mathbf{y}} \mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) \mathbf{x}_1^\dagger \mathbf{P}_1^\dagger\right] \tag{4.69}$$

Substitute (4.69) into (4.63), we get:

$$\frac{\partial \mathbf{I}(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{\partial \mathbf{H}_1^\dagger} = \int \left(1 + \log\left(\mathbf{p}_{\mathbf{y}}(\mathbf{y})\right)\right) \mathbb{E}_{x_1,x_2}\left[\nabla_{\mathbf{y}} \mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) \mathbf{x}_1^\dagger \mathbf{P}_1^\dagger\right] \mathbf{dy} \tag{4.70}$$

$$= \mathbb{E}_{x_1,x_2}\left[\left(\int \left(1 + \log\left(\mathbf{p}_{\mathbf{y}}(\mathbf{y})\right)\right) \nabla_{\mathbf{y}} \mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) \mathbf{dy}\right) \mathbf{x}_1^\dagger \mathbf{P}_1^\dagger\right] \tag{4.71}$$

Using integration by parts applied to the real and imaginary parts of $\mathbf{y}$ we have:

$$\int \left(1 + \log\left(\mathbf{p}_{\mathbf{y}}(\mathbf{y})\right)\right) \frac{\partial \mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})}{\partial t} \mathbf{dt} =$$
$$\int \left(1 + \log\left(\mathbf{p}_{\mathbf{y}}(\mathbf{y})\right)\right) \mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{\mathbf{1}}{\mathbf{p}_{\mathbf{y}}(\mathbf{y})} \frac{\partial \mathbf{p}_{\mathbf{y}}(\mathbf{y})}{\partial \mathbf{t}} \mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) \mathbf{dt} \tag{4.72}$$

The first term in (4.72) goes to zero as $\|\mathbf{y}\| \to \infty$. Therefore,

$$\frac{\partial \mathbf{I}(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{\partial \mathbf{H}_1^\dagger} = \mathbb{E}_{x_1,x_2}\left[-\int \left(\frac{\mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})}{\mathbf{p}_{\mathbf{y}}(\mathbf{y})} \nabla_{\mathbf{y}} \mathbf{p}_{\mathbf{y}}(\mathbf{y}) \mathbf{dy}\right) \mathbf{x}_1^\dagger \mathbf{P}_1^\dagger\right] \tag{4.73}$$

$$\frac{\partial \mathbf{I}(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{\partial \mathbf{H}_1^\dagger} = -\int \nabla_{\mathbf{y}} \mathbf{p}_{\mathbf{y}}(\mathbf{y}) \mathbb{E}_{\mathbf{x_1},\mathbf{x_2}}\left[\frac{\mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})}{\mathbf{p}_{\mathbf{y}}(\mathbf{y})} \mathbf{x}_1^\dagger \mathbf{P}_1^\dagger\right] \mathbf{dy} \tag{4.74}$$

$$\frac{\partial \mathbf{I}(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{\partial \mathbf{H}_1^\dagger} = -\int \nabla_{\mathbf{y}} \mathbf{p}_{\mathbf{y}}(\mathbf{y}) \mathbb{E}_{\mathbf{x_1},\mathbf{x_2}}\left[\mathbf{x_1}|\mathbf{y}\right]^\dagger \mathbf{P}_1^\dagger \mathbf{dy} \tag{4.75}$$

However,

$$\begin{aligned}
\nabla_{\mathbf{y}} \mathbf{p}_{\mathbf{y}}(\mathbf{y}) &= \nabla_{\mathbf{y}} \mathbb{E}_{\mathbf{x_1},\mathbf{x_2}}\left[\mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\right] \\
&= \mathbb{E}_{x_1,x_2}\left[\nabla_{\mathbf{y}} \mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\right] \\
&= -\mathbb{E}_{x_1,x_2}\left[\mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2}) \left(\mathbf{y} - \mathbf{H_1}\mathbf{P_1}\mathbf{x_1} - \mathbf{H_2}\mathbf{P_2}\mathbf{x_2}\right)\right] \\
&= -\mathbb{E}_{x_1,x_2}\left[\mathbf{p}_{\mathbf{y}}(\mathbf{y}) \left(\mathbf{y} - \mathbf{H_1}\mathbf{P_1}\mathbf{x_1} - \mathbf{H_2}\mathbf{P_2}\mathbf{x_2}\right)|\mathbf{y}\right] \\
&= -\mathbf{p}_{\mathbf{y}}(\mathbf{y}) \left(\mathbf{y} - \mathbf{H_1}\mathbf{P_1}\mathbb{E}[\mathbf{x_1}|\mathbf{y}] - \mathbf{H_2}\mathbf{P_2}\mathbb{E}[\mathbf{x_2}|\mathbf{y}]\right)
\end{aligned} \tag{4.76}$$

Substitute (4.76) into (4.75) we get:

$$\frac{\partial \mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y})}{\partial \mathbf{H}_1^\dagger} = \int \mathbf{p_y}(\mathbf{y}) \left(\mathbf{y} - \mathbf{H_1 P_1} \mathbb{E}[\mathbf{x_1}|\mathbf{y}] - \mathbf{H_2 P_2} \mathbb{E}[\mathbf{x_2}|\mathbf{y}]\right) \mathbb{E}_{\mathbf{x_1},\mathbf{x_2}} \left[\mathbf{x_1}|\mathbf{y}\right]^\dagger \mathbf{P}_1^\dagger \mathbf{dy} \tag{4.77}$$

$$\frac{\partial \mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y})}{\partial \mathbf{H}_1^\dagger} = \mathbb{E}[\mathbf{y}\mathbf{x}_1^\dagger]\mathbf{P}_1^\dagger - \mathbb{E}[\mathbf{H_1 P_1} \mathbb{E}[\mathbf{x_1}|\mathbf{y}] \mathbb{E}[\mathbf{x_1}|\mathbf{y}]^\dagger \mathbf{P}_1^\dagger - \mathbb{E}[\mathbf{H_2 P_2} \mathbb{E}[\mathbf{x_2}|\mathbf{y}] \mathbb{E}[\mathbf{x_1}|\mathbf{y}]^\dagger] \mathbf{P}_1^\dagger \tag{4.78}$$

Therefore,

$$\begin{aligned} \frac{\partial \mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y})}{\partial \mathbf{H}_1^\dagger} = \mathbf{H_1 P_1} \mathbb{E}[\mathbf{x_1}\mathbf{x}_1^\dagger] \mathbf{P}_1^\dagger - \mathbf{H_1 P_1} \mathbb{E}[\mathbb{E}[\mathbf{x_1}|\mathbf{y}] \mathbb{E}[\mathbf{x_1}|\mathbf{y}]^\dagger \mathbf{P}_1^\dagger \\ - \mathbf{H_2 P_2} \mathbb{E}[\mathbb{E}[\mathbf{x_2}|\mathbf{y}] \mathbb{E}[\mathbf{x_1}|\mathbf{y}]^\dagger] \mathbf{P}_1^\dagger \end{aligned} \tag{4.79}$$

$$\frac{\partial \mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y})}{\partial \mathbf{H}_1^\dagger} = \mathbf{H_1 P_1 E_1} \mathbf{P}_1^\dagger - \mathbf{H_2 P_2} \mathbb{E}[\mathbb{E}[\mathbf{x_2}|\mathbf{y}] \mathbb{E}[\mathbf{x_1}|\mathbf{y}]^\dagger] \mathbf{P}_1^\dagger \tag{4.80}$$

Similarly, we can derive the gradient of the mutual information in terms of the channel matrix that corresponds to the second user in the two-user MAC.

$$\frac{\partial \mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y})}{\partial \mathbf{H}_2^\dagger} = \mathbf{H_2 P_2 E_2} \mathbf{P}_2^\dagger - \mathbf{H_1 P_1} \mathbb{E}[\mathbb{E}[\mathbf{x_1}|\mathbf{y}] \mathbb{E}[\mathbf{x_2}|\mathbf{y}]^\dagger] \mathbf{P}_2^\dagger \tag{4.81}$$

Therefore, the gradient of the mutual information with respect to per user channel and the per user MMSE and input estimates (or covariances) is as follows:

$$\nabla_{\mathbf{H_1}} \mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y}) = \mathbf{H_1 P_1 E_1} \mathbf{P}_1^\dagger - \mathbf{H_2 P_2} \mathbb{E}[\widehat{\mathbf{x_2}} \widehat{\mathbf{x_1}}^\dagger] \mathbf{P}_1^\dagger \tag{4.82}$$

$$\nabla_{\mathbf{H_2}} \mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y}) = \mathbf{H_2 P_2 E_2} \mathbf{P}_2^\dagger - \mathbf{H_1 P_1} \mathbb{E}[\widehat{\mathbf{x_1}} \widehat{\mathbf{x_2}}^\dagger] \mathbf{P}_2^\dagger \tag{4.83}$$

Therefore, Theorem 1 has been proved.

Appendix B: Proof of Theorem 2

The gradient of the mutual information with respect to the precoding matrix of user 1 on the two-user MAC is as follows:

$$\frac{\partial \mathbf{I}(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{\partial \mathbf{P}_\mathbf{1}^\dagger} = -\frac{\partial}{\partial \mathbf{P}_\mathbf{1}^\dagger}\int \mathbf{p_y}(\mathbf{y})\mathbf{log}\left(\mathbf{p_y}(\mathbf{y})\right)\mathbf{dy} \tag{4.84}$$

$$= -\int \left(\mathbf{p_y}(\mathbf{y})\frac{\mathbf{1}}{\mathbf{p_y}(\mathbf{y})} + \mathbf{log}\left(\mathbf{p_y}(\mathbf{y})\right)\right)\frac{\partial \mathbf{p_y}(\mathbf{y})}{\partial \mathbf{P}_\mathbf{1}^\dagger}\mathbf{dy} \tag{4.85}$$

$$= -\int \left(1 + \mathbf{log}\left(\mathbf{p_y}(\mathbf{y})\right)\right)\frac{\partial \mathbf{p_y}(\mathbf{y})}{\partial \mathbf{P}_\mathbf{1}^\dagger}\mathbf{dy} \tag{4.86}$$

The derivative of the conditional output can be written as:

$$\frac{\partial \mathbf{p_{y|x_1,x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})}{\partial \mathbf{P}_\mathbf{1}^\dagger} = - \mathbf{p_{y|x_1,x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\frac{\partial}{\partial \mathbf{P}_\mathbf{1}^\dagger}\left(\mathbf{y} - \mathbf{H_1}\mathbf{P_1}\mathbf{x_1} - \mathbf{H_2}\mathbf{P_2}\mathbf{x_2}\right)^\dagger\left(\mathbf{y} - \mathbf{H_1}\mathbf{P_1}\mathbf{x_1} - \mathbf{H_2}\mathbf{P_2}\mathbf{x_2}\right) \tag{4.87}$$

$$= \mathbf{p_{y|x_1,x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\mathbf{H}_\mathbf{1}^\dagger\left(\mathbf{y} - \mathbf{H_1}\mathbf{P_1}\mathbf{x_1} - \mathbf{H_2}\mathbf{P_2}\mathbf{x_2}\right)\mathbf{x}_\mathbf{1}^\dagger \tag{4.88}$$

$$= -\mathbf{H}_\mathbf{1}^\dagger\nabla_\mathbf{y}\mathbf{p_{y|x_1,x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\mathbf{x}_\mathbf{1}^\dagger \tag{4.89}$$

Therefore, we have:

$$\mathbb{E}_{x_1,x_2}\left[\nabla_{\mathbf{P}_\mathbf{1}^\dagger}\mathbf{p_{y|x_1,x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\right] = \mathbb{E}_{x_1,x_2}\left[-\mathbf{H}_\mathbf{1}^\dagger\nabla_\mathbf{y}\mathbf{p_{y|x_1,x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\mathbf{x}_\mathbf{1}^\dagger\right] \tag{4.90}$$

Substitute (4.90) into (4.86), we get:

$$\frac{\partial \mathbf{I}(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{\partial \mathbf{P}_\mathbf{1}^\dagger} = \int \left(1 + \mathbf{log}\left(\mathbf{p_y}(\mathbf{y})\right)\right)\mathbb{E}_{x_1,x_2}\left[\mathbf{H}_\mathbf{1}^\dagger\nabla_\mathbf{y}\mathbf{p_{y|x_1,x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\mathbf{x}_\mathbf{1}^\dagger\right]\mathbf{dy} \tag{4.91}$$

$$= \mathbb{E}_{x_1,x_2}\left[\mathbf{H}_\mathbf{1}^\dagger\left(\int \left(\mathbf{1} + \mathbf{log}\left(\mathbf{p_y}(\mathbf{y})\right)\right)\nabla_\mathbf{y}\mathbf{p_{y|x_1,x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\mathbf{dy}\right)\mathbf{x}_\mathbf{1}^\dagger\right] \tag{4.92}$$

Repeating the same steps as in (4.72), we have:

$$\frac{\partial \mathbf{I}(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{\partial \mathbf{P}_\mathbf{1}^\dagger} = \mathbb{E}_{x_1,x_2}\left[-\int \mathbf{H}_\mathbf{1}^\dagger\left(\frac{\mathbf{p_{y|x_1,x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})}{\mathbf{p_y}(\mathbf{y})}\nabla_\mathbf{y}\mathbf{p_y}(\mathbf{y})\mathbf{dy}\right)\mathbf{x}_\mathbf{1}^\dagger\right] \tag{4.93}$$

$$\frac{\partial \mathbf{I}(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{\partial \mathbf{P}_\mathbf{1}^\dagger} = -\int \mathbf{H}_\mathbf{1}^\dagger\nabla_\mathbf{y}\mathbf{p_y}(\mathbf{y})\mathbb{E}_{\mathbf{x_1},\mathbf{x_2}}\left[\frac{\mathbf{p_{y|x_1,x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})}{\mathbf{p_y}(\mathbf{y})}\mathbf{x}_\mathbf{1}^\dagger\right]\mathbf{dy} \tag{4.94}$$

$$\frac{\partial \mathbf{I}(\mathbf{x_1},\mathbf{x_2};\mathbf{y})}{\partial \mathbf{P}_\mathbf{1}^\dagger} = -\int \mathbf{H}_\mathbf{1}^\dagger\nabla_\mathbf{y}\mathbf{p_y}(\mathbf{y})\mathbb{E}_{\mathbf{x_1},\mathbf{x_2}}\left[\mathbf{x_1}|\mathbf{y}\right]^\dagger\mathbf{dy} \tag{4.95}$$

Substitute (4.76) into (4.95) we get:

$$\frac{\partial \mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y})}{\partial \mathbf{P}_\mathbf{1}^\dagger} = \int \mathbf{H}_\mathbf{1}^\dagger \mathbf{p_y}(\mathbf{y}) \left(\mathbf{y} - \mathbf{H_1 P_1} \mathbb{E}[\mathbf{x_1}|\mathbf{y}] - \mathbf{H_2 P_2} \mathbb{E}[\mathbf{x_2}|\mathbf{y}]\right) \mathbb{E}_{\mathbf{x_1},\mathbf{x_2}} \left[\mathbf{x_1}|\mathbf{y}\right]^\dagger \mathbf{dy} \tag{4.96}$$

$$\frac{\partial \mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y})}{\partial \mathbf{P}_\mathbf{1}^\dagger} = \mathbf{H}_\mathbf{1}^\dagger \mathbb{E}[\mathbf{y}\mathbf{x}_\mathbf{1}^\dagger] - \mathbf{H}_\mathbf{1}^\dagger \mathbb{E}[\mathbf{H_1 P_1} \mathbb{E}[\mathbf{x_1}|\mathbf{y}] \mathbb{E}[\mathbf{x_1}|\mathbf{y}]^\dagger - \mathbf{H}_\mathbf{1}^\dagger \mathbb{E}[\mathbf{H_2 P_2} \mathbb{E}[\mathbf{x_2}|\mathbf{y}] \mathbb{E}[\mathbf{x_1}|\mathbf{y}]^\dagger] \tag{4.97}$$

Therefore,

$$\begin{aligned}\frac{\partial \mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y})}{\partial \mathbf{P}_\mathbf{1}^\dagger} = \mathbf{H}_\mathbf{1}^\dagger \mathbf{H_1 P_1} \mathbb{E}[\mathbf{x_1}\mathbf{x}_\mathbf{1}^\dagger] - \mathbf{H}_\mathbf{1}^\dagger \mathbf{H_1 P_1} \mathbb{E}[\mathbb{E}[\mathbf{x_1}|\mathbf{y}] \mathbb{E}[\mathbf{x_1}|\mathbf{y}]^\dagger \\ - \mathbf{H}_\mathbf{1}^\dagger \mathbf{H_2 P_2} \mathbb{E}[\mathbb{E}[\mathbf{x_2}|\mathbf{y}] \mathbb{E}[\mathbf{x_1}|\mathbf{y}]^\dagger]\end{aligned} \tag{4.98}$$

$$\frac{\partial \mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y})}{\partial \mathbf{P}_\mathbf{1}^\dagger} = \mathbf{H}_\mathbf{1}^\dagger \mathbf{H_1 P_1 E_1} - \mathbf{H}_\mathbf{1}^\dagger \mathbf{H_2 P_2} \mathbb{E}[\mathbb{E}[\mathbf{x_2}|\mathbf{y}] \mathbb{E}[\mathbf{x_1}|\mathbf{y}]^\dagger] \tag{4.99}$$

Similarly, we can derive the gradient of the mutual information in terms of the channel matrix that corresponds to the second user in the two-user MAC.

$$\frac{\partial \mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y})}{\partial \mathbf{P}_\mathbf{2}^\dagger} = \mathbf{H}_\mathbf{2}^\dagger \mathbf{H_2 P_2 E_2} - \mathbf{H}_\mathbf{2}^\dagger \mathbf{H_1 P_1} \mathbb{E}[\mathbb{E}[\mathbf{x_1}|\mathbf{y}] \mathbb{E}[\mathbf{x_2}|\mathbf{y}]^\dagger] \tag{4.100}$$

Therefore, the gradient of the mutual information with respect to per user precoding and the per user MMSE and input estimates (or covariances) is as follows:

$$\nabla_{\mathbf{P_1}} \mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y}) = \mathbf{H}_\mathbf{1}^\dagger \mathbf{H_1 P_1 E_1} - \mathbf{H}_\mathbf{1}^\dagger \mathbf{H_2 P_2} \mathbb{E}[\widehat{\mathbf{x_2}} \widehat{\mathbf{x_1}}^\dagger] \tag{4.101}$$

$$\nabla_{\mathbf{P_2}} \mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y}) = \mathbf{H}_\mathbf{2}^\dagger \mathbf{H_2 P_2 E_2} - \mathbf{H}_\mathbf{2}^\dagger \mathbf{H_1 P_1} \mathbb{E}[\widehat{\mathbf{x_1}} \widehat{\mathbf{x_2}}^\dagger] \tag{4.102}$$

Therefore, Theorem 2 has been proved.

Appendix C: Proof of Theorem 3

From the steps in Theorem 1, we can see that:

$$\begin{aligned}\mathbf{p_y}(\mathbf{y}) \left(\mathbf{y} - \mathbf{H_1 P_1} \mathbb{E}[\mathbf{x_1}|\mathbf{y}] - \mathbf{H_2 P_2} \mathbb{E}[\mathbf{x_2}|\mathbf{y}]\right) \\ = \mathbb{E}_{x_1,x_2} \left[\mathbf{p_{y|x_1,x_2}}(\mathbf{y}|\mathbf{x_1}, \mathbf{x_2}) \left(\mathbf{y} - \mathbf{H_1 P_1 x_1} - \mathbf{H_2 P_2 x_2}\right)\right]\end{aligned} \tag{4.103}$$

Therefore,

$$\begin{aligned}\mathbf{H_1P_1}\mathbb{E}[\mathbf{x_1}|\mathbf{y}] - \mathbf{H_2P_2}\mathbb{E}[\mathbf{x_2}|\mathbf{y}] \\ &= \mathbf{y} - \frac{\mathbb{E}_{\mathbf{x_1},\mathbf{x_2}}\left[\mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\left(\mathbf{y}-\mathbf{H_1P_1x_1}-\mathbf{H_2P_2x_2}\right)\right]}{\mathbf{p}_\mathbf{y}(\mathbf{y})} \\ &= \mathbf{y} + \frac{\mathbb{E}_{\mathbf{x_1},\mathbf{x_2}}\left[\nabla_\mathbf{y}\mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\right]}{\mathbf{p}_\mathbf{y}(\mathbf{y})} \\ &= \mathbf{y} + \frac{\nabla_\mathbf{y}\mathbb{E}_{\mathbf{x_1},\mathbf{x_2}}\left[\mathbf{p}_{\mathbf{y}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y}|\mathbf{x_1},\mathbf{x_2})\right]}{\mathbf{p}_\mathbf{y}(\mathbf{y})} \end{aligned} \tag{4.104}$$

Thus,

$$\mathbf{H_1P_1}\mathbb{E}[\mathbf{x_1}|\mathbf{y}] - \mathbf{H_2P_2}\mathbb{E}[\mathbf{x_2}|\mathbf{y}] = \mathbf{y} + \frac{\nabla_\mathbf{y}\mathbf{p}_\mathbf{y}(\mathbf{y})}{\mathbf{p}_\mathbf{y}(\mathbf{y})} \tag{4.105}$$

Therefore, Theorem 3 has been proved.

Appendix D: Proof of Theorem 4

The MMSE Wiener filters are known to be MSE minimizers. In fact, this kind of receive filters inherently include the linear MMSE matrix. Therefore, to start the proof of the theorem, lets first define the linear MMSE as:

$$mmse(snr) = \underbrace{\mathbb{E}\left[\|\mathbf{H_1P_1}(\mathbf{x_1}-\widehat{\mathbf{x_1}})\|^\mathbf{2}\right]}_{\text{MMSE1}} + \underbrace{\mathbb{E}\left[\|\mathbf{H_2P_2}(\mathbf{x_2}-\widehat{\mathbf{x_2}})\|^\mathbf{2}\right]}_{\text{MMSE2}}, \tag{4.106}$$

with the linear estimates should be found by multiplying the received vector $\mathbf{y}$ by the receive filter $\mathbf{H_r}$,

$$\widehat{\mathbf{x_1}} = \mathbf{H_{r1}y} \tag{4.107}$$

$$\widehat{\mathbf{x_2}} = \mathbf{H_{r2}y} \tag{4.108}$$

We break down the problem into two parts for each per user MMSE. Notice that we call it MMSE since our target is the minimum MSE, therefore, we first take the first part, i.e., MSE1 and break it down; it follows that:

$$\tilde{\mathbf{E1}} = \mathbb{E}\left[\|\mathbf{H_1P_1}(\mathbf{x_1}-\widehat{\mathbf{x_1}})\|^\mathbf{2}\right] \tag{4.109}$$

$$= \mathbb{E}\left[\mathbf{H_1P_1}(\mathbf{x_1} - \widehat{\mathbf{x_1}})(\mathbf{x_1} - \widehat{\mathbf{x_1}})^\dagger(\mathbf{H_1P_1})^\dagger\right] \tag{4.110}$$

$$= \mathbb{E}\left[(\mathbf{H_1P_1x_1} - \mathbf{H_1P_1}\widehat{\mathbf{x_1}})(\mathbf{H_1P_1x_1} - \mathbf{H_1P_1}\widehat{\mathbf{x_1}})^\dagger\right] \tag{4.111}$$

$$\begin{aligned} = &\underbrace{\mathbb{E}\left[\mathbf{H_1P_1x_1x_1}^\dagger(\mathbf{H_1P_1})^\dagger\right]}_{\text{first term}} - \underbrace{\mathbb{E}\left[\mathbf{H_1P_1x_1}\widehat{\mathbf{x_1}}^\dagger(\mathbf{H_1P_1})^\dagger\right]}_{\text{second term}} \\ &- \underbrace{\mathbb{E}\left[\mathbf{H_1P_1}\widehat{\mathbf{x_1}}\mathbf{x_1}^\dagger(\mathbf{H_1P_1})^\dagger\right]}_{\text{third term}} + \underbrace{\mathbb{E}\left[\mathbf{H_1P_1}\widehat{\mathbf{x_1}}\widehat{\mathbf{x_1}}^\dagger(\mathbf{H_1P_1})^\dagger\right]}_{\text{forth term}} \end{aligned} \tag{4.112}$$

Lets work on each term in (4.112) separately. Notice that $\mathbb{E}[\mathbf{x_1x_1^\dagger}] = \mathbb{E}[\mathbf{x_2x_2^\dagger}] = \mathbf{I}$, $\mathbb{E}[\mathbf{x_1x_1^T}] = \mathbb{E}[\mathbf{x_2x_2^T}] = \mathbf{0}$, $\mathbb{E}[\mathbf{x_2x_1^\dagger}] = \mathbb{E}[\mathbf{x_1x_2^\dagger}] = \mathbf{0}$, $\mathbb{E}[\mathbf{nn}^\dagger] = \mathbf{I}$, and $\mathbb{E}[\mathbf{nn^T}] = \mathbf{0}$.

The first term is:

$$\mathbb{E}\left[(\mathbf{H_1P_1x_1x_1}^\dagger(\mathbf{H_1P_1})^\dagger\right] = \mathbb{E}\left[\mathbf{H_1P_1}(\mathbf{H_1P_1})^\dagger\right], \tag{4.113}$$

The second term is:

$$\mathbb{E}\left[\mathbf{H_1P_1x_1}\widehat{\mathbf{x_1}}^\dagger(\mathbf{H_1P_1})^\dagger\right] = \mathbb{E}\left[\mathbf{H_1P_1x_1}(\mathbf{H_{r1}y})^\dagger(\mathbf{H_1P_1})^\dagger\right] \tag{4.114}$$

$$\mathbb{E}\left[\mathbf{H_1P_1x_1y}^\dagger\mathbf{H_{r1}^\dagger}(\mathbf{H_1P_1})^\dagger\right] = \mathbb{E}\left[\mathbf{H_1P_1x_1}(\mathbf{H_1P_1x_1} + \mathbf{H_2P_2x_2} + \mathbf{n})^\dagger\mathbf{H_{r1}^\dagger}(\mathbf{H_1P_1})^\dagger\right] \tag{4.115}$$

$$= \mathbb{E}\left[\mathbf{H_1P_1P_1^\dagger H_1^\dagger H_{r1}^\dagger}(\mathbf{H_1P_1})^\dagger\right] \tag{4.116}$$

The third term is:

$$\mathbb{E}\left[\mathbf{H_1P_1}\widehat{\mathbf{x_1}}\mathbf{x_1^\dagger}(\mathbf{H_1P_1})^\dagger\right] = \mathbb{E}\left[\mathbf{H_1P_1H_{r1}yx_1^\dagger}(\mathbf{H_1P_1})^\dagger\right] \tag{4.117}$$

$$= \mathbb{E}\left[\mathbf{H_1P_1H_{r1}H_1P_1}(\mathbf{H_1P_1})^\dagger\right] \tag{4.118}$$

The forth term is:

$$\mathbb{E}\left[\mathbf{H_1P_1}\widehat{\mathbf{x_1}}\widehat{\mathbf{x_1}}^{\dagger}(\mathbf{H_1P_1})^{\dagger}\right] = \mathbb{E}\left[\mathbf{H_1P_1H_{r1}yy}^{\dagger}\mathbf{H}_{\mathbf{r1}}^{\dagger}(\mathbf{H_1P_1})^{\dagger}\right] \tag{4.119}$$

$$\begin{aligned}\mathbb{E}\left[\mathbf{H_1P_1H_{r1}yy}^{\dagger}\mathbf{H}_{\mathbf{r1}}^{\dagger}(\mathbf{H_1P_1})^{\dagger}\right] &= \mathbb{E}\left[\mathbf{H_1P_1H_{r1}H_1P_1P}_{\mathbf{1}}^{\dagger}\mathbf{H}_{\mathbf{1}}^{\dagger}\mathbf{H}_{\mathbf{r1}}^{\dagger}(\mathbf{H_1P_1})^{\dagger}\right] \\ &+ \mathbb{E}\left[\mathbf{H_1P_1H_{r1}H_2P_2P}_{\mathbf{2}}^{\dagger}\mathbf{H}_{\mathbf{2}}^{\dagger}\mathbf{H}_{\mathbf{r1}}^{\dagger}(\mathbf{H_1P_1})^{\dagger}\right] + \mathbb{E}\left[\mathbf{H_1P_1H_{r1}H}_{\mathbf{r1}}^{\dagger}(\mathbf{H_1P_1})^{\dagger}\right]\end{aligned} \tag{4.120}$$

Therefore, the MSE1 can be given as follows:

$$\begin{aligned}&MSE1(snr) = Tr\left(\mathbb{E}\left[\mathbf{H_1P_1}(\mathbf{H_1P_1})^{\dagger}\right]\right) - Tr\left\{\mathbb{E}\left[\mathbf{H_1P_1P}_{\mathbf{1}}^{\dagger}\mathbf{H}_{\mathbf{1}}^{\dagger}\mathbf{H}_{\mathbf{r1}}^{\dagger}(\mathbf{H_1P_1})^{\dagger}\right]\right\} \\ &- Tr\left\{\mathbb{E}\left[\mathbf{H_1P_1H_{r1}H_1P_1}(\mathbf{H_1P_1})^{\dagger}\right]\right\} + Tr\left\{\mathbb{E}\left[\mathbf{H_1P_1H_{r1}H_1P_1P}_{\mathbf{1}}^{\dagger}\mathbf{H}_{\mathbf{1}}^{\dagger}\mathbf{H}_{\mathbf{r1}}^{\dagger}(\mathbf{H_1P_1})^{\dagger}\right]\right\} \\ &+ Tr\left\{\mathbb{E}\left[\mathbf{H_1P_1H_{r1}H_2P_2P}_{\mathbf{2}}^{\dagger}\mathbf{H}_{\mathbf{2}}^{\dagger}\mathbf{H}_{\mathbf{r1}}^{\dagger}(\mathbf{H_1P_1})^{\dagger}\right]\right\} + Tr\left\{\mathbb{E}\left[\mathbf{H_1P_1H_{r1}H}_{\mathbf{r1}}^{\dagger}(\mathbf{H_1P_1})^{\dagger}\right]\right\}\end{aligned} \tag{4.121}$$

We need now to find the minimum MSE1, and apply the KKT conditions to derive the optimal MMSE1 Wiener receive filter $\mathbf{H_{r1}}$ as follows:

$$\frac{\partial MSE1}{\partial H_{r1}} = 0 \tag{4.122}$$

Capitalizing on the derivative rules for the trace of matrices, $\frac{\partial Tr(ABA^{\dagger})}{\partial A} = AB^{\dagger}$, $\frac{\partial Tr(AB)}{\partial A} = B^{\dagger}$, and $\frac{\partial Tr(BA^{\dagger})}{\partial A} = B$. We have:

$$\mathbf{H}_{\mathbf{r1}}^{\star} = \mathbf{P}_{\mathbf{1}}^{\dagger}\mathbf{H}_{\mathbf{1}}^{\dagger}(\mathbf{I} + \mathbf{P}_{\mathbf{1}}^{\dagger}\mathbf{H}_{\mathbf{1}}^{\dagger}\mathbf{P_1H_1} + \mathbf{P}_{\mathbf{2}}^{\dagger}\mathbf{H}_{\mathbf{2}}^{\dagger}\mathbf{P_2H_2})^{-\mathbf{1}} \tag{4.123}$$

Similar steps to break down MSE2,

$$\tilde{\mathbf{E2}} = \mathbb{E}\left[\|\mathbf{H_2P_2}(\mathbf{x_2} - \widehat{\mathbf{x_2}})\|^{\mathbf{2}}\right], \tag{4.124}$$

leads to the optimal MMSE2 Wiener filter $\mathbf{H_{r2}}$ as follows:

$$\mathbf{H}_{\mathbf{r2}}^{\star} = \mathbf{P}_{\mathbf{2}}^{\dagger}\mathbf{H}_{\mathbf{2}}^{\dagger}(\mathbf{I} + \mathbf{P}_{\mathbf{1}}^{\dagger}\mathbf{H}_{\mathbf{1}}^{\dagger}\mathbf{P_1H_1} + \mathbf{P}_{\mathbf{2}}^{\dagger}\mathbf{H}_{\mathbf{2}}^{\dagger}\mathbf{P_2H_2})^{-\mathbf{1}} \tag{4.125}$$

Therefore, to estimate the user inputs linearly we do the following:

$$\widehat{\mathbf{x_1}} = \mathbf{P}_{\mathbf{1}}^{\dagger}\mathbf{H}_{\mathbf{1}}^{\dagger}(\mathbf{I} + \mathbf{P}_{\mathbf{1}}^{\dagger}\mathbf{H}_{\mathbf{1}}^{\dagger}\mathbf{P_1H_1} + \mathbf{P}_{\mathbf{2}}^{\dagger}\mathbf{H}_{\mathbf{2}}^{\dagger}\mathbf{P_2H_2})^{-\mathbf{1}}\mathbf{y} \tag{4.126}$$

$$\widehat{\mathbf{x_2}} = \mathbf{P}_{\mathbf{2}}^{\dagger}\mathbf{H}_{\mathbf{2}}^{\dagger}(\mathbf{I} + \mathbf{P}_{\mathbf{1}}^{\dagger}\mathbf{H}_{\mathbf{1}}^{\dagger}\mathbf{P_1H_1} + \mathbf{P}_{\mathbf{2}}^{\dagger}\mathbf{H}_{\mathbf{2}}^{\dagger}\mathbf{P_2H_2})^{-\mathbf{1}}\mathbf{y} \tag{4.127}$$

Therefore, Theorem 4 has been proved.

Appendix E: Proof of Theorem 5

The possible solutions to (4.2) subject to subject to (4.3) and (4.4) are characterized by the KKT conditions, which give necessary conditions for the matrix to be a critical point, known as the KKT or first-order conditions. To compute the KKT conditions, we first build the Lagrangian:

$$\mathcal{L}(P_1, P_2, \lambda_1, \lambda_2) = -I(x_1, x_2; y) - \lambda_1(\mathbf{Q_1} - \mathbf{P_1}) - \lambda_2(\mathbf{Q_2} - \mathbf{P_2}) - \mu_1\mathbf{P_1} - \mu_2\mathbf{P_2} \quad (4.128)$$

With primal feasibility conditions, $\lambda_1(Q_1 - P_1) = 0$, $\mu_1 P_1 = 0$, $\lambda_2(Q_2 - P_2) = 0$, and $\mu_2 P_2 = 0$, and dual feasibility conditions, $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$. In which the Lagrange multipliers λ_1 and λ_2 accounting for the in-equality constraint, has to be non-negative. The first-order conditions are given by:

$$\nabla_{P_1}\mathcal{L}(P_1, P_2, \lambda_1, \lambda_2) = -\nabla_{P_1} I(x_1, x_2; y) + \lambda_1\mathbf{P_1} = \mathbf{0} \quad (4.129)$$

$$\nabla_{P_2}\mathcal{L}(P_1, P_2, \lambda_1, \lambda_2) = -\nabla_{P_2} I(x_1, x_2; y) + \lambda_2\mathbf{P_2} = \mathbf{0} \quad (4.130)$$

Therefore, the solution of the optimal precoders satisfies:

$$\mathbf{P_1}^\star = \nu_1^{-1}\mathbf{H_1}^\dagger\mathbf{H_1}\mathbf{P_1}^\star\mathbf{E_1} - \nu_1^{-1}\mathbf{H_1}^\dagger\mathbf{H_2}\mathbf{P_2}^\star\mathbb{E}[\widehat{\mathbf{x_2}}\widehat{\mathbf{x_1}}^\dagger] \quad (4.131)$$

$$\mathbf{P_2}^\star = \nu_2^{-1}\mathbf{H_2}^\dagger\mathbf{H_2}\mathbf{P_2}^\star\mathbf{E_2} - \nu_2^{-1}\mathbf{H_2}^\dagger\mathbf{H_1}\mathbf{P_1}^\star\mathbb{E}[\widehat{\mathbf{x_1}}\widehat{\mathbf{x_2}}^\dagger] \quad (4.132)$$

where the solution follows from the gradient of the mutual information with respect to the precoding matrices. We eliminate λ_1 and λ_2 by introducing $\nu_1 = \frac{\lambda_1}{snr}$, and $\nu_2 = \frac{\lambda_2}{snr}$. Therefore, Theorem 5 has been proved.

Appendix F: Proof of Theorem 6

Digging into the depth of equations (4.23) and (4.24), we can do a singular value decomposition of the channel matrix $\mathbf{H_1} = \mathbf{U_{H_1}}\mathbf{\Lambda_{H_1}}\mathbf{V}_{\mathbf{H_1}}^\dagger$, and $\mathbf{H_2} = \mathbf{U_{H_2}}\mathbf{\Lambda_{H_2}}\mathbf{V}_{\mathbf{H_2}}^\dagger$, and the eigen value decomposition of the MMSE matrices, $\mathbf{E_1} = \mathbf{U_{E_1}}\mathbf{\Lambda_{E_1}}\mathbf{V}_{\mathbf{E_1}}^\dagger$, and $\mathbf{E_2} = \mathbf{U_{E_2}}\mathbf{\Lambda_{E_2}}\mathbf{V}_{\mathbf{E_2}}^\dagger$, we assume that the second terms in equations (4.21) and (4.22) are zeros. Therefore, multiplying the first term of (4.21) by $\mathbf{P_1}^{\star\dagger}$, and of (4.22) by $\mathbf{P_2}^{\star\dagger}$,

and taking the trace, which corresponds to each of the users power constraints. It follows that:

$$\begin{aligned} Tr\left\{\mathbf{P_1}^{\star}\mathbf{P_1}^{\star\dagger}\right\} &= Tr\left\{\mathbf{H_1}\mathbf{H_1^{\dagger}}\mathbf{P_1^{\star}}\mathbf{E_1}\mathbf{P_1}^{\star\dagger}\right\} \\ &= Tr\left\{\mathbf{P_1}^{\star\dagger}\mathbf{H_1}\mathbf{H_1^{\dagger}}\mathbf{P_1^{\star}}\mathbf{E_1}\right\} \\ &= Tr\left\{\mathbf{E_1}\mathbf{P_1}^{\star\dagger}\mathbf{H_1}\mathbf{H_1^{\dagger}}\mathbf{P_1^{\star}}\right\} \\ &= Tr\left\{\mathbf{P_1^{\star}}\mathbf{E_1}\mathbf{P_1}^{\star\dagger}\mathbf{H_1}\mathbf{H_1^{\dagger}}\right\} \\ &= Tr\left\{\mathbf{\Lambda_{P_1}}\mathbf{\Lambda_{E_1}}\mathbf{\Lambda_{P_1}}\mathbf{\Lambda_{H_1}^2}\right\} \\ &= Tr\left\{\mathbf{\Lambda_{H_1}^2}\mathbf{\Lambda_{P_1}}\mathbf{\Lambda_{E_1}}\mathbf{\Lambda_{P_1}}\right\} \\ &= Tr\left\{\mathbf{\Lambda_{H_1}^2}\mathbf{\Lambda_{P_1}^2}\mathbf{\Pi}\mathbf{\Lambda_{E_1}}\right\} \end{aligned} \tag{4.133}$$

With:

$$\mathbf{U_1} = \mathbf{V_{H_1}}, \tag{4.134}$$

$$\mathbf{U_2} = \mathbf{V_{H_2}}, \tag{4.135}$$

$$\mathbf{D_1} = \mathbf{diag}(\sqrt{\mathbf{p_{1,1}}},\ \cdots,\sqrt{\mathbf{p_{1,n_t}}}), \tag{4.136}$$

$$\mathbf{D_2} = \mathbf{diag}(\sqrt{\mathbf{p_{2,1}}},\ \cdots,\sqrt{\mathbf{p_{2,n_t}}}), \tag{4.137}$$

$$\mathbf{R_1} = \mathbf{\Pi}\mathbf{U_{E_1}}, \tag{4.138}$$

$$\mathbf{R_2} = \mathbf{\Pi}\mathbf{U_{E_2}}. \tag{4.139}$$

Similar steps to prove the setup of the fixed point equation of the second user. Therefore, the fixed point equation of the per user optimal precoder precluding the other users interference is given as:

$$\mathbf{P_1} = \mathbf{U_1}\mathbf{D_1}\mathbf{R_1^{\dagger}} \tag{4.140}$$

$$\mathbf{P_2} = \mathbf{U_2}\mathbf{D_2}\mathbf{R_2^{\dagger}} \tag{4.141}$$

Therefore, Theorem 6 is proved. Note that we provide another proof for this theorem that will be discussed later in Theorem 11.

Appendix G: Proof of Theorem 7

The Lagrangian for the optimization problem (4.43) subject to (4.44) to (4.46) is given by:

$$\mathcal{L}(P_1, P_2, \lambda_1, \lambda_2) = -I(x_1, x_2; y) - \lambda_1(\mathbf{1} - \sum_{\mathbf{j=1}}^{\mathbf{n_t}} \mathbf{p_{1j}}) - \lambda_{\mathbf{2}}(\mathbf{1} - \sum_{\mathbf{j=1}}^{\mathbf{n_t}} \mathbf{p_{2j}}) - \sum_{\mathbf{j=1}}^{\mathbf{n_t}} \mu_{\mathbf{1}} \mathbf{p_{1j}} - \sum_{\mathbf{j=1}}^{\mathbf{n_t}} \mu_{\mathbf{2}} \mathbf{p_{2j}} \tag{4.142}$$

With primal feasibility conditions, $\lambda_1(1 - \sum_{j=1}^{n_t} \mathbf{p_{1j}}) = \mathbf{0}$, $\mu_1 \sum_{j=1}^{n_t} \mathbf{p_{1j}} = \mathbf{0}$, $\lambda_2(1 - \sum_{j=1}^{n_t} \mathbf{p_{2j}}) = \mathbf{0}$, and $\mu_2 \sum_{j=1}^{n_t} \mathbf{p_{2j}} = \mathbf{0}$, and dual feasibility conditions, $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$. The KKT conditions are given by:

$$\frac{\partial \mathcal{L}(P_1, P_2, \lambda_1, \lambda_2)}{\partial \mathbf{p_1}} = -\frac{\partial I(x_1, x_2; y)}{\partial \mathbf{p_1}} + \lambda_1 - \mu_1 = 0 \tag{4.143}$$

$$\frac{\partial \mathcal{L}(P_1, P_2, \lambda_1, \lambda_2)}{\partial \mathbf{p_2}} = -\frac{\partial I(x_1, x_2; y)}{\partial \mathbf{p_2}} + \lambda_2 - \mu_2 = 0 \tag{4.144}$$

Note that $\mu_1 = 0$ for $\mathbf{p_1} > \mathbf{0}$ and $\mu 2 = 0$ for $\mathbf{p_1} > \mathbf{0}$. Therefore, the derivative with respect to $\mathbf{p_1}$, and $\mathbf{p_2}$ respectively, is given by:

$$\frac{\partial \mathcal{L}(P_1, P_2, \lambda_1, \lambda_2)}{\partial \mathbf{p_1}} = \frac{snr}{\sqrt{\mathbf{p_1}}} \left(\mathbf{P_1}^{\star} \mathbf{H_1^{\dagger}} \mathbf{H_1} \mathbf{P_1}^{\star} \mathbf{E_1} - \mathbf{P_1}^{\star} \mathbf{H_1^{\dagger}} \mathbf{H_2} \mathbf{P_2}^{\star} \mathbb{E}[\widehat{\mathbf{x_2}} \widehat{\mathbf{x_1}}^{\dagger}] \right)_{1j} \tag{4.145}$$

$$\frac{\partial \mathcal{L}(P_1, P_2, \lambda_1, \lambda_2)}{\partial \mathbf{p_2}} = \frac{snr}{\sqrt{\mathbf{p_2}}} \left(\mathbf{P_2}^{\star} \mathbf{H_2^{\dagger}} \mathbf{H_2} \mathbf{P_2}^{\star} \mathbf{E_2} - \mathbf{P_2}^{\star} \mathbf{H_2^{\dagger}} \mathbf{H_1} \mathbf{P_1}^{\star} \mathbb{E}[\widehat{\mathbf{x_1}} \widehat{\mathbf{x_2}}^{\dagger}] \right)_{2j} \tag{4.146}$$

Therefore, the optimal power allocation satisfies:

$$\mathbf{p}_1{}^{\star} = \gamma_1^{-1} \left(\mathbf{P_1}^{\star} \mathbf{H_1^{\dagger}} \mathbf{H_1} \mathbf{P_1}^{\star} \mathbf{E_1} - \mathbf{P_1}^{\star} \mathbf{H_1^{\dagger}} \mathbf{H_2} \mathbf{P_2}^{\star} \mathbb{E}[\widehat{\mathbf{x_2}} \widehat{\mathbf{x_1}}^{\dagger}] \right)_{1j} \tag{4.147}$$

$$\mathbf{p}_2{}^{\star} = \gamma_2^{-1} \left(\mathbf{P_2}^{\star} \mathbf{H_2^{\dagger}} \mathbf{H_2} \mathbf{P_2}^{\star} \mathbf{E_2} - \mathbf{P_2}^{\star} \mathbf{H_2^{\dagger}} \mathbf{H_1} \mathbf{P_1}^{\star} \mathbb{E}[\widehat{\mathbf{x_1}} \widehat{\mathbf{x_2}}^{\dagger}] \right)_{2j} \tag{4.148}$$

where $\gamma_1 = \frac{\lambda_1}{snr}$, and $\gamma_2 = \frac{\lambda_2}{snr}$. Therefore, Theorem 7 has been proved.

Chapter 5
Optimal Power Allocation and Optimal Precoding with Multi-Cell Processing

5.1 Introduction

Multi-cell cooperative processing is well acknowledged for significantly improving spectral efficiency and fairness amongst users. Exploiting the concepts of cooperation via transmit diversity and virtualizing the networks can move networks into upper bounds despite the fundamental limits of cooperation [109]. Therefore, cooperation in small size clustered network frameworks from one side can boost network performance. However, from another side, demonstrate that adaptivity and feedback can have a dramatic effect on the data rates when transmitter adapt to the channel experience. In this Chapter we consider a cooperative framework using mutli-cell processing [110]. Firstly, we consider the problem of joint cooperative optimal power allocation. Therefore, with a prior knowledge of each channel state and the data from each UT, the base stations will cooperate to jointly design the optimal power allocation that maximizes the joint reliable information rates; i.e., the clustered network MIMO capacity. Each BS will then communicate the optimal power via feedback DL links to each user in order to use in their UL transmissions considering that the process is adaptive, and the processing time is very small, such that the CSI doesn't change. The UL/DL reciprocity/duality is not assumed in solving the problem and there is no special setup like TDD considered, [41], [111]. Secondly, we consider the problem of joint cooperative precoding for the DL scenario through which both BSs can jointly design the optimal precoding vectors. Besides, we provide insights into the way the interference can be mitigated from a precoding perspective and on the other hand, how a studied interference can be thought of as a positive factor in the MCP framework, instead of dealing with it as a limiting factor to the network capacity.

In this Chapter, we investigate the optimal power allocation and optimal precoding for a cluster of two BSs which cooperate to jointly maximize the achievable rate for two users connecting to each BS in a MCP framework. This framework is modeled by a virtual network MIMO channel due to the framework of full cooperation. In particular, due to sharing the CSI and data between the two BSs over the backhaul link. We provide a generalized fixed point equation of the optimal precoder in the asymptotic regimes of the low- and high-snr. We introduce a new iterative approach that leads to a closed-form expression for the optimal precoding matrix in the high-snr regime which is known to be an NP-hard problem. Two MCP distributed algorithms have been introduced, a power allocation algorithm for the UL, and a precoding algorithm for the DL. We have also quantified the losses in the mutual information ecountered due to the interference, by capitalizing on the connections between information measures and estimation measures.

5.2 The MCP System Model

Consider the scenario shown in Figure 5.1 where MCP is implemented in clusters of two base stations. The base stations no longer tune their physical and link/MAC layer parameters separately (power level, time slots, sub-carrier usage, precoding coefficients etc.), but instead coordinate their coding and decoding operations on the basis of channel state information and user data information exchanged over a backhaul link [110]. As illustrated in Figure 5.1, we suppose that the base stations will share their CSI and will exchange the information received by the user terminals, UT1 and UT2 who are roaming under the coverage of BS1 and BS2, respectively. Therefore, BS1 and BS2 will receive from UT1 and UT2 respectively,

$$\mathbf{y_1} = \sqrt{snr}\ h_{11}\sqrt{P_1}\mathbf{x_1} + \sqrt{snr}\ h_{21}\sqrt{P_2}\mathbf{x_2} + \mathbf{n_1} \tag{5.1}$$

$$\mathbf{y_2} = \sqrt{snr}\ h_{12}\sqrt{P_1}\mathbf{x_1} + \sqrt{snr}\ h_{22}\sqrt{P_2}\mathbf{x_2} + \mathbf{n_2} \tag{5.2}$$

$\mathbf{y_1} \in \mathbb{C}^{\mathbf{n}}$ and $\mathbf{y_2} \in \mathbb{C}^{\mathbf{n}}$ represent the received vectors of complex symbols at BS1 and BS2 respectively, $\mathbf{x_1} \in \mathbb{C}^{\mathbf{n}}$ and $\mathbf{x_2} \in \mathbb{C}^{\mathbf{n}}$ represent the vectors of complex transmit symbols with zero mean $\mathbb{E}[\mathbf{x_1}\mathbf{x_1^T}] = \mathbb{E}[\mathbf{x_2}\mathbf{x_2^T}] = \mathbf{0}$, and identity covariance $\mathbb{E}[\mathbf{x_1}\mathbf{x_1^\dagger}] = \mathbb{E}[\mathbf{x_2}\mathbf{x_2^\dagger}] = \mathbf{I}$, $\mathbf{n_1} \in \mathbb{C}^{\mathbf{n}}$ and $\mathbf{n_2} \in \mathbb{C}^{\mathbf{n}}$ represent vectors of circularly symmetric complex

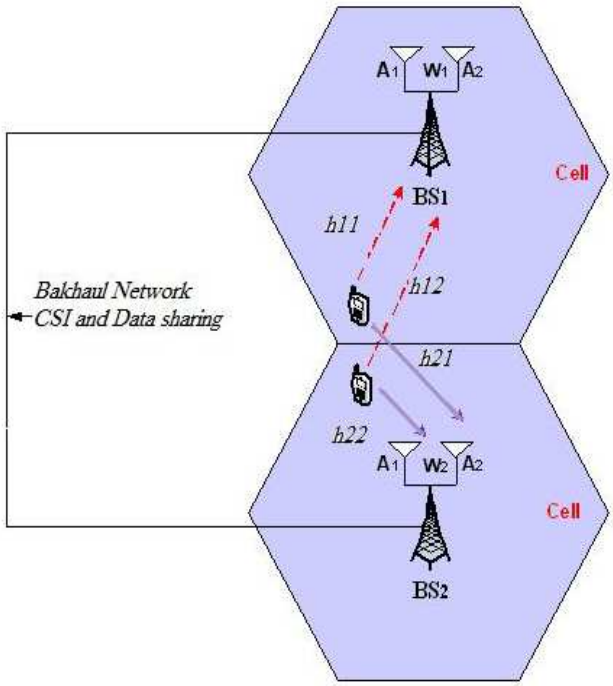

Figure 5.1: MCP in a cluster of two base stations.

Gaussian random noise with zero mean and identity covariance, i.e., with $\mathcal{CN}(0, I)$. $\mathbf{h_{ij}}$ represent the complex gains of the sub-channels between transmitter $\mathbf{i}$ and receiver $\mathbf{j}$, where the main links are the ones with $\mathbf{i} = \mathbf{j}$, and the interference links are the ones with $\mathbf{i} \neq \mathbf{j}$. $\sqrt{P1}$ and $\sqrt{P2}$ represent the amplitude of the transmitted signals from UT1 and UT2, respectively. And *snr* is the received signal to noise power ratio.

The cooperation between the two base stations is incorporated via using the upper bound of the achievable rates in MIMO channels [31], ICs [101], and BCs [112], as well as the MAC [101]. The achievable rates are:

$$R_1 \leq I(\mathbf{x_1}; \mathbf{y_1} | \mathbf{x_2}) \tag{5.3}$$

$$R_2 \leq I(\mathbf{x_2}; \mathbf{y_2} | \mathbf{x_1}) \tag{5.4}$$

$$R_1 + R_2 \leq min\ [I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}), I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_2})] \tag{5.5}$$

$$\leq I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}, \mathbf{y_2}) \tag{5.6}$$

Therefore, the optimization will be performed over the joint mutual information

subject to the users power constraints, as follows:

$$max\ I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}, \mathbf{y_2}) \tag{5.7}$$

Subject to:

$$P_1 \leq Q_1,\ P_2 \leq Q_2,\ P_1 \geq 0\ and\ P_2 \geq 0 \tag{5.8}$$

Where P_1 and P_2 are the transmitted power corresponding to each UT, Q_1 and Q_2 is the total and maximum power each UT can use, the channels considered are scalar channels, and precoding is precluded in this UL scenario.

5.3 Optimal Power Allocation with MCP

5.3.1 Gaussian Inputs

For Gaussian inputs, the mutual information is defined as:

$$I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}, \mathbf{y_2}) = log\ \left|\left(\mathbf{HPP}^{\dagger}\mathbf{H}^{\dagger} + \mathbf{I}\right)\right| \tag{5.9}$$

The following theorem provides a closed-form of the optimal power allocation solving the optimization problem of maximizing the mutual information in (5.9), subject to per user power constraints in (5.8).

Theorem 8. *The optimal power allocation for two UTs in the MCP framework* $(P_1^{\star}, P_2^{\star})$ *with Gaussian inputs follows the following form:*

$$\begin{cases} P_1^{\star} = Q_1, \\ P_2^{\star} = Q_2, \end{cases} \tag{5.10}$$

Proof. Theorem 8 follows the solution of the KKT conditions of (5.7) subject to (5.8), and due to the fact that the function is increasing with respect to the power, this also follows the fact that the matrix within the log is a positive definite matrix. Thus, we notice that the solution of the derivative with respect to P_1 leads to $P_2^{\star}$, and the derivative with respect to P_2 leads to $P_1^{\star}$. □

It can be easily verified that (5.9) is concave with respect to each user main power since the second derivative is always negative, and also through the positive definiteness of

the matrix in (5.9). Capitalizing on the relation between the gradient of the mutual information and the MMSE in [31], we re-investigate the result in the context of the MCP cooperative framework. The relation between the gradient of the mutual information in (5.7) and the MMSE is as follows:

$$\nabla_{\mathbf{P}}\ I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}, \mathbf{y_2}) = \mathbf{H}^{\dagger}\mathbf{HPE} \tag{5.11}$$

$$= \begin{bmatrix} \nabla_{P_{11}} I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}, \mathbf{y_2}) & \nabla_{P_{12}} I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}, \mathbf{y_2}) \\ \nabla_{P_{21}} I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}, \mathbf{y_2}) & \nabla_{P_{22}} I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}, \mathbf{y_2}) \end{bmatrix}, \tag{5.12}$$

where:

$$\nabla_{P11} I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}, \mathbf{y_2}) = {h_{11}}^* h_{11}\sqrt{P_1}E_{11} + {h_{21}}^* h_{21}\sqrt{P_1}E_{11} + {h_{11}}^* h_{12}\sqrt{P_2}E_{21} + {h_{21}}^* h_{22}\sqrt{P_2}E_{21} \tag{5.13}$$

$$\nabla_{P12} I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}, \mathbf{y_2}) = {h_{11}}^* h_{11}\sqrt{P_1}E_{12} + {h_{21}}^* h_{21}\sqrt{P_1}E_{12} + {h_{11}}^* h_{12}\sqrt{P_2}E_{22} + {h_{21}}^* h_{22}\sqrt{P_2}E_{22} \tag{5.14}$$

$$\nabla_{P21} I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}, \mathbf{y_2}) = {h_{12}}^* h_{11}\sqrt{P_1}E_{11} + {h_{22}}^* h_{21}\sqrt{P_1}E_{11} + {h_{12}}^* h_{12}\sqrt{P_2}E_{21} + {h_{22}}^* h_{22}\sqrt{P_2}E_{21} \tag{5.15}$$

$$\nabla_{P22} I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}, \mathbf{y_2}) = {h_{12}}^* h_{11}\sqrt{P_1}E_{12} + {h_{22}}^* h_{21}\sqrt{P_1}E_{12} + {h_{12}}^* h_{12}\sqrt{P_2}E_{22} + {h_{22}}^* h_{22}\sqrt{P_2}E_{22}. \tag{5.16}$$

The MMSE matrix $\mathbf{E}$ defines the elements of the gradient of the mutual information with respect to the main links and interference links powers, as follows:

$$\mathbf{E} = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix}, \tag{5.17}$$

with the expansion of $\mathbf{E}$ is given by:

$$E_{11} = \mathbb{E}[(\mathbf{x_1} - \mathbb{E}(\mathbf{x_1}|\mathbf{y_1}, \mathbf{y_2}))(\mathbf{x_1} - \mathbb{E}(\mathbf{x_1}|\mathbf{y_1}, \mathbf{y_2}))^{\dagger}] \tag{5.18}$$

$$E_{12} = \mathbb{E}[(\mathbf{x_1} - \mathbb{E}(\mathbf{x_1}|\mathbf{y_1}, \mathbf{y_2}))(\mathbf{x_2} - \mathbb{E}(\mathbf{x_2}|\mathbf{y_1}, \mathbf{y_2}))^{\dagger}] \tag{5.19}$$

$$E_{21} = \mathbb{E}[(\mathbf{x_2} - \mathbb{E}(\mathbf{x_2}|\mathbf{y_1}, \mathbf{y_2}))(\mathbf{x_1} - \mathbb{E}(\mathbf{x_1}|\mathbf{y_1}, \mathbf{y_2}))^{\dagger}] \tag{5.20}$$

$$E_{22} = \mathbb{E}[(\mathbf{x_2} - \mathbb{E}(\mathbf{x_2}|\mathbf{y_1}, \mathbf{y_2}))(\mathbf{x_2} - \mathbb{E}(\mathbf{x_2}|\mathbf{y_1}, \mathbf{y_2}))^{\dagger}] \tag{5.21}$$

E_{11} and E_{22} correspond to MMSE1 and MMSE2, respectively; that is the per-user MMSE which defines the error in each main link, and their sum is the total error. However, E_{12} and E_{21} are covariance functions of the estimates of the decoded symbols for each UT. Note that the non-linear estimates of each user input is given as follows:

$$\widehat{\mathbf{x_1}} = \mathbb{E}[\mathbf{x_1}|\mathbf{y_1}, \mathbf{y_2}] = \sum_{\mathbf{x_1}, \mathbf{x_2}} \frac{\mathbf{x_1} \mathbf{p}_{\mathbf{y_1}, \mathbf{y_2}|\mathbf{x_1}, \mathbf{x_2}}(\mathbf{y_1}, \mathbf{y_2}|\mathbf{x_1}, \mathbf{x_2}) \mathbf{p}_{\mathbf{x_1}}(\mathbf{x_1}) \mathbf{p}_{\mathbf{x_2}}(\mathbf{x_2})}{\mathbf{p}_{\mathbf{y}}(\mathbf{y_1}, \mathbf{y_2})} \tag{5.22}$$

$$\widehat{\mathbf{x_2}} = \mathbb{E}[\mathbf{x_2}|\mathbf{y_1}, \mathbf{y_2}] = \sum_{\mathbf{x_1}, \mathbf{x_2}} \frac{\mathbf{x_2} \mathbf{p}_{\mathbf{y_1}, \mathbf{y_2}|\mathbf{x_1}, \mathbf{x_2}}(\mathbf{y_1}, \mathbf{y_2}|\mathbf{x_1}, \mathbf{x_2}) \mathbf{p}_{\mathbf{x_1}}(\mathbf{x_1}) \mathbf{p}_{\mathbf{x_2}}(\mathbf{x_2})}{\mathbf{p}_{\mathbf{y_1}, \mathbf{y_2}}(\mathbf{y_1}, \mathbf{y_2})}. \tag{5.23}$$

The non-linear estimates give a statistical intuition to the problem. However, in practical setups, such estimates can be found via the linear MMSE. In particular, the inputs estimates can be found by deriving the optimal Wiener receive filters solving a minimization optimization problem of the MMSE, the following theorem provides the linear estimates.

Theorem 9. *The linear estimates of the inputs of the vector* $\mathbf{x} = [\mathbf{x_1}, \mathbf{x_2}]$ *for each user in the virtual MIMO MCP framework given the output vector* $\mathbf{y} = [\mathbf{y_1}, \mathbf{y_2}]$ *can be expressed as:*

$$\widehat{\mathbf{x}} = \mathbf{P}^{\dagger}\mathbf{H}^{\dagger}(\mathbf{I} + \mathbf{P}^{\dagger}\mathbf{H}^{\dagger}\mathbf{P}\mathbf{H})^{-\mathbf{1}}\mathbf{y} \tag{5.24}$$

Proof. The proof of this theorem follows the same steps of the proof of Theorem 4 in Appendix D. □

5.3.2 Arbitrary Inputs

There are no closed-form expressions for the mutual information with arbitrary inputs; therefore, we need to capitalize on the relation between the gradient of the mutual information and the MMSE to derive the optimal power allocation for the generalized inputs. The following theorem provides a generalized form to the optimal power allocation solving the optimization problem defined in (5.7) subject to (5.8).

Theorem 10. *The optimal power allocation for two UTs in the MCP framework* $(P_1^{\star}, P_2^{\star})$ *with arbitrary inputs -in terms of channel coefficients, and the MMSE- takes*

the following form:

$$\lambda_1{}^{\star}\sqrt{P_1} = (h_{11}{}^{*}h_{11} + h_{21}{}^{*}h_{21})\sqrt{P_1}E_{11} + (h_{11}{}^{*}h_{12} + h_{21}{}^{*}h_{22})\sqrt{P_2}E_{21} \tag{5.25}$$

$$\lambda_2{}^{\star}\sqrt{P_2} = (h_{12}{}^{*}h_{11} + h_{22}{}^{*}h_{21})\sqrt{P_1}E_{12} + (h_{12}{}^{*}h_{12} + h_{22}{}^{*}h_{22})\sqrt{P_2}E_{22} \tag{5.26}$$

Proof. See Appendix H. □

Theorem 10 can be solved numerically to search the optimal power allocation of both users, where $\boldsymbol{\lambda_1}$ and $\boldsymbol{\lambda_2}$ are the Lagrange multipliers, it assimilates a mercury/waterfilling for the arbitrary inputs that compensate for the non-Gaussianess of the binary constellations, and a waterfilling for the Gaussian inputs, where more power is allotted to less noisy channels. Moreover, when both user powers are non-zero, we can re-write Theorem 10 with respect to the MMSE and the covariance as follows:

$$P_1^{\star} = \frac{1}{snr1|h_{11}^{*}h_{11} + h_{21}^{*}h_{21}|}\ mmse\ (snr1|h_{11}^{*}h_{11} + h_{21}^{*}h_{21}|P_1^{\star}) + \frac{1}{snr2|h_{11}^{*}h_{12} + h_{21}^{*}h_{22}|}\ cov\ (snr2|h_{11}^{*}h_{12} + h_{21}^{*}h_{22}|P_2^{\star}), \tag{5.27}$$

and,

$$P_2^{\star} = \frac{1}{snr2|h_{12}^{*}h_{12} + h_{22}^{*}h_{22}|}\ mmse\ (snr2|h_{12}^{*}h_{12} + h_{22}^{*}h_{22}|P_2^{\star}) + \frac{1}{snr1|h_{12}^{*}h_{11} + h_{22}^{*}h_{21}|}\ cov\ (snr1|h_{12}^{*}h_{11} + h_{22}^{*}h_{21}|P_1^{\star}). \tag{5.28}$$

Its straightforward to see that for the case when the inputs are time-division multiplexed, the optimal power allocation takes the form: $P_2^{\star} = Q_2$ when $P_1 = 0$, and $P_1^{\star} = Q_1$ when $P_2 = 0$. In addition, we can easily specialize the result of (5.25) and (5.26) to the one in (5.10) for Gaussian inputs. In particular, we substitute the linear MMSE for Gaussian inputs given by:

$$\mathbf{E} = (\mathbf{P}^{\dagger}\mathbf{H}^{\dagger}\mathbf{H}\mathbf{P} + \mathbf{I})^{-\mathbf{1}} \tag{5.29}$$

into (5.25) and (5.26), it can be easily shown that the optimal power allocation in Theorem 10 matches the one in Theorem 8. However, it is worth to notice other solutions that can be derived from (5.27) and (5.28) for other setups, like the two-user MAC-channel, [113].

5.4 Optimal Precoding with MCP

We consider the MCP cooperation in the DL where both BSs jointly cooperate to design the optimal precoding vectors that maximize their achievable rates. The optimization problem stays the same, this choice is convenient due to the fact that the joint mutual information upper bounds the broadcast framework, see [112]. The following theorem gives a generalization of the optimal precoder structure at the low- and high-snr regime. In particular, we will show, in the following theorem, that the precoder should admit a structure that performs matching of the strongest source modes to the weakest noise modes, and this alignment enforces a permutation process to appear in the power allocation.

Theorem 11. *The non-unique first-order optimal precoder that maximizes the mutual information with the MCP - that substitutes a MIMO setup - subject to an average power constraint can be written as follows:*

$$\mathbf{P}^{\star} = \mathbf{U}\mathbf{D}\mathbf{R}^{\dagger} \tag{5.30}$$

Where $\mathbf{U}$ is a unitary matrix, $\mathbf{D}$ is a diagonal matrix, and $\mathbf{R}$ is a rotation matrix.

Proof. Theorem 11 follows the relation between the gradient of the mutual information and the MMSE, with respect to the precoding matrix and the decomposition of its matrix components, see Appendix I. □

Note that each row vector of the precoding matrix $\mathbf{P}^{\star}$ corresponds to the optimal precoding weight that each BS should assign to each transmission in the MCP setup.

5.5 The Asymptotic Regimes

The following sections will specialize the study of the MCP framework to key asymptotic regimes of the SNR, particularly, the low-snr and the high-snr. One of the interesting observations that follows an in depth analysis of both regimes is that the optimal designs for both the low- and high-snr perform a diagonalization operation to at least one of the system elements that is causing correlation among different system variables; whether the correlation is among the sub-channels and so diagonalizing the

channel matrix, or among the inputs and so diagonalizing the error matrix. However, the optimal precoder is not necessarily diagonal. In fact, for Gaussian inputs, the optimal precoder is a diagonal matrix. However, the optimal precoder for arbitrary inputs is a non-diagonal matrix.

5.5.1 The Low-SNR Regime

Consider the analysis of the optimal power allocation and optimal precoding with MCP capitalizing on the low-snr expansions of the conditional probability distribution of the Gaussian noise defined as:

$$\mathbf{p}_{\mathbf{y_1},\mathbf{y_2}|\mathbf{x_1},\mathbf{x_2}}(\mathbf{y_1},\mathbf{y_2}|\mathbf{x_1},\mathbf{x_2}) = \frac{1}{\pi^{n_r}} e^{-\left\| \begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \end{bmatrix} - \sqrt{snr}\ \mathbf{HP} \begin{bmatrix} \mathbf{x_1} \\ \mathbf{x_2} \end{bmatrix} \right\|^2} \tag{5.31}$$

And the MMSE defined as:

$$mmse(snr) = \mathbb{E}\left[\left\| \mathbf{HP} \left(\begin{bmatrix} \mathbf{x_1} \\ \mathbf{x_2} \end{bmatrix} - \mathbb{E}\begin{bmatrix} \mathbf{x_1}|\mathbf{y_1},\mathbf{y_2} \\ \mathbf{x_2}|\mathbf{y_1},\mathbf{y_2} \end{bmatrix} \right) \right\|^{\mathbf{2}} \right] \tag{5.32}$$

The low-snr expansion of the MMSE matrix can be expressed as follows:

$$\mathbf{E} = \mathbf{I} - (\mathbf{HP})^{\dagger}\mathbf{HP}.snr + \mathcal{O}(snr^2) \tag{5.33}$$

Consequently, the low-snr expansion of the non-linear MMSE is given by the following theorem.

Theorem 12. *The low-snr expansion of the non-linear MMSE in* (5.32) *as* $snr \to 0$ *is given by:*

$$mmse(snr) = Tr\left\{\mathbf{HP}(\mathbf{HP})^{\dagger}\right\} - Tr\left\{(\mathbf{HP}(\mathbf{HP}^{\dagger}))^{\mathbf{2}}\right\}.snr + \mathcal{O}(snr^2) \tag{5.34}$$

Proof. See Appendix J. □

Now, by virtue of the relationship between mutual information and MMSE, the Taylor low-snr expansion of the mutual information is given by:

$$I(snr) = Tr\left\{\mathbf{HP}(\mathbf{HP})^{\dagger}\right\}.snr - Tr\left\{(\mathbf{HP}(\mathbf{HP}^{\dagger}))^{\mathbf{2}}\right\}.\frac{snr^2}{2} + \mathcal{O}(snr^3) \tag{5.35}$$

According to (5.35), for first-order optimality, the form of the optimal precoder - in the DL - follows from the low-snr expansions the form of optimal precoder for the complex Gaussian inputs settings. To prove this claim, we will re-define our optimization problem as follows:

$$max\ Tr\left\{\mathbf{HP}(\mathbf{HP})^{\dagger}\right\}.snr \tag{5.36}$$

Subject to:

$$Tr\left\{\mathbf{PP}^{\dagger}\right\} \leq 1 \tag{5.37}$$

Then, we do an eigen value decomposition such that: $\mathbf{H}^{\dagger}\mathbf{H} = \mathbf{U}\mathbf{\Omega}\mathbf{U}^{\dagger}$. Let $\tilde{\mathbf{P}} = \mathbf{U}^{\dagger}\mathbf{P}$, and let $\mathbf{Z} \succeq \mathbf{0}$ is a positive semi-definite matrix, such that: $\mathbf{Z} = \tilde{\mathbf{P}}\tilde{\mathbf{P}}^{\dagger}$. Then, the optimization problem can be re-written as: $max\ \ Tr\left\{\mathbf{Z}\mathbf{\Omega}\right\}$ subject to: $Tr\left\{\mathbf{Z}\right\} \leq 1$, which leads to the solution, $\mathbf{Z} = \lambda^{-\mathbf{1}}\mathbf{\Omega}snr$. In the DL, this result proves that the optimal precoder in the low-snr performs mainly two operations: Firstly, it aligns the transmit directions with the eigenvectors of each user sub-channel. Secondly, it performs power allocation over the user sub-channels; i.e., the main and the interference links. Moreover, in the UL, it can be easily shown that specializing the low-snr results to the Gaussian inputs case by deriving the Taylor expansion of (5.9) as $snr \to 0$ will follow the one in (5.35) for the general inputs. It follows that the optimal power allocation as $snr \to 0$, for any inputs regardless of their signaling will follow the one for the Gaussian inputs in (5.10). Consequently, the mutual information is insensitive to the distribution of the inputs signaling in the low-snr. To prove this claim, we substitute all channel gains and powers into (5.36), and the optimization problem will be as follows:

$$max\left\{{h_{11}}^2P_1 + {h_{12}}^2P_2 + {h_{21}}^2P_1 + {h_{22}}^2P_2\right\}.snr \tag{5.38}$$

Subject to:

$$P_1 \leq Q_1 \tag{5.39}$$

$$P_2 \leq Q_2 \tag{5.40}$$

It follows that the gradient of the mutual information in (5.38) with respect to the input powers is only a function of the channel states and the snr, that is:

$$\nabla_{P_1} I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}, \mathbf{y_2}) = \left\{h_{11}^2 + h_{12}^2\right\} snr = \lambda_1\sqrt{P_1} \tag{5.41}$$

and,

$$\nabla_{P_2} I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}, \mathbf{y_2}) = \left\{h_{21}^2 + h_{22}^2\right\} snr = \lambda_2\sqrt{P_2} \tag{5.42}$$

Therefore, when $snr \rightarrow 0$, we can write the result (5.41) and (5.42) in a matrix formulation, see [Eq.48, [47]], as follows:

$$D_{\mathbf{P}} I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}, \mathbf{y_2}) = vec\left(\mathbf{H}^{\dagger}\mathbf{H}\right) \tag{5.43}$$

And this proves our claim.

5.5.2 The High-SNR Regime

The characterization of the optimal precoder at the high-snr is known to be an NP-hard problem. In fact, the lack of explicit expressions for the capacity of binary input constellations makes the goal more difficult. In [30], they provide an explicit expression of the capacity of SISO Gaussian channels with BPSK inputs and the MMSE counter part. They verified the fundamental relation between the mutual information and the MMSE for the special case of Gaussian inputs to the BPSK signaling, therefore, its proved for general inputs. The MMSE and the mutual information for BPSK signaling over SISO AWGN channel, respectively, are obtained as:

$$mmse(snr) = 1 - \int_{-\infty}^{\infty} \frac{e^{-(\zeta - \sqrt{\mathbf{snr}})^{\mathbf{2}}}}{\sqrt{\pi}} tanh(2\sqrt{snr}\zeta) d\zeta \tag{5.44}$$

$$I(snr) = snr - \int_{-\infty}^{\infty} \frac{e^{-(\zeta - \sqrt{\mathbf{snr}})^{\mathbf{2}}}}{\sqrt{\pi}} log\ cosh(2\sqrt{snr}\zeta) d\zeta \tag{5.45}$$

The authors in [30] didn't provide a detailed proof of their result, therefore, we present the derivation in Appendix K. The tanh term corresponds to the conditional mean estimate of the input or the non-linear estimate $\mathbb{E}[\mathbf{x}|\mathbf{y}]$. Moreover, we can also see how the mmse in (5.44) relates to the error function, or in other words, to the probability of bit error rate of BPSK inputs under AWGN channel as follows:

$$mmse(snr) = 1 - \int_{-\infty}^{\infty} \frac{e^{-(\zeta - \sqrt{\mathbf{snr}})^{\mathbf{2}}}}{\sqrt{\pi}} tanh(2\sqrt{snr}\zeta) d\zeta \tag{5.46}$$

$$\geq 1 - \int_{0}^{\infty} \frac{e^{-(\zeta - \sqrt{\mathbf{snr}})^{\mathbf{2}}}}{\sqrt{\pi}} d\zeta \tag{5.47}$$

$$= 1 - \int_{\sqrt{snr}}^{\infty} \frac{e^{-\zeta^{\mathbf{2}}}}{\sqrt{\pi}} d\zeta = \frac{1}{2} erfc\left(\sqrt{snr}\right) \tag{5.48}$$

In addition, its worth to observe the geometric properties of the solution in (5.45). The cosh hyperbolic function defines the decision regions of the constellation points detection multiplied by a two sided error function, through which the mutual information

reaches a saturation limit of $log_2(2) = 1$; a normalized snr when the second term goes to zero, which makes any binary constellation matches the Gaussian one in terms of mutual information at the low-snr. Nonetheless, (5.44) and (5.45) can be multiplied by 2 to gain the closed-form expressions of QPSK inputs since the decision regions of the hyperbolic function extends over the real and imaginary parts with constellation points $\mathbf{x_{QPSK}} = \{\mathbf{1+1j}, \mathbf{1-1j}, \mathbf{-1-1j}, \mathbf{-1+1j}\}$ instead of $\mathbf{x_{BPSK}} = \{\mathbf{1}, \mathbf{-1}\}$. This geometric interpretation of the solution may help advancing future research to find closed form expressions for the mutual information of other types of binary constellations as well as multi-user setups. Moreover, the mutual information for SISO Gaussian channels with BPSK input distribution is expanded for high-snr and upper- and lower-bounded in terms of the minimum transmit lattice distance d_{min} and maximum receive lattice distance d_{max} between the constellation points, see [Theorem 4, [37]]. Therefore, capitalizing on the result in [37], we can derive the structure of the optimal precoder for each user terminal in the MCP setup. We can define the optimization problem using the upper bound as follows:

$$max\ log\ M - \frac{e^{-{d_{min}}^2 \frac{snr}{4}}}{M d_{min} snr}\big(\sqrt{\pi} - \frac{4.37 + 2\sqrt{\pi}}{{d_{min}}^2 snr}\big) \tag{5.49}$$

Subject to:

$$Tr\left\{\mathbf{P}\mathbf{P}^{\dagger}\right\} = 1 \tag{5.50}$$

With M is the product of the constellation cardinality and d_{min} is the the minimum distance between the M possible realizations of the input vector of the constellation, therefore, its defined as:

$$d_{min} = min_{i \neq j} \|\mathbf{HP}\left(\mathbf{x_i} - \mathbf{x_j}\right)\| \tag{5.51}$$

However, due to (5.51) and due to the fact that there is no explicit form for the optimal precoder in the high-snr regime, and by virtue of [Eq.16, [114]], which has been identified as an NP-hard problem, we define the initial value in the numerical solution of d_{min} as follows:

$$d_{min} = (\mathbf{HP})^{\dagger}\mathbf{HP} \tag{5.52}$$

Theorem 13. *The optimal precoder matrix in the high-snr for a BPSK constellation in the MCP setup is the solution of:*

$$\begin{aligned} D_{\mathbf{P}} I(x_1, x_2; y_1, y_2) = \\ - a\big((\mathbf{G^{-1}VW})^{\mathbf{T}} \otimes (\mathbf{G^{-1}P^{\dagger}C})\big) + \mathbf{a}\big((\mathbf{GW})^{\mathbf{T}} \otimes (\mathbf{UVP^{\dagger}C})\big) \\ + a\big(\mathbf{W^T}(\mathbf{UVGP^{\dagger}C})\big) + \mathbf{ab}\big((\mathbf{G^{-2}})^{\mathbf{T}} \otimes (\mathbf{UVG^{-1}P^{\dagger}C})\big) \\ + ab\big((\mathbf{G^{-1}})^{\mathbf{T}} \otimes (\mathbf{UVG^{-1}P^{\dagger}C})\big) = \mathbf{0} \end{aligned} \tag{5.53}$$

With $a = \frac{\sqrt{pi}}{2M}$, $b = \frac{4.37+2\sqrt{pi}}{4\sqrt{pi}}$, $\mathbf{C} = \frac{\mathbf{snr}}{\mathbf{4}}\mathbf{H}^{\dagger}\mathbf{H}$, $\mathbf{G} = \mathbf{P}^{\dagger}\mathbf{C}\mathbf{P}$, $\mathbf{U} = \mathbf{G}^{-\mathbf{1}}$, $\mathbf{V} = \mathbf{e}^{-\mathbf{G}^{\mathbf{2}}}$, *and* $\mathbf{W} = \mathbf{I} - \mathbf{b}\mathbf{G}^{-\mathbf{2}}$.

Proof. See Appendix L. □

Solving (5.53) numerically, we can see that the optimal precoding matrix is always a non-diagonal matrix.

5.6 MCP Distributed Algorithms

Comparing the cooperative framework using MCP - which models a MIMO channel - to the non-cooperative framework - which models an interference channel - we can analytically understand the benefits and drawbacks of each framework. In particular, the achievable rates via the cooperation is higher than that without cooperation. However, the processing overload and the CSI and data exchange overhead is another tradeoff. If the interference is orthogonal to the main channel, then we can preclude the interference effect, which can be dealt with through receive antenna diversity, and therefore we can maximize the information rates in the UL. However, we can attack the interference problem in the DL via adding a studied interference, i.e., via aligning the interference, or via precoding; as proposed in Theorem 11. We will introduce the MCP distributed algorithms, the first algorithm gives the optimal power allocation for the UL, and the second algorithm gives the optimal precoding for the DL.

5.7 Numerical Analysis

We shall now introduce a set of illustrative results that cast insight into the problem. The results for the Gaussian inputs setup is straightforward with the mutual information closed form. However, we used Monte-Carlo method to generate the achievable rates for arbitrary inputs. Its of particular relevance to notice that the Gaussian inputs distribution is optimal compared to the arbitrary inputs distribution from the rate achievability sense, as shown in Figure 5.2, and Figure 5.3. We can easily verify that for the same transmit power, higher achievable rates are possible with Gaussian inputs. Moreover, the arbitrary inputs may lie at a certain point at the null space of the channel causing a decay in the achievable rates. For the Gaussian inputs, the

Algorithm 1:
Optimum Power Allocation with MCP-Uplink
Full cooperation: CSI and data sharing

BS1 Input: $CSI1$, $\mathbb{E}(\mathbf{x_1}|\mathbf{y_1})$, $\mathbb{E}(\mathbf{x_2}|\mathbf{y_1})$

BS2 Input: $CSI2$, $\mathbb{E}(\mathbf{x_1}|\mathbf{y_2})$, $\mathbb{E}(\mathbf{x_2}|\mathbf{y_2})$

if BW Backhaul $\geq Threshold\ \tau$ **then**
  BS1 and BS2 declare congestion and minimal cooperation message
else
  BS1 sends decoded $\mathbf{x_1}:\ \mathbb{E}(\mathbf{x_1}|\mathbf{y_1})$ **and** $CSI1$ **to BS2**
  BS2 sends decoded $\mathbf{x_2}:\ \mathbb{E}(\mathbf{x_2}|\mathbf{y_2})$ **and** $CSI2$ **to BS1**
  BS1 and BS2 check resources$\rightarrow$ **handshaking**$\rightarrow$ $BS1/BS2$ **will do the processing.**

Output: The optimum power allocation in the UL is the solution for:

$$\mathbf{P_{k+1}} = \alpha_\mathbf{k}\mathbf{P_k} + \alpha_\mathbf{k}\lambda\mathbf{H}^\dagger\mathbf{H}\mathbf{P_k}\mathbf{E_k}$$

For the two UT case,

$$P_1{}^\star = \mathbf{P}(\mathbf{1},\mathbf{1})$$
$$P_2{}^\star = \mathbf{P}(\mathbf{2},\mathbf{2})$$

BS1 and BS2 share $P_1{}^\star$ **and** $P_2{}^\star$ **and feedback to** $UT1$ **and** $UT2$**.**

optimal power allocation chosen by each user is to use their own maximum power, as illustrated in Figure 5.4, therefore, this serves to maximize the data rates in the UL and DL scenarios. However, for the case of arbitrary inputs, the optimality is to search for a set where both inputs don't lie in the null space of the channel - Voronoi region - therefore, they don't cancel each other. Hence, optimal power allocation is not a sufficient solution, therefore, we can improve the decay in the mutual information either by orthogonalizing the inputs, or via precoding them.

In addition, Figure 5.5 illustrates the main ideas of interference and interference free channels with respect to the mutual information and the errors. Notice that the channel gains are chosen to be unity for main and interference links when interference is considered. This will assure that the channel will not amplify nor attinuate the transmitted signals. Therefore, we can see that the loss in the achievable rate is 0.5 bits, such that without interference the achievable rate is 2 bits, and with interference

Algorithm 2:
Optimum Precoding with MCP-Downlink
Full cooperation: CSI and data sharing

BS1 Input: $CSI1$, $\mathbf{x_1}$, $\mathbf{x_2}$

BS2 Input: $CSI2$, $\mathbf{x_1}$, $\mathbf{x_2}$

BS1 and BS2 perform SVD(H): $\mathbf{H} = \mathbf{U_H \Lambda_H V_H^{\dagger}}$

$$BS1 \textbf{ sends } (h_{21}\nu_{h_{11}}\sqrt{P_1} + h_{22}\nu_{h_{21}}\sqrt{P_1})\mathbf{x_1} \textbf{ to } BS2$$
$$BS2 \textbf{ sends } (h_{11}\nu_{h_{12}}\sqrt{P_2} + h_{12}\nu_{h_{22}}\sqrt{P_2})\mathbf{x_2} \textbf{ to } BS1$$

Output: The optimum precoding in the DL is done via each BS solving,

$$\mathbf{P_k} = \mathbf{V_H diag}(\sqrt{P_1}, \sqrt{P_2})$$
$$\mathbf{P_{k+1}} = \alpha_\mathbf{k}\mathbf{P_k} + \alpha_\mathbf{k}\lambda\mathbf{H}^{\dagger}\mathbf{H}\mathbf{P_k}\mathbf{E_k}$$

BS1 transmits :

$$(h_{11}\nu_{h_{11}}\sqrt{P_1} + h_{12}\nu_{h_{21}}\sqrt{P_1})\mathbf{x_1} + (h_{11}\nu_{h_{12}}\sqrt{P_2} + h_{12}\nu_{h_{22}}\sqrt{P_2})\mathbf{x_2}$$

BS2 transmits :

$$(h_{21}\nu_{h_{11}}\sqrt{P_1} + h_{22}\nu_{h_{22}}\sqrt{P_1})\mathbf{x_1} + (h_{21}\nu_{h_{12}}\sqrt{P_2} + h_{22}\nu_{h_{22}}\sqrt{P_2})\mathbf{x_2}$$

The process will be iteratively repeated for each simultaneous transmission of $BS1$ and $BS2$.

the achievable rate is 1.5 bits. However, the 0.5 bit loss is induced through E_{12} and E_{21} causing $E_{11} + E_{22}$ to saturate at 0.5 instead of zero.

Table 2 presents few quantified results for the mutual information with and without interference for different cooperation levels, i.e., for different MIMO setups. The achievable rates and losses are quantified via Monte-Carlo method, for higher constellations like $16-$QAM the number of permutations for: a 2 x 2 MIMO setup are 256, 3 x 3 MIMO setup are 4096, 4 x 4 MIMO setup are 65536; therefore, we limit the presentation to results that are computationally less demanding.

Finally, it is worth to present a result that confirms the value of cooperation; i.e., the importance of the MCP network MIMO. Figure 5.6 not only illustrates the gain via

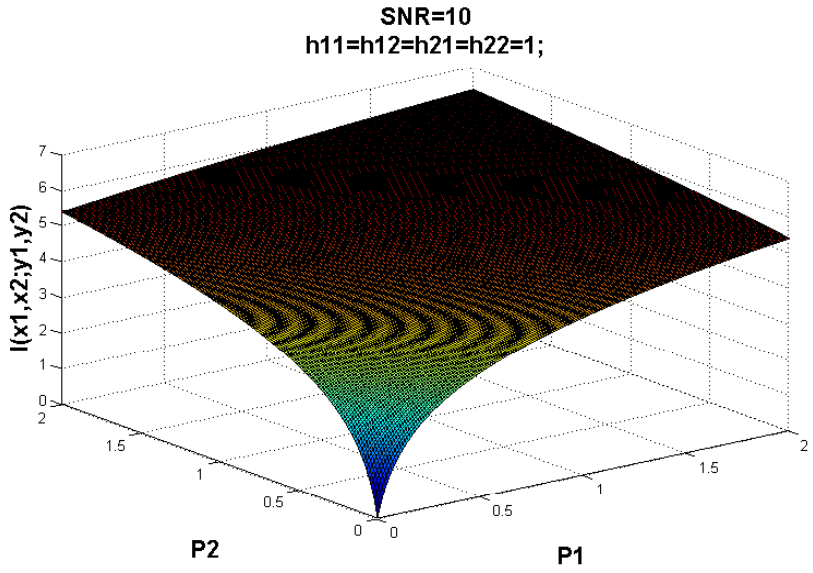

Figure 5.2: The achievable rate for the MCP with respect to UTs main power with Gaussian inputs.

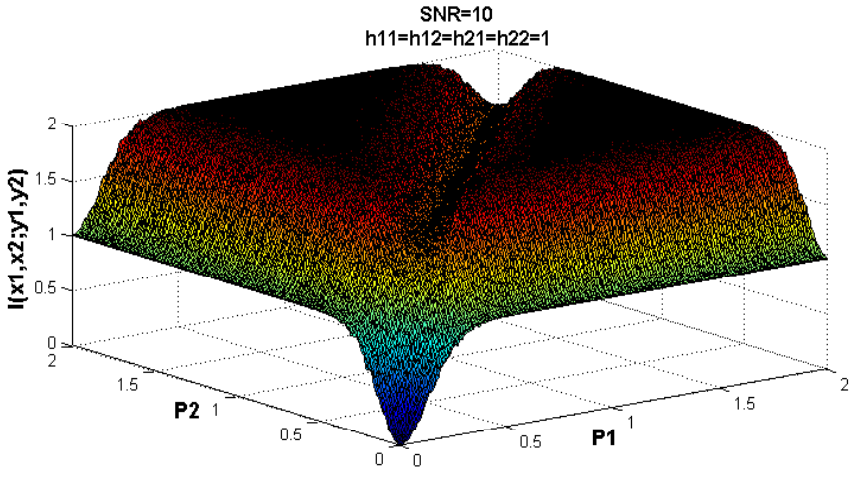

Figure 5.3: The achievable rate for the MCP with respect to UTs main power with BPSK inputs.

precoding in comparison to power allocation techniques, but it shows also that even for a diagonal channel without cooperation, i.e., without interference, a non-diagonal

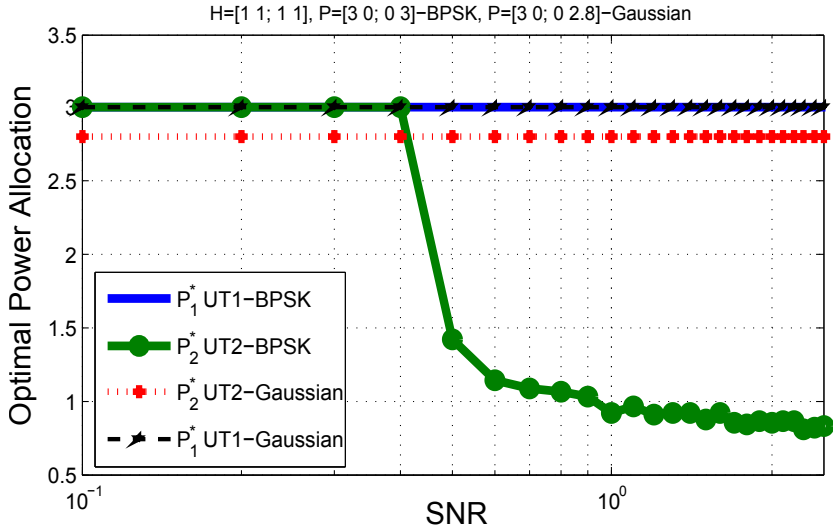

Figure 5.4: Optimal power allocation for Gaussian inputs and BPSK inputs.

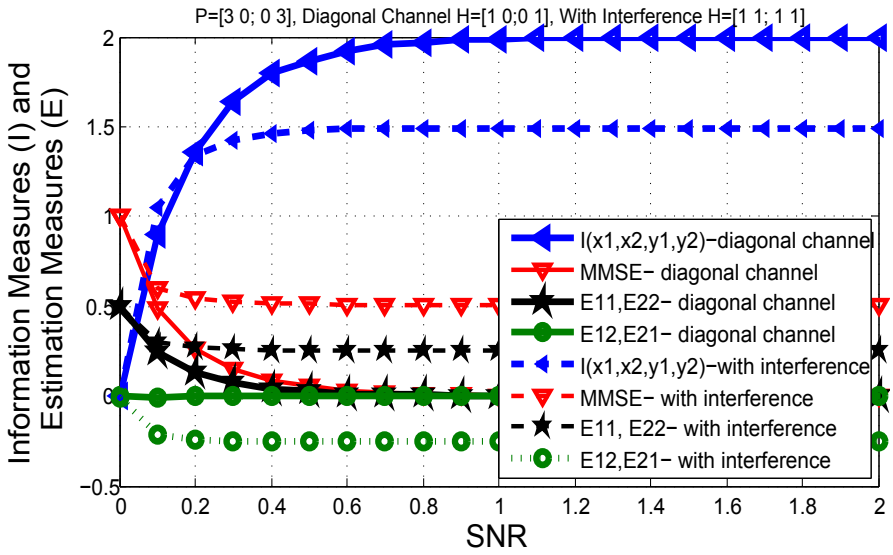

Figure 5.5: Information measures and estimation measures.

precdoing matrix $\mathbf{P}^{\star}$ is a rate maximizer and better than the mercury waterfilling power allocation $\mathbf{P_{TPC}}$ with a total power constraint alone and minimum distance of $\sqrt{6}$. In addition, the precoding which inherently includes power allocation and minimum distance maximization is also better than the power allocation with per user power constraint $\mathbf{P_{UTPC}}$ with a minimum distance of $\sqrt{8}$, for both the precoder and power allocation. The matrices used for comparison are as follows: $\mathbf{H}$=[$\sqrt{3}$ 0; 0 1], $\mathbf{P_{TPC}}$=[$1/\sqrt{2}$ 0; 0 $\sqrt{3/2}$], $\mathbf{P_{UTPC}}$=[1 0; 0 1], and $\mathbf{P}^{\star}$=[$1/\sqrt{2}$ $1/\sqrt{2}$;-

Table 1: Mutual information with and without interference

Signaling	MI without Int.(bits)	MI with Int.(bits)	Losses (bits)	MIMO setup
BPSK	2	1.5	0.5	2×2
QPSK	4	3	1	2×2
BPSK	3	1.8	1.2	3×3
QPSK	6	3.623	2.377	3×3
BPSK	4	2	2	4×4
QPSK	8	4	4	4×4

$1/\sqrt{2}$ $1/\sqrt{2}]$. Therefore, from a precoding perspective, the result illustrates a new look into interference through which a studied one can be a rate maximizer. Moreover, this casts insights that from a network level perspective, coding across packets could be of particular relevance to achieve the network capacity. Such framework is similar to network coding over coefficients drawn from smaller sets; called GFs, or as a special case of analog network coding, on the physical layer. NC is the focus of the last two Chapters of this PhD thesis, through which we addressed the delay problem with scenarios and technologies in practice.

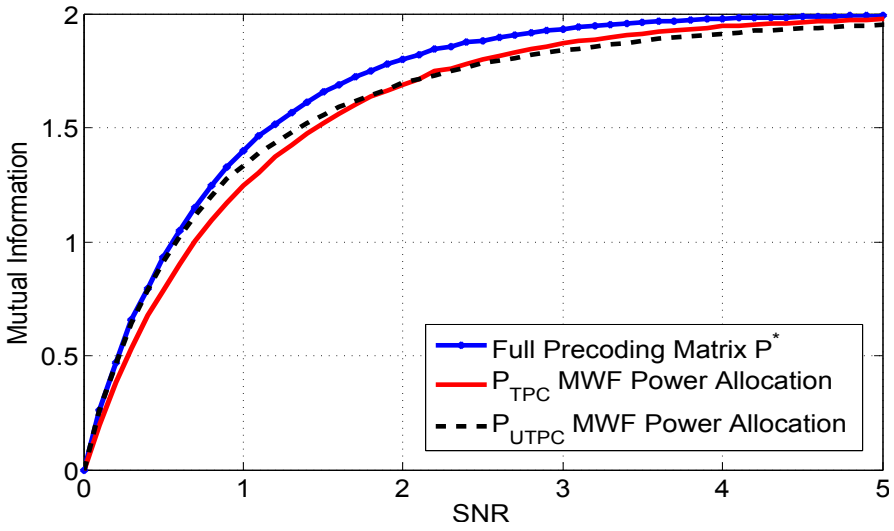

Figure 5.6: Mutual information for BPSK inputs with precoding and power allocation.

5.8 Conclusion

In this Chapter, we studied a cooperative framework where multi-cell processing is used between a cluster of two base stations. We derived the optimal power allocation and the optimal precoding structure which have been found to constitute the optimal setups for MIMO channels. We generalize a non-unique fixed point equation for the optimal precoding and power allocation in the two asymptotic regimes of the high- and low-snr. We provide an iterative approach for the design of the optimal precoding matrix for BPSK constellations at the high-snr. We build upon two distributed algorithms for the optimal solutions in the uplink and the downlink. It has been shown that the cooperation introduces a new look to interference through which a studied addition of interference can add positively to the spectral efficiency of the network. We have also highlighted the coupling between the information rates and the error rates through which the error - or particularly the covariance - caused by the interference links substitutes the drop in the information rates in the main links, this casts insights into having an interpretation of the interference with respect to the channel, transmitted power, and the error. In addition, this has explained why a non-studied interference is a capacity limiting factor in communication channels. Besides the implications of our designs, which are discussed in the conclusion of the previous Chapter, on defining fundamental limits of cooperation and providing new optimal designs that mitigate the effect of possible and known interferers. The impact of our studied framework extends to more generalized models that include in its structure more information about the system. In particular, the framework introduced in this Chapter casts insights into investigating the connections between information-theoretic measures and estimation-theoretic measures on a network level. Therefore, the system model can include the geometrical properties of the nodes in the network, and this structure exploits a network coding framework. In such framework, it is instrumental to revisit our derivations and design optimal setups that are adapted to the network in conjunction with the knowledge about the physical system. Moreover, of particular relevance are the implications of our derivations and optimal designs on other applications of measurement systems. For instance, systems that are not only interested in reconstructing the original data with lowest error rates, from an estimation perspective, but also aims to do a classification of the data into certain classes. More specifically, the optimal percoding matrix fixed point equation performs a pre-processing over the original data before it is contaminated by noise, and so, it acquires in its structure a maximization of the data rates or information obtained. Similarly, such structure provides studied projections that

can be of importance to validate in reconstructing the signal from a compressive measurement. If the optimal precoder for arbitrary inputs distributions is a one which has a non-diagonal structure and a minimum distance maximizer, we could expect that the sparsity of compressed measurements can be designed with similar setups in order to be reconstructed correctly. Furthermore, our formulation of an iterative solution to solve the NP-hard problem can be used to solve similar problems, minimizing the search space into smaller dimensions, via smaller search spaces that are practically relevant to the physical systems under study.

Appendix H: Proof of Theorem 10

Theorem 10 follows the KKT conditions solving (5.7), subject to (5.8) the relation between the gradient of the mutual information and the MMSE. First, we define the Lagrangian of the optimization problem as follows:

$$\mathcal{L}(P_1, P_2, \lambda_1, \lambda_2) = -I(x_1, x_2; y_1, y_2) - \lambda_1(Q_1 - P_1) - \lambda_2(Q_2 - P_2) - \mu_1 P_1 - \mu_2 P_2 \quad (5.54)$$

The relation between the gradient of the mutual information with respect to the diagonal power allocation matrix $\mathbf{P} = \mathbf{diag}\left(\sqrt{P_1}, \sqrt{P_2}\right)$ and the MMSE for linear vector Gaussian channels (MIMO) is given by:

$$\nabla_{\mathbf{P}} I(x_1, x_2; y_1, y_2) = \mathbf{H}^{\dagger}\mathbf{HPE} \quad (5.55)$$

and,

$$\nabla_{\mathbf{PP}^{\dagger}} I(x_1, x_2; y_1, y_2)\mathbf{P} = \mathbf{H}^{\dagger}\mathbf{HPE} \quad (5.56)$$

Given that the inputs covariance and the noise covariance are identities. To define the conditions of the theorem, lets re-write the gradient of the Lagrangian:

$$\nabla_{P_1} \mathcal{L}(P_1, P_2, \lambda_1, \lambda_2) = -\nabla_{P_1} I(x_1, x_2; y_1, y_2) + \lambda_1 - \mu_1, \quad (5.57)$$

and,

$$\nabla_{P_2} \mathcal{L}(P_1, P_2, \lambda_1, \lambda_2) = -\nabla_{P_2} I(x_1, x_2; y_1, y_2) + \lambda_2 - \mu_2, \quad (5.58)$$

with primal feasibility condition, $\lambda_1(Q_1 - P_1) = 0$, $\mu_1 P_1 = 0$, $\lambda_2(Q_2 - P_2) = 0$, and $\mu_2 P_2 = 0$, and dual feasibility condition, $\lambda_1 \geq 0$, and $\lambda_2 \geq 0$. It follows that:

$$\lambda_1{}^{\star}\sqrt{P_1} = (h_{11}{}^{*}h_{11} + h_{21}{}^{*}h_{21})\sqrt{P_1}E_{11} + (h_{11}{}^{*}h_{12} + h_{21}{}^{*}h_{22})\sqrt{P_2}E_{21} \quad (5.59)$$

$$\lambda_2{}^{\star}\sqrt{P_2} = (h_{12}{}^{*}h_{11} + h_{22}{}^{*}h_{21})\sqrt{P_1}E_{12} + (h_{12}{}^{*}h_{12} + h_{22}{}^{*}h_{22})\sqrt{P_2}E_{22} \quad (5.60)$$

Case 1: $P_1 = 0$, and $P_2 > 0$. It follows that:
$\mu_1 \geq 0$, and $\mu_2 = 0$, taking the gradient with respect to P_1 for the Lagrangian and applying the KKT conditions follows that: $P_2 = Q_2$ when $\lambda_2 \leq (h_{12}^{*}h_{12} + h_{22}^{*}h_{22})E_{22}$.

Case 2: $P_1 > 0$, and $P_2 = 0$. It follows that:
$\mu_1 = 0$, and $\mu_2 \geq 0$, taking the gradient with respect to P_2 for the Lagrangian and

applying the KKT conditions follows that: $P_1 = Q_1$ when $\lambda_1 \leq (h_{11}^* h_{11} + h_{21}^* h_{21})E_{11}$.

Case 3: $P_1 > 0$, and $P_2 > 0$. It follows that:
$\mu_1 = 0$, and $\mu_2 = 0$, and the generalized power allocation for both UTs follows:

$$\sqrt{P_1} = \frac{1}{\lambda_1^\star}({h_{11}}^* h_{11} + {h_{21}}^* h_{21})\sqrt{P_1}E_{11} + \frac{1}{\lambda_1^\star}({h_{11}}^* h_{12} + {h_{21}}^* h_{22})\sqrt{P_2}E_{21} \tag{5.61}$$

$$\sqrt{P_1} = \frac{1}{\lambda_1^\star}({h_{11}}^* h_{11} + {h_{21}}^* h_{21})\sqrt{P_1}\mathbb{E}[(x_1 - \mathbb{E}(x_1|y_1, y_2))(x_1 - \mathbb{E}(x_1|y_1, y_2))^\dagger] + \frac{1}{\lambda_1^\star}({h_{11}}^* h_{12} + {h_{21}}^* h_{22})\sqrt{P_2}\mathbb{E}[(x_2 - \mathbb{E}(x_2|y_1, y_2))(x_1 - \mathbb{E}(x_1|y_1, y_2))^\dagger]. \tag{5.62}$$

$$P_1^\star = \frac{1}{snr1|h_{11}^* h_{11} + h_{21}^* h_{21}|}\ mmse\ (snr1|h_{11}^* h_{11} + h_{21}^* h_{21}|P_1^\star) + \frac{1}{snr2|h_{11}^* h_{12} + h_{21}^* h_{22}|}\ cov\ (snr2|h_{11}^* h_{12} + h_{21}^* h_{22}|P_2^\star). \tag{5.63}$$

and,

$$\sqrt{P_2} = \frac{1}{\lambda 2^*}({h_{12}}^* h_{11} + {h_{22}}^* h_{21})\sqrt{P_1}E_{12} + \frac{1}{\lambda_2^\star}({h_{12}}^* h_{12} + {h_{22}}^* h_{22})\sqrt{P_2}E_{22} \tag{5.64}$$

$$\sqrt{P_2} = \frac{1}{\lambda_2^\star}({h_{12}}^* h_{11} + {h_{22}}^* h_{21})\sqrt{P_1}\mathbb{E}[(x_1 - \mathbb{E}(x_1|y_1, y_2))(x_2 - \mathbb{E}(x_2|y_1, y_2))^\dagger] + \frac{1}{\lambda_2^\star}({h_{12}}^* h_{12} + {h_{22}}^* h_{22})\sqrt{P_2}\mathbb{E}[(x_2 - \mathbb{E}(x_2|y_1, y_2))(x_2 - \mathbb{E}(x_2|y_1, y_2))^\dagger]. \tag{5.65}$$

$$P_2^\star = \frac{1}{snr2|h_{12}^* h_{12} + h_{22}^* h_{22}|}\ mmse\ (snr2|h_{12}^* h_{12} + h_{22}^* h_{22}|P_2^\star) + \frac{1}{snr1|h_{12}^* h_{11} + h_{22}^* h_{21}|}\ cov\ (snr1|h_{12}^* h_{11} + h_{22}^* h_{21}|P_1^\star). \tag{5.66}$$

Therefore, Theorem 10 has been proved.

Appendix I: Proof of Theorem 11

We can show that the optimum precoding matrix for a MIMO setup satisfies the following fixed point equation:

$$\mathbf{P}^\star = \nu^{-1}\mathbf{H}^\dagger\mathbf{H}\mathbf{P}^\star\mathbf{E} \tag{5.67}$$

$$= \nu^{-1}\mathbf{H}^{\dagger}\mathbf{H}\mathbf{P}^{\star}\mathbb{E}[XX^{\dagger} - \mathbb{E}[X|Y]\mathbb{E}[X|Y]^{\dagger}] \tag{5.68}$$

$$= \nu^{-1}\mathbf{H}^{\dagger}\mathbf{H}\mathbf{P}^{\star}\mathbb{E}[\mathbf{C} - \hat{\mathbf{C}}] = \nu^{-\mathbf{1}}\mathbf{H}^{\dagger}\mathbf{H}\mathbf{P}^{\star}\mathbb{E}[\mathbf{U}_{\mathbf{C}}^{\dagger}\mathbf{\Lambda}\mathbf{\Pi}\mathbf{U}_{\hat{\mathbf{C}}}], \tag{5.69}$$

via Wielandt-Hoffman theorem[2] [115]. With $\nu = ||\mathbf{H}\mathbf{H}^{\dagger}\mathbf{P}^{\star}\mathbf{E}||$, $X = [\mathbf{x_1}\ \mathbf{x_2}]^{\mathbf{T}}$, and $Y = [\mathbf{y_1}\ \mathbf{y_2}]^{\mathbf{T}}$. Therefore, digging into the depth of equation (6.16), we can do a singular value decomposition of the channel matrix $\mathbf{H} = \mathbf{U_H}\mathbf{\Lambda_H}\mathbf{V_H^{\dagger}}$ and the MMSE matrix $\mathbf{E} = \mathbf{U_E}\mathbf{\Lambda_E}\mathbf{V_E^{\dagger}}$, such that the optimal precoder is: $\mathbf{P} = \mathbf{U}\mathbf{D}\mathbf{R}^{\dagger}$, with $\mathbf{U} = \mathbf{V_H}$ corresponds to the channel right singular vectors, $\mathbf{D} = \mathbf{diag}\left(\sqrt{P_1}, \sqrt{P_2}\right)$, is a power allocation matrix; i.e., corresponds to the mercury/waterfilling [44]. $\mathbf{R} = \mathbf{\Pi}\mathbf{U_E}$ contains in its structure the eigenvectors of the MMSE matrix which can be permuted and/or projected with Π based on the correlation of the inputs and their estimates. Therefore, the rotation matrix insures firstly, allocation of power into the strongest channel singular vectors, and secondly, diagonalizes the MMSE matrix to insure uncorrelating the error or in other words independence between inputs, see also [113], [37], and [107]. Therefore, Theorem 11 has been proved.

Appendix J: Proof of Theorem 12

First we will find the low-snr expansion of the MMSE matrix $\mathbf{E}$ in (6.19) as $snr \to 0$, then we will prove (5.34). The low-snr expansion of the conditional probability exponent is as follows:

$$\begin{aligned}
&\left\|\mathbf{y} - \sqrt{snr}\ \mathbf{H}\mathbf{P}\mathbf{x}\right\|^2 \\
&= \left(\mathbf{y} - \sqrt{snr}\ \mathbf{H}\mathbf{P}\mathbf{x}\right)^{\dagger}\left(\mathbf{y} - \sqrt{snr}\ \mathbf{H}\mathbf{P}\mathbf{x}\right) \\
&= \|\mathbf{y}\|^2 - \sqrt{snr}\ \left(\mathbf{y}^{\dagger}\mathbf{H}\mathbf{P}\mathbf{x} + \left(\mathbf{y}^{\dagger}\mathbf{H}\mathbf{P}\mathbf{x}\right)^{\dagger}\right) + snr\|\mathbf{H}\mathbf{P}\mathbf{x}\|^2 \\
&= \|\mathbf{y}\|^2 - 2\sqrt{snr}\ \mathcal{R}\left(\mathbf{y}^{\dagger}\mathbf{H}\mathbf{P}\mathbf{x}\right) + snr\ \ \|\mathbf{H}\mathbf{P}\mathbf{x}\|^2
\end{aligned} \tag{5.70}$$

Hence,

$$\mathbf{p}_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}|\mathbf{x}\right) = \frac{\mathbf{1}}{\pi^{\mathbf{n_r}}}\exp\left(-\left\|\mathbf{y} - \sqrt{\mathbf{snr}}\ \mathbf{H}\mathbf{P}\mathbf{x}\right\|^{\mathbf{2}}\right) \tag{5.71}$$

[2]The Wielandt-Hoffman Theorem states that if $\mathbf{A}$ and $\mathbf{B}$ are normal matrices and $\mathbf{C} = \mathbf{A} - \mathbf{B}$, and if a_i and b_i are the eigen values of $\mathbf{A}$ and $\mathbf{B}$ arranged so that $\sum_{i=1}^{n}|a_i - b_i|^2$ is the minimum over all possible orderings, then: $\sum_{i=1}^{n}|a_i - b_i|^2 \leq \|\mathbf{C}\|_F$.

$$= \frac{1}{\pi^{n_r}} \exp\left(-\|\mathbf{y}\|^2\right) \exp\left(2\sqrt{snr}\ \mathcal{R}\left(\mathbf{y}^\dagger \mathbf{HPx}\right) - snr\ \|\mathbf{HPx}\|^2\right) \tag{5.72}$$

However, due to:

$$\begin{aligned} &\exp\left(a\sqrt{snr} - bsnr\right) \\ &= \frac{\exp\left(a\sqrt{snr}\right)}{\exp\left(bsnr\right)} \\ &= \frac{1 + a\sqrt{snr} + \mathcal{O}\left(snr\right)}{1 + \mathcal{O}\left(snr\right)} \\ &= 1 + a\sqrt{snr} + \mathcal{O}\left(snr\right), \end{aligned} \tag{5.73}$$

the low-snr expansion of the conditional probability distribution of the Gaussian noise is defined as:

$$\mathbf{p}_{\mathbf{y}|\mathbf{x}}\left(\mathbf{y}|\mathbf{x}\right) = \frac{1}{\pi^{n_r}} \exp\left(-\|\mathbf{y}\|^2\right) \left(1 + 2\sqrt{snr}\ \mathcal{R}\left(\mathbf{y}^\dagger \mathbf{HPx}\right) + \mathcal{O}\left(snr\right)\right). \tag{5.74}$$

The first term of the MMSE matrix $\mathbf{E}$ is $\mathbb{E}[\mathbf{x}\mathbf{x}^\dagger] = \mathbf{I}$. However, to find the second term of $\mathbf{E}$ we need to substitute (5.74) as follows:

$$\mathbb{E}_y[\mathbb{E}_{\mathbf{x}|\mathbf{y}}[\mathbf{x}|\mathbf{y}]\left(\mathbb{E}_{\mathbf{x}|\mathbf{y}}[\mathbf{x}|\mathbf{y}]\right)^\dagger] = \int_{\mathbf{y}\in\mathbb{C}^{n_r}} \frac{1}{\pi^{n_r}} \exp\left(-\|\mathbf{y}\|^2\right) \times \frac{\sum_x \mathbf{x}\left(1 + 2\sqrt{snr}\mathcal{R}\left(\mathbf{y}^\dagger \mathbf{HPx}\right) + \mathcal{O}\left(snr\right)\right) p_x\left(\mathbf{x}\right) \left(\sum_x \mathbf{x}\left(1 + 2\sqrt{snr}\mathcal{R}\left(\mathbf{y}^\dagger \mathbf{HPx}\right) + \mathcal{O}\left(snr\right)\right) p_x\left(\mathbf{x}\right)\right)^\dagger}{\sum_{x'}\left(1 + 2\sqrt{snr}\mathcal{R}\left(\mathbf{y}^\dagger \mathbf{HPx}'\right) + \mathcal{O}\left(snr\right)\right) p_{\mathbf{x}}\left(\mathbf{x}'\right)} d\mathbf{y} \tag{5.75}$$

$$= \int_{\mathbf{y}\in\mathbb{C}^{n_r}} \frac{1}{\pi^{n_r}} \exp\left(-\|\mathbf{y}\|^2\right) \times \frac{\left(\sqrt{snr}\left(\mathbf{HP}\right)^\dagger \mathbf{y} + \mathcal{O}\left(snr\right)\right)\left(\sqrt{snr}\left(\mathbf{HP}\right)^\dagger \mathbf{y} + \mathcal{O}\left(snr\right)\right)^\dagger}{1 + \mathcal{O}\left(snr\right)} d\mathbf{y} \tag{5.76}$$

However,

$$\begin{aligned}
&\sum_{x} x\left(1+2\sqrt{snr}\mathcal{R}\left(\mathbf{y}^{\dagger}\mathbf{HPx}\right)+\mathcal{O}\left(snr\right)\right)p_{\mathbf{x}}\left(\mathbf{x}\right)\\
&=\mathbb{E}_{x}\{x\}+\sqrt{snr}\sum_{\mathbf{x}}\mathbf{x}\left(\mathbf{y}^{\dagger}\mathbf{HPx}+\left(\mathbf{y}^{\dagger}\mathbf{HPx}\right)^{\dagger}\right)p_{\mathbf{x}}\left(\mathbf{x}\right)+\mathcal{O}\left(snr\right)\\
&=\sqrt{snr}\left(\sum_{\mathbf{x}}\mathbf{x}\mathbf{x}^{\mathbf{T}}\left(\mathbf{y}^{\dagger}\mathbf{HP}\right)^{\mathbf{T}}p_{\mathbf{x}}\left(\mathbf{x}\right)+\sum_{\mathbf{x}}\mathbf{x}\mathbf{x}^{\dagger}\left(\mathbf{HP}\right)^{\dagger}\mathbf{y}p_{\mathbf{x}}\left(\mathbf{x}\right)\right)+\mathcal{O}\left(snr\right)\\
&=\sqrt{snr}\left(\mathbb{E}_{\mathbf{x}}\{\mathbf{x}\mathbf{x}^{\mathbf{T}}\}\left(\mathbf{y}^{\dagger}\mathbf{HP}\right)^{\mathbf{T}}+\mathbb{E}_{\mathbf{x}}\{\mathbf{x}\mathbf{x}^{\dagger}\}\left(\mathbf{HP}\right)^{\dagger}\mathbf{y}\right)+\mathcal{O}\left(snr\right)\\
&=\sqrt{snr}\left(\mathbf{HP}\right)^{\dagger}\mathbf{y}+\mathcal{O}\left(snr\right),
\end{aligned}\tag{5.77}$$

and,

$$\begin{aligned}
&\sum_{\mathbf{x}'}\left(1+2\sqrt{snr}\ \mathcal{R}\left(\mathbf{y}^{\dagger}\mathbf{HPx}'\right)+\mathcal{O}\left(snr\right)\right)p_{\mathbf{x}}\left(\mathbf{x}'\right)\\
&=1+\sqrt{snr}\ \sum_{\mathbf{x}'}\left(\mathbf{y}^{\dagger}\mathbf{HPx}'+\left(\mathbf{y}^{\dagger}\mathbf{HPx}'\right)^{\dagger}\right)p_{\mathbf{x}}\left(\mathbf{x}\right)+\mathcal{O}\left(snr\right)\\
&=1+\sqrt{snr}\left(\mathbf{y}^{\dagger}\mathbf{HP}\mathbf{E}_{\mathbf{x}}\{\mathbf{x}\}+\mathbf{E}_{\mathbf{x}}\{\mathbf{x}^{\dagger}\}\left(\mathbf{y}^{\dagger}\mathbf{HP}\right)^{\dagger}\right)+\mathcal{O}\left(snr\right)\\
&=1+\mathcal{O}\left(snr\right)
\end{aligned}\tag{5.78}$$

Therefore,

$$\begin{aligned}
\mathbb{E}_{\mathbf{y}}[\mathbb{E}_{\mathbf{x}|\mathbf{y}}[\mathbf{x}|\mathbf{y}]\left(\mathbb{E}_{\mathbf{x}|\mathbf{y}}[\mathbf{x}|\mathbf{y}]\right)^{\dagger}]=&\int_{\mathbf{y}\in\mathbb{C}^{n_r}}\frac{1}{\pi^{n_r}}\exp\left(-\|\mathbf{y}\|^{2}\right)\\
&\times\frac{snr\ \ \left(\mathbf{HP}\right)^{\dagger}\mathbf{y}\mathbf{y}^{\dagger}\mathbf{HP}+\mathcal{O}\left(snr^{2}\right)}{1+\mathcal{O}\left(snr\right)}d\mathbf{y}
\end{aligned}\tag{5.79}$$

It follows that:

$$\mathbb{E}[\mathbb{E}[\mathbf{x}|\mathbf{y}]\mathbb{E}[\mathbf{x}|\mathbf{y}]^{\dagger}]=\left(\mathbf{HP}\right)^{\dagger}\mathbf{HP}.snr+\mathcal{O}\left(snr^{2}\right)\tag{5.80}$$

Consequently, the low-snr expansion of the MMSE matrix $\mathbf{E}$ is given as follows:

$$\mathbf{E}=\mathbf{I}-\left(\mathbf{HP}\right)^{\dagger}\mathbf{HP}.snr+\mathcal{O}(snr^{2}).\tag{5.81}$$

Therefore, we can express the MMSE in terms of the snr as follows:

$$\begin{aligned} & mmse\left(snr\right) \\ & = Tr\left\{\mathbf{HPE}\left(\mathbf{HP}\right)^{\dagger}\right\} \\ & = Tr\left\{\mathbf{HP}\left(I-\left(\mathbf{HP}\right)^{\dagger}\mathbf{HP}.snr+\mathcal{O}\left(snr^{2}\right)\right)\left(\mathbf{HP}\right)^{\dagger}\right\} \\ & = Tr\left\{\mathbf{HP}\left(\mathbf{HP}\right)^{\dagger}\right\}-Tr\left\{\left(\mathbf{HP}\left(\mathbf{HP}\right)^{\dagger}\right)^{2}\right\}snr+\mathcal{O}\left(snr^{2}\right) \end{aligned} \tag{5.82}$$

Therefore, Theorem 12 has been proved.

Appendix K: Derivation of mmse(snr) and I(snr) for BPSK Inputs

The non-linear MMSE matrix $\mathbf{E}$ is defined as:

$$\mathbf{E}=\mathbb{E}[(\mathbf{x}-\mathbb{E}[\mathbf{x}|\mathbf{y}])(\mathbf{x}-\mathbb{E}[\mathbf{x}|\mathbf{y}])^{\dagger}]=\mathbb{E}[\mathbf{x}\mathbf{x}^{\dagger}]-\mathbb{E}\mathbb{E}[\mathbf{x}|\mathbf{y}]\mathbb{E}[\mathbf{x}|\mathbf{y}], \tag{5.83}$$

with,

$$\mathbb{E}[\mathbf{x}|\mathbf{y}]=\frac{\sum_{\mathbf{x}}\mathbf{x}\mathbf{p}_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})\mathbf{p}_{\mathbf{x}}(\mathbf{x})}{\mathbf{p}_{\mathbf{y}}(\mathbf{y})} \tag{5.84}$$

$$=\frac{\sum_{\mathbf{x}}\mathbf{x}\mathbf{p}_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})\mathbf{p}_{\mathbf{x}}(\mathbf{x})}{\sum_{\mathbf{x}'}\mathbf{p}_{\mathbf{y}|\mathbf{x}'}(\mathbf{y}|\mathbf{x}')\mathbf{p}_{\mathbf{x}}(\mathbf{x}')} \tag{5.85}$$

For the BPSK inputs, the values of $\mathbf{x}=\{\mathbf{1},-\mathbf{1}\}$. Therefore, the non-linear estimate with respect to all possible permutations of the possible inputs is as follows:

$$\mathbb{E}[\mathbf{x}|\mathbf{y}]=\frac{\mathbf{e}^{-(\mathbf{y}-\sqrt{\mathbf{snr}})^{\mathbf{2}}}-\mathbf{e}^{-(\mathbf{y}+\sqrt{\mathbf{snr}})^{\mathbf{2}}}}{\mathbf{e}^{-(\mathbf{y}-\sqrt{\mathbf{snr}})^{\mathbf{2}}}+\mathbf{e}^{-(\mathbf{y}+\sqrt{\mathbf{snr}})^{\mathbf{2}}}} \tag{5.86}$$

However,

$$\mathbb{E}\left[\mathbb{E}(\mathbf{x}|\mathbf{y})\mathbb{E}(\mathbf{x}|\mathbf{y})^{\dagger}\right]=\int\left(\frac{\sum_{\mathbf{x}}\mathbf{x}\mathbf{p}_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})\mathbf{p}_{\mathbf{x}}(\mathbf{x})}{p_{y}(\mathbf{y})}\right)^{2}p_{y}(\mathbf{y})\mathbf{d}\mathbf{y} \tag{5.87}$$

$$\mathbb{E}\left[\mathbb{E}(\mathbf{x}|\mathbf{y})\mathbb{E}(\mathbf{x}|\mathbf{y})^{\dagger}\right]=\int\frac{\left(\sum_{\mathbf{x}}\mathbf{x}\mathbf{p}_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})\mathbf{p}_{\mathbf{x}}(\mathbf{x})\right)^{2}}{p_{y}(\mathbf{y})}\mathbf{d}\mathbf{y} \tag{5.88}$$

$$\mathbb{E}\left[\mathbb{E}(\mathbf{x}|\mathbf{y})\mathbb{E}(\mathbf{x}|\mathbf{y})^{\dagger}\right]=\int\frac{\sum_{\mathbf{x}}\mathbf{x}\mathbf{p}_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})\mathbf{p}_{\mathbf{x}}(\mathbf{x})}{p_{y}(\mathbf{y})}\sum_{\mathbf{x}}\mathbf{x}\mathbf{p}_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x})\mathbf{p}_{\mathbf{x}}(\mathbf{x})\mathbf{d}\mathbf{y} \tag{5.89}$$

Therefore,

$$\mathbb{E}\left[\mathbb{E}(\mathbf{x}|\mathbf{y})\mathbb{E}(\mathbf{x}|\mathbf{y})^{\dagger}\right] = \frac{1}{2\pi}\int \frac{e^{-(\mathbf{y}-\sqrt{\mathbf{snr}})^{\mathbf{2}}} - e^{-(\mathbf{y}+\sqrt{\mathbf{snr}})^{\mathbf{2}}}}{e^{-(\mathbf{y}-\sqrt{\mathbf{snr}})^{\mathbf{2}}} + e^{-(\mathbf{y}+\sqrt{\mathbf{snr}})^{\mathbf{2}}}}\left(e^{-(\mathbf{y}-\sqrt{\mathbf{snr}})^{\mathbf{2}}} - e^{-(\mathbf{y}+\sqrt{\mathbf{snr}})^{\mathbf{2}}}\right) d\mathbf{y} \tag{5.90}$$

Digging into the depth of the right hand side of equation (5.90), we have:

$$(\mathbf{y} - \sqrt{snr})^2 = \mathbf{y^2} - 2\sqrt{snr}\mathbf{y} + snr \tag{5.91}$$

$$(\mathbf{y} + \sqrt{snr})^2 = \mathbf{y^2} + 2\sqrt{snr}\mathbf{y} + snr \tag{5.92}$$

Thus,

$$\frac{e^{-(\mathbf{y}-\sqrt{\mathbf{snr}})^{\mathbf{2}}} - e^{-(\mathbf{y}+\sqrt{\mathbf{snr}})^{\mathbf{2}}}}{e^{-(\mathbf{y}-\sqrt{\mathbf{snr}})^{\mathbf{2}}} + e^{-(\mathbf{y}+\sqrt{\mathbf{snr}})^{\mathbf{2}}}} = \frac{e^{\mathbf{2}\sqrt{\mathbf{snr}}\mathbf{y}} - e^{-\mathbf{2}\sqrt{\mathbf{snr}}\mathbf{y}}}{e^{\mathbf{2}\sqrt{\mathbf{snr}}\mathbf{y}} + e^{-\mathbf{2}\sqrt{\mathbf{snr}}\mathbf{y}}} \tag{5.93}$$

$$= tanh\left(2\sqrt{snr}\mathcal{R}(\mathbf{y})\right) \tag{5.94}$$

It follows that:

$$\mathbb{E}\left[\mathbb{E}(\mathbf{x}|\mathbf{y})\mathbb{E}(\mathbf{x}|\mathbf{y})^{\dagger}\right] = \frac{1}{2\pi}\int tanh\left(2\sqrt{snr}\mathcal{R}(\mathbf{y})\right)\left(e^{-(\mathbf{y}-\sqrt{snr})^2} - e^{-(\mathbf{y}+\sqrt{snr})^2}\right) d\mathbf{y} \tag{5.95}$$

Therefore,

$$\mathbb{E}\left[\mathbb{E}(\mathbf{x}|\mathbf{y})\mathbb{E}(\mathbf{x}|\mathbf{y})^{\dagger}\right] = \frac{1}{2\pi}\int_{y\in C} tanh\left(2\sqrt{snr}\mathcal{R}(\mathbf{y})\right) e^{-\left(\mathbf{y}-\sqrt{\mathbf{snr}}\right)^2} - \frac{1}{2\pi}\int_{y\in C} tanh\left(2\sqrt{snr}\mathcal{R}(\mathbf{y})\right) e^{-\left(\mathbf{y}+\sqrt{\mathbf{snr}}\right)^2} d\mathbf{y} \tag{5.96}$$

However, it is known that:

$$tanh(-x) = -tanh(x) \tag{5.97}$$

and the expectation remains the same if $\mathbf{y} \sim\mathcal{N}(\sqrt{snr}, 1)$ replaced by $\mathbf{y} \sim\mathcal{N}(-\sqrt{snr}, 1)$, due to symmetry, therefore, we have:

$$\mathbb{E}\left[\mathbb{E}(\mathbf{x}|\mathbf{y})\mathbb{E}(\mathbf{x}|\mathbf{y})^{\dagger}\right] = \frac{1}{\pi}\int_{y\in C} tanh\left(2\sqrt{snr}\mathbf{y}\right) e^{-\left(\mathbf{y}-\sqrt{snr}\right)^2} d\mathbf{y} \tag{5.98}$$

Therefore, due to marginalization of the complex domain into the real domain, substituting ζ into (5.98), and $\mathbb{E}[\mathbf{x}\mathbf{x}^{\dagger}] = \mathbf{1}$ into (5.83), the mmse(snr) of a BPSK input over a SISO channel is given by:

$$mmse(snr) = 1 - \frac{1}{\sqrt{\pi}}\int_{\zeta\in R} tanh\left(2\sqrt{snr}\zeta\right) e^{-\left(\zeta-\sqrt{\mathbf{snr}}\right)^2}\mathbf{d}\zeta \tag{5.99}$$

Due to the fundamental relation between mmse(snr) and the mutual information, we will integrate (5.99) with respect to the snr to get the close form expression of the mutual information I(snr) of a BPSK input over a SISO channel, as follows:

$$I(snr) = snr - \frac{1}{\sqrt{\pi}} \int_{\zeta \in R} log\ cosh\left(2\sqrt{snr\zeta}\right) e^{-\left(\zeta - \sqrt{\mathbf{snr}}\right)^2} d\zeta \tag{5.100}$$

Therefore, we proved the relation for BPSK inputs as in (5.44), and (5.45).

Appendix L: Proof of Theorem 13

Substitute (5.52) into (5.49) and capitalizing on the matrix differentiation theories in [116], directly leads to the following:

$$\begin{aligned}
&\frac{\exp\left(-d^2\frac{snr}{4}\right)}{Md\sqrt{snr}}\left(\sqrt{\pi}\mathbf{I} - \frac{4.37 + 2\sqrt{\pi}}{d^2 snr}\right) = \frac{\sqrt{\pi}\exp\left(-d^2\frac{snr}{4}\right)}{\sqrt{4}Md\sqrt{\frac{snr}{4}}}\left(\mathbf{I} - \frac{4.37 + 2\sqrt{\pi}}{4\sqrt{\pi}d^2\frac{snr}{4}}\right) \\
&= a\frac{\exp\left(-d^2\frac{snr}{4}\right)}{d\sqrt{\frac{snr}{4}}}\left(\mathbf{I} - b\frac{1}{d^2\frac{snr}{4}}\right) \\
&= a\frac{\exp\left(-\left(\mathbf{P}^\dagger\mathbf{H}^\dagger\mathbf{H}\mathbf{P}\right)^2\frac{snr}{4}\right)}{\mathbf{P}^\dagger\mathbf{H}^\dagger\mathbf{H}\mathbf{P}\sqrt{\frac{\mathbf{snr}}{\mathbf{4}}}}\left(\mathbf{I} - b\frac{1}{\left(\mathbf{P}^\dagger\mathbf{H}^\dagger\mathbf{H}\mathbf{P}\right)^2\frac{snr}{4}}\right) \\
&= a\frac{\exp\left(-\left(\mathbf{P}^\dagger\mathbf{C}\mathbf{P}\right)^2\right)}{\mathbf{P}^\dagger\mathbf{C}\mathbf{P}}\left(\mathbf{I} - b\frac{1}{\left(\mathbf{P}^\dagger\mathbf{C}\mathbf{P}\right)^2}\right) \\
&= a\left(\mathbf{P}^\dagger\mathbf{C}\mathbf{P}\right)^{-1}\exp\left(-\left(\mathbf{P}^\dagger\mathbf{C}\mathbf{P}\right)^2\right)\left(\mathbf{I} - b\left(\mathbf{P}^\dagger\mathbf{C}\mathbf{P}\right)^{-2}\right)
\end{aligned} \tag{5.101}$$

where in the second equality we define:

$$a = \frac{\sqrt{\pi}}{\sqrt{4}M} \tag{5.102}$$

$$b = \frac{4.37 + 2\sqrt{\pi}}{4\sqrt{\pi}} \tag{5.103}$$

in the third equality we replaced d by $\mathbf{P}^\dagger\mathbf{H}^\dagger\mathbf{H}\mathbf{P}$ and in the fourth equality we define:

$$C = \sqrt{\frac{snr}{4}}\mathbf{H}^\dagger\mathbf{H} \tag{5.104}$$

Let,

$$F\left(\mathbf{P},\mathbf{P}^{*}\right)=a\left(\mathbf{P}^{\dagger}\mathbf{C}\mathbf{P}\right)^{-1}\exp\left(-\left(\mathbf{P}^{\dagger}\mathbf{C}\mathbf{P}\right)^{2}\right)\left(I-b\left(\mathbf{P}^{\dagger}\mathbf{C}\mathbf{P}\right)^{-2}\right) \tag{5.105}$$

$$U\left(\mathbf{P},\mathbf{P}^{*}\right)=\left(\mathbf{P}^{\dagger}\mathbf{C}\mathbf{P}\right)^{-1} \tag{5.106}$$

$$V\left(\mathbf{P},\mathbf{P}^{*}\right)=\exp\left(-\left(\mathbf{P}^{\dagger}\mathbf{C}\mathbf{P}\right)^{2}\right) \tag{5.107}$$

$$W\left(\mathbf{P},\mathbf{P}^{*}\right)=\left(\mathbf{I}-b\left(\mathbf{P}^{\dagger}\mathbf{C}\mathbf{P}\right)^{-2}\right) \tag{5.108}$$

$$G\left(\mathbf{P},\mathbf{P}^{*}\right)=\mathbf{P}^{\dagger}\mathbf{C}\mathbf{P} \tag{5.109}$$

We wish to determine the derivatives: $\mathcal{D}_{\mathbf{P}}F$ and $\mathcal{D}_{\mathbf{P}^{*}}F$. Our strategy is first to find the differentials of (5.106), (5.107) and (5.108) from which we will obtain the differential of (5.105) given by:

$$dF=a\left(\left(dU\right)VW+U\left(dV\right)W+UV\left(dW\right)\right) \tag{5.110}$$

which determines $\mathcal{D}_{\mathbf{P}}F$ and $\mathcal{D}_{\mathbf{P}^{*}}F$ via:

$$dvec\left(F\right)=\left(\mathcal{D}_{\mathbf{P}}F\right)dvec\left(\mathbf{P}\right)+\left(\mathcal{D}_{\mathbf{P}^{*}}F\right)dvec\left(\mathbf{P}^{*}\right) \tag{5.111}$$

$$\begin{aligned}
dU&=d\left(\left(\mathbf{P}^{\dagger}\mathbf{C}\mathbf{P}\right)^{-1}\right)\\
&=-\left(\mathbf{P}^{\dagger}\mathbf{C}\mathbf{P}\right)^{-1}d\left(\mathbf{P}^{\dagger}\mathbf{C}\mathbf{P}\right)\left(\mathbf{P}^{\dagger}\mathbf{C}\mathbf{P}\right)^{-1}\\
&=-\left(\mathbf{P}^{\dagger}\mathbf{C}\mathbf{P}\right)^{-1}\left(\left(d\mathbf{P}^{*}\right)^{T}\mathbf{C}\mathbf{P}+\mathbf{P}^{\dagger}\mathbf{C}\left(\mathbf{dP}\right)\right)\left(\mathbf{P}^{\dagger}\mathbf{C}\mathbf{P}\right)^{-1}\\
&=-\left(\mathbf{P}^{\dagger}\mathbf{C}\mathbf{P}\right)^{-1}\left(d\mathbf{P}^{*}\right)^{T}\mathbf{C}\mathbf{P}\left(\mathbf{P}^{\dagger}\mathbf{C}\mathbf{P}\right)^{\mathbf{-1}}-\left(\mathbf{P}^{\dagger}\mathbf{C}\mathbf{P}\right)^{\mathbf{-1}}\mathbf{P}^{\dagger}\mathbf{C}\left(\mathbf{dP}\right)\left(\mathbf{P}^{\dagger}\mathbf{C}\mathbf{P}\right)^{\mathbf{-1}}\\
&=-\mathbf{G}^{\mathbf{-1}}\left(\mathbf{dP}^{*}\right)^{\mathbf{T}}\mathbf{C}\mathbf{P}\mathbf{G}^{\mathbf{-1}}-\mathbf{G}^{\mathbf{-1}}\mathbf{P}^{\dagger}\mathbf{C}\left(\mathbf{dP}\right)\mathbf{G}^{\mathbf{-1}}
\end{aligned} \tag{5.112}$$

$$
\begin{aligned}
dV &= d\exp\left(-\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)^{2}\right)\\
&= Vd\left(-\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)^{2}\right)\\
&= -Vd\left(\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)\right)\\
&= -V\Big(d\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)+\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)d\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)\Big)\\
&= -V\Big(\left((d\mathbf{P}^{*})^{T}\mathbf{CP}+\mathbf{P}^{\dagger}\mathbf{C}\left(\mathbf{dP}\right)\right)\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)+\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)\left((d\mathbf{P}^{*})^{T}\mathbf{CP}+\mathbf{P}^{\dagger}\mathbf{C}\left(\mathbf{dP}\right)\right)\Big)\\
&= -V\Big(\left(d\mathbf{P}^{*}\right)^{T}\mathbf{CP}\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)+\mathbf{P}^{\dagger}\mathbf{C}\left(\mathbf{dP}\right)\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)\\
&\qquad+\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)\left(d\mathbf{P}^{*}\right)^{T}\mathbf{CP}+\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)\mathbf{P}^{\dagger}\mathbf{C}\left(\mathbf{dP}\right)\Big)\\
&= -V\Big(\left(d\mathbf{P}^{*}\right)^{T}\mathbf{CPG}+\mathbf{P}^{\dagger}\mathbf{C}\left(\mathbf{dP}\right)\mathbf{G}+\mathbf{G}\left(\mathbf{dP}^{*}\right)^{\mathbf{T}}\mathbf{CP}+\mathbf{GP}^{\dagger}\mathbf{C}\left(\mathbf{dP}\right)\Big)
\end{aligned}
\tag{5.113}
$$

$$
\begin{aligned}
dW &= d\left(\mathbf{I}-b\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)^{-2}\right)\\
&= -bd\left(\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)^{-2}\right)\\
&= -bd\left(\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)^{-1}\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)^{-1}\right)\\
&= -bd\Big(d\left(\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)^{-1}\right)\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)^{-1}+\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)^{-1}d\left(\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)^{-1}\right)\Big)\\
&= b\Big(\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)^{-1}\left(d\mathbf{P}^{*}\right)^{T}\mathbf{CP}\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)^{\mathbf{-2}}+\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)^{\mathbf{-1}}\mathbf{P}^{\dagger}\mathbf{C}\left(\mathbf{dP}\right)\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)^{\mathbf{-2}}\\
&\qquad+\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)^{-2}\left(d\mathbf{P}^{*}\right)^{T}\mathbf{CP}\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)^{\mathbf{-1}}+\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)^{\mathbf{-2}}\mathbf{P}^{\dagger}\mathbf{C}\left(\mathbf{dP}\right)\left(\mathbf{P}^{\dagger}\mathbf{CP}\right)^{\mathbf{-1}}\Big)\\
&= b\Big(\mathbf{G^{-1}}\left(\mathbf{dP}^{*}\right)^{\mathbf{T}}\mathbf{CPG^{-2}}+\mathbf{G^{-1}P}^{\dagger}\mathbf{C}\left(\mathbf{dP}\right)\mathbf{G^{-2}}\\
&\qquad+\mathbf{G^{-2}}\left(\mathbf{dP}^{*}\right)^{\mathbf{T}}\mathbf{CPG^{-1}}+\mathbf{G^{-2}P}^{\dagger}\mathbf{C}\left(\mathbf{dP}\right)\mathbf{G^{-1}}\Big)
\end{aligned}
\tag{5.114}
$$

Combining (5.110), dU, dV, and dW, we obtain:

$$
\begin{aligned}
dF = &-a\mathbf{G^{-1}}\left(\mathbf{dP}^{*}\right)^{\mathbf{T}}\mathbf{CPG^{-1}VW}-\mathbf{aG^{-1}P}^{\dagger}\mathbf{C}\left(\mathbf{dP}\right)\mathbf{G^{-1}VW}\\
&-a\mathbf{U}V\left(d\mathbf{P}^{*}\right)^{T}\mathbf{CPGW}-\mathbf{aUVP}^{\dagger}\mathbf{C}\left(\mathbf{dP}\right)\mathbf{GW}\\
&-a\mathbf{U}V\mathbf{G}\left(d\mathbf{P}^{*}\right)^{T}\mathbf{CPW}-\mathbf{aUVGP}^{\dagger}\mathbf{C}\left(\mathbf{dP}\right)\mathbf{W}\\
&+ab\mathbf{U}V\mathbf{G^{-1}}\left(d\mathbf{P}^{*}\right)^{T}\mathbf{CPG^{-2}}+\mathbf{abUVG^{-1}P}^{\dagger}\mathbf{C}\left(\mathbf{dP}\right)\mathbf{G^{-2}}\\
&+ab\mathbf{U}V\mathbf{G^{-2}}\left(d\mathbf{P}^{*}\right)^{T}\mathbf{CPG^{-1}}+\mathbf{abUVG^{-2}P}^{\dagger}\mathbf{C}\left(\mathbf{dP}\right)\mathbf{G^{-1}}
\end{aligned}
\tag{5.115}
$$

By applying $vec\left(\cdot\right)$ to both sides of (5.115) and capitalizing on:

$$vec\left(\mathbf{ABC}\right)=\left(\mathbf{C^{T}}\otimes\mathbf{A}\right)vec\left(\mathbf{B}\right) \tag{5.116}$$

$$\mathbf{K}vec\left(\mathbf{A}\right)=vec\left(\mathbf{A^{T}}\right) \tag{5.117}$$

where $\mathbf{K}$ is the commutation matrix we obtain:

$$\begin{aligned}
dvec\left(F\right)=&-a\left(\left(\mathbf{CPG^{-1}}V\mathbf{W}\right)^{T}\otimes\mathbf{G^{-1}}\right)\mathbf{K}dvec\left(\mathbf{P}^{*}\right)\\
&-a\left(\left(\mathbf{G^{-1}}V\mathbf{W}\right)^{T}\otimes\left(\mathbf{G^{-1}P}^{\dagger}\mathbf{C}\right)\right)dvec\left(\mathbf{P}\right)-a\left(\left(\mathbf{CPGW}\right)^{T}\otimes\left(\mathbf{U}V\right)\right)\mathbf{K}dvec\left(\mathbf{P}^{*}\right)\\
&-a\left(\left(\mathbf{GW}\right)^{T}\otimes\left(\mathbf{U}V\mathbf{P}^{\dagger}\mathbf{C}\right)\right)dvec\left(\mathbf{P}\right)-a\left(\left(\mathbf{CPW}\right)^{T}\otimes\left(\mathbf{U}V\mathbf{G}\right)\right)\mathbf{K}dvec\left(\mathbf{P}^{*}\right)\\
&-a\left(\mathbf{W^{T}}\otimes\left(\mathbf{UVGP}^{\dagger}\mathbf{C}\right)\right)dvec\left(\mathbf{P}\right)+ab\left(\mathbf{CPG^{-2}}\right)^{T}\otimes\left(\mathbf{U}V\mathbf{G^{-1}}\right)\mathbf{K}dvec\left(\mathbf{P}^{*}\right)\\
&+ab\left(\left(\mathbf{G^{-2}}\right)^{T}\otimes\left(\mathbf{U}V\mathbf{G^{-1}P}^{\dagger}\mathbf{C}\right)\right)dvec\left(\mathbf{P}\right)\\
&+ab\left(\left(\mathbf{CPG^{-1}}\right)^{T}\otimes\left(\mathbf{U}V\mathbf{G^{-2}}\right)\right)\mathbf{K}dvec\left(\mathbf{P}^{*}\right)\\
&+ab\left(\left(\mathbf{G^{-1}}\right)^{T}\otimes\left(\mathbf{U}V\mathbf{G^{-2}P}^{\dagger}\mathbf{C}\right)\right)dvec\left(\mathbf{P}\right)
\end{aligned} \tag{5.118}$$

Hence, via (5.111):

$$\begin{aligned}
\mathcal{D}_{\mathbf{P}}F=&-a\left(\left(\mathbf{G^{-1}}V\mathbf{W}\right)^{T}\otimes\left(\mathbf{G^{-1}P}^{\dagger}\mathbf{C}\right)\right)\\
&-a\left(\left(\mathbf{GW}\right)^{T}\otimes\left(\mathbf{U}V\mathbf{P}^{\dagger}\mathbf{C}\right)\right)\\
&-a\left(\mathbf{W^{T}}\otimes\left(\mathbf{UVGP}^{\dagger}\mathbf{C}\right)\right)\\
&+ab\left(\left(\mathbf{G^{-2}}\right)^{T}\otimes\left(\mathbf{U}V\mathbf{G^{-1}P}^{\dagger}\mathbf{C}\right)\right)\\
&+ab\left(\left(\mathbf{G^{-1}}\right)^{T}\otimes\left(\mathbf{U}V\mathbf{G^{-2}P}^{\dagger}\mathbf{C}\right)\right)\\
\mathcal{D}_{\mathbf{P}^{*}}F=&-a\left(\left(\mathbf{CPG^{-1}}V\mathbf{W}\right)^{T}\otimes\mathbf{G^{-1}}\right)\mathbf{K}\\
&-a\left(\left(\mathbf{CPGW}\right)^{T}\otimes\left(\mathbf{U}V\right)\right)\mathbf{K}\\
&-a\left(\left(\mathbf{CPW}\right)^{T}\otimes\left(\mathbf{U}V\mathbf{G}\right)\right)\mathbf{K}\\
&+ab\left(\mathbf{CPG^{-2}}\right)^{T}\otimes\left(\mathbf{U}V\mathbf{G^{-1}}\right)\mathbf{K}\\
&+ab\left(\left(\mathbf{CPG^{-1}}\right)^{T}\otimes\left(\mathbf{U}V\mathbf{G^{-2}}\right)\right)\mathbf{K}
\end{aligned} \tag{5.119}$$

Therefore, Theorem 13 has been proved.

Chapter 6
Optimal Power Allocation and Optimal Precoding in Multi-Cell Processing with Minimal Cooperation

6.1 Introduction

Cooperative MCP systems has the potential to significantly increase the maximum reliable achievable rates compared to non-cooperative ones [117]. A system where multiple collaborating base stations jointly serving multiple UTs is referred to as network MIMO system. The MAC channel stands as a special case of interfering channels. In particular, we can model an interference channel with distinct MAC channels [101]. In the downlink, interference between UTs is already handled at the transmitter side by means of precoding, where the user data is pre-equalized according to the instantaneous CSI. In [110], the authors exploit a new look into interference via cooperation, where a studied interference can be thought of as a positive factor when precoding is considered, [113]. In [118], a MCP with data and CSI sharing has been considered idealizing the setup. However, in practice, CSI is typically impaired by channel estimation errors, lossy compression for feedback transmission and feedback delays. In addition, in practice, cooperative systems incur latency and rate restrictions of backhaul links. In [109], the authors show that there are fundamental limits for cooperation, where clusters of limited size should be used. Therefore, it was of particular relevance to consider scenarios of practical relevance where processing overhead need to be minimal and backhaul congestion need to be taken into consideration. In this Chapter, we are interested in solving the problem of optimal precoding and power allocation in a MCP with minimal cooperation. In particular, when only CSI is shared and CSI imperfections due to estimation exist.

In this Chapter we will capitalize on our results in Chapter 4 and Chapter 5 to investigate the optimal power allocation and optimal precoding for a MCP framework with minimal cooperation between the base stations. In particular, we consider a cluster of two BSs which maximize the achievable rate for two users connecting to each BS and only sharing their CSI. We provide a generalized fixed point equation of the optimal precoder with respect to the estimated channel, power, and the MMSE. We provide the optimal power allocation with respect to the estimated channel and MMSE. The designs introduced are optimal for MAC Gaussian coherent time-varying fading channels with arbitrary inputs and can be specialized to MIMO channels by decoding interference. The impact of interference on the capacity is quantified by the gradient of the mutual information with respect to the power, channel, and error of the interferer. We provide two novel distributed MCP algorithms that provide the solutions for the optimal power allocation and optimal precoding for the UL and DL with a two way channel estimation to keep track over the channel variations over time in a block by block basis.

6.2 The System Model

Consider the scenario in Figure 6.1 where a cluster of two base stations implement minimal cooperation levels via sharing their CSI in order to insure higher data rates received. It follows that the base stations BS1 and BS2 will share their CSI of the links of user terminals UT1 and UT2 who are roaming under their coverage, respectively. In a cooperative framework with CSI and data sharing the two direct links and the two interference links with the BS cooperation can be modeled by a cooperative interference channel or a MIMO channel. However, here we use a divide and concur approach to have minimal cooperation by sharing only the CSI between the two base stations. Therefore, BS1 and BS2 will receive from UT1 and UT2 the following receive vectors respectively,

$$\mathbf{y_1} = \sqrt{snr}\ h_{11}\sqrt{P_1}\mathbf{x_1} + \sqrt{snr}\ h_{21}\sqrt{P_2}\mathbf{x_2} + \mathbf{n_1} \tag{6.1}$$

$$\mathbf{y_2} = \sqrt{snr}\ h_{12}\sqrt{P_1}\mathbf{x_1} + \sqrt{snr}\ h_{22}\sqrt{P_2}\mathbf{x_2} + \mathbf{n_2} \tag{6.2}$$

$\mathbf{y_1} \in \mathbb{C}^{\mathbf{n}}$ and $\mathbf{y_2} \in \mathbb{C}^{\mathbf{n}}$ represent the received vectors of complex symbols at BS1 and BS2 respectively, $\mathbf{x_1} \in \mathbb{C}^{\mathbf{n}}$ and $\mathbf{x_2} \in \mathbb{C}^{\mathbf{n}}$ represent the vectors of complex transmit symbols with zero mean $\mathbb{E}\left[\mathbf{x_1}\mathbf{x_1^T}\right] = \mathbb{E}\left[\mathbf{x_2}\mathbf{x_2^T}\right] = \mathbf{0}$, and identity covariance

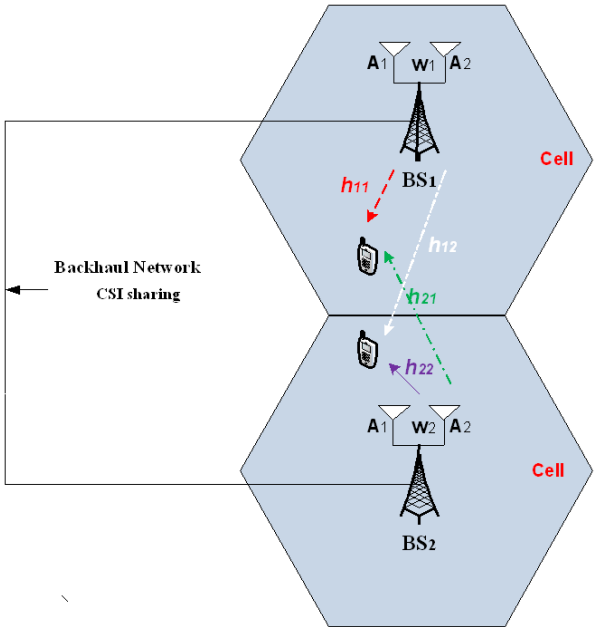

Figure 6.1: A cluster of two base stations with two distinct two-user MACs.

$\mathbb{E}\left[\mathbf{x_1}\mathbf{x_1}^{\dagger}\right]$ $=\mathbb{E}\left[\mathbf{x_2}\mathbf{x_2}^{\dagger}\right] = \mathbf{I}$, respectively, $\mathbf{n_1} \in \mathbb{C}^{\mathbf{n}}$, and $\mathbf{n_2} \in \mathbb{C}^{\mathbf{n}}$ represent vectors of circularly symmetric complex Gaussian random noise with zero mean and identity covariance, i.e., with $\mathcal{CN}(0, I)$. $\mathbf{h_{kl}}$ represent the complex gains of the Rayleigh fading distributed sub-channels between transmitter k and receiver l, where the main links are with $k = l$, and the interference links are with $k \neq l$.$\sqrt{P_1}$ and $\sqrt{P_2}$ represent the amplitudes of the transmitted signal from UT1 and UT2, respectively. And *snr* is the received signal to noise power ratio. Therefore, we can write (6.1) and (6.2) as: $\mathbf{y} = \sqrt{\mathbf{snr}}\,\mathbf{HPx} + \mathbf{n}$. For a backhaul with finite bandwidth, sharing CSI and data may cause processing overhead on the BSs; in turn, minimal cooperation is required [110]. Therefore, we can model the framework presented in Figure 6.1 by two MACs.

6.2.1 Problem Formulation

Instantaneous capacity is meaningful only when instantaneous CSI is available at the transmitter and the receiver, this takes place either when the channel remains fixed, or when it exhibits slow variation over time, so that it can be considered fixed during a number of transmissions. It is then possible to adapt the transmitter signal to each channel realization to achieve such instantaneous capacity. However,

obtaining CSI at the transmitter requires either a feedback channel or the application of the channel reciprocity when transmission and reception operate at the same carrier frequency (TDD) and the time variation of the channel is sufficiently slow. In many scenarios of practical relevance, the channel estimates at the transmitter may suffer imperfections mainly due to a fast time-varying nature of the channel, therefore, the transmitter owns a delayed version of the CSI. The natural extension of the capacity, given that there are no delay constraints in the communication system when the CSI is random, is called the ergodic capacity defined as the maximum mutual information averaged over all channel states, [41]. For the sake of problem formulation, we will first present the achievable rate regions of a two user MAC Gaussian fading channel [41], [101]. In fact, the upper bounds of the achievable rates for each two-user MAC Gaussian fading channel shown in Figure 6.1 are as follows:

$$R_1 \leq \mathbb{E}_{\widehat{\mathbf{H}}} I(\mathbf{x_1}; \mathbf{y_1} | \mathbf{x_2}) \tag{6.3}$$

$$R_2 \leq \mathbb{E}_{\widehat{\mathbf{H}}} I(\mathbf{x_2}; \mathbf{y_2} | \mathbf{x_1}) \tag{6.4}$$

$$R_1 + R_2 \leq min \;\; \mathbb{E}_{\widehat{\mathbf{H}}} \big[\mathbf{I}(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}), I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_2}) \big], \tag{6.5}$$

with,

$$I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}) = I(\mathbf{x_1}; \mathbf{y_1}) + I(\mathbf{x_2}; \mathbf{y_1} | \mathbf{x_1}), \tag{6.6}$$

and,

$$I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_2}) = I(\mathbf{x_2}; \mathbf{y_2}) + I(\mathbf{x_1}; \mathbf{y_2} | \mathbf{x_2}). \tag{6.7}$$

Its worth to note that the averaging over multiple channel estimates can be dropped when the CSI is considered known at both the transmitter and receiver, this way of thinking will later lead us to propose algorithms that can keep track of channel fast variation over time and so, the averaging term can be dropped. In turn, this leads to the optimal designs introduced in Chapter 4 and Chapter 5 for the case of known CSI. Therefore, we consider that BS1 will maximize the mutual information for a constrained capacity, achieved by coding over multiple fading blocks, (for MAC1) as follows:

$$max \;\; \frac{K - ML}{K} \mathbb{E}_{\widehat{\mathbf{H}}} I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1} | \widehat{\mathbf{H}}) \tag{6.8}$$

Subject to:

$$P_1 \leq Q1, \;\; P_2 \leq Q2, \;\; P_1 \geq 0 \; and \; P_2 \geq 0 \tag{6.9}$$

And BS2 will maximize the mutual information for a constrained capacity, achieved by coding over multiple fading blocks, (for MAC2) as follows:

$$max \ \ \frac{K-ML}{K}\mathbb{E}_{\widehat{\mathbf{H}}}I(\mathbf{x_1},\mathbf{x_2};\mathbf{y_2}|\widehat{\mathbf{H}}) \tag{6.10}$$

Subject to:

$$P_1 \leq Q1, \ P_2 \leq Q2, \ P_1 \geq 0 \ and \ P_2 \geq 0 \tag{6.11}$$

Where K is the number of symbols per fading block. M is the number of transmit antennas, L is the number of pilot symbols in the channel estimation process, P_1 and P_2 are the transmitted power corresponding to each UT, $Q1$ and $Q2$ is the total and maximum power each UT can use, respectively. We consider scalar channels, therefore, $\widehat{\mathbf{H}}$ is the estimated channel matrix at a certain time, at the receiver side, via pilot-assisted estimation, or at the transmitter side, via auto-regression, to provide timely designs for the UL and DL. However, considering that $M = 1$, and $K >> L = 1$, i.e., $(K - ML)/K \rightarrow 1$ for our analysis, the achievable rates are:

$$min\big[max \ \mathbb{E}_{\widehat{\mathbf{H}}}I(\mathbf{x_1},\mathbf{x_2};\mathbf{y_1}|\widehat{\mathbf{H}}), max \ \mathbb{E}_{\widehat{\mathbf{H}}}I(\mathbf{x_1},\mathbf{x_2};\mathbf{y_2}|\widehat{\mathbf{H}})\big]. \tag{6.12}$$

Therefore, the optimal power allocation set is the solution of the optimization problem that solves the minimum of the maximum of (6.8) or (6.10) subject to the two power constraints (6.9) and (6.11); see the Han-Kobayashi achievable rates for ICs [101].

6.2.2 Channel Estimation

Here we consider channel estimation is done at the receiver side with a pilot assisted mechanism. However, to solve the time mismatch in the design provided at the transmitter, we propose that the transmitter perform estimation of the future channel over the time varying block - if coherence is not guaranteed - via auto-regressive (AR) models. AR models provide a tool to estimate the dynamics of fading channels, the auto-regressive model can be of first order or higher, an auto-regressive moving-average model with order L, see [119], is expressed as:

$$\widehat{\mathbf{H}}(t) = -\sum_{i=1}^{L}\rho\mathbf{H}(\mathbf{t}-\mathbf{i}) + \omega(\mathbf{t}), \tag{6.13}$$

where ω is a zero mean unit variance white Gaussian process $\omega \sim \mathcal{CN}(0,1)$, and the AR correlation coefficient bounded as $0 \leq \ \rho \ \leq 1$ corresponds to slow fading at $\rho = 1$ and very fast fading at $\rho = 0$. In general, the AR coefficient is given by:

$$\rho_{k,l} = \frac{\mathbb{E}[h_{k,l}(t)h_{k,l}(t-i)]}{\sigma^2}, \tag{6.14}$$

$\mathbb{E}[h(t)h(t-i)]$ corresponds to the PDP of the channel indexed by k, l. Therefore, if $\sigma = 1$ then, ρ corresponds to the PDP of the channel. A special case of this model is the first-order auto-regressive (AR1) model, here we consider that each BS tracks the channel variation over time via a first-order AR model. The AR1 model represents the fading process in discrete time as:

$$\widehat{\mathbf{H}}(t) = -\rho\mathbf{H}(\mathbf{t}-\mathbf{1}) + \omega(\mathbf{t}) \tag{6.15}$$

Therefore, after receiving the pilots, each BS will estimate the time varying future channel of the main user and interferer and share the CSI information. It is worth to note that, due to CSI imperfections, the noise covariance will be the covariance of the effective noise in the detection given by: $\mathbb{E}_{\mathbf{n}|\widehat{\mathbf{H}}}[\mathbf{n}(\mathbf{t})\mathbf{n}(\mathbf{t})^\dagger] = \mathbf{I} + \mathbb{E}_{\omega|\widehat{\mathbf{H}}}[\omega(\mathbf{t})\omega(\mathbf{t})^\dagger] = \mathbf{I} + \mathbb{E}[\widehat{\mathbf{H}}(\mathbf{t})\widehat{\mathbf{H}}(\mathbf{t})^\dagger]$ for the Rayleigh fading case. Notice that we drop the time index in the rest of the Chapter.

6.3 Optimal Power Allocation with Minimal Cooperation

This section will present the characterization of optimal power allocation for the MCP framework with minimal cooperation modeled by two distinct MAC Gaussian fading channels and driven by Gaussian, arbitrary, and mixed inputs. However, since the noise is AWGN, it helps in the estimation process via auto-regression at the transmitter side, making it possible to estimate the probability distribution of the channel estimate conditioned on the prior-knowledge of the pilot assisted estimates at the receiver side. Hence, the probability of bit error rate can be also estimated and can be another design criterion to use in order to derive optimal designs, by minimizing the BER. However, the scope of this PhD thesis focuses on the maximization of the mutual information as the main design criterion to consider.

6.3.1 Gaussian Inputs

The mutual information for BS1 and BS2, respectively, is defined as:

$$I\left(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1} \middle| \widehat{\mathbf{H}}\right) = log\ \left(|\widehat{h}_{11}|^2 P_1 + |\widehat{h}_{12}|^2 P_2 + 1\right) \tag{6.16}$$

$$I\left(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_2} \middle| \widehat{\mathbf{H}}\right) = log\ \left(|\widehat{h}_{21}|^2 P_1 + |\widehat{h}_{22}|^2 P_2 + 1\right) \tag{6.17}$$

Therefore, solving the optimization problem for maximizing each MAC achievable rate subject to the users power constraints will lead to the optimal power allocation as it will be shown in the following theorem.

Theorem 14. *The optimal power allocation for two UTs in the MCP framework with minimal cooperation* $(P_1^\star, P_2^\star)$ *with Gaussian inputs follows the following form:*

$$\begin{cases} P_1^\star = Q1, \\ P_2^\star = Q2, \end{cases} \tag{6.18}$$

Proof. Theorem 14 follows the solution of the KKT conditions of (6.8) subject to (6.9) and (6.10) subject to (6.11), and due to the fact that the function is increasing with respect to the power. Thus, we notice that the solution of the derivative with respect to P_1 leads to $P_2^\star$, and the derivative with respect to P_2 leads to $P_1^\star$. □

Notice that when the inputs are subject to an average total power constraint, i.e., not per-user power constraints, the interpretation of the solution is a waterfilling, where the power distribution between the two users is such that: The optimal power allocation of the main channel m with respect to the interference i, is the water level $\frac{1}{\lambda}$, minus the noise $\frac{1}{|\widehat{h}_m|^2}$ and the interference levels, which can be written as follows:

$$\begin{cases} P_m^\star = \frac{1}{\lambda} - \frac{|\widehat{h}_i|^2}{|\widehat{h}_m|^2 P_i} - \frac{1}{|\widehat{h}_m|^2}, & \lambda \leq \frac{|\widehat{h}_m|^2}{|\widehat{h}_i|^2 P_i + 1} \\ P_m^\star = 0, & \lambda > \frac{|\widehat{h}_m|^2}{|\widehat{h}_i|^2 P_i + 1} \end{cases} \tag{6.19}$$

The solutions can be specialized to the single-user waterfilling interpretation noticing that we can substitute $P_1^\star$ for $P_2^\star$ in (6.19), and vice-versa, see [120]. Yet, the optimum solution set $(P_1^\star, P_2^\star)$ satisfies (6.12). Moreover, to tackle the optimal power allocation and optimal precoding problem with arbitrary inputs, we require the fundamental relation between the mutual information and the MMSE, see [30], [31], and [113]. The average system mmse(snr) is given by:

$$\overline{mmse}(snr) = \mathbb{E}_{\widehat{\mathbf{H}}}[mmse(snr, \widehat{\mathbf{H}})] = \mathbb{E}_{\widehat{\mathbf{H}}}\left[\mathbb{E}\left[\left\|\widehat{\mathbf{H}}\mathbf{P}(\mathbf{x} - \mathbb{E}[\mathbf{x}|\mathbf{y}])\right\|^2\right] |\widehat{H}\right] \tag{6.20}$$

Therefore, we can write the system MMSE matrix $\mathbf{E}$ using the elements of the gradient of the mutual information with respect to the main and interference links power. In particular, for our problem, when each UT uses SISO links, the MMSE Matrix is given by:

$$\mathbf{E} = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix}, \tag{6.21}$$

with E_{mm} is the error in each link, and E_{mi} is the covariance induced due to the interferer link, given by:

$$E_{11} = \mathbb{E}_{\widehat{\mathbf{H}}} \left[\mathbb{E}[(\mathbf{x_1} - \mathbb{E}(\mathbf{x_1}|\mathbf{y_1}, \widehat{\mathbf{H}}))(\mathbf{x_1} - \mathbb{E}(\mathbf{x_1}|\mathbf{y_1}, \widehat{\mathbf{H}}))^{\dagger}] \right] \tag{6.22}$$

$$E_{12} = \mathbb{E}_{\widehat{\mathbf{H}}} \left[\mathbb{E}[(\mathbf{x_1} - \mathbb{E}(\mathbf{x_1}|\mathbf{y_1}, \widehat{\mathbf{H}}))(\mathbf{x_2} - \mathbb{E}(\mathbf{x_2}|\mathbf{y_1}, \widehat{\mathbf{H}}))^{\dagger}] \right] \tag{6.23}$$

$$E_{21} = \mathbb{E}_{\widehat{\mathbf{H}}} \left[\mathbb{E}[(\mathbf{x_2} - \mathbb{E}(\mathbf{x_2}|\mathbf{y_2}, \widehat{\mathbf{H}}))(\mathbf{x_1} - \mathbb{E}(\mathbf{x_1}|\mathbf{y_2}, \widehat{\mathbf{H}}))^{\dagger}] \right] \tag{6.24}$$

$$E_{22} = \mathbb{E}_{\widehat{\mathbf{H}}} \left[\mathbb{E}[(\mathbf{x_2} - \mathbb{E}(\mathbf{x_2}|\mathbf{y_2}, \widehat{\mathbf{H}}))(\mathbf{x_2} - \mathbb{E}(\mathbf{x_2}|\mathbf{y_2}, \widehat{\mathbf{H}}))^{\dagger}] \right] \tag{6.25}$$

The input estimates are given by:

$$\mathbb{E}(\mathbf{x_k}|\mathbf{y_l}, \widehat{\mathbf{H}}) = \sum_{x_k} \frac{\mathbf{x_k} \mathbf{p}_{\mathbf{y_l}|\mathbf{x_k}, \widehat{\mathbf{H}}}(\mathbf{y_l}|\mathbf{x_k}, \widehat{\mathbf{H}}) \mathbf{p}(\mathbf{x_k})}{\mathbf{p}_{\mathbf{y_l}}(\mathbf{y_l})} \tag{6.26}$$

6.3.2 Arbitrary Inputs

For the Gaussian MAC time-varying fading channels with arbitrary inputs with timely channel estimation, we can derive the optimal power allocation for the generalized inputs capitalizing on the relation between the mutual information and the MMSE, [113], [31].

Theorem 15. *The relation between the gradient of the mutual information in* (6.8) *and* (6.10) *and with respect to the estimated channel, precoder, and the MMSE matrix for the two user Gaussian MAC channel is given by:*
For MAC1:

$$\nabla_{P_1} I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}|\widehat{\mathbf{H}}) = |\widehat{h}_{11}|^2 \sqrt{P_1} E_{11} + \widehat{h}_{11}^* \widehat{h}_{12} \sqrt{P_2} E_{12} \tag{6.27}$$

$$\nabla_{P_2} I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}|\widehat{\mathbf{H}}) = |\widehat{h}_{12}|^2 \sqrt{P_2} E_{12} + \widehat{h}_{12}^* \widehat{h}_{11} \sqrt{P_1} E_{11} \tag{6.28}$$

For MAC2:

$$\nabla_{P_1} I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_2}|\widehat{\mathbf{H}}) = |\widehat{h}_{21}|^2\sqrt{P_1}E_{12} + \widehat{h}_{21}^*\widehat{h}_{22}\sqrt{P_2}E_{22} \tag{6.29}$$

$$\nabla_{P_2} I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_2}|\widehat{\mathbf{H}}) = |\widehat{h}_{22}|^2\sqrt{P_2}E_{22} + \widehat{h}_{22}^*\widehat{h}_{21}\sqrt{P_1}E_{21} \tag{6.30}$$

Proof. Theorem 15 is a direct consequence to the derivation of the gradient of the mutual information with respect to the precoding matrix, see [113], and [31]. □

Theorem 15 shows how much rate loss induced due to interference links, this is due to the fact that some terms in the gradient of the mutual information preclude the effect of the mutual interference of the main links. Therefore, this way, we can quantify how much rate is lost due to interference which can be accounted for via optimal power allocation, which will be presented in the following theorem.

Theorem 16. *The optimal power allocation for the two user Gaussian MAC time-varying fading channel with arbitrary inputs - in terms of estimated channels, the interfer power, and the MMSE matrix - takes the following form:*
For MAC1:

$$\lambda_1^\star\sqrt{P_1^\star} = |\widehat{h}_{11}|^2\sqrt{P_1^\star}E_{11} + \widehat{h}_{11}^*\widehat{h}_{12}\sqrt{P_2^\star}E_{12} \tag{6.31}$$

$$\lambda_2^\star\sqrt{P_2^\star} = |\widehat{h}_{12}|^2\sqrt{P_2^\star}E_{12} + \widehat{h}_{12}^*\widehat{h}_{11}\sqrt{P_1^\star}E_{11} \tag{6.32}$$

For MAC2:

$$\lambda_1^\star\sqrt{P_1^\star} = |\widehat{h}_{21}|^2\sqrt{P_1^\star}E_{21} + \widehat{h}_{21}^*\widehat{h}_{22}\sqrt{P_2^\star}E_{22} \tag{6.33}$$

$$\lambda_2^\star\sqrt{P_2^\star} = |\widehat{h}_{22}|^2\sqrt{P_2^\star}E_{22} + \widehat{h}_{22}^*\widehat{h}_{21}\sqrt{P_1^\star}E_{21} \tag{6.34}$$

$(P_1^\star, P_2^\star)$ *is the solution set that satisfies* (6.12)*, and* λ_1*,* λ_2 *are Lagrange multipliers normalized by the received snr. Therefore, the numerical solution satisfies the one either for MAC1* (6.31) *and* (6.32)*, or MAC2* (6.33) *and* (6.34)*.*

Proof. Proof follows the KKT conditions solving (6.8) subject to (6.9) and (6.10) subject to (6.11). □

Its straightforward to see that, $P_2^\star = Q2$, when $P_1 = 0$, and $P_1^\star = Q1$ when $P_2 = 0$. In addition, we can specialize the result of Theorem 16 to the one for Gaussian inputs. In particular, we substitute the elements of the linear MMSE for Gaussian

inputs into (6.31) to (6.34), then it can be easily shown that the optimal power allocation in Theorem 16 matches the one in Theorem 14. Theorem 16 assimilates a mercury/waterfilling for the arbitrary inputs - which compensates for the non-Gaussianess of the binary constellations - and a waterfilling for the Gaussian inputs. Moreover, we can re-write Theorem 15 with respect to the MMSE and the covariance, for the main and interference of each MAC as follows:

$$P_m^\star = \frac{1}{snr|\widehat{h}_m|^2}\, mmse_m(snr|\widehat{h}_m|^2 P_m^\star) + \frac{1}{snr\widehat{h}_m^*\widehat{h}_i}\, cov(snr|\widehat{h}_m^*\widehat{h}_i|P_i^\star) \tag{6.35}$$

$$P_i^\star = \frac{1}{snr|\widehat{h}_i^*\widehat{h}_m|}\, mmse_m(snr|\widehat{h}_i^*\widehat{h}_m|P_m^\star) + \frac{1}{snr|\widehat{h}_i|^2}\, cov(snr|\widehat{h}_i|^2 P_i^\star), \tag{6.36}$$

6.3.3 Mixed Inputs

In the case where one input is Gaussian and the other is BPSK, for instance, this assumption is of practical relevance when multiple technologies are applied to different neighboring cells, or when an unfriendly jammer is trying to inject interference in the form of Gaussian or arbitrary noise, see [121]. The implementation of such scenario can be tackled via the chain rule of the mutual information as follows:

$$I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}|\widehat{\mathbf{H}}) = I(\mathbf{x_1}; \mathbf{y_1}|\widehat{\mathbf{H}}) + I(\mathbf{x_2}; \mathbf{y_1}|\mathbf{x_1}, \widehat{\mathbf{H}}) \tag{6.37}$$

If $\mathbf{x_1}$ corresponds to the BPSK input, and $\mathbf{x_2}$ corresponds to the Gaussian input, then we can find the achievable rate of BS1; assuming that BS1 satisfies (6.12), then we can re-write (6.37) as follows:

$$I(\mathbf{x_1}, \mathbf{x_2}; \mathbf{y_1}|\widehat{\mathbf{H}}) = I(\mathbf{x_1}; \mathbf{y_1}|\widehat{\mathbf{H}}) + log(|\widehat{h}_{12}|^2 P_2 + 1), \tag{6.38}$$

with,

$$I(\mathbf{x_1}; \mathbf{y_1}|\widehat{\mathbf{H}}) = \mathbb{E}\left[\frac{\mathbf{p}_{\mathbf{y_1}|\mathbf{x_1},\widehat{\mathbf{H}}}(\mathbf{y_1}|\mathbf{x_1}, \widehat{\mathbf{H}})}{p_{\mathbf{y_1}}(\mathbf{y_1})}\right] \tag{6.39}$$

$$p_{\mathbf{y_1}|\mathbf{x_1},\widehat{\mathbf{H}}}(\mathbf{y_1}|\mathbf{x_1}, \widehat{\mathbf{H}}) = \frac{1}{|\widehat{h}_{12}|^2 P_2 + 1} e^{-\frac{\|\mathbf{y_1} - \widehat{h}_{11} P_1 \mathbf{x_1}\|^2}{2(|\widehat{h}_{12}|^2 P_2 + 1)}} \tag{6.40}$$

$$p_{\mathbf{y_1}}(\mathbf{y_1}) = \sum_{\mathbf{x_1}} p_{\mathbf{y_1}|\mathbf{x_1},\widehat{\mathbf{H}}}(\mathbf{y_1}|\mathbf{x_1}, \widehat{\mathbf{H}}) p_{\mathbf{x_1}}(\mathbf{x_1}) \tag{6.41}$$

We can easily verify that the optimal power allocation for the mixed inputs case satisfies the one for Gaussian inputs. That is, the optimal power allocation is to select the maximum power available.

6.4 Optimal Precoding with Minimal Cooperation

Consider the MCP with minimal cooperation in the DL where both BSs jointly cooperate sharing their CSI estimated versions to design the optimal precoding vectors that maximize their achievable rates. The following theorem gives a generalized optimal precoder for the MCP with minimal cooperation. In particular, this theorem provides an optimal precoding set for BS1 and BS2.

Theorem 17. *The non-unique optimal precoding set that maximizes the achievable rates for the MCP framework with minimal cooperation subject to per user power constraint is the numerical solution of the following form:*
For MAC1:

$$\mathbf{P}_1{}^{\star} = \nu_1^{-1}\widehat{\mathbf{H}}_{11}^{\dagger}\widehat{\mathbf{H}}_{11}\mathbf{P_1}^{\star}\mathbf{E_{11}} + \nu_1^{-1}\widehat{\mathbf{H}}_{11}^{\dagger}\widehat{\mathbf{H}}_{12}\mathbf{P_2}^{\star}\mathbf{E_{12}} \tag{6.42}$$

$$\mathbf{P}_2{}^{\star} = \nu_1^{-1}\widehat{\mathbf{H}}_{12}^{\dagger}\widehat{\mathbf{H}}_{12}\mathbf{P_2}^{\star}\mathbf{E_{12}} + \nu_1^{-1}\widehat{\mathbf{H}}_{12}^{\dagger}\widehat{\mathbf{H}}_{11}\mathbf{P_1}^{\star}\mathbf{E_{11}} \tag{6.43}$$

For MAC2:

$$\mathbf{P}_1{}^{\star} = \nu_1^{-1}\widehat{\mathbf{H}}_{21}^{\dagger}\widehat{\mathbf{H}}_{21}\mathbf{P_1}^{\star}\mathbf{E_{21}} + \nu_1^{-1}\widehat{\mathbf{H}}_{21}^{\dagger}\widehat{\mathbf{H}}_{22}\mathbf{P_2}^{\star}\mathbf{E_{22}} \tag{6.44}$$

$$\mathbf{P}_2{}^{\star} = \nu_2^{-1}\widehat{\mathbf{H}}_{22}^{\dagger}\widehat{\mathbf{H}}_{22}\mathbf{P_2}^{\star}\mathbf{E_{22}} + \nu_2^{-1}\widehat{\mathbf{H}}_{22}^{\dagger}\widehat{\mathbf{H}}_{21}\mathbf{P_1}^{\star}\mathbf{E_{21}} \tag{6.45}$$

$(\mathbf{P}_2{}^{\star}, \mathbf{P}_2{}^{\star})$ *is the solution set that satisfies* (6.12)*, and* ν_1 *and* ν_2 *are the per MAC received snr normalized by the Lagrange multipliers.*

Proof. Theorem 17 follows from the relation between the gradient of the mutual information and the MMSE and the decomposition of its matrix components. The optimal precoding set is $(\mathbf{P}_1{}^{\star}, \mathbf{P}_2{}^{\star})$ that satisfies (6.12). The solution can be found in an iterative approach similar to the distributed MCP algorithms in [118]. □

In this minimal cooperation scenario, it is worth to consider that each BS try to decode the interference rather than sharing the global data with the other BS. Therefore, for

the case of the two-user disjoint MAC channels, we first consider the optimal received MMSE filters[3] are as follows:

$$\mathbf{R_{11}} = \mathbf{P_1^\dagger \widehat{H}_{11}^\dagger}(\mathbf{I} + \mathbf{P_1^\dagger \widehat{H}_{11}^\dagger P_1 \widehat{H}_{11}} + \mathbf{P_2^\dagger \widehat{H}_{12}^\dagger P_2 \widehat{H}_{12}})^{-1} = \mathbf{P_1^\dagger \widehat{H}_{11}^\dagger E_{11}} \tag{6.46}$$

$$\mathbf{R_{12}} = \mathbf{P_1^\dagger \widehat{H}_{12}^\dagger}(\mathbf{I} + \mathbf{P_2^\dagger \widehat{H}_{11}^\dagger P_2 \widehat{H}_{11}} + \mathbf{P_1^\dagger \widehat{H}_{12}^\dagger P_1 \widehat{H}_{12}})^{-1} = \mathbf{P_1^\dagger \widehat{H}_{12}^\dagger E_{11}} \tag{6.47}$$

$$\mathbf{R_{21}} = \mathbf{P_2^\dagger \widehat{H}_{21}^\dagger}(\mathbf{I} + \mathbf{P_1^\dagger \widehat{H}_{22}^\dagger P_1 \widehat{H}_{22}} + \mathbf{P_2^\dagger \widehat{H}_{21}^\dagger P_2 \widehat{H}_{21}})^{-1} = \mathbf{P_2^\dagger \widehat{H}_{21}^\dagger E_{22}} \tag{6.48}$$

$$\mathbf{R_{22}} = \mathbf{P_2^\dagger \widehat{H}_{22}^\dagger}(\mathbf{I} + \mathbf{P_2^\dagger \widehat{H}_{22}^\dagger P_2 \widehat{H}_{22}} + \mathbf{P_1^\dagger \widehat{H}_{21}^\dagger P_1 \widehat{H}_{21}})^{-1} = \mathbf{P_2^\dagger \widehat{H}_{22}^\dagger E_{22}} \tag{6.49}$$

Thus, each input can be linearly estimated at the processing BS as $\widehat{\mathbf{x}}_\mathbf{1} = \mathbf{R_{11}y_1}$, $\widehat{\mathbf{x}}_\mathbf{2} = \mathbf{R_{21}y_1}$ at BS1 and $\widehat{\mathbf{x}}_\mathbf{1} = \mathbf{R_{12}y_2}$, and $\widehat{\mathbf{x}}_\mathbf{2} = \mathbf{R_{22}y_2}$ at BS2. Moreover, we can preserve the optimal properties of the iterative solution in Theorem 17 and through decoding interference using MMSE filters, we can preclude the effect of the covariance terms. Therefore, the optimal precoder takes the fixed point equation optimal solution of MIMO channels as: $\mathbf{P}^\star = \nu^{-1}\widehat{\mathbf{H}^\dagger}\widehat{\mathbf{H}}\mathbf{P}^\star\mathbf{E}$, where, $\mathbf{P}^\star = \mathbf{U}_{\widehat{\mathbf{H}}}\mathbf{D\Pi U_E}$, such that, $\mathbf{U}_{\widehat{\mathbf{H}}}$, $\mathbf{U_E}$ are unitary matrices due to the decomposition of the estimated channel matrix and the MMSE matrix, respectively, $\mathbf{D}$ is a power allocation diagonal matrix, and $\mathbf{\Pi}$ is a permutation matrix. This rotation structure insures firstly, allocation of power into the strongest channel singular vectors, and secondly, diagonalizes the MMSE matrix to insure un-correlating the error or in other words independence between inputs. It is worth to notice that the second covariance term in (6.42) to (6.45) quantifies the power cost we need to pay due to interference. However, the second term in equations (6.27) to (6.30) in the gradient quantifies the losses in terms of information rates due to interference.

6.5 MCP with Minimal Cooperation Distributed Algorithms

We introduce the MCP with minimal cooperation distributed algorithms, the first algorithm gives the optimal power allocation for the UL, and the second algorithm gives the optimal precoding for the DL.

[3]Notice that we abuse the notation, since we provide the general solution in Theorem 17 and the MMSE filters for user inputs with full channel matrix, while the model presents scalar channels.

Algorithm 3:
Optimum Power Allocation with MCP-UL
Minimal cooperation: CSI sharing

BS1 Input: CSI1: $\widehat{h}_{11}(t=1),...,\widehat{h}_{11}(t=L)$**,** $\widehat{h}_{21}(t=1),...,\widehat{h}_{21}(t=L)$**,** $\mathbb{E}[\mathbf{x_1}|\mathbf{y_1}]$

BS2 Input: CSI2: $\widehat{h}_{22}(t=1),...,\widehat{h}_{22}(t=L)$**,** $\widehat{h}_{12}(t=1),...,\widehat{h}_{12}(t=L)$**,** $\mathbb{E}[\mathbf{x_2}|\mathbf{y_2}]$

if BW Backhaul $\geq Threshold\ \tau$ **then**
  BS1 and BS2 declare congestion and no cooperation
else
  BS1 sends $CSI1$ **to BS2, BS2 sends** $CSI2$ **to BS1**
Output: BS1 and BS2 find the optimum power allocation $(P_1^\star, P_2^\star)$ **in the UL as the solution for:**

$$P_{mk+1} = \alpha_k P_{mk} + \alpha_k \lambda \widehat{h}_m^\dagger \widehat{h}_m P_{mk} E_{m_k} + \alpha_k \lambda \widehat{h}_m^\dagger \widehat{h}_i P_{ik} E_{i_k}$$

BS1 and BS2 check resources $\rightarrow$ **handshaking** $\rightarrow$ **BS1 and BS2 jointly decide**

$$(P_1^\star, P_2^\star) \textbf{ that satisfies } (6.12)$$

BS1 and BS2 feedback $P_1{}^\star$ **and** $P_2{}^\star$ **to UT1 and UT2, respectively.**

6.6 Numerical Analysis

We shall now introduce a set of illustrative results that cast insight into the problem. We analyze the results for the non-cooperative MAC for three different types of inputs. The results for the Gaussian inputs setup is straightforward with the mutual information closed form. However, we used Monte-Carlo method to generate the achievable rates for arbitrary inputs. Figure 6.2, Figure 6.3, and Figure 6.4 show the achievable rate for one BS in the MAC setup - supposing that this achievable rate is the one that solves the minimum of the maximum of the achievable rates of both BSs - using Gaussian inputs, BPSK inputs, and mixed inputs, respectively. In the case where both inputs are arbitrary, for example; BPSK, the achievable rate faces decay at equal input powers when both user channels are equal and real; for instance, are unity, this is due to the fact when both user inputs stay in the null space of the channel. One way to overcome this is by orthogonalizing the inputs in the UL (and/or precoding in the DL). As we have shown in Chapter 4 and Chapter 5, for Gaussian

Algorithm 4:
Optimum Precoding with MCP-DL
Minimal cooperation: CSI sharing

BS1 Input: $CSI1$**,** $\mathbf{x_1}$

BS2 Input: $CSI2$**,** $\mathbf{x_2}$

AR function: BS1 and BS2 estimates channel variation for the block length K **using auto-regression of order** L**,** $AR_L(ML, ..., K - ML)$

SVD function: BS1 and BS2 performs SVD($\widehat{\mathbf{H}}$): $\widehat{\mathbf{H}} = \mathbf{U}_{\widehat{\mathbf{H}}}\mathbf{\Lambda}_{\widehat{\mathbf{H}}}\mathbf{V}^{\dagger}_{\widehat{\mathbf{H}}}$

BS1 sends $CSI1$ **to BS2 and BS2 sends** $CSI2$ **to BS1**

Output: BS1 and BS2 find the optimum precoding $(P_1^{\star}, P_2^{\star})$ **in the DL as the solution for:**

$$\mathbf{P_{mk}} = \mathbf{V}_{\widehat{\mathbf{H}}_\mathbf{m}}\sqrt{P_m}$$
$$\mathbf{P}_{\mathbf{m}k+1} = \alpha_k\mathbf{P_{mk}} + \alpha_\mathbf{k}\lambda\widehat{\mathbf{H}}^{\dagger}_{\mathbf{m}}\widehat{\mathbf{H}}_\mathbf{m}\mathbf{P_{mk}}\mathbf{E}_{\mathbf{m_k}}$$

BS1 and BS2 select jointly the optimal set that satisfies (6.12)**,**

BS1 transmits : $(h_{11}\nu_{\widehat{h}_{11}}\sqrt{P_1} + h_{12}\nu_{\widehat{h}_{21}}\sqrt{P_1})\mathbf{x_1}$

BS2 transmits : $(h_{21}\nu_{\widehat{h}_{12}}\sqrt{P_2} + h_{22}\nu_{\widehat{h}_{22}}\sqrt{P_2})\mathbf{x_2}$
The process will be iteratively repeated for each simultaneous transmission of BS1 and BS2.

UT_1 **and** UT_2 **receives the block of** K **symbols, estimate** L **pilots and feedback to BS1 and BS2**

inputs, the optimal power allocation is to use the total power. While for arbitrary inputs, the numerical results introduce the fact that the optimal power allocation is such that the inputs don't stay in the null space of the channel, therefore, it provides an unbalanced power allocation. However, for the mixed input case, the optimal power allocation is similar to the Gaussian case, where users allocate their maximum power. It is worth to note also that the optimal power allocation and optimal precoding due to the two way channel tracking will be the optimal designs for channels with known CSI.
Another set of illustrative results for the average and instantaneous mutual information as well as the average and instantaneous MMSE are presented in Figure 6.5,

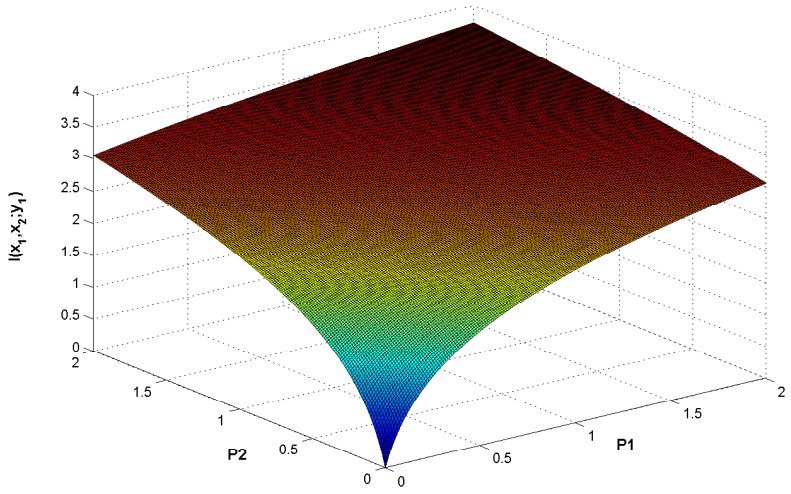

Figure 6.2: The achievable rate for MAC1 with respect to UTs main power with Gaussian inputs.

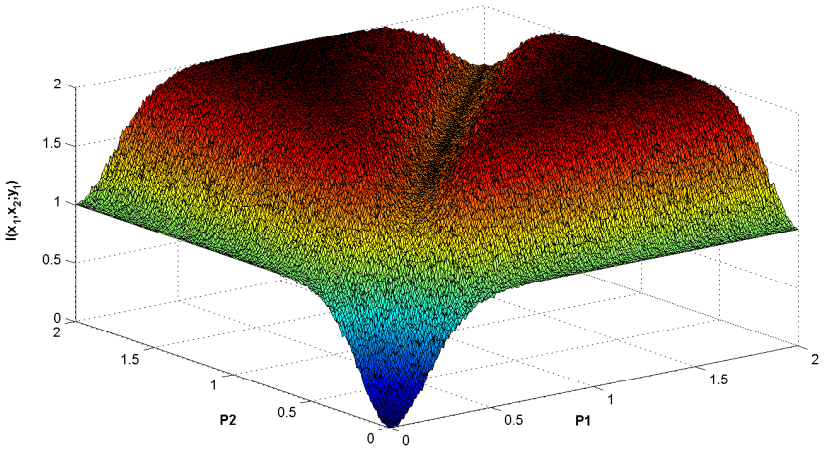

Figure 6.3: The achievable rate for MAC1 with respect to UTs main power with BPSK inputs.

and Figure 6.6. It is of particular relevance to observe: First, that diagonalizing the channel and so the MMSE matrix is not an optimal solution for binary constellations. Second, it is worth to note that the average behavior doesn't lead to optimal designs. Therefore, an instantaneous knowledge about the channel variation over time is of particular relevance. Third, through the instantaneous knowledge, it is possible, not only to design, but also to quantify the losses incurred in the data rates due to a sudden fade or interference. As we can see from Figure 6.5, there are gains instead

of losses in the mutual information due to averaging and the phase induced with the channel gain when the channel distribution is complex like in Rayleigh. Therefore, the average behavior doesn't allow a real vision of the way how to mitigate the interference or real losses in the channel.

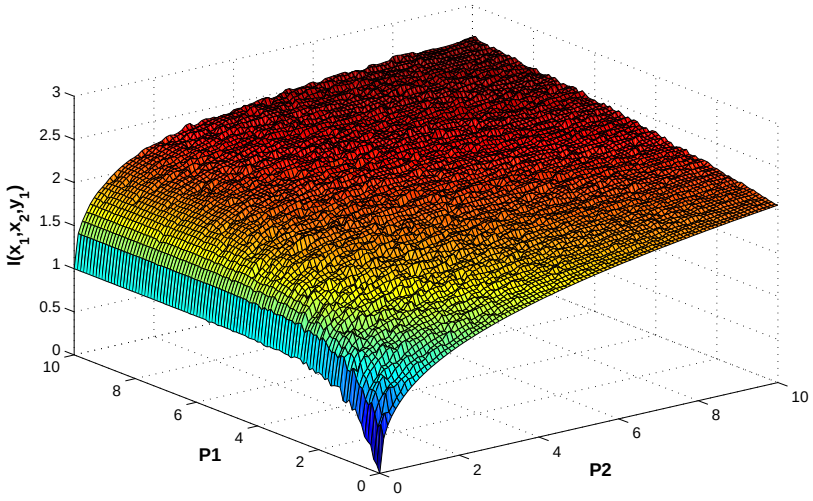

Figure 6.4: The achievable rate for MAC1 with respect to UTs main power with mixed inputs, x_1 BPSK, x_2 Gaussian.

Finally, the results in Figure 6.5 and Figure 6.6 illustrate also that a studied interference can be thought of as a positive factor instead of a capacity limiting factor, where a case of precoding with a predetermined coefficients via CSI and data sharing can increase the spectral efficiency of the network moving it into capacity bounds via virtual MIMOs.

On the other hand, this has been illustrated in the case where we can carefully say that the random fading coefficients with the interference case can play the role of precoding. In other words, long-term averages are great, but practically, wireless is unpredictable over short time-scales.

Therefore, by precoding at the transmitter side or exploiting the inherent mixing of signals offered by the wireless medium, further performance gains can be achieved.

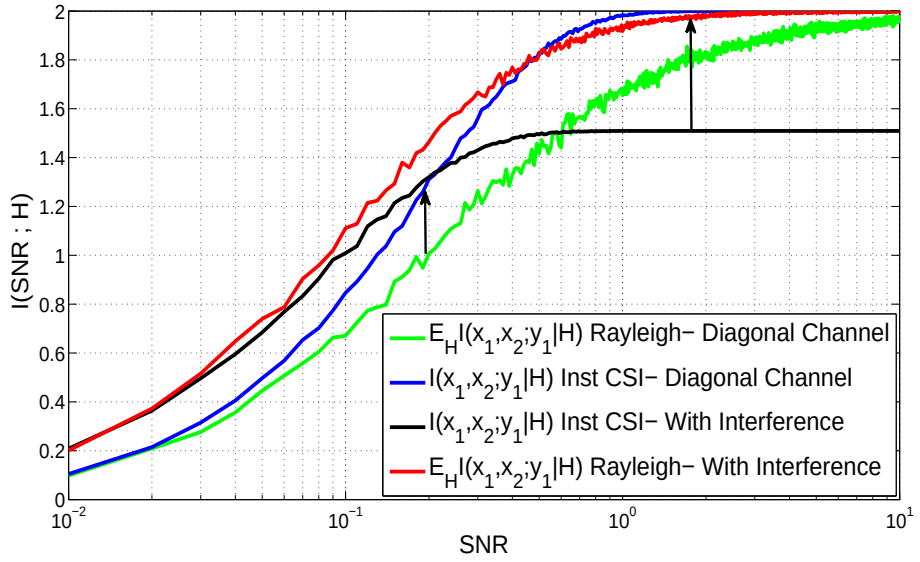

Figure 6.5: Average and instantaneous mutual information for BPSK inputs under Rayleigh fading and real (for diagonal and interference channels), with $Q1 = Q2 = 2$.

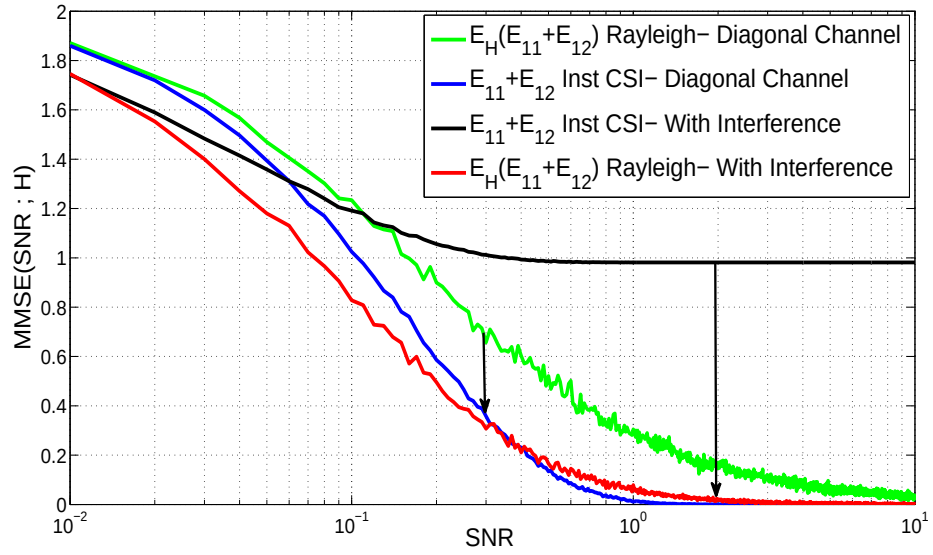

Figure 6.6: Average and instantaneous MMSE for BPSK inputs under Rayleigh fading and real (for diagonal and interference channels), with $Q1 = Q2 = 2$.

6.7 Conclusion

We have addressed the problem of optimal power allocation and optimal precoding in a minimal cooperation framework. We propose a two way channels estimation process that allows the transmitter to foresee the channel variation over the block length of transmission via a prior-knowledge of the channel distribution and a pilot-assisted channel estimation at the receiver side, aiming to minimize the imperfections of the CSI at the transmitter which cause untimely designs. Therefore, the solutions indeed break down into the ones with instantaneous knowledge of the channel. Consequently, we derive the optimal power allocation and the optimal precoding in the UL and DL. We provide a fixed point equation for the optimal precoder that follows the one for MIMO channels when interference is decoded. We highlighted a very important aspect in the precoding framework which shows that the wireless channel do a natural mixing or precoding over the transmitted data that can be invested if - from a reverse engineering perspective - added to the transmission process in a systematic way.

Chapter 7
The MAC Poisson Channel: Capacity and Optimal Power Allocation

7.1 Introduction

The majority of worldwide data and voice traffic is transported using optical communication channels. As the demand for bandwidth continues to increase, it is of great importance to find closed-form expressions of the information capacity for the optical communications applications at the backbone as well as the access networks. Information theory provides one of its strongest developments via the notion of maximum bit rate or channel capacity. Determining an ultimate limit to the rate at which we can reliably transmit information over a physical medium in a given environment is an earnest attempt of fundamental and practical consideration. Such a limit is referred to as the channel capacity and the process of evaluating this limit leads to an understanding of the technical solutions required to approach it. Therefore, if the capacity can be found, then the goal of the engineer is to design an architecture which achieves that capacity. Capacity evaluations require information theory which must be adapted to the specific characteristics of the channel under study. The seminal work of Shannon published in 1948 [122] gave birth to information theory. Shannon determined the capacity of memoryless channels, including channels impaired by AWGN for a given SNR. However, applying concepts of information theory to the optical communications channels encounters major challenges. The most important difficulty is dealing with the simultaneous interaction of specifically: The noise, filtering, and Kerr non-linearity phenomena in the optical channel. These phenomena are distributed along the propagation path, and influence each other leading to deterministic as well as stochastic impairments [123]. Several contributions have been done using information-theoretic approaches to derive the capacity of Poisson channels

under constant and time varying noise via martingale [4] processes [124], [125], or via approximations using Bernoulli processes [126], to define upper and lower bounds for the capacity and the rate regions of different models [127], [128], to define relations between information measures and estimation measures [61], in addition to deriving optimal power allocation for such channels [129], [125] [130]. In this Chapter, we introduce a simple framework for deriving the capacity of Poisson channels for the model of consideration - The MAC Poisson channel - with the assumption of constant stochastic martingale noise, i.e., for the sake of simplicity, we didn't model the noise as Gaussian within the stochastic intensity rate process. In addition, we build upon derivations for the optimal power allocation.

In Poisson channels, the shot noise is the dominant noise whenever the power received at the photodetector is high; such noise is modeled as a Poisson random process. In fact, such framework has been investigated in many researches, see [123], [124], [125], [127], [128], [61], [130], and [131]. Capitalizing on the expressions derived on [124], [132], [129], [125] and on the results by [129], [125], [127], we investigate the derivation process of the channel capacity in a straightforward way; we then determine the optimal power allocation that maximizes the information rates. To derive the optimal power allocation for different channel frameworks, it's worth to notice that different optimization criteria could be relevant. In particular, the optimization criteria could be the peak power, the average optical power, or the average electrical power. The average electrical power is the standard power measure in digital and wireless communications and it helps in assessing the power consumption in optical communications, while the average optical power is an important measure for safety considerations and helps in quantifying the impact of shot noise in wireless optical channels. In addition, the peak power, whether electrical or optical, gives a measure of tolerance against the non-linearities in the system, for example the Kerr non-linearity which is identified by a non-linear phase delay in the optical intensity or in other words as the change in the refractive index of the medium as a function of the electric field intensity.

In this Chapter, we introduce an information-theoretic derivation of the capacity expressions of Poisson channels that model the application. The closed-form expression for the capacity of the SISO Poisson channel - derived by Kabanov in 1978, and Davis in 1980 - will be revisited. Similarly, we will derive closed-form expressions

[4] In probability theory, a martingale is a stochastic process, i.e., a sequence of random variables, such that the conditional expected value of an obsevation at some time t over all observations up to some earlier time s, is equal to the observation at that time s. Therefore, $\lambda(s) = \mathbb{E}[\lambda(t)|N_s]$.

for the capacity of the MAC Poisson channel under the assumption of constant shot noise. This provides a framework for an empirical form of the k-users MAC Poisson channel capacity with average powers that are not necessarily equal. Moreover, we interestingly observed that the capacity of the MAC Poisson channel is a function of the SISO Poisson channel and upper bounded by this capacity plus some quadratic non-linear terms. We have also observed that the optimal power allocation in the case of Poisson channels follows a waterfilling alike interpretation to the one in Gaussian channels, where power is allotted to less noisy channels. Therefore, we establish a comparison between Gaussian channels and Poisson optical channels in the context of information theory and optical communications.

7.2 The Communication Framework

In a communication framework, the information source inputs a message to a transmitter. The transmitter couples the message onto a transmission channel in the form of a signal which matches the transfer properties of the channel. The channel is the medium that bridges the distance between the transmitter and the receiver. This can be either a guided transmission such as a wire or a wave guide, or it can be an unguided free space channel. A signal traverses the channel will suffer from attenuation and distortion. For example, electric power can be lost due to heat generation along a wire, and optical power can be attenuated due to scattering and absorption by air molecules in a free space. Therefore, channels are characterized by a transfer function which models the input-output process. The input-output process statistics is dominated by the noise characteristics the modulated input experiences during its propagation along the communication medium, in addition to the detection procedure experienced at the channel output. In particular, when the noise $n_G(t)$ is a zero-mean Gaussian process with double-sided power spectral density $N0/2$, the channel is called an AWGN. However, when the electrical input is modulated by a light source, like a laser diode, the channel will be an optical channel with the dominant shot noise $n_d(t)$ arising from the statistical nature of the production and collection of photoelectrons when an optical signal is incident on a photodetector, such statistics characterized by a Poisson random process. Figure 7.1 illustrates both the AWGN and the Poisson optical channels.

In this Chapter, we focus on the Poisson optical communication channel and we

derive capacity closed-form expression for the MAC Poisson channel capitalizing on the framework of derivation of the SISO Poisson channel capacity under a constant shot noise. Then, we build upon characterizations of the optimal power allocation. We have also shed light into the fundamental relation between the derivative of the mutual information with respect to the estimation errors as a counter part to the one for Gaussian channels.

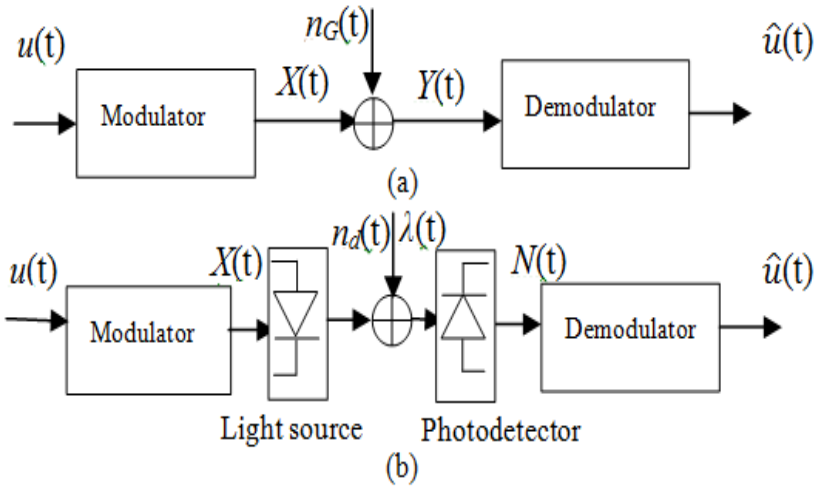

Figure 7.1: (a) The AWGN channel. (b) The Poisson optical channel.

7.3 The SISO Poisson Channel

Consider a SISO Poisson channel $\mathcal{P}$. Let $N(t)$ represent the channel output, which is the number of photoelectrons counted by a direct detection device (photodetector) in the time interval $[0, T]$. $N(t)$ has been shown to be a doubly stochastic Poisson process with instantaneous average rate $\lambda(t) + n$. The input $\lambda(t)$ is the rate at which photoelectrons are generated at time t in units of photons per second. And n is a constant representing the photodetector dark current and background noise.

7.3.1 Derivation of the Capacity of SISO Poisson Channels

Let $p(N_T)$ be the sample function density of the compound regular point process $N(t)$ and $p(N_T|S_T)$ be the conditional sample function of $N(t)$ given the message signal

process $S(t)$ in the time interval $[0, T]$. Then, we have:

$$p(N_T|S_T) = e^{-\int_0^T(\lambda(t)+n)dt+\int_0^T log(\lambda(t)+n)dN(t)}, \tag{7.1}$$

$$p(N_T) = e^{-\int_0^T(\widehat{\lambda(t)}+n)dt+\int_0^T log(\widehat{\lambda(t)}+n)dN(t)}, \tag{7.2}$$

where $\widehat{\lambda(t)}$ is the conditional mean estimate of the input $\lambda(t)$ given the output N_T, and $\mathbb{E}[.]$ is the expectation operation over time. Therefore, the mutual information is defined as follows:

$$I(S_T; N_T) = \mathbb{E}\left[log\left(\frac{p(N_T|S_T)}{p(N_T)}\right)\right] \tag{7.3}$$

Substitute (7.1) and (7.2) into (7.3), we have:

$$I(S_T; N_T) = \mathbb{E}\left[-\int_0^T(\lambda(t) - \widehat{\lambda(t)})dt + \int_0^T log\left(\frac{\lambda(t)+n}{\widehat{\lambda(t)}+n}\right)dN(t)\right] \tag{7.4}$$

Since $\mathbb{E}[\widehat{\lambda(t)}] = \mathbb{E}\left[\mathbb{E}[\lambda(t)|N_T]\right] = \mathbb{E}[\lambda(t)]$, it follows that:

$$I(S_T; N_T) = \mathbb{E}\left[\int_0^T log\left(\frac{\lambda(t)+n}{\widehat{\lambda(t)}+n}\right)dN(t)\right] \tag{7.5}$$

However, $N(t) - \int_0^T(\lambda(t)+n)dt$ is a martingale from theorems of stochastic integrals, see [129], [61]. Therefore,

$$I(S_T; N_T) = \mathbb{E}\left[\int_0^T(\lambda(t)+n)log\left(\frac{\lambda(t)+n}{\widehat{\lambda(t)}+n}\right)dt\right] \tag{7.6}$$

Therefore, we can break down (7.6) as follows:

$$I(S_T; N_T) = \int_0^T \mathbb{E}\left[(\lambda(t)+n)log\left(\lambda(t)+n\right)\right] - \mathbb{E}\left[(\lambda(t)+n)log\left(\widehat{\lambda(t)}+n\right)\right]dt \tag{7.7}$$

$$I(S_T; N_T) = \int_0^T \mathbb{E}\left[(\lambda(t)+n)log\left(\lambda(t)+n\right)\right] - \mathbb{E}\left[\mathbb{E}\left[(\lambda(t)+n)\right]log\left(\widehat{\lambda(t)}+n\right)|N_T\right]dt \tag{7.8}$$

$$I(S_T; N_T) = \int_0^T \mathbb{E}\left[(\lambda(t)+n)log\left(\lambda(t)+n\right)\right] - \mathbb{E}\left[\mathbb{E}\left[(\lambda(t)+n)|N_T\right]log\left(\widehat{\lambda(t)}+n\right)\right]dt \tag{7.9}$$

$$I(S_T; N_T) = \int_0^T \mathbb{E}\left[(\lambda(t)+n)log\left(\lambda(t)+n\right)\right] - \mathbb{E}\left[(\widehat{\lambda(t)}+n)log\left(\widehat{\lambda(t)}+n\right)\right]dt \tag{7.10}$$

It has been shown in [61] that the derivative of the input output mutual information of a Poisson channel with respect to the intensity of the dark current is equal to the expected error between the logarithm of the actual input and the logarithm of its conditional mean estimate, it follows that:

$$\frac{dI(S_T;N_T)}{d\ n(t)} = \mathbb{E}\left[log\left(\frac{\lambda(t)+n}{\widehat{\lambda(t)}+n}\right)\right] \tag{7.11}$$

The right hand side term of (7.11) is the derivative of the mutual information corresponding to the integration of the estimation errors. This plays as a counter part to the well known fundamental relation between the mutual information and the MMSE in Gaussian channels, [30]. The capacity of the SISO Poisson channel is defined as the maximum of (7.10) solving the following optimization problem:

$$max \quad I(S_T;N_T) \tag{7.12}$$

Subject to average and peak power constraints:

$$\frac{1}{T}\mathbb{E}\left[\int_0^T \lambda(t)dt\right] \leq \sigma P, 0 \leq \lambda(t) \leq P \tag{7.13}$$

Where P is the maximum power and the ratio of average to peak power σ is such that $0 \leq \sigma \leq 1$. We can easily check that the mutual information is strictly convex via its second derivative with respect to $\lambda(t)$ as follows:

$$\frac{d^2I(S_T;N_T)}{d\lambda(t)^2} = log\left(\frac{\lambda(t)+n}{\widehat{\lambda(t)}+n}\right) \succ 0 \tag{7.14}$$

Therefore, the mutual information is convex with respect to $\lambda(t)$. Solving the optimization problem in (7.12), subject to (7.13) as:

$$max \int_0^T \mathbb{E}\left[(\lambda(t)+n)log\left(\lambda(t)+n\right)\right] - \mathbb{E}\left[(\widehat{\lambda(t)}+n)log\left(\widehat{\lambda(t)}+n\right)\right]dt \ - \frac{\zeta}{T}\mathbb{E}[\lambda(t)], \tag{7.15}$$

where ζ is the Lagrange multiplier. The possible values of $\mathbb{E}\left[(\lambda(t)+n)log\left(\lambda(t)+n\right)\right]$ must lie in the set of all y-coordinates of the closed convex hull of the graph $y = (x+n)log(x+n)$. Hence, the maximum mutual information achieved using the distribution:

$$p(\lambda(t)=P) = 1 - p(\lambda(t)=0) = \alpha, \quad 0 \leq \alpha \leq 1 \tag{7.16}$$

So that the average power is $\mathbb{E}[\lambda(t)] = K$. Therefore, we have $\mathbb{E}[\lambda(t)] = \mathbb{E}[\sum \lambda(t) p\left(\lambda(t)\right)]$, and it follows that: $K = Pp(\lambda(t) = P) = P\alpha$. Thus, $\alpha = \frac{K}{P}$ and the capacity of the SISO Poisson channel is therefore as given in the following theorem.

Theorem 18. *The capacity of the SISO Poisson channel derived by (Kabanov 78 [124]−Davis 80 [132]) is given by:*

$$C = \frac{K}{P}\left(P+n\right) log\left(P+n\right) + \left(1-\frac{K}{P}\right) nlog\left(n\right) - \left(K+n\right) log\left(K+n\right) \tag{7.17}$$

Kabanov [124] and Davis [132] derived the capacity of the SISO Poisson channel as shown in Theorem 18. Therefore, we have first revisited their result and establish an information-theoretic estimation-theoretic approach to derive the closed-form expression for the capacity of the SISO Poisson channel. This builds the foundations for our derivation of the capacity of the $k-$user MAC Poisson channel in the forthcoming section, under the assumption of direct detection or photon counting receiver and constant shot noise.

7.3.2 Optimal Power Allocation for SISO Poisson Channels

The optimal power allocation for SISO Poisson channels determines the optimal average power need to be used in order to maximize the capacity, or in other words the utilization of the channel. Note that we perform a maximization over the capacity to have a decision on the optimal peak to average power threshold that can be encountered without losing the optimality of the solution. Therefore, we need to solve the following optimization problem:

$$max \;\; \frac{K}{P}\left(P+n\right) log\left(P+n\right) + \left(1-\frac{K}{P}\right) nlog\left(n\right) - \left(K+n\right) log\left(K+n\right) - \frac{\zeta}{T}K \tag{7.18}$$

Note that (7.18) is concave with respect to the average power K, i.e., the second derivative of (7.18) with respect to K is negative. Therefore, using the Lagrangian of the optimization problem; taking the derivative of the objective with respect to K and applying the KKT conditions, the optimal power allocation with respect to the peak power and the shot noise takes the following form:

$$K^{\star} = (P+n)e^{-(1+\frac{\zeta}{T})+\frac{n}{P}log(1+\frac{n}{P})} - n \tag{7.19}$$

Worth to note that the optimal power allocation satisfies the average and total power constraints defined previously. This indicates that its with probability distribution $\frac{K}{P}$, the arrival should have an average power of $K^\star$ to be restored. However, a peak to average power of unity which corresponds to $K = P$ leads to zero capacity. This fact will be in depth analyzed in the comparison between Poisson and Gaussian channels. Moreover, its straightforward to generalize the results of the capacity and power allocation for the SISO Poisson channel to the one with parallel independent Poisson channels like in OFDM, see [125].

7.4 The MAC Poisson Channel

Consider a two-user MAC Poisson channel $\mathcal{P}$. Let $N_1(t)$ represents the channel output, which is the number of photoelectrons counted by a direct detection device (photodetector) in the time interval $[0,\ T]$. $N_1(t)$ is a doubly stochastic Poisson process with instantaneous average rate $\lambda_1(t) + \lambda_2(t) + n$.

7.4.1 Derivation of the Capacity of MAC Poisson Channels

Let $p(N_1)$ be the sample function density of the compound regular point process $N_1(t)$ and $p(N_1|S_1,\ S_2)$ be the conditional sample function of $N_1(t)$ given the message signal processes $S_1(t)$ and $S_2(t)$ in the time interval $[0, T]$. Then, we have:

$$p(N_1|S_1,\ S_2) = e^{-\int_0^T(\lambda_1(t)+\lambda_2(t)+n)dt+\int_0^T log(\lambda_1(t)+\lambda_2(t)+n)dN_1(t)}, \tag{7.20}$$

$$p(N_1) = e^{-\int_0^T(\widehat{\lambda_1(t)}+\widehat{\lambda_2(t)}+n)dt+\int_0^T log(\widehat{\lambda_1(t)}+\widehat{\lambda_2(t)}+n)dN_1(t)}, \tag{7.21}$$

with $\widehat{\lambda_1(t)}$ and $\widehat{\lambda_2(t)}$ are the conditional mean estimates of the inputs $\lambda_1(t)$ and $\lambda_2(t)$, respectively, given the output N_1, and $\mathbb{E}[.]$ is the expectation operation over time. Therefore, the mutual information is defined as follows:

$$I(S_1, S_2; N_1) = \mathbb{E}\left[log\left(\frac{p(N_1|S_1,\ S_2)}{p(N_1)}\right)\right] \tag{7.22}$$

Substitute (7.20) and (7.21) into (7.22), we have:

$$I(S_1,S_2;N_1)=\mathbb{E}\left[-\int_0^T(\lambda_1(t)-\widehat{\lambda_1(t)})dt-\int_0^T(\lambda_2(t)-\widehat{\lambda_2(t)})dt\right] + \mathbb{E}\left[\int_0^T log\left(\frac{\lambda_1(t)+\lambda_2(t)+n}{\widehat{\lambda_1(t)}+\widehat{\lambda_2(t)}+n}\right)dN_1(t)\right] \quad (7.23)$$

Since: $\mathbb{E}[\widehat{\lambda_1(t)}+\widehat{\lambda_2(t)}]=\mathbb{E}\left[\mathbb{E}[\lambda_1(t)+\lambda_2(t)|N_1]\right]=\mathbb{E}[\lambda_1(t)+\lambda_2(t)]$, it follows that:

$$I(S_1,S_2;N_1)=\mathbb{E}\left[\int_0^T log\left(\frac{\lambda_1(t)+\lambda_2(t)+n}{\widehat{\lambda_1(t)}+\widehat{\lambda_2(t)}+n}\right)dN_1(t)\right] \quad (7.24)$$

However, $N_1(t)-\int_0^T(\lambda_1(t)+\lambda_2(t)+n)\,dt$ is a martingale from theorems of stochastic integrals, see [129], [61]. Therefore,

$$I(S_1,S_2;N_1) \quad = \quad \mathbb{E}\left[\int_0^T(\lambda_1(t)+\lambda_2(t)+n)log\left(\frac{\lambda_1(t)+\lambda_2(t)+n}{\widehat{\lambda_1(t)}+\widehat{\lambda_2(t)}+n}\right)dt\right] \quad (7.25)$$

Therefore, we can break down (7.25) as follows:

$$I(S_1,S_2;N_1)=\int_0^T\mathbb{E}\left[(\lambda_1(t)+\lambda_2(t)+n)log\left(\lambda_1(t)+\lambda_2(t)+n\right)\right] - \mathbb{E}\left[(\lambda_1(t)+\lambda_2(t)+n)log\left(\widehat{\lambda_1(t)}+\widehat{\lambda_2(t)}+n\right)\right]dt \quad (7.26)$$

$$I(S_1,S_2;N_1)=\int_0^T\mathbb{E}\left[(\lambda_1(t)+\lambda_2(t)+n)log\left(\lambda_1(t)+\lambda_2(t)+n\right)\right] - \mathbb{E}\left[\mathbb{E}\left[(\lambda_1(t)+\lambda_2(t)+n)\right]log\left(\widehat{\lambda_1(t)}+\widehat{\lambda_2(t)}+n\right)|N_1\right]dt \quad (7.27)$$

$$I(S_1,S_2;N_1)=\int_0^T\mathbb{E}\left[(\lambda_1(t)+\lambda_2(t)+n)log\left(\lambda_1(t)+\lambda_2(t)+n\right)\right] - \mathbb{E}\left[\mathbb{E}\left[(\lambda_1(t)+\lambda_2(t)+n)|N_1\right]log\left(\widehat{\lambda_1(t)}+\widehat{\lambda_2(t)}+n\right)\right]dt \quad (7.28)$$

$$I(S_1,S_2;N_1)=\int_0^T\mathbb{E}\left[(\lambda_1(t)+\lambda_2(t)+n)log\left(\lambda_1(t)+\lambda_2(t)+n\right)\right] - \mathbb{E}\left[(\widehat{\lambda_1(t)}+\widehat{\lambda_2(t)}+n)log\left(\widehat{\lambda_1(t)}+\widehat{\lambda_2(t)}+n\right)\right]dt \quad (7.29)$$

It has been shown in [61] that the derivative of the input-output mutual information of a Poisson channel with respect to the intensity of the dark current is equal to the expected error between the logarithm of the actual input and the logarithm of its conditional mean estimate, it follows that:

$$\frac{dI(S_1, S_2; N_1)}{d\ n(t)} = \mathbb{E}\left[log\left(\frac{\lambda_1(t) + \lambda_2(t) + n}{\widehat{\lambda_1(t)} + \widehat{\lambda_2(t)} + n}\right)\right], \tag{7.30}$$

The right hand side term of (7.30) is the derivative of the mutual information corresponding to the integration of the estimation errors. This plays as a counter part to the known fundamental relation between the mutual information and the MMSE with respect to the SNR in Gaussian channels, see [30]. The capacity of the MAC Poisson channel is defined as the maximum of (7.29) solving the following optimization problem:

$$max \quad I(S_1, S_2; N_1) \tag{7.31}$$

Subject to average and peak power constraints:

$$\frac{1}{T}\mathbb{E}\left[\int_0^T \lambda_1(t) + \lambda_2(t)dt\right] \leq \sigma P, 0 \leq \lambda_1(t) \leq P_1, 0 \leq \lambda_2(t) \leq P_2 \tag{7.32}$$

Where P_1 and P_2 are the maximum power and the ratio of average to peak power σ is such that $0 \ \leq \ \sigma \ \leq 1$. Solving the optimization problem in (7.31), subject to (7.32) as:

$$\begin{aligned} max \ & \int_0^T \mathbb{E}\left[(\lambda_1(t) + \lambda_2(t) + n)log\left(\lambda_1(t) + \lambda_2(t) + n\right)\right] \\ & - \mathbb{E}\left[(\widehat{\lambda_1(t)} + \widehat{\lambda_2(t)} + n)log\left(\widehat{\lambda_1(t)} + \widehat{\lambda_2(t)} + n\right)\right] dt \ - \frac{\zeta}{T}\mathbb{E}[\lambda_1(t) + \lambda_2(t)], \end{aligned} \tag{7.33}$$

where ζ is the Lagrange multiplier.
The possible values of $\mathbb{E}\left[(\lambda_1(t) + \lambda_2(t) + n)log\left(\lambda_1(t) + \lambda_2(t) + n\right)\right]$ must lie in the set of all y-coordinates of the closed convex hull of the graph $y = (x1 + x2 + n)log(x1 + x2 + n)$. Suppose that the maximum power for both inputs is $P_1 + P_2 = \sigma P$. Hence, the maximum mutual information achieved using the distribution:

$$p(\lambda_1(t)) = 1 - p(\lambda_2(t)) = \alpha, \quad 0 \ \leq \ \alpha \ \leq \ 1 \tag{7.34}$$

So that the average power per user are $\mathbb{E}[\lambda_1(t)] = K_1$ and $\mathbb{E}[\lambda_2(t)] = K_2$. Therefore, we have $\mathbb{E}[\lambda_1(t) + \lambda_2(t)] = \mathbb{E}[\sum\left(\lambda_1(t)p\left(\lambda_1(t)\right) + \lambda_2(t)p\left(\lambda_2(t)\right)\right)]$, and it follows that,

$K_1 = P_1 p(\lambda_1(t) = P_1) = P\alpha$ and $K_2 = P_2 p(\lambda_2(t) = P_2) = P(1-\alpha)$. Thus, $\alpha = \frac{K_1}{P}$ and $1 - \alpha = \frac{K_2}{P}$. Therefore, the capacity of the MAC Poisson channel is as given in the following theorem.

Theorem 19. *The capacity of the two-user MAC Poisson channel is given by:*

$$C = \left(\frac{K_1}{P} + \frac{K_2}{P}\right)(P+n)\,log\,(P+n) + \left(1 - \left(\frac{K_1}{P} + \frac{K_2}{P}\right)\right) nlog\,(n) - (K_1 + K_2 + n)\,log\,(K_1 + K_2 + n) \quad (7.35)$$

Notice that the capacity in Theorem 19 can be maximized when $\frac{K_1}{P} = \frac{K_2}{P}$. Moreover, it is worth to notice that there is another average power term $K_3 = P_1 p\,(0 \leq \lambda_1(t) \leq \sigma P) + P_2 p\,(0 \leq \lambda_2(t) \leq \sigma P) = P_1\alpha + P_2\,(1-\alpha)$. However, K_3 is not considered in the capacity equations since we only need the maximum and the minimum powers for both $\lambda_1(t)$ and $\lambda_2(t)$ to get the maximum expected value. Therefore, our framework of derivation differs from [127] by solving the problem geometrically.

7.4.2 Optimal Power Allocation for MAC Poisson Channels

The optimal power allocation for MAC Poisson channels determines the optimal average power need to be used in order to maximize the capacity, or in other words the utilization of the channel. Note that we perform a maximization over the capacity to have a decision on the optimal peak to average power threshold that can be encountered without losing the optimality of the solution. Therefore, we need to solve the following optimization problem:

$$max \quad \left(\frac{K_1}{P} + \frac{K_2}{P}\right)(P+n)\,log\,(P+n) + \left(1 - \left(\frac{K_1}{P} + \frac{K_2}{P}\right)\right) nlog\,(n) - (K_1 + K_2 + n)\,log\,(K_1 + K_2 + n) - \frac{\zeta}{T}(K_1 + K_2) \quad (7.36)$$

Using the Lagrangian of the optimization problem; taking the derivative of the objective with respect to $K = K_1 + K_2$ and applying the KKT conditions, the optimal power allocation with respect to the peak power and the shot noise takes the following form:

$$K_1^\star + K_2^\star = (P+n)e^{-(1+\frac{\zeta}{T})+\frac{n}{P}log(1+\frac{n}{P})} - n \quad (7.37)$$

The optimal power allocation solution introduces the fact that orthogonalizing the inputs via time or frequency sharing will achieve the capacity. Therefore, it follows the importance of interface solutions to aggregate different inputs to the Poisson channel.

7.5 MAC Poisson Channel Capacity and Rate Regions

We dedicate this section to analyze the result of Theorem 19. We will introduce the two-user MAC Poisson channel rate regions. We will then define the MAC capacity with respect to the SISO capacity and to bounds found mainly in [127]. The rate regions for the two-user MAC Poisson channel is given by:

$$R1 \leq I(S_1; N_1|S_2) \tag{7.38}$$

$$R2 \leq I(S_2; N_1|S_1) \tag{7.39}$$

$$R1 + R2 \leq I(S_1, S_2; N_1) \tag{7.40}$$

The mutual information that defines the sum of the rates $I(S_1, S_2; N_1)$ is defined in [Eq. 3.21, [127]] under the condition that the average inputs for the two users are equal; in particular when both inputs are equiprobable. Here, we can manipulate this result into a sum rate upper bound with the two users having different average input powers as follows:

$$\begin{aligned} I(S_1, S_2; N_1) = &\left(\frac{K_1}{P} + \frac{K_2}{P}\right)(P+n)\,log\,(P+n) + \left(1 - \left(\frac{K_1}{P} + \frac{K_2}{P}\right)\right) n log\,(n) \\ &- (K_1 + K_2 + n)\,log\,(K_1 + K_2 + n) - 2\left(\frac{K_1^2}{P^2} + \frac{K_2^2}{P^2}\right)(P+n)\,log\,(P+n) \\ &+ \left(\frac{K_1 K_2}{P^2}\right)(2P+n)\,log\,(2P+n) + \left(\frac{K_1 K_2}{P^2}\right) n log\,(n) \end{aligned} \tag{7.41}$$

Where $I(S_1, S_2; N_1)$ is maximized when $\frac{K_1}{P} = \frac{K_2}{P}$. It is important to notice two things in equation (7.41): Firstly, the first non-quadratic terms of $I(S_1, S_2; N_1)$ is the capacity of the SISO Poisson channel with the input as $\lambda_1(t) + \lambda_2(t)$. Therefore, we can see through Theorem 19 that the capacity is approximately defined by this first term plus a quadratic term as follows:

$$I(S_1, S_2; N_1) = C_{SISO}(\lambda_1(t) + \lambda_2(t)) + \beta, \tag{7.42}$$

with,

$$\beta = -2\left(\frac{K_1^2}{P^2}+\frac{K_2^2}{P^2}\right)(P+n)\,log\,(P+n) + \left(\frac{K_1K_2}{P^2}\right)(2P+n)\,log\,(2P+n) + \left(\frac{K_1K_2}{P^2}\right)nlog\,(n) \quad (7.43)$$

Therefore, we can deduce that the rate region - as defined in [127] - is an upper bound for the capacity. Thus, we can write an empirical form for the k-user MAC Poisson capacity, using the first non-quadratic terms of the above equation as follows:

$$C_{k-userMAC} = C_{SISO}(\lambda_1(t) + ... + \lambda_k(t)), \quad (7.44)$$

We can also verify Theorem 19 comparing it to the results in [127] for different setups, for example:
case 1: consider the case when $K_1 = K_2 = K$, the capacity will be:

$$C = \left(\frac{2K}{P}\right)(P+n)\,log\,(P+n) + \left(1-\left(\frac{2K}{P}\right)\right)nlog\,(n) - (2K+n)\,log\,(2K+n)\,. \quad (7.45)$$

case 2: consider the case when $K_1 = K_2 = K = P$, the negative terms indicates a zero capacity, therefore, $C = 0$.

case 3: consider the case when $K_1 = K_2 = K \neq P$, and the dark shot noise $n = 0$, the capacity will be:

$$C = 2Klog_e\left(\frac{P}{2K}\right), \quad (7.46)$$

and the sum rate will be:

$$I(S_1, S_2; N_1) = 2Klog_e\left(\frac{P}{2K}\right) + \frac{2K^2}{P}log_e\,(2)\,. \quad (7.47)$$

Therefore, $I(S_1, S_2; N_1)$ given in [Eq. 3.21, [127]] upper bounds the capacity by the term $\frac{2K^2}{P}log_e\,(2)$, and via the constraints over the average power, $\frac{2K^2}{P}log_e\,(2) \leq 2log_e\,(2)$, it follows that this upper bounds the capacity with a value always less than or equal to 1.4 nats/sec for the two-user MAC. In a more generalized way, the empirical form differs from the upper bound by less than or equal to $k\ log_e\,(k)$, where k corresponds to the number of inputs/users to the MAC Poisson channel. We can also verify that the maximum capacity achieved by orthogonalizing the inputs such that the capacity approaches $\frac{P}{e}$ nats/sec for each user. Therefore, non-orthogonalizing the inputs incurs a maximum of around $0.5256P$ power loss in the two-user MAC case. This well explains the limitation in the number of users for the MAC Poisson channel.

7.6 Simulation Results

We shall now present a set of illustrative results that cast insights into the problem. Figure 7.2 shows the capacity of different Poisson channels under a total power constraint of $P = 5$ on the SISO channel and each user's input of the parallel channel and the MAC channel, an equal average input power $K_1 = K_2 = K$, and shot noise $n = 0.1$. When the average input power is around one quarter the total power $K = P/4$, the rate is the maximum achievable rate, this explains the power loss in the two-user MAC case explained before. We can notice that the maximum mutual information presented by Lapidoth et al. in [Eq. 3.21, [127]] upper bounds the rate region of all given channels. However, we can see that the maximum achievable rate is always $C \leq P/e$ nats/sec. In particular, for the MAC channel, the maximum achievable rate with total power $P = 10$ is 3.425 nats/sec which is $C \leq 10/e \leq 3.7037$ nats/sec, i.e., the capacity for the $k-$user MAC is always $C \leq kP/e$. We can further see that in the low average power regime, both the upper bound and the empirical capacity of the MAC matches, while it logarithmically differs at the high average power regime, this is due to the quadratic part that is missed in the empirical capacity formula denoted by β. Notice also that the MAC channel capacity under the given conditions upper bounds the capacity of the parallel channel. In other words, the parallel channel capacity lower bounds the MAC channel capacity when both inputs are active.

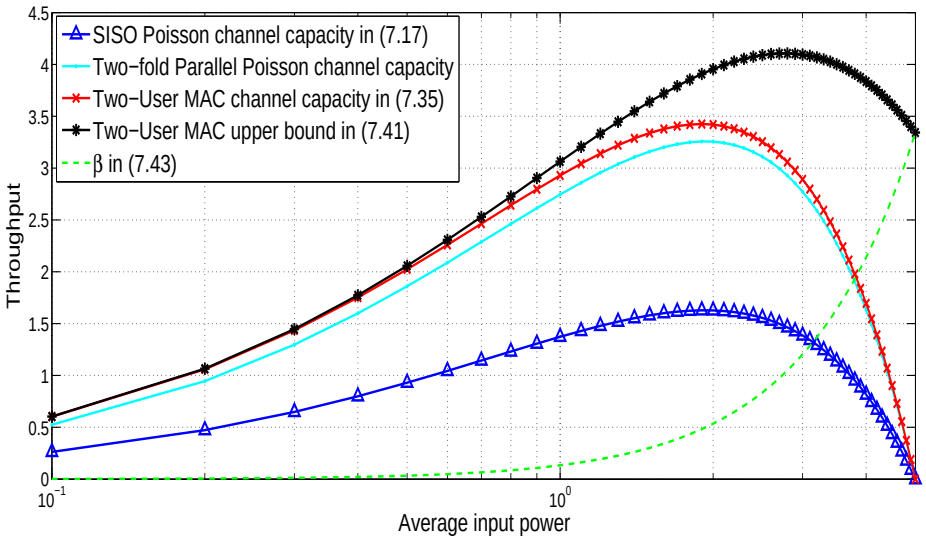

Figure 7.2: Capacity of the Poisson channels versus the average power.

Figure 7.3 shows the capacity of different Poisson channels with respect to the noise

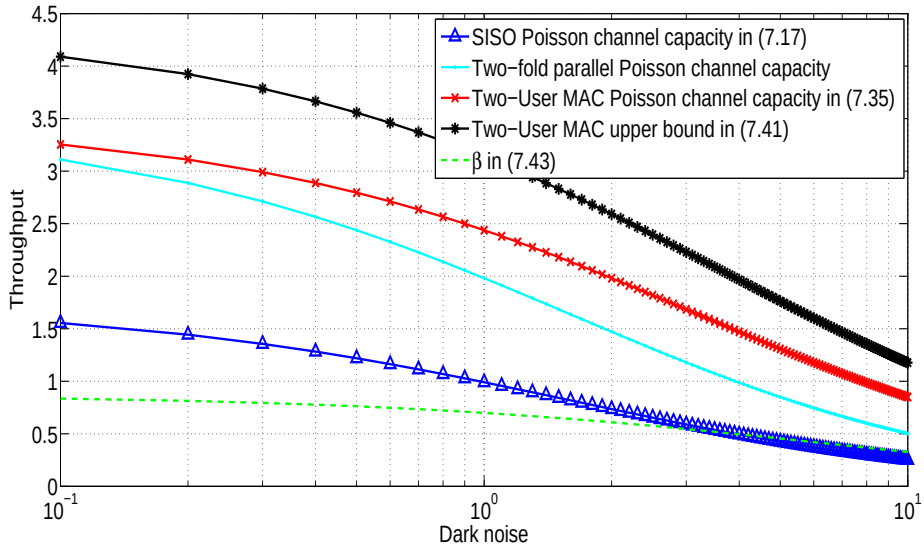

Figure 7.3: Capacity of the Poisson channels versus the shot noise.

where naturally the capacity decreases with respect to the increase in the shot noise. However, it is of particular relevance to notice that in the low noise power regime, Lapidoth upper bound for the MAC maximum achievable rate [127] indeed cannot be achieved due to existence of the quadratic terms. This gives rise of the achieved capacity over the right one, $C \leq kP/e$. However, our empirical form of the MAC capacity shows consistency regarding this relation and can be generalized to k-users. Finally, Figure 7.4 shows the optimal power allocation for the SISO, Parallel, and MAC Poisson channels. The total power used for the three channels is $P = 3$. It can be deduced, via the mathematical formulas as well as the simulations, that the power allocation is a decreasing value with respect to the dark current for all Poisson channels.

7.7 Analytical Results

The solutions provided in this Chapter show that the capacity of Poisson channels is a function of the average and peak power of the input. As a natural consequence to the expressions of the SISO Poisson channel, the Poisson parallel channels capacity is the sum of their independent SISO channels, proof is provided in [125]. For the MAC Poisson channel, the capacity expression derived here gives a generalization of a closed-

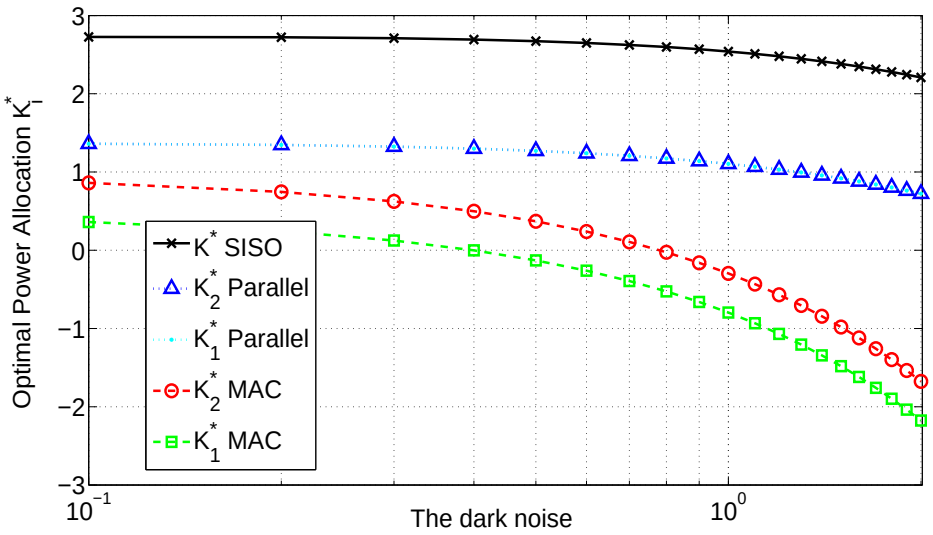

Figure 7.4: Optimal power allocation of the Poisson channels versus the shot noise.

form expression for the k-user MAC Poisson channel. The authors in [127] studied the capacity regions of the two-user MAC Poisson channels. They also pointed out an interesting observation that we can also emphasize and verify via Theorem 19; that is; in contrary to the Gaussian MAC, in the Poisson MAC the maximum throughput is bounded in the number of inputs, and similar to the Gaussian MAC in terms of achieving the capacity via orthogonalizing the inputs or via the usage of a limited average input power for each user that is equal to one quarter the total power in the two-user MAC case. In fact, for the Poisson MAC, when equal input powers up to half the total power for each are used, the capacity faces a decay to zero, while when they differ, i.e., inputs are orthogonal, the capacity is again maximized. In addition, we can also verify that the two main factors in the MAC capacity is the orthogonalization and the maximum power, while increasing the average power for one or the two inputs above a certain limit will not add positively to the capacity, see [125]. We can also see that the maximum power is a function of the average power through which both can be optimized to maximize the capacity. Moreover, it can be deduced via the mathematical formulas that the power allocation is a decreasing value with respect to the dark current for all Poisson channels. It means that the power allocation for the Poisson channels in some way or another follows a waterfilling alike interpretation to the one for the Gaussian setup where less power is allotted to the more noisy channels [125], [43]. However, it's well known that the optimal power allocation is an increasing function in terms of the maximum power.

7.7.1 Gaussian Channels versus Poisson Channels

Here, we summarize some important points about the capacity of Poisson channels in comparison to Gaussian channels within the context of this work. Firstly, in comparison to the Gaussian capacity, the channel capacity of the Poisson channel is maximized with binary inputs, i.e. $[0, 1]$, while the distribution that achieves the Gaussian capacity is a Gaussian input distribution. Secondly, the maximum achievable rates for the Poisson channel is a function of its maximum and average powers due to the nature of the Poisson process which follows a stochastic random process with martingale characteristics, while in Gaussian channels, the processes are random and modeled by the normal distribution. Thirdly, the optimal power allocation for the Poisson channels is very similar for different models depending on the defined power constraints, and in comparison to the Gaussian optimal power allocation; it follows a similar interpretation to the waterfilling, at which more power is allocated to stronger channels, i.e., power allocation is inversely proportional to the more noisy channel. However, although the optimal inputs distribution for the Poisson channel is a binary input distribution, the optimal power allocation is a waterfilling alike, i.e., unlike the Gaussian channels with arbitrary inputs where it follows a mercury-waterfilling interpretation to compensate for the non-Gaussianess in the binary input [44]. Finally, it is worth to emphasize two more important differences that were already shown in [128], which can be straight forward to proof here: Unlike the Gaussian channels, in Poisson channels, due to the characteristics of the Poisson distribution, we cannot implement interference cancellation techniques, since it is not possible to construct the probability of $p(N_1 = \lambda_1 + n)$ from the probability $p(N_1 = \lambda_1 + \lambda_2 + n)$, if λ_2 is considered as an interferer to λ_1. Besides, unlike Gaussian channels, Poisson channels are scale-invariant, since $p(N_1 = \lambda_1 + n/a) \neq p(N_1 = a\lambda_1 + n)$, if a scaling factor of $a \neq 1$ is multiplied to the inputs, the mutual information $I(S_1, S_2; N_1 = a\lambda_1 + a\lambda_2 + n) \neq I(S_1, S_2; N_1 = \lambda_1 + \lambda_2 + n/a)$.

7.8 Conclusion

In this Chapter, we revisited the derivation of the closed-form expression for the capacity of the SISO Poisson channel - derived by Kabanov in 1978, and Davis in 1980. We have derived a new closed-form expression for the capacity of MAC Poisson channels that provides a tighter bound than that derived by Lapidoth et al. [127].

In particular, we provide an empirical form of the k-users MAC Poisson channel capacity with average powers that are not necessarily equal, i.e., not equiprobable, and under the assumption of constant shot noise. The significance of such form is on the evaluation of practical setups like in optical communications where its relevant to have multiple users/channels aggregated over one fiber link. Moreover, we interestingly observed that the capacity of the MAC Poisson channel is a function of the SISO Poisson channel and upper bounded by this capacity plus some quadratic non-linear terms. This provides a tighter bound and hence more feasible form than the one derived by Lapidoth et al. in [127] for equal average user powers. We shed light on the existence of fundamental connections between information-theoretic and estimation-theoretic measures for Poisson channels of different types. It is shown − through the limitation on users within the capacity of the MAC Poisson channel − that the interface solutions for the aggregation of multiple users/channels over a single Poisson channel are of great importance. However, a technology like OFDM for optical communications stands as one interface solution. While it introduces attenuation via narrow filtering, etc. it therefore follows the importance of optimal power allocation which can mitigate such effects, hence, we build upon optimal power allocation derivations that aim to maximize the capacity. We establish a comparison between Gaussian channels and Poisson optical channels, where we have observed that the optimal power allocation in the case of Poisson channels follows a waterfilling alike interpretation to the one in Gaussian channels.

Chapter 8
Network Coding Mechanisms for Time Varying Channel

8.1 Introduction

Network coding can be thought of as one kind of precoding techniques through which coding occurs across the packets by selecting random linear coding coefficients from a GF. It can be thought of as a counter part to selecting precoding coefficients in a bit level or packet level. The first exploits the nature of the network architecture from the physical to the network layer. The later inherently exploits the space through its design. Therefore, to address the delay problem in future communications systems, network coding offers a solution which is more adaptive to practical requirements.

NC mechanisms have been adopted for the last few years since the pioneering work of Alswede et al. [49]. There have been several contributions since then that aim to study the performance gains of using network coding in different applications. The study of NC mechanisms in wireless networks need to take into consideration different aspects in the wireless medium like noise, interference, and fading. The diversity benefits of network coding that can mitigate the wireless fading was shown in [50].

Moreover, from a decodability perspective, the authors in [74] proposed ZigZag decoding that is based on interference cancellation. Hence, it requires a precise estimation of channel coefficients for each packet involved in a collision. In [75] an opportunistic network coding approach was introduced. In particular, the codewords are adapted according to the received information from the neighbors. In [76], the authors show that fixed network codes without CSI cannot achieve instantaneous min-cut. However, they proved that adaptive network codes with one bit global CSI have lower erasure probability than the codes without CSI. Network codes adaptation strategies that are

based on channel state information are limited to packet erasure channel model, that is, a two-state Markov model of Gilbert-Elliott channel, in which a packet is whether dropped with a certain probability or received without error, see [77], [78]. In [77], the authors developed a rate-controlled, multipath strategy using network coding. They showed that such strategy can provide throughput performance comparable to multipath flooding of the network while utilizing bandwidth nearly as efficiently as single-path routing. In [78], the authors studied the delay and energy performance under bursty erasures. They proved that channel-aware policies reduce delay by up to a factor of 3 and significantly increase the network's stable throughput region compared to a simple queue-length driven policy.

Further, the authors in [71] explored the timing nature of coding across packets over TDD channels, capitalizing on the time to absorption, see [99]. However, despite the novelty of the model proposed, the model considers channels with fixed erasure probabilities, which correspond to channels at steady state taking apart the channel variation over time due to different fading processes. Therefore, they consider time invariant channels, i.e., with fixed erasure probability.

Despite the fact that random linear network coding inherently adapts its rate to variations of the channel, however, in fading channels, the packet erasures become time dependent. Therefore, it was of particular relevance to tackle the network coding mechanisms over time varying channels and to describe on a packet level the statistical model of time-varying channels; aiming to describe, design, and optimize network coding mechanisms on top of scenarios where fading causes time variation and attenuation over the transmitted signals.

The modeling process provides the physical processes taking part in the propagation and reception process of a packet [133]. Therefore, statistical models are more customizable to real-world than empirical ones which are based on fitting of measured data. It is well known that the log-normal distribution is used to describe the variations of the modulated signal amplitude due to attenuation caused by obstacles impeding the line of site path.

Therefore, we try in this Chapter to put forth a statistical model for packet transmission over time varying fading channels via a Markov chain. The channel model is analyzed for the log-normal distributions. However, the same model can be analyzed for other small scale fading processes like Rayleigh, Ricean, or mixed ones. The variation of the channel states over time induces a time variation in each packet transmission. Therefore, we exploit the channel delay profile and the dependency

between channel states via a first order auto-regressive model (AR1). The auto-regressive model can be of first or higher order, however, AR1 is an appropriate choice for satellite channels, see [119]. This auto-regressive structure casts insights to the channel variations under fading and consequently, we derive an approximate closed-form for the delay induced due to the fading process. Furthermore, the auto-regressive model provides a general form to the proposed Markov model where each log-normally generated channel state(s) per Markov state - with or without sub-carrier correlation - puts forth the general distribution of the next states. We capitalize on the statistical properties of the Markov process to derive an approximate closed-form expression for the time to absorption corresponding to the time to deliver N coded packets assuming finite number of time slots to transmit a given number of packets.

This Chapter focuses on the study of the end-to-end delay of packet flows by tracking the CSI variation over time and their associated erasures, over Ka-band satellite systems. Our goal is to find network coding mechanisms that can adapt to the channel variability on the fly. Therefore, defining a strategy of when to stop rate adaptation is a decision of particular interest for the delay-throughput performance gains.

The contributions of this Chapter are three fold:

First, we propose a novel model for uncoded and coded packet transmission over time varying channels.

Second, we propose a network coding scheme, provide delay approximations for the mean completion time to deliver the coded packets over time varying channels.

Third, we propose a novel adaptive transmission strategy that accounts for the lost degrees of freedom due to channel erasures.

Our results show that network coding non-adaptive mechanism for time variant channels has around 2 times throughput and delay performance gains for small size packets over network coding mechanisms with fixed channel erasures and similar performance gains for large size packets. In addition, its shown that network coding non-adaptive mechanism for time variant channels has similar performance to the SR with ARQ, and better performance when packet error probability is high, while due to better utilization of channel resources SR performance is similar or moderately better at very low erasures, i.e., at high SNR.

However, our adaptive transmission scheme outperforms the network coding non-adaptive mechanism and SR with more than 7 times in throughput and delay performance gains.

8.2 Channel Model

Consider a downlink transmission over a wireless channel, the received vector will be modeled by:

$$y(t) = h(t)P_T x(t) + n(t) \tag{8.1}$$

$x(t)$ and $y(t)$ correspond to the transmit and receive symbols respectively. Assuming a generalized fading model where fading is time varying and log-normally distributed process, with the channel gain between transmitter and receiver modeled by $h \sim log - \mathcal{N}(m, \sigma^2)$, P_T is the transmitted power, and n is the zero mean complex white Gaussian noise, with $\mathcal{CN}(0, 1)$. The mean of h is $\mathbb{E}[h(t)] = e^{(m+\sigma^2)/2}$, and its variance is $\mathbb{E}[h(t)^2] = e^{(2m+\sigma^2)}(e^{\sigma^2} - 1)$. Such channels are commonly found in the microwave region; such as satellite channels in the Ka-band. We take the Ka-band as an example for the studies of this Chapter. In such band, the dynamics of rain fading are generally described by an auto-regressive moving-average model with order p [134], [135], [136] as follows:

$$h(t) = -\sum_{i=1}^{p} a_i h(t-i) + \omega(t), \tag{8.2}$$

$\omega(t)$ is a zero mean unit variance white Gaussian process, and the AR correlation coefficient bounded as $0 \leq a_i \leq 1$ corresponding to slow fading at $a_i = 1$ and very fast fading at $a_i = 0$. In general, the AR coefficient of an AR model is given by:

$$a_i = \frac{\mathbb{E}[h(t)h(t-i)]}{\sigma^2} \tag{8.3}$$

With $\mathbb{E}[h(t)h(t-i)]$ corresponds to the PDP of the channel. Therefore, if $\sigma = 1$ then a corresponds to the PDP of the channel. A special case of this model is the first-order AR1 model, which was found in [119] to provide a good match with experimental data. The AR1 model represents the fading process in discrete time as:

$$h(t) = -a_1 h(t-1) + \omega(t) \tag{8.4}$$

Thus, a_1 is a one-step autocorrelation coefficient, defined as:

$$a_1 = \frac{\mathbb{E}[h(t)h(t-1)]}{\sigma^2} = \frac{R_h(T_s)}{R_h(0)} \tag{8.5}$$

Where $R_h(t)$ corresponds to the autocorrelation function of the underlying continuous-time process. The sampling period is T_s. The corresponding PSD of the AR1 model has the following rational form:

$$S_h(f) = \frac{\sigma^2}{|1 + a_1 e^{-j2\pi f}|^2} = \frac{\sigma^2/\pi f_d}{1 + (f/f_d)^2} \tag{8.6}$$

Where $a_1 = e^{-\beta}$, and $\beta = 2\pi f_d T_s$ is the normalized fading rate through which the power delay profile decays exponentially with f_d the Doppler spread of the fading process. After generating the first channel state h corresponding to the log-normally distributed channel at a certain Markov state, the estimation of the channel gains at one-step forward transition can be found by solving the Yule-Walker equations to find the AR filter coefficients, multiplying equation (8.4) by $h(t-1)$ and taking the expectation of both sides. However, we assume a pre-knowledge of the starting channel state, and a pre-knowledge of the channel correlation coefficient a.

8.2.1 Probability of Error

Assume that there is no channel coding within the received packets. For a packet to be received every symbol need to be received. Therefore, the packet erasure probability at channel state h_j with respect to the bit error probability is given by:

$$P_e(h_j) = 1 - (1 - P_b(h_j))^B \tag{8.7}$$

B is the total number of bits in a packet. P_b is the bit error probability which is dependent on the channel model, which is considered to be log-normally fading channel. We consider a pre-knowledge of the first channel state h_0. Therefore, for a given channel gain $G_0 = \mathbb{E}[h_0(t)^2]$ that is log-normally distributed, the probability of error is given by:

$$P_b(G_0) = \int_0^\infty Q\left(\sqrt{SNRG_0}\right) p_G(G_0), \tag{8.8}$$

With the one-dimensional Gaussian Q-function defined as:

$$Q\left(x\right) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{u^2}{2}} du, \tag{8.9}$$

and the channel gain probability density function is given by:

$$p_G(G_0) = \frac{4.3429}{G_0 \sigma \sqrt{2\pi}} e^{-(10\log_{10} G_0 - m)/2\sigma^2}, \tag{8.10}$$

where $10\log_{10} G(t)$ follows a Gaussian distribution. Hence, choosing $m = -\sigma^2/2$ makes the average power loss due to channel fading equals unity. To ensure that the fading does not attenuate or amplify the average power, h is normalized such that

the MGF[5] $M_h(2) = \mathbb{E}[h(t)^2] = 1$. Therefore, a typical value is to choose $m = -0.5$, and $\sigma^2 = 1$ [137]. Therefore, with a prior knowledge of the first channel state h_0, we can find the rest of the channel states by the first order auto-regressive model with one time lag such that, the distribution of the new channel gain given the previous $p(h_k(t)|h_{k-1}(t-1))$ is Gaussian with mean $ah_{k-1}(t-1)$, and variance equals to the noise variance which is 1.

Therefore, the bit error probability of the predicted states can be given by:

$$P_b(h_k|h_{k-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} Q\left(\sqrt{SNR|h_k|^2}\right) p_{h_k|h_{k-1}}(h_k|h_{k-1}) dh_k, \tag{8.11}$$

and due to the moving average $h_k \sim \mathcal{N}(ah_{k-1}, 1)$, the conditional Gaussian fading channel gain probability is given by:

$$p_{h_k|h_{k-1}}(h_k|h_{k-1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(h_k - ah_{k-1})^2}{2\sigma^2}} du \tag{8.12}$$

8.3 Time-Varying Channel Model of Packet Transmission

We model an N packet transmission over a time-varying channel, i.e., we consider a model with slotted time. Therefore, the single packet transmission will occur over one or more number of time slots based on the channel conditions. Each packet transmission corresponds to a vector of transmitted symbols that suffer from log-normal fading over time, each vector is represented by one state in the Markov chain. We assume a finite number of time slots[6] required to transmit the packets successfully, and we consider pre-knowledge about the first channel state h_0 the first packet will encounter. Therefore, we first consider one channel coefficient corresponding to one packet transmission.

Figure 8.1 illustrates the auto-regressive Markov model with order $p = 1$ through which the current channel state is only dependent on one-time lag past state. Note that this is a fair assumption when considering practical systems like satellite wideband systems [134]. In addition, it illustrates the model of 3 packet transmission over time-varying channels. To transmit 3 packets, first packet transmission at state

[5]The moment generating function of h is given by $M_h(s) = exp(m_h s + \frac{\sigma_h^2 s^2}{2})$.

[6]To transmit N packets, we assume a finite number of time slots $(\sum(N)+1)$ per packet transmission, creating a transition matrix P of size equals to $((\sum(N)+1) \times N) \times ((\sum(N)+1) \times N)$.

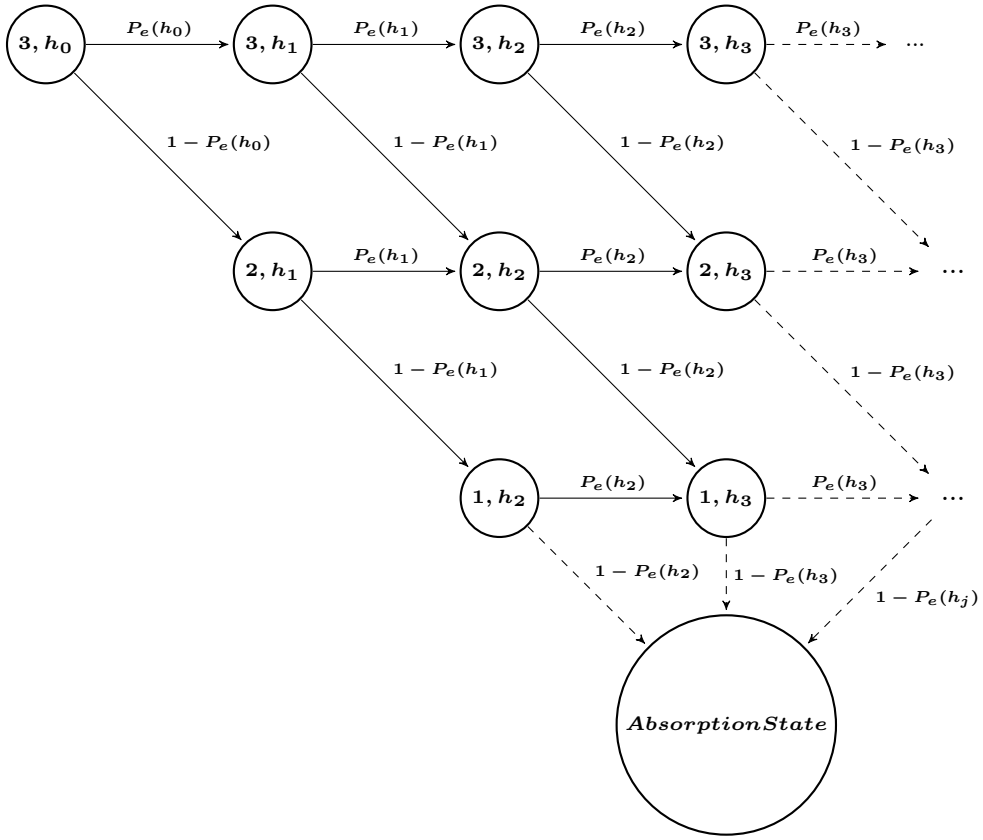

Figure 8.1: Channel State auto-regressive Model with First Order Dependency: 3 Packets Transmission.

$(3, h_0)$ will be either successfully delivered at channel state h_0 with probability $1 - P_e(h_0)$. Therefore, the transition will occur to state $(2, h_1)$, or fail to be delivered with probability $P_e(h_0)$ causing a transition to state $(3, h_1)$, and then the transition will occur to states $(3, h_3)$ or $(2, h_2)$ whether the packet fail or delivered successfully, respectively. If the last packet at state $(1, h_k)$, $k \geq 3$ is delivered successfully after the channel evolution over time, i.e., after a certain number of time slots, the chain will be absorbed. Worth to note that each packet will be transmitted at a new channel state, and each channel state is estimated via the previous state creating a correlated structure. Therefore, digging into the physical layer and the bit - or symbol - level modeling, the sum of log-normally distributed channel states can be represented by one log-normally distributed random variable which corresponds to the packet level, see [138].

8.3.1 The Expected Time to Transmit i Packets

We can write the expected time to deliver i packets in the following closed-form:

$$T(i,h_j) = T_d(i,h_j) + \sum_{\forall l,k} p_{(i,h_j)\to(l,h_k)} T(l,h_k) \tag{8.13}$$

Therefore, the expected time to deliver i packets becomes as follows:

$$T(i,h_j) = T_d(i,h_j) + p_{(i,h_j)\to(i-1,h_{j+1})} T(i-1,h_{j+1}) + p_{(i,h_j)\to(i,h_{j+1})} T(i,h_{j+1}) \tag{8.14}$$

The time to deliver a packet at a given channel state equals the packet length $T_d(i,h_j) = T_p$, and the transition probability $p_{(i,h_j)\to(i-1,h_{j+1})}$ is given by:

$$p_{(i,h_j)\to(i-1,h_{j+1})} = 1 - P_e(h_j), \tag{8.15}$$

and the probability of failure in transmitting one coded or uncoded packet $p_{(i,h_j)\to(i,h_{j+1})}$ is given by:

$$p_{(i,h_j)\to(i,h_{j+1})} = P_e(h_j), \tag{8.16}$$

where $P_e(h_j)$ is the packet erasure probability when the channel $h(t) = h_j$ for the duration of the packet transmission. Therefore, we can write the time evolution of the packet transmission over the channel fading states in a matrix form T, where the matrix T is not necessarily symmetric, as follows:

$$T = \begin{pmatrix} T_{1,h_1} & T_{1,h_2} & \cdots & T_{1,h_n} \\ T_{2,h_1} & T_{2,h_2} & \cdots & T_{2,h_n} \\ \vdots & \vdots & \ddots & \vdots \\ T_{N,h_1} & T_{N,h_2} & \cdots & T_{N,h_n} \end{pmatrix} \tag{8.17}$$

It is straightforward to observe that the expected time to deliver N packets can be the discrete sum of the packet transmission over different channel states. However, given a packet transmission and retransmission, an N packet transmission will be a function of the previous $N-1$ packet transmissions and so the channel states differ in the transmission-retransmission parts. Emphasizing that the model considers that a packet cannot be re/transmitted at the same channel state that it was transmitted with, therefore, we account for the time variations in the fading process.

8.4 Network Coding Scheme

Our goal is to transmit N packets or degrees of freedom over the time varying channel. Therefore, we propose a network coding scheme that accounts for channel variability over time. In particular, we propose a network coding scheme where we transmit N_i coded packets that would account for the channel variations and the degrees of freedom of the receiver. If the N_i coded packets are transmitted under certain channel variability, received and decoded successfully, the receiver will send an acknowledgment, asking for the lost degrees of freedom in the first transmission. The new coded packets will be transmitted with new channel states, the process is repeated until all coded packets that are adaptively not transmitted due to channel variability are compensated with encoding across the packets. In particular, the process will be finished when all the degrees of freedom are successfully delivered.

Therefore, looking into Figure 8.1, the transmission of N_i coded packets is underlined by all the states in the Markov chain. In fact, the transition probability $p_{(i,h_j)\rightarrow(l,h_k)}$ is decided by one or more number of paths the ending state is reached through. It follows that we have a transition probability matrix P, defined up to N_i slots, with two transition probability components, $p_{(i,h_j)\rightarrow(i,h_{j+N_i})}$ and $p_{(i,h_j)\rightarrow(l,h_{j+N_i})}$, $\forall l < i$. The one step transition probabilities are as given in (8.15) and (8.16), and the probability of transitioning from the channel state and back to itself equals zero. Moreover, the transition probability $p_{(i,h_j)\rightarrow(l,h_k)}$ between the states over all the possible paths has $m = k - j$ steps in time slots for a transmission to take place.

From a timing perspective, if the transmission of N_i coded packets over $h_j, ..., h_{j+N_i}$ was not successfully decoded for all N_i coded packets; we account for the extra time via the probability of transitioning and being absorbed after first transmission, in addition to the time accounting for the lost packets or lost DoF of the receiver. Therefore, the expected time required to deliver N_i coded packets is a sum of the first transmission time, the waiting time to have acknowledgment, and the transmission time of the lost DoF; where the ACK packet includes the information of the lost DoF, during which the channel process evolves until a new transmission will take place at the new channel state. Therefore, the expected time to deliver N_i coded packets will be a function of the expected time to deliver the packets at their given - or predicted - states as follows:

$$T(i,h_j) = T_d(N_i,h_j) + \sum_{l=1}^{i} P^{N_i}_{(i,h_j)\rightarrow(l,h_{j+N_i})} T(l,h_{j+N_i+1}), \tag{8.18}$$

with $T_d(N_i,h_j) = (N_i+1)T_p$ where the acknowledgment time appears as an addition

of one into the time slot indices in the equation above, and the round trip time is considered to be zero. The matrix P is the transition matrix of the proposed model, however,

$$\left(\prod_{i=1}^{N_i} P\right)_{(i,h_j)\to(l,h_{j+N_i})} = P^{N_i}_{(i,h_j)\to(l,h_{j+N_i})}, \tag{8.19}$$

corresponds to all transition probabilities over the time slots from the initial given log-normally distributed channel state h_j until the estimated state h_{j+N_i} at $j+N_i$ time slot. The $j+N_i+1$ appears in the timing consideration due to the acknowledgment.

8.4.1 Minimum Time to Deliver N_i Coded Packets

The minimum expected time to deliver N_i coded packets can be optimized to find the optimal number of coded packets to transmit over each transmission, as follows:

$$\min_{N_1,\ldots,N_i} T(i,h_j) = \min_{N_1,\ldots,N_i} T_d(N_i,h_j) + \min_{N_1,\ldots,N_i} \sum_{l=1}^{i} P^{N_i}_{(i,h_j)\to(l,h_{j+N_i})} T(l,h_{j+N_i+1}) \tag{8.20}$$

However, its not possible to find a closed-form expression for the optimal number of coded packets, through this approach as it become combinatorial problem. Therefore, such optimization can be performed numerically, otherwise, we can capitalize on the special case where the probability of erasure is fixed over all transmissions, the authors in [71] provide the optimal $N_1^\star$. It follows that, its very important to find a strategy that can optimize the number of coded packets on the fly.

8.4.2 Adaptive Transmission Scheme of Coded Packets

In this section, we propose a novel adaptive transmission scheme of coded packets, the strategy will rely on the knowledge of the channel. Due to the variation of packet erasure probability over each packet transmission, given by $P_e(h_j)$, the receiver can successfully decode $1-P_e(h_j)$ packets. Therefore, the adaptive transmission strategy will be to account for the lost packets or lost DoF, via the transmission of coded packets. The following equation presents the proposed strategy:

$$\sum_{s=j}^{N_i^\star} (1-P_e(h_s)) = i, \tag{8.21}$$

where $j = 0$ corresponds to the initial state the timing starts with, which is, $T(N_i, h_0)$. If N_i packets transmitted starting from h_0 to receive successfully i DoF, then for each other transmission we account for the lost DoF by extra coded packets. For example, if the receiver need to receive successfully i packets, i.e., i DoF, then the transmitter need to account for the packet erasures. This is done by successfully transmitting N_i coded packets; at the point the sum of $(1 - P_e(h_s))$ is equal to the DoF, and so on and so forth until the last DoF is delivered to the receiver. The adaptive strategy will produce for each SNR a set of optimal number of coded packets to transmit $N_i^\star$. The implication of this process, on the mean completion time, is that the transition matrix should span up to $N_i^\star$ in time slots. Therefore, we adaptively transmit accounting for packet erasures caused by channel variation over time as well as the DoF of the receiver. Algorithm 5 explains the process of transition matrix generation, delay calculation, and the adaptive optimization process. Notice that the adaptation is dependent on the SNR level. Therefore, the time when to switch on or off adaptation is of particular importance to the process. In fact, adaptation is more required at the low-SNR regime.

8.5 Simulation Results

We shall now present a set of results to provide further insight into the solution. We consider a packet length $T_p = 1/150$ sec and the log-normally generated channel with mean $m = -0.5$ and variance $\sigma = 1$, we first analyze the delay and throughput with respect to the SNR under different correlation coefficients. Figure 8.2 shows the throughput vs. the SNR, and Figure 8.3 shows the delay vs. the SNR. Figure 8.2 clearly depicts the maximum throughput reached in each case which is equal to $1/T_p$. We see that in the case when $a = 0$ and the channel variation is fast, and so the delay is highest, throughput is lowest. However, when $a = 1$ the channel variation is slow, so the delay is lowest, and throughput is highest. Intuitively, we can understand that the main factor influencing the process is the correlation coefficient. In fact, the change in the knowledge in the first channel state has less influence. We can also see that the assumption of fixed erasure probability is relevant as it gives an average or near to very slow variation behavior of the channel. However, it doesn't give an accurate view of the channel variation and the decay in throughput or increase in delay caused due to fading, particularly at low SNRs.

Moreover, its worth to notice how the model considers the delay introduced due to acknowledging each batch of transmitted packets updating for the lost DoF. Figure 8.4

Algorithm 5: Adaptive Transmission Scheme

Inputs:

degrees of freedom N=[1, 2, 3,..., dof]

pre-knowledge of channel state: h_0

estimated channel state(s) $h_1, h_2, ..., h_L$

For a given $SNR(k)$

Find the probability of erasure vector: $P_e(h_0), P_e(h_1|h_0), ..., P_e(h_L|h_{L-1})$

function1: Generate transition matrix $P(SNR)$

The transition matrix elements is such that:

$p_{(i,h_j)\to(i,h_j)} = 0,$

$p_{(i,h_j)\to(i,h_{j+1})} = P_e(h_j),$

$p_{(i,h_j)\to(i-1,h_j)} = 0,$

$p_{(l,h_j)\to(i,h_j)} = 0, \ \forall l > i,$

$p_{(i,h_j)\to(l,h_j)} = 0,$ if $\ l > i-1,$

$p_{(i,h_j)\to(l,h_{j+1})} = 1 - P_e(h_j),$ if $l = i-1,$

$p_{(l,h_j)\to(l,h_j)} = 0, \ \forall \ l,$

$p_{(l,h_j)\to ABS} = 0, \ \forall \ l \neq 1,$

$p_{(1,h_j)\to ABS} = 1 - P_e(h_j),$ and,

$p_{ABS\to ABS} = 1.$

function2: for $z = 1 : dof$

Find: $N^\star(z)$ such that $\sum_{s=j}^{N^\star(i)} (1 - P_e(h_s)) = dof(z)$, Then: $N(dof) \longleftarrow N^\star(z)$

Find: $P^{N^\star}(i)$

Find: $T(i, h_j) = T_d(N_i, h_j) + \sum_{l=1}^{i} P^{N_i}_{(i,h_j)\to(l,h_{j+N_i})} T(l, h_{j+N_i+1})$

Output :

Delay: $T(i, h_j, SNR)$

Optimal Number of Coded packets: $N^\star(SNR) = [N^\star(1), ..., N^\star(dof)]$

and Figure 8.5 illustrate network coded non-adaptive scheme over time varying channel under different RTT, where high RTT is known to be a major issue in satellite communications. However, its straightforward to notice that the time of the ACK have significant impact on the delay-throughput performance. In fact, if we design our system to acknowledge only specific packets like odd or even ones, instead of

acknowledging each packet, we can choose the strategy that encounters less delays based on our transmissions if odd or even.

8.5.1 Comparison with SR-ARQ

We consider 3 packet transmission and a window size of 3 to compare our schemes to half-duplex SR with ARQ as in [Eq.17, [71]] with $P_e = P_e(h_0)P_e(h_1|h_0)P_e(h_2|h_1)$, each packet is 1000 bits length, $T_w = 0$, and correlation coefficient $a = 1$. We use an upper bound based on the vector of erasure probabilities found via our AR model. Therefore, we account only for the delays due to the first 3 packets transmission. Notice that we can also make a comparison between our novel schemes and selective repeat with channel state knowledge on the steady state capitalizing on the outage probability of the sum of log-normally correlated random variables, [138]. Higher throughputs can be realized in Figure 8.6 and Figure 8.7 using the adaptive scheme that mitigates the channel variation. In addition, we can see that NC over time variant channels schemes outperform the selective repeat with ARQ at the low-snr when error probability is high. However, we can see that at the high-snr all schemes meet at a certain point when error probability is almost zero. It is worth to know that the upper bound for NC with 1 or 2 packets addition increases the throughput. However, this doesn't account for the extreme channel variations under the low-SNR. Therefore, the analysis via our novel adaptive scheme introduces a new upper bound for the optimal or sub-optimal number of coded packets that need to be transmitted before stopping to listen an ACK. This works in favor to lessen the number of ACK packets and therefore, optimizes better the delay performance in real scenarios. Therefore, with the adaptation scheme, whether a NC or a SR, could inherently exist when running the system on the fly. Such hybrid nature of this adaptive scheme would be sufficient for mitigating the channel variations; boosting the throughput, and improving the delay performance. It is also easy to quantify the performance gains of our schemes through the results. For instance, at medium-snr, the adaptive scheme outperforms the non-adaptive schemes with more than 7 times. At the high-SNR regime, the number of packets to transmit will be equal to the received packets $N_i^\star = N_{rx}$. Notice that for the case where fixed erasure is considered as shown in Figure 8.2 and Figure 8.3, the optimal number of packets are $N_i^\star = N_{rx}/(1 - P_e)$, therefore, if a transmission occurs over a channel with 0.2 erasure probability, the transmitter need to account for this by transmitting 1.2 packets over the channel. Similarly, we are accounting for the channel spread of different erasures via our adaptive transmission strategy.

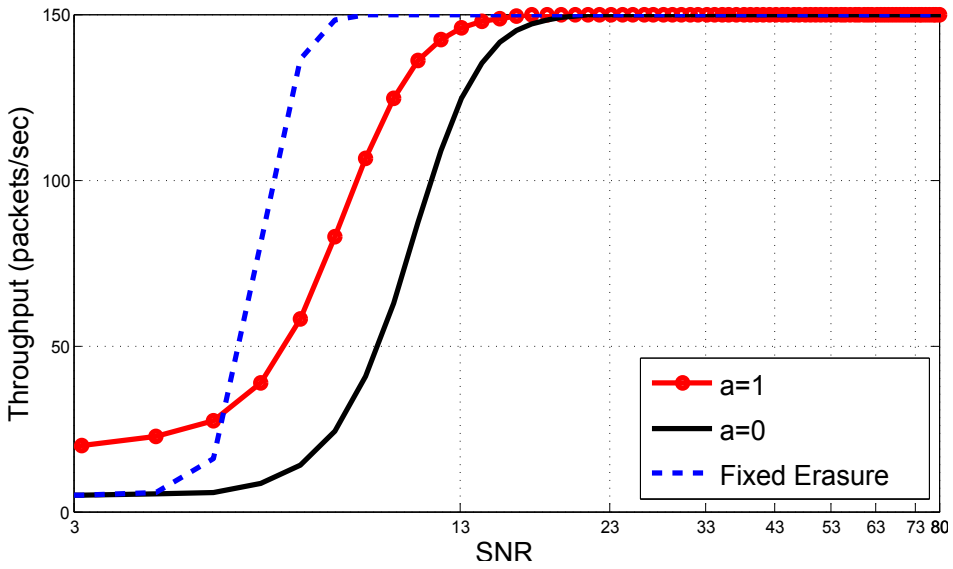

Figure 8.2: Throughput vs. SNR and channel correlation a.

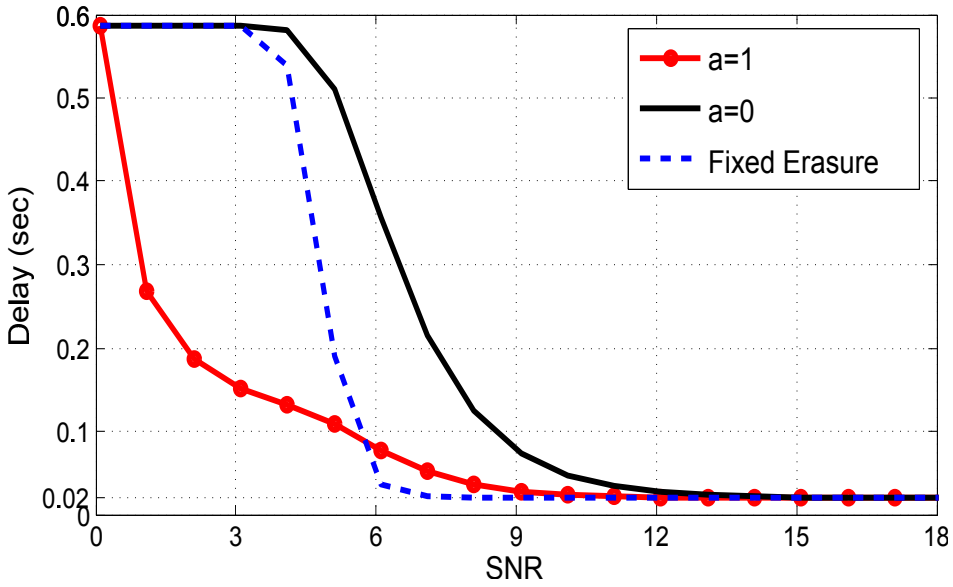

Figure 8.3: Delay vs. SNR and channel correlation a.

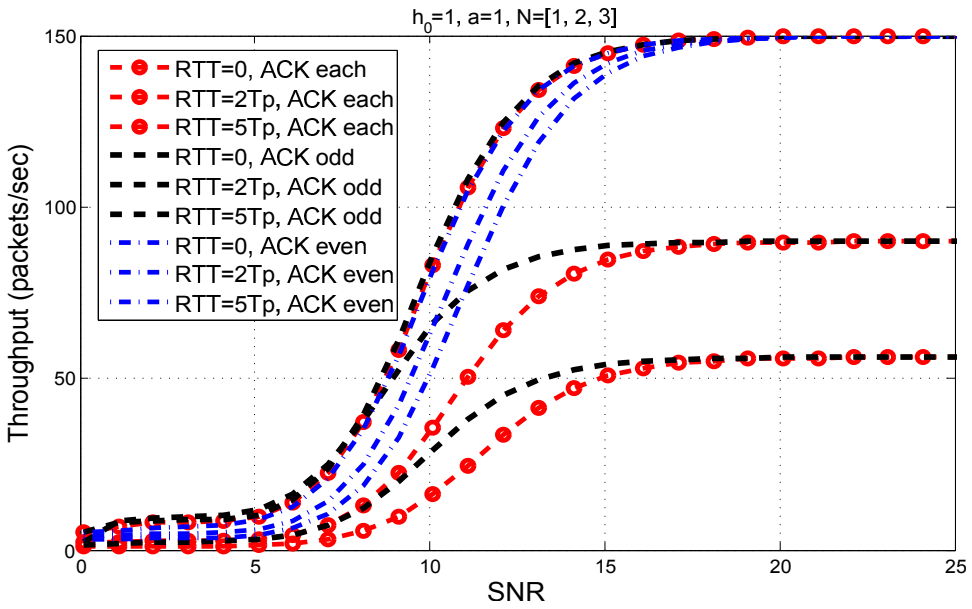

Figure 8.4: Throughput vs. SNR under different RTT, with odd and even ACK.

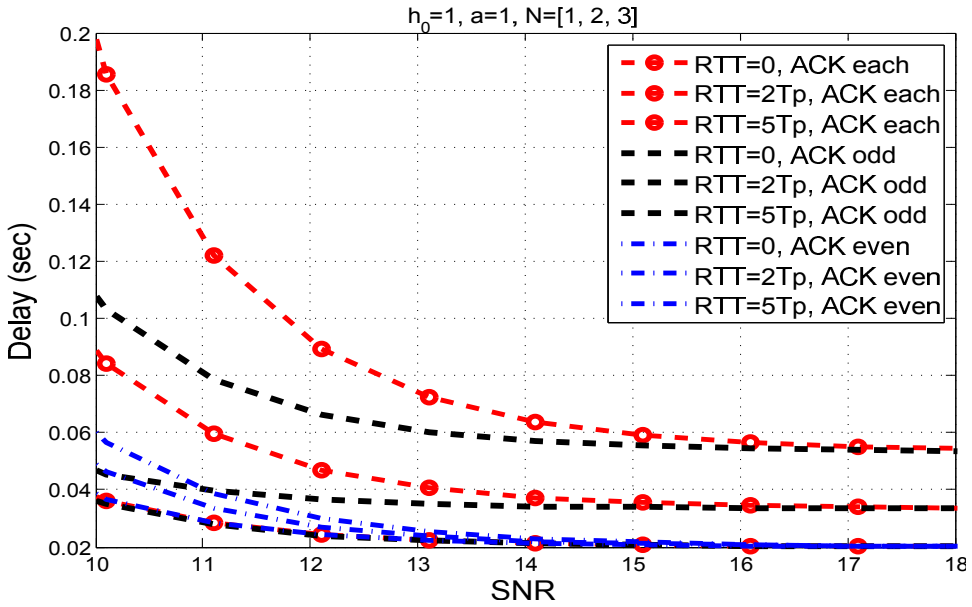

Figure 8.5: Delay vs. SNR under different RTT, with odd and even ACK.

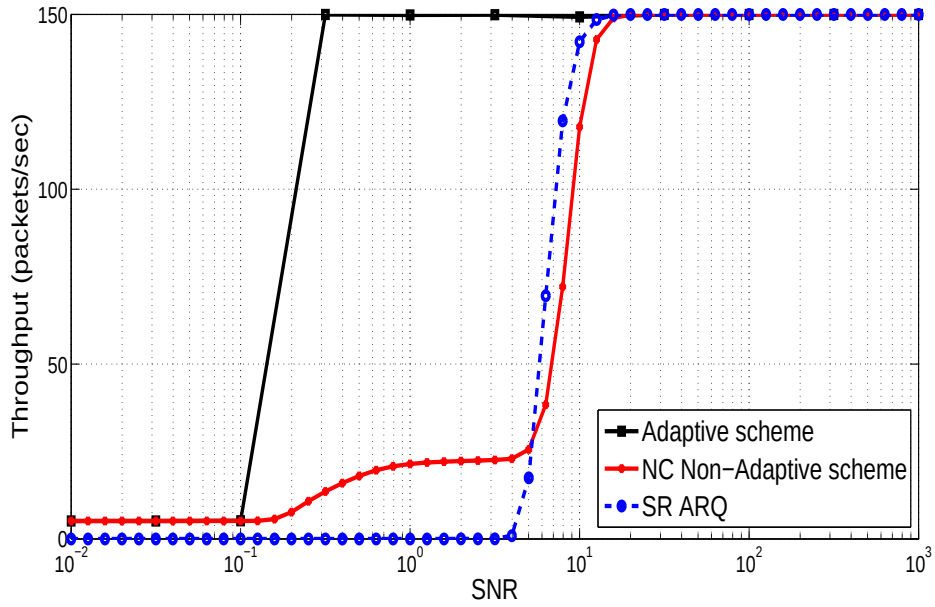

Figure 8.6: Throughput for NC scheme, adaptive scheme and compared to SR-ARQ.

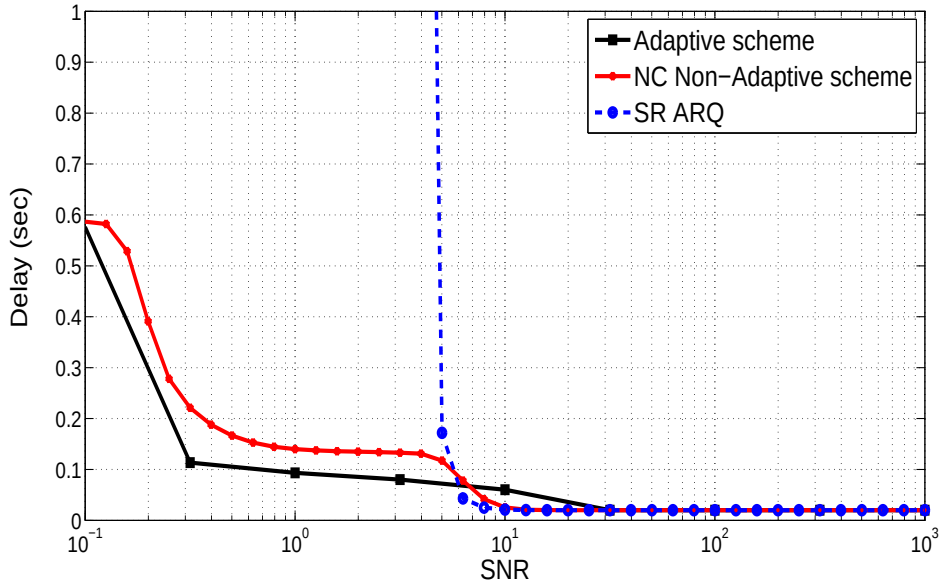

Figure 8.7: Delay for NC scheme, adaptive scheme and compared to SR-ARQ.

8.6 Conclusion

In this Chapter, we address network coding over time varying channels. We consider a small scale fading channel over Ka-band satellite communications. We propose a novel model for packet transmission over time variant channels that exploits the channel delay profile and the dependency between channel states via first order auto-regression. We provide an approximation of the delay induced assuming finite number of time slots to transmit a given number of packets. Due to the fact that the problem of maximizing the data rates to derive the optimal number of packets to transmit become combinatorial, and due to the the time dependency between the erasures, a NC scheme which is not only aware of the channel behavior but also adapts its behavior based on its knowledge need to be thought of. Therefore, we propose a novel adaptive transmission scheme that compensates for the lost degrees of freedom by tracking the packet erasures over time. Therefore, we provide a characterization of the optimal number of coded packets that need to be transmit back to back before the transmitter stop to wait for acknowledgment. This optimal number of coded packets are found based on the channel erasures variation and the degrees of freedom of the receiver to decode the received packets. In fact, this stands as an upper bound on the state of the art optimal number of coded packets identified for single transmission with fixed erasures. This setup, sheds light on the importance of the time of switching on/off the coded adaptive scheme to achieve the maximum capacity. Therefore, this setup demonstrates the relevance of defining adaptive transmission and coding strategies that are time dependent. Furthermore, we show via a set of illustrative results that our network coding non-adaptive scheme and the adaptive scheme for time variant channels outperform existing schemes like selective repeat ARQ under communication channels which encounter high erasures, making it a robust solution for existing systems. In particular, our results show that network coding non-adaptive mechanism for time variant channels has around 2 times throughput and delay performance gains for small size packets over network coding mechanisms with fixed channel erasures and similar performance gains for large size packets. In addition, its shown that network coding non-adaptive mechanism for time variant channels has similar performance to the SR with ARQ, and better performance when packet error probability is high, while due to better utilization of channel resources SR performance is similar or moderately better at very low erasures, i.e., at high SNR. However, our adaptive transmission scheme outperforms the network coding non-adaptive mechanism and SR with more than 7 times in throughput and delay performance gains. Notice that the proposed NC schemes for time variant channels consider a pre-knowledge of the first channel state

through which we can estimate the channel variation over time. However, if the first channel state is unknown, its hard to estimate the probability of error and therefore, it requires different methods to measure this probability, and so the time to deliver the packets. In turn, this might limit the usage of the proposed NC adaptive scheme, through which the adapted optimal number of packets to transmit as a function of the foreseen erasure over time will also change. Notice also that the proposed schemes have been analyzed for the log-normal distributions which are applicable to scenarios where shadowing exists. However, such schemes can be analyzed for other small scale fading processes like Rayleigh, Ricean, or mixed ones. In the broad sense, this assumption of small scale fading works well for slow fading channels. However, tracking the channel over fast fading scenarios, and so tracking the probability of erasures over fast time variations is another challenging process which may depart from the 1 time lag dependency known for satellite applications. Therefore, its of particular relevance to revisit the scenario considered for other kinds of fading, like Rayleigh or Rician, which takes into consideration a more detailed sense of the channel variations. In addition, thinking about the proposed NC adaptive scheme for time variant channels, over uncoded transmissions may depart from optimality due to the fact that the order of the transmitted packets may be unknown. Therefore, such adaptive scheme may be more suitable on top of coded transmissions. This is due to the fact that in a coded transmission, we will always have sufficient coded packets to restore the lost ones regardless of their order. All the proposed schemes with the limitations and/or extensions discussed are worth to verify in real systems for evaluation purposes. Future research will also consider the problem of energy consumption of the proposed scheme and an extension of the principles proposed (which were analyzed for one link in a network) to the general problem of wireless networks like multi-cast.

Chapter 9
IEEE 802.11 Network-Coded Handover

9.1 Introduction

Mobility is an essential key feature which requires special care in network design. Nowadays, there are several approaches attempting to provide optimal mobility management, from different stack perspectives, starting in TCP/IP Layer 2 with wireless and cellular technologies up to TCP/IP Layer 5 with SIP. However, all these approaches introduce some disruption time while the handover is performed. If different technologies are available, several other approaches provide the ability for the mobile terminals to be connected to the different heterogeneous technologies while moving and decreasing the disruption time. Even for the same technology, it is possible to make use of the "make before break" paradigm, where the new connection and configuration in the mobile terminal is prepared before the handover takes place. However, there is still some disconnection while a handover is performed. In a different perspective, NC constitutes a disruptive paradigm that relies on the mixing (coding) of packets at intermediate nodes in the network. NC is based on the simple concept that intermediate nodes are allowed to re-mix information flows in addition to routing them, being able to improve network capacity. From a receiver's perspective, it is no longer crucial to focus on gathering specific packets, but to gather enough coded packets to recover the original information. From the mobility perspective, this enables the support of recovering from dropped packets during handover through the transmission of some combinations of packets. Many analytical models that describe the operational properties in IEEE 802.11 WLANs have been derived the last few years, particularly, they try to model the primary MAC techniques: the DCF and/or PCF. DCF is a CSMA/CA scheme which implements a binary exponential backoff mechanism. Different models have been devised based on Bianchi's model of the DCF mechanism [65], and different proposed schemes to optimize the network capacity

or to enhance the QoS have been conducted. In [79] an analytical model to study the throughput of a p-persistent IEEE 802.11 protocol that selects a backoff window size that balance collision and idle period costs, other contributions built on top can be found in [80], [81], etc. Recently, an EDCF that employs a radically different contention window size as compared to DCF of IEEE 802.11 was introduced into the IEEE 802.11e. Therefore, an accurate theoretically based understanding is crucial to guide the design and improvement of effective schemes. A set of proposals to improve the QoS via introducing minor changes in the mechanisms like fixing the maximum contention window to be equal to the first contention, or limiting the number of retrials for the sake of decreasing the collisions, increasing fairness, or to mitigate the hidden terminal problem are given in, [82] and [83]. In other works, they tried to improve the delay performance in the handover scenario focusing on providing mechanisms that minimize the most contributor to the delay, which is the probing delay, [84], [85].

Therefore, in this Chapter we propose a new model of the IEEE 802.11 WiFi DCF, we propose a fixed average contention window, and we analyze the delay and throughput under different transmission modes, in particular the unicast and different broadcast transmission modes. We analyze the station mobility under a quasi-static Rayleigh fading scenario where we connect the probability of erasure to the signal to noise ratio for the coded and uncoded modulation and to the distance of the mobile station via the free space path loss model. We finally, make a comparison between the broadcast transmission to the unicast one for the uncoded case like in the DCF to the coded case using NC, i.e., coding across the packets.

It has been shown that for the coded case, the delay is decreased by taking into consideration the degrees of freedom the station has to be able to decode the coded packets. Therefore, we reduce the effect of packet erasures and so reduce the number of retransmissions required in case of failure. Moreover, we reduce delay as well as increase reliability in comparison with uncoded broadcast. This proposed framework adds to the listen then talk DCF a mechanism to stop talking after a certain time [71]. This serves in scenarios where streaming applications exist and particularly under handover scenarios where decreasing the delay is an endeavor. If cashing is also considered in the new AP, the station who is performing a handover to a new AP will not suffer real time recognizable delays.

In this Chapter we provide a solution for future communication systems that can decrease the disruption of the active sessions running during the handover by the use of network coding. In our approach, network coding will be the key enabler towards ensuring soft handover, providing the correct delivery of all packets in the running

session without duplicating the packets in both networks. In our approach, the idea is to use network coding to send additional coded packets, i.e., linear combination of the original packets, towards the mobile terminal performing handover. Firstly, this enhances information transmission before the handover, when the channel quality for both data and feedback is severely degraded. Secondly, it improves the probability of correct reception of the information after the handover, when the exact knowledge of the status of the terminals is unknown. These two steps shall be instrumental in achieving a seamless handover.

The choice of the number of coded packets to be sent before and after the handover needs to take into account several issues, such as the mobility pattern of the mobile terminal, signal variation in both the previous and the new access networks, handover time and the expected performance degradation while handover is taking place.

Moreover, the decision of handover with network coding needs also to be addressed, since it is required to evaluate the cost of handover - in terms of signaling and performance degradation - and the additional cost of introducing coding, both in the extra amount of information delivered and its delay in real-time communications.

In this Chapter, we focus on the IEEE 802.11 WiFi networks and we consider a handover scenario of one station between two WiFi APs. We particularly focus on the modeling issue, the delay issue, and the evaluation of current technologies with respect to the usage of NC in such scenario, and we provide a novel formulation to characterize the optimal time to switch to the new AP.

Therefore, we propose new models for the DCF of the IEEE 802.11, we analyze the delay over the unicast and broadcast transmission for a network topology of one AP and one station. We provide a closed-form expression for the expected time to deliver the N packets for the DCF mechanism, with unicast, the general broadcast, broadcast with ACK for the uncoded and coded transmission. We have shown that coding across packets in an acknowledged broadcast scenario encounters less delays, higher reliability, and higher throughout than for the uncoded broadcast or unicast cases. We propose a new protocol that utilizes network coding to broadcast coded packets to the station performing handover. This new proposed network-coded handover framework will immensely serve if implemented in the current standardized IEEE 802.11 systems. We build upon constraints that take into consideration the distance of the station and the degrees of freedom it owns to decode the received packets before it switches the connection to the next AP. Therefore, we provide a framework under which the QoS over delay sensitive streaming applications can be radically improved.

9.2 The DCF

The DCF of the 802.11 is based on a contention based mechanism. Each station contends to access the medium and succeeds in its access after a time the medium is sensed to be idle, called the DIFS. This will let the station generate a random backoff window in the range of $[0, .., CW_{min}]$, then perform fragmentation of the MSDU into a set of MPDUs to be transmitted; this indeed serves in increasing the reliability of transmission via per MPDU acknowledgment (ACK). Therefore, after the first contention window, the $MPDU_1$ transmitted, if it receives an ACK, the second $MPDU_2$ can be directly transmitted after a SIFS, if no ACK is received, the station waits for another DIFS to confirm that still the medium is idle and then generates a random backoff window in the range of $[0, .., 2CW_{min}]$, then it generates $MPDU_1$ again. Therefore, this backoff mechanism dictates a new contention if a failure in transmission or if no ACK is received until the last level of the backoff $[0, .., CW_{max}]$. If the transmission was successfully established over all the N packets $[MPDU_1, .., MPDU_N]$, the backoff is only done at the beginning and is not repeated along the process, such that the station transmits as follows, DIFS, CW1, $MPDU_1$, SIFS, ACK, SIFS, $MPDU_2$, SIFS, ACK, SIFS,.., SIFS, ACK, SIFS, $MPDU_N$, SIFS, ACK.

9.2.1 Modeling a Single-Packet Unicast Transmission in the DCF

Consider the absorption Markov chain shown in Figure 9.1, this models the DCF mechanism for a single packet unicast transmission. Of particular relevance to note that reliability in the unicast transmission dictates the usage of a contention mechanism by the station. The station after a DIFS and first backoff window $CW1$ will transmit its first packet with probability of success $p_s = (1 - p_e)(1 - p_{ack})$, this fact will let the process to be absorbed at the state of transmission TX, where the station can continue to transmit the rest of the packets. However, if the first transmission was not successfully established due to erasure probability p_e or to a loss in the ACK with probability p_{ack}, the station will choose to move to the next backoff stage with probability $1 - p_s$, such that it can only access the medium after a new contention with a backoff equal to $CW2$, then it retransmits the first message, if successfully, the Markov chain will be absorbed into the transmission state, if not, the process continues until the last backoff trial at level ℓ until its finally absorbed, i.e., the first packet is

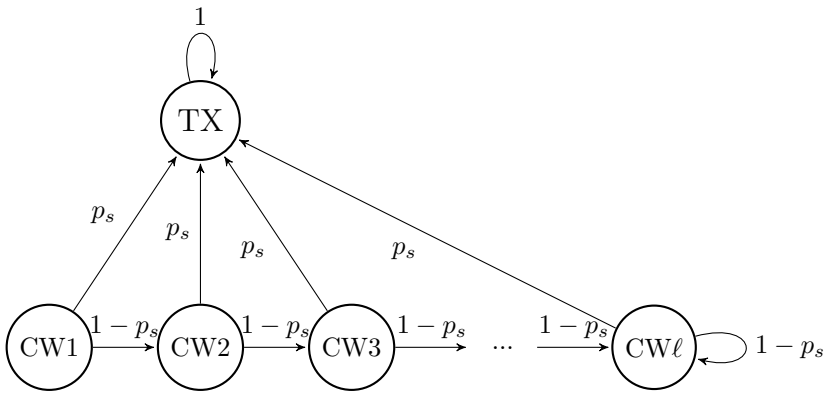

Figure 9.1: Single-packet unicast transmission in the IEEE 802.11 DCF.

successfully transmitted. Accordingly, the second packet will be transmitted with the same process until the Nth packet is transmitted from the station. For the sake of simplicity, First: we considered that if the transmission failed with probability $1 - p_s$ at the last backoff stage, the state stays as is - with a self loop - and is not absorbed to let the station keep trails without being absorbed into a distinct fail state, hence there is no limit defined in the number of retrials, and the maximum contention window is fixed at the last retry to its maximum size. Second: we didn't consider the frozen backoff case, i.e., the case when there is another station that can access the medium before the last decrement of the current station backoff takes place, i.e., if the current station senses a DIFS first, it will have the priority to continue accessing the medium, as long as its all MPDUs - within an MSDU - are to be transmitted; this is similar to the broadcast transmission with almost zero backoff but with higher reliability.

9.2.2 The Expected Time to Deliver First Packet

The model in Figure 9.1 illustrates the delay the packet encountered until it is transmitted at the absorption state TX. We provide a closed-form expression for the average time to deliver the first packet capitalizing on the time to absorption per transient state. The starting time is the MPDU period T_p plus the first contention window $CW1 \in [0, ..., CW_{min}]$ generated by the station after sensing the medium idle for a DIFS period. The expected time to deliver one packet can be written as the sum over all the expected times to deliver the packet at each contention

stage. If the time to deliver the packet at first trial with probability of successful transmission p_s is $T_d(\omega_1)$, then this is the time to deliver the first packet. However, if the packet is not transmitted (absorbed) at first trial, a random contention window $CW2 \in [0, ..., 2CW_{min}]$ is generated, and the expected time to deliver this first packet at second contention stage with probability $p_s(1-p_s)$ is $T_d(\omega_2)$, if the packet is not successfully transmitted at this stage, a random contention window $CW3 \in [0, ..., 3CW_{min}]$ is generated, and the expected time to deliver this first packet at second contention stage with probability $p_s(1-p_s)^2$ is $T_d(\omega_3)$. As far as the first packet transmission is not successfully established, the backoff mechanism will be repeated such that at the last backoff stage, the station generates a random contention window $CW\ell \in [0, ..., CW_{max}]$, and successful transmission at this stage occurs with an expected time to deliver this first packet with probability $p_s(1-p_s)^{\ell-1}$ is $T_d(\omega_\ell)$, the self loop with probability of failure $(1-p_s)$ at the last contention stage corresponds to the finite number of retrials the station will do until it transmits the packet, so no fail state is considered. Therefore, the average expected time to deliver the first packet starting from the first contention stage is as follows:

$$\begin{aligned}\mathbb{E}\left[T_{deliver\ \ first\ \ packet}\right] = T_d(\omega_1) + (1-p_s)T_d(\omega_2) + \\ (1-p_s)^2 T_d(\omega_3) + \ \ldots + \\ (1-p_s)^{\ell-2} T_d(\omega_{\ell-1}) + \frac{(1-p_s)^{\ell-1}}{p_s} T_d(\omega_\ell) \quad (9.1)\end{aligned}$$

However, digging into the depths of the equation by breaking down the components of the time to deliver the first packet at each contention stage, will be as follows:

$$T_d(\omega_1) = \bar{T}_p + \mathbb{E}\left[\omega_1\right] \quad (9.2)$$

$$T_d(\omega_2) = \bar{T}_p + \mathbb{E}\left[\omega_2\right] \quad (9.3)$$

$$T_d(\omega_3) = \bar{T}_p + \mathbb{E}\left[\omega_3\right] \quad (9.4)$$

$$\vdots$$

$$T_d(\omega_{\ell-1}) = \bar{T}_p + \mathbb{E}\left[\omega_{\ell-1}\right] \quad (9.5)$$

$$T_d(\omega_\ell) = \bar{T}_p + \mathbb{E}\left[\omega_\ell\right] \quad (9.6)$$

Therefore, we can write the expected time to deliver the first packet in the following closed-form:

$$\mathbb{E}\left[T_{deliver\ \ first\ \ packet}\right] = \sum_{i=0}^{\ell-2} \bar{T}_p(1-p_s)^i + \sum_{i=0}^{\ell-2} \mathbb{E}\left[\omega_{i+1}\right] T_{slot}(1-p_s)^i + \frac{(1-p_s)^{\ell-1}}{p_s}(\bar{T}_p + \mathbb{E}\left[\omega_\ell\right] T_{slot}) \quad (9.7)$$

It is worth to note that the expected backoff time (in T_{slot}) is:

$$\mathbb{E}\left[\omega_\ell\right] = \frac{1}{\ell+1}\sum_{i=0}^{\ell} i \quad (9.8)$$

And that,

$$\bar{T}_p = DIFS + T_p + SIFS + ACK \quad (9.9)$$

9.3 Modeling the Transmission of N Packets in the DCF

We shall now present a set of transmission schemes and their corresponding models. In particular, we will study different transmission schemes used on top of the DCF at the MAC Layer of the IEEE 802.11. We will first study the unicast without fragmentation, unicast with fragmentation, the uncoded broadcast without ACK, and the uncoded broadcast with ACK. Finally, we will dedicate a new section to discuss the network-coded broadcast with ACK.

9.3.1 Unicast without Fragmentation

Further, we will adapt the process of the single-packet unicast transmission in an iterative fashion to emulate the DCF process in transmitting N packets without fragmentation. Figure 9.2 illustrates the unicast transmission of N packets without fragmentation. Worth to note that each stage corresponds to a single-packet transmission, and the last absorption state of each stage corresponds to the initial state to the

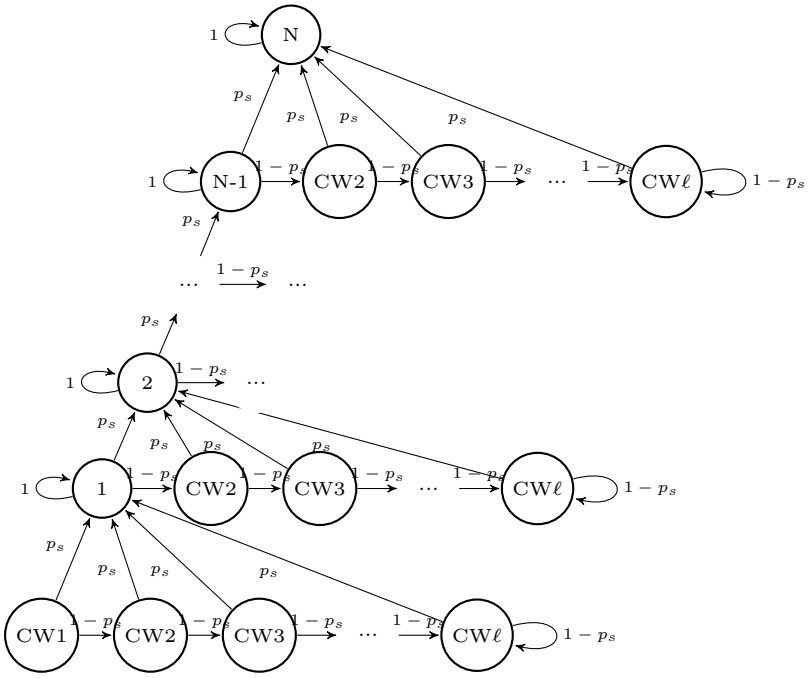

Figure 9.2: N packet unicast transmission without fragmentation in the IEEE 802.11 DCF.

next-packet transmission until the Nth packet is delivered, this is due to the fact that after a successful packet transmission, the backoff mechanism is reset and the station have to contend again, so after a DIFS and CW1 it can establish its second packet transmission with the backoff mechanism if any failure detected, if second packet is successfully transmitted, the station needs again to contend waiting a DIFS and CW1, and so on and so forth until all N packets are delivered. The expected time to deliver the N packets will be an accumulated sum over the average time to deliver each packet. In fact, this calculation is feasible since we are dealing with an average time. Therefore, the expected time to deliver N packets is the number of packets times the expected time to deliver the first packet, as follows:

$$\mathbb{E}\left[T_{deliver\ \ N\ \ packets}\right] = N\mathbb{E}\left[T_{deliver\ \ first\ \ packet}\right] \tag{9.10}$$

Note that the time to deliver N packets is a non-linear relation in the probabilistic sense. However, it is a linear relation between the times of delivering each individual packet.

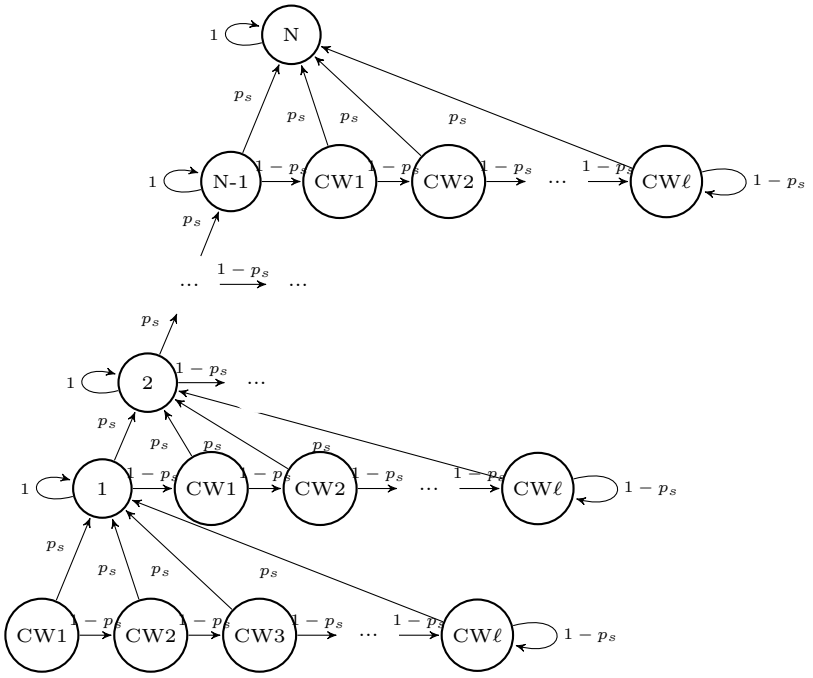

Figure 9.3: N packet unicast transmission with fragmentation in the IEEE 802.11 DCF.

9.3.2 Unicast with Fragmentation

When the MSDU size is bigger than a certain limit, the MAC layer do a fragmentation mechanism through which the MSDU is fragmented into a set of MPDUs where the timing between each packet and another is a $SIFS+ACK+SIFS = 2SIFS+ACK$. Unlike the Unicast without fragmentation; after a successful MPDU transmission, with $SIFS + ACK$ received, the station continues the $N-1$ MPDUs transmission with $SIFS + ACK + SIFS$ in between. Therefore, the station doesn't have to backoff unless there is a loss in an MPDU or its ACK. This could speed up a long packet transmission with higher reliability. Figure 9.3 illustrates the unicast transmission of N packets with fragmentation. It is worth to note the shift in timing in the states due to the mechanism discussed. Here the time to deliver the first packet (MPDU) stays the same, while the time to deliver the rest $N-1$ packets will be a little bit different. Therefore, the time to deliver the N packets with fragmentation is the expected time to deliver the first packet plus $N-1$ times the expected time to deliver the second packet, and given by the following closed-form:

$$\mathbb{E}\left[T_{deliver\ N\ packets}\right] = \mathbb{E}\left[T_{deliver\ first\ packet}\right] + (N-1)\mathbb{E}\left[T_{deliver\ second\ packet}\right] \quad (9.11)$$

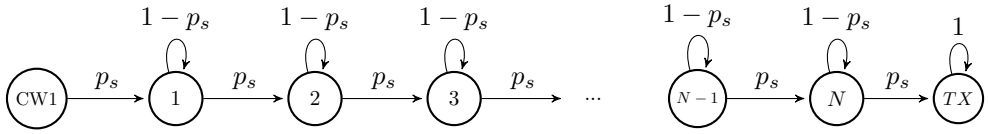

Figure 9.4: N packets broadcast transmission in the IEEE 802.11

With,

$$\mathbb{E}\left[T_{deliver\ \ second\ \ packet}\right] = \sum_{i=1}^{\ell-1} \bar{T}_p (1-p_s)^i + \sum_{i=1}^{\ell-1} \mathbb{E}\left[\omega_i\right] T_{slot} (1-p_s)^i + \frac{(1-p_s)^\ell}{p_s} (\bar{T}_p + \mathbb{E}\left[\omega_\ell\right] T_{slot}) + 2SIFS + T_p + ACK \quad (9.12)$$

9.3.3 Uncoded Broadcast Transmission without ACK

Figure 9.4 illustrates the Markov chain of an N packets uncoded broadcast transmission based on the IEEE 802.11. Broadcast frames neither protected by RTS/CTS, nor acknowledged. Therefore, correct reception cannot be guaranteed. And so, most of the applications uses unicast traffic while broadcast is usually used for routing update messages and beacon messages. Unicast favors reliability, while broadcast favors speed.

In a similar analytical way to the previous one for the unicast case, we can derive the expected time to deliver N packets via uncoded broadcast transmission for the model in Figure 9.4. Therefore, the expected time to deliver N packets via broadcast transmission is as follows:

$$\mathbb{E}\left[T_{deliver\ \ N\ \ packets}\right] = \sum_{i=1}^{N} \frac{T_d(i)}{p_s} \quad (9.13)$$

However,

$$p_s = (1 - p_e) \quad (9.14)$$

Similarly, due to the fact that no backoff mechanism in the broadcast transmission, therefore $T_d(1) = DIFS + CW1 + T_p$, and the spacing in time between each packet and the other is the $SIFS$ period without any ACK, we can rewrite the expected

time to deliver N packets for the uncoded broadcast without ACK in a more compact form as follows:

$$\mathbb{E}\left[T_{deliver\ N\ packets}\right] = \frac{NT_p + (N-1)SIFS + DIFS + CW1}{(1-p_e)} \tag{9.15}$$

In a best case scenario where no erasures happen and all transmission is established after first contention, the expected time to deliver N packets by a station that have already win the medium via broadcast, unicast without fragmentation, and unicast with fragmentation respectively are as follows:

$$\mathbb{E}\left[T_{deliver\ N\ packets\ broadcast}\right] = NT_p + (N-1)SIFS + DIFS + CW1 \tag{9.16}$$

$$\mathbb{E}\left[T_{deliver\ N\ packets\ unicast}\right] = NT_p + NSIFS + NACK + DIFS + CW1 \tag{9.17}$$

$$\mathbb{E}\left[T_{deliver\ N\ fragmented\ packets\ unicast}\right] = NT_p + (2N-1)SIFS + NACK + DIFS + CW1 \tag{9.18}$$

It is quite clear the difference between the delay using the different modes of transmission in the IEEE 802.11, in a best case scenario the difference is at least $NACK+SIFS$ for the non-fragmented case. Notice that we compare between the fragmented and the non-fragmented unicast in a common bases; however, it is worth to note that T_p of an $MPDU$ is a fraction of a T_p of a non-fragmented packet, but, we used the same notation for simplicity. In addition, in a real world scenario with a backoff mechanism with random (non-fixed) contention window, the unicast adds huge differences into the delay. Therefore, it is worth to propose a hybrid approach where reliability and speed can be taken into consideration. Therefore, we can consider one example, like using acknowledged broadcast as will be shown in the next section.

9.3.4 Uncoded Broadcast with ACK

Figure 9.4 partially models a broadcast transmission with an ACK at the end of the transmission, such that two states models the framework. One state representing the batch of packets that need to be broadcasted and acknowledged and the other is the absorption state. However, the probability to successfully deliver N packets differs in the ACK component; where acknowledging a batch of N packets will let the probability of success $p_s = (1-p_e^N)(1-p_{ack})$. The steps that have been done to

derive the time to absorption are the same. Therefore, the expected time to deliver N packets using broadcast transmission with ACK and without considering the time required for retransmission is as follows:

$$\mathbb{E}\left[T_{deliver\ N\ packets}\right] = \frac{NT_p + NSIFS + DIFS + CW1 + T_w}{(1 - p_e^N)(1 - p_{ack})} \tag{9.19}$$

Where $T_w = ACK + T_{rt}$ corresponds to the time for acknowledgment and round trip time (RTT).

9.4 Network-Coded Broadcast

Mixing the two approaches of reliable transmission and fast transmission available in uncoded unicast and uncoded broadcast, we can develop a network-coded broadcast model to have coding across packets and an ACK at the end of the N coded packets; such that if an erasure was detected on one or more of the packets we can retransmit the packet(s). In a similar but more efficient approach to a packet repetition framework, we will propose a network-coded approach where coding across packets will be considered. We can transmit equal number of coded packets per transmission with linear combinations between the packets over some defined GF. This way the number of transmissions required to restore more packets at the receiver side is less. For example, if we transmit coded packet $aMPDU_1 \oplus bMPDU_2 \oplus cMPDU_3$ between state 1 and state 2, then we transmit $dMPDU_1 \oplus eMPDU_2 \oplus fMPDU_3$ between state 2 and state 3, this means that we can save one transmission in the uncoded broadcast case such that we can transmit three packets in two transmissions, such that the receiver needs only to decode the three packets solving the two linear equations by a simple Gaussian elimination process. As well, the probability of successful decoding will be equally-likely for all the packets. Therefore, for a network-coded scheme, the probability of successfully transmitting N_c coded (or uncoded) packets is as follows:

$$p_s = \sum_{j=1}^{N_c} \binom{N_c}{j} (1 - p_e)^j p_e^{N_c - j} (1 - p_{ack}) \tag{9.20}$$

In fact, its worth to note that the expected time to deliver N_c coded packets is exactly similar to the one for broadcast with an ACK at the end of N transmitted packets. If we are transmitting N_i coded packets the probability of successful transmission will

be $(1-{p_e}^{N_i})(1-p_{ack})$. However, there are practical considerations that the previous model lacks: On the one hand, the previous model doesn't consider the time required to retransmit the uncoded or coded packets that have not been successfully acknowledged. On the other hand, from a network coding perspective, the previous model doesn't take into consideration the degrees of freedom the receiver owns to be able to decode successfully the received packets. In addition, for a repeated packets framework, the redundancy and resource wasting incurred is huge since there is a possibility to receive and decode successfully all packets from the first few trials. Therefore, we develop the network-coded broadcast model similar to the one introduced in [71] which can be integrated to the DCF functionality of the MAC layer of the IEEE 802.11. Figure 9.5 illustrates a network-coded broadcast through which the transmission process is adaptive to the receiver experience. First, we transmit a linear combination of N_c coded packets, where those packets correspond to a linear combination of the $MPDU_1 \oplus ... \oplus MPDU_c$. If the N_i coded packets are received successfully with probability $(1-{p_e}^{N_i})(1-p_{ack})$, the chain is absorbed. If a failure in transmission occurs at state i, a self transition will occur with probability $1-(1-{p_e}^{N_i})(1-p_{ack})$. If not, a transition will occur with probability $p_{i\to j}$. Thus, if a packet erasure occurs and the receiver received less number of packets; the receiver will send back to the transmitter - via the ACK message - an information about the degrees of freedom it owns such that the chain transition between state i and j corresponds to the number of received packets $i-j$, and the state to where the transition happens corresponds to the coded packets need to be re-transmitted to the receiver, therefore, the transition probability is not fixed over the Markov chain and the probability of success corresponds to the successful transmission of the N_i coded packets at the current state i.

The probability of transition from state i to state j for the Markov chain in Figure 9.5 is given by:

$$p_{i\to j} = \binom{N_i}{i-j}(1-p_e)^{i-j}p_e^{N_i-i+j}(1-p_{ack}) \tag{9.21}$$

And the probability of failure in transmission of N_i coded packets at state i represented by a self loop is given by:

$$p_{i\to i} = (1-p_{ack})p_e^{N_i} + p_{ack} \tag{9.22}$$

Therefore, the expected time to deliver the N_c coded packets for the model in Figure 9.5 can be written as follows:

$$\mathbb{E}\left[T_{deliver\ N_c-packets}\right] = T_i + \frac{\sum_{j=1}^{i-1}\binom{N_i}{i-j}(1-p_e)^{i-j}p_e^{N_i+i-j}T_j}{(1-{p_e}^{N_i})(1-p_{ack})} \tag{9.23}$$

With,

$$T_i = \frac{N_i T_p + CW1 + DIFS + T_w + N_i SIFS}{(1 - {p_e}^{N_i})(1 - p_{ack})}, \tag{9.24}$$

and,

$$T_j = \frac{N_j T_p + CW1 + DIFS + T_w + N_j SIFS}{(1 - {p_e}^{N_j})(1 - p_{ack})} \tag{9.25}$$

Note that if the transmission is established successfully from the first transmission then $N_i = N_c$, if after the second transmission then $N_j + N_i = N_c$, and so on and so forth until the receiver receives and decodes successfully all coded packets[7]. Therefore, if the optimal number of packets were transmitted, we can guarantee the minimum time to deliver the packets; and therefore the maximum throughput received, as we will introduce in the following subsection. Notice also that in the framework of the novel adaptive transmission scheme we have proposed in the previous Chapter, we can optimize the number of coded packets. However, this is given that we have a pre-knowledge - or estimation of the channel coefficients - and therefore, their corresponding erasures. However, in this setup we consider a network with fixed erasure probability.

9.4.1 Maximizing Throughput with Optimal Number of Coded Packets

Our objective is to maximize the throughput $1/\mathbb{E}\left[T_{deliver\ \ N_c - packets}\right]$ over all the number of packets to be transmitted, which is equivalent to minimizing the expected time to deliver the packets we have:

$$\max_{N_1,\ldots,N_c} \frac{1}{\mathbb{E}\left[T_{deliver\ \ N_c\ \ packets}\right]} \tag{9.26}$$

This can be solved by optimizing jointly over all coded packets, in that case the optimal number of packets in each transmission can be any number that maximizes the throughput. Otherwise, this can be solved iteratively such that the optimal N_1 that solves $max_{N_1} 1/T_1$ is used to compute $max_{N_1,N_2} 1/T_2$, and so on and so forth. However, if we constrained the number of coded packets need to be transmitted, lets say $N_1 + N_2 + ... + N_c = N_K$, this will lead to optimal set with optimized K, i.e., with a limited number of re-transmissions to deliver N_c coded packets. Notice that a

[7]Note that the condition $N_j + N_i = N_c$ is sufficient for the transmission of the N_c but not necessary since the station can adaptively account for the erased packets with another packets that matches the degrees of freedom of the receiver.

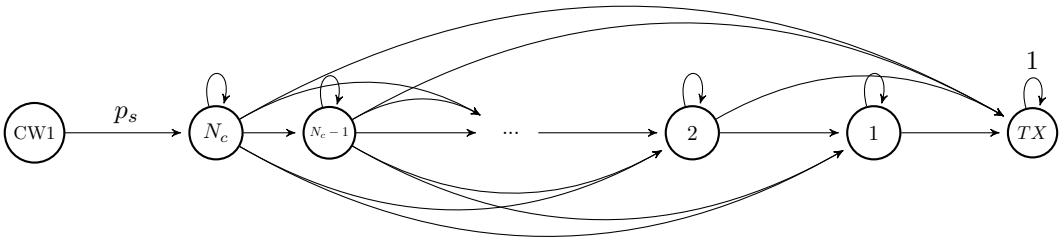

Figure 9.5: N coded packets broadcast transmission.

closed-form solution of the optimal number of packets is not possible to find, so the iterative solution is the only way, were a closed-form expression is only possible for ${N_1}^\star$.

9.5 Slow Fading Handover

One source of degradation of the signal power over the transmission path is due to fading. Fading occurs due to the path loss of the signal as a function of the distance and shadowing. When the mobile moves through a large distance within the coverage of an AP or at the point of handover, this is characterized by a slow fading process with an average path loss between the AP and the station in fixed locations. However, if the station moves through small distances within the coverage of an AP, this is characterized by fast fading process where rapid variations would occur to the signal levels due to interference of multipaths. In fact, we focus here in the slow fading process to our interest in the scenario where the station is approaching the cell edge and just at the time before initiating the process of handover to another AP. In particular, if the time to deliver the N packets to the mobile terminal at such point is related to the SNR and so to the distance, we can characterize an optimal time when to switch to the new AP, and also when the new AP would start to send the mobile the required coded packets to keep its streaming with the least delay effects. In the simple framework of an uncoded modulated signal (i.e., no forward error correction is considered), a detection of one bit error in the received packet - in the MAC level terminology - corresponds to one packet loss. Therefore, we capitalize on the relation between the probability of erasure of a packet p_e and the bit error probability p_b, which is defined

as:

$$p_e = 1 - (1 - p_b)^B \tag{9.27}$$

Where B denotes the length (in bits) of a single packet.

Moreover, the bit error probability p_b is well defined in the literature with respect to the SNR or mainly to the energy per bit to the noise ratio for coherent as well as the non-coherent detection. For the uncoded BPSK modulation with coherent detection and without fading, the bit error probability p_b is given by:

$$p_b = Q(\sqrt{2snr}) \tag{9.28}$$

For the uncoded BPSK with coherent detection under flat (one path) Rayleigh fading with random channel gain $h \sim \mathcal{CN}(0, 1)$, see [139], is given by:

$$p_b = \mathbb{E}\left[Q\left(\sqrt{2\left|h\right|^2 snr}\right)\right] = \frac{1}{2}\left(1 - \sqrt{\frac{snr}{1 + snr}}\right) \tag{9.29}$$

However, for the coded BPSK modulation with coherent detection, the bit error probability p_b depends on the coding, and the error correction capability of the code, as well as the type of decoding with soft or hard decision. Several upper bounds have been presented for the bit error probability for convolutional codes in [140], [141], [142], [104], and [143]. However, we used the upper bound of [140] since it is more consistent than other bounds. In the case of IEEE 802.11, where 1/2 rate convolutional codes are used, the pairwise error probability $p_2(\delta)$ for coherent detection without fading is given by:

$$p_2(\delta) = Q(\sqrt{2\delta r_c snr}) \tag{9.30}$$

For the convolutional coded BPSK with coherent detection and under flat (one path) Rayleigh fading with random channel gain, see [140], is given by:

$$p_2(\delta) = \int_h p_2(\delta|h)p(h)dh = \mathbb{E}\left[Q\left(\sqrt{2\left|h\right|^2 \delta snr}\right)\right] = \frac{1}{2}\left(1 - \sqrt{\frac{\delta snr}{1 + \delta snr}}\right) \tag{9.31}$$

Where δ is the hamming distance of the convolutional code and $r_c = k_c/n_c$ is the code rate. Therefore, the bit error probability is upper bounded as in [140]:

$$p_b \leq \sum_{\delta = d_{free}}^{\infty} \frac{c(\delta)}{k_c} \int_h p_2(\delta|h)p(h)dh = \sum_{\delta = d_{free}}^{\infty} \frac{c(\delta)}{k_c} p_2(\delta) \tag{9.32}$$

With,

$$c(\delta) = \sum_{w_i=1}^{\infty} w_i a_i(\delta), \tag{9.33}$$

and,

$$a(\delta) = \sum_{w_i=1}^{\infty} a_i(\delta) \tag{9.34}$$

And $p(h)$ is the probability density function of the Rayleigh fading channel distribution. The distance spectrum of a convolutional code is defined by $c(\delta)$ which corresponds to the error weight in information bits, and $a_i(\delta)$ is the number of error events with length δ, and w_i bit errors. Note that the free distance d_{free}[8] provides a first order asymptotic approximation of the error performance. Therefore, as much as we sum over higher order terms, the upper bound would approach the error performance of the uncoded case, and as the constraint length is higher for good convolutional codes, the number of possible codewords grows exponentially, therefore, its somehow sufficient to sum over terms up to $d_{free}+K$, where K is the constraint length, and this would be a condition of truncation of a path in the decoding process, see [141], and [144]. In the case of IEEE 802.11, where $r_c = 1/2$ rate convolutional code with generator matrix [133, 171] in octets, the minimum hamming weight of the codewords is $\delta = d_{free} = 10$, and the constraint length $K = 7$ corresponding to $2^{K-1} = 64$ states, see [145], the distance spectrum is as follows, $a(\delta) = [11, 0, 38, 0, 193, 0, 1331, 0, ...]$ and $c(\delta) = [36, 0, 211, 0, 1404, 0, 11633, 0, ...]$, notice that the vector starts at $\delta = d_{free}$, see Table 1 in [141]. Notice also that the meaning of such setup leads to a formulation of a transfer function of the convolutional code; that presents that at $\delta = d_{free}$, 11 error events of weight 36 occurs and at $\delta = d_{free} + 1$, 0 error events of weight 0 occurs, and so on and so forth. This can be easily found by Matlab built-in function (distspec).

9.6 Free Path Loss

The average received signal to noise power ratio $snr = P_r/N_0$ per symbol, and $Q(.)$ is the complementary cumulative distribution function of a Gaussian random variable. Meanwhile, we will use the free-space propagation model to relate the distance of the mobile to the AP. Let d denote the distance in meters between the AP and the station. η is the path loss exponent and $\eta = 2$ for free space; i.e., it is environment dependent.

[8]The free distance d_{free} is the minimal hamming distance between different encoded sequences, through which the correcting capability of a convolutional code t is upper bounded by $\frac{d_{free}-1}{2}$.

In fact, free space loss is a simple model of propagation. Moreover, it is a convenient choice to consider the channel between the AP and the station as quasi-static Rayleigh fading channel in a handover scenario where the station is approaching the edge of the AP coverage. This type of channel exhibits slow fading and so the fading coefficients remain constant during the transmission of the entire N packets, while changes randomly and independently between different transmission/retransmission according to a complex Gaussian distribution with variance equals to the $snr = cd^{-2}$, therefore, $h \sim \mathcal{CN}(0, cd^{-2})$. c corresponds to a constant that can be chosen to maintain a given snr(dB) at a given distance. The expected time to deliver the N packets for the unicast, broadcast without ACK, and broadcast with ACK transmission modes under the previous assumptions can be directly derived with respect to the SNR and/or the distance by substituting into the probability of erasure. In addition, it is straightforward to relate the time to deliver the N packets and the throughput to the mobile station velocity via the basic distance-velocity-time relation; $d = tv$, so if we know that the mobile station is moving with velocity v meters/sec, we will know that at time t, it will be at location d, which means that the expected time required to deliver N packets for the station at this location is T_{AP1} and so, we can search for the optimal number of packets to transmit before switching to the other AP, as well as the prospective new AP can compare its T_{AP2}. Therefore, both APs can minimize the time the station requires for the scanning in the handover process by establishing the connection and transmitting cooperatively the coded packets required (cashed) and/or the native (uncoded) ones.

9.7 Network-Coded Handover

In this section we will propose an optimal handover scheme based on reliable broadcast with network coding. Optimal handover decision is usually based on the signal strength of the AP, however at the border contours of the different levels of the received signal to noise ratio where the decision can only be taken based on the path loss, we cannot guarantee that the station's service will stay with the same quality. Therefore, introducing another objective is worth to think about. We will introduce a framework that decides when and where is the optimal point(s) to do a network-coded handover. We need to optimize the handover decision to guarantee maximum received sum rates from AP1 the station is already accessing to AP2 the one the mobile station will handover to. Suppose that probing, authentication, and re-association times can be

minimized, see [84], [85] where the authors of the first showed that probing contribute to the main delay in a handover process, while the later proposed a selective scanning and cashing mechanism to reduce the probing delay. This is in fact a forthcoming result to a decision based on the maximum sum rate objective, therefore, the mobile station can be at anytime receiving from AP1 or AP2 before its fully served by the AP its fully associated with and under its coverage far from the border. This is a relevant assumption due to the fact that the station can experience ups and downs in the received SNR from different APs due to mobility and distance changes, while probing is done via broadcast messages which completes the setup introduced. The objective will be formulated as a LP to maximize the rate cost functions to find the optimal switching time as follows:

$$max_t \int_{t1}^{t} R_1(x(t), y(t))\,\mathrm{d}t + \int_{t+\tau}^{t2} R_2(x(t), y(t))\,\mathrm{d}t \tag{9.35}$$

With,

$$R_1 = \frac{1}{\mathbb{E}\left[T_{deliver\ N_c\ packets\ AP1}\right]}, \tag{9.36}$$

and,

$$R_2 = \frac{1}{\mathbb{E}\left[T_{deliver\ N_c\ packets\ AP2}\right]} \tag{9.37}$$

And $x(t), y(t)$ correspond to the coordinates of the station at a given time. And t_1 is the initial time at the beginning of the mobile station path, τ is the time to re-associate to AP2, and t_2 is the time of measurement at the end of the mobile station path.

9.7.1 Station Associated to Both APs

Worth to notice that we are interested to find optimal time $t^\star$ when the sum-rate can be maximized at which both stations can start to code across the packets via a broadcast transmission mode at the same time to let the handover process being performed with no disturbance of the user service, and the user will always find the required packet flows to decode, cashing can also be useful in implementing the setup introduced here to reduce the time of probing and full re-association. In particular, the network-coded handover introduced include the following steps:

1. When the signal strength received at a certain point is almost equivalent from both APs, the station starts sending probe messages to neighbor access points, and through cashing this process delay will be minimal, therefore, it will find AP2 as the first option in the neighboring list.

2. The station will transmit to both APs via ACK messages the degrees of freedom it requires to continue the service without service disruption.

3. AP1 will activate network-coded broadcast and AP2 will authenticate and associate the station before it fully disconnects from the previous AP1 and directly sends the network-coded broadcast data required.

4. The station will continue to acknowledge its degrees of freedom to both APs, and by the mobility considerations, the station will be reassociated to AP2 while still receiving its coded-packets, when the association is completed to AP2. The station is not anymore connected to AP1.

Therefore, we introduce a framework that would allow the station to be connected to two stations at a certain point receiving from both of them. The station will be announced at the optimal distance d and so at the optimal time $t^\star$ to receive coded broadcast before it fully re-associate to the new AP, and so the network-coded handover will be reliably performed over broadcast while the station is moving; guaranteeing a near optimal packet flow without disruptions. In fact, the optimal time to switch to network-coding broadcast mode is an optimal set of points starting few meters back and few meters forth in the borders of the SNR contours of both APs coverage, which are also path dependent. In other words, this point in time can be moved back if AP cooperation is implemented over backhaul link - with CSI and data sharing -; i.e., similar to the framework introduced in Chapter 5, while the exact optimal time when to switch to the new AP can be found numerically. Such framework is practically feasible in FiWi networks with optical fiber backhaul connecting the two APs to a central unit. Notice also that, one possible optimal point to start the transmission of coded packets from AP1 is at the distance corresponding to $d(t^\star) - d(\tau)$ meters from the optimal distance to switch to AP2, $d(t^\star)$, where by the time $t^\star$, the station will switch to AP2, receiving the required/lost DoF or innovative (new) packets.

9.7.2 Station Associated to One AP at a Time

Here we consider that the path of the mobile station is known before hand, and so, given the location of the station $(x(t), y(t))$ at a given time, we can measure the distance from AP1 located at (x_1, y_1) to the mobile station as follows:

$$d_1(t) = \sqrt{(x_1 - x(t))^2 + (y_1 - y(t))^2} \tag{9.38}$$

Similarly, we can measure the distance from AP2 located at (x_2, y_2) to the mobile station as follows:

$$d_2(t) = \sqrt{(x_2 - x(t))^2 + (y_2 - y(t))^2} \tag{9.39}$$

Substituting the distances d_1 and d_2 of each AP from the mobile station into the objective that aims to maximize the sum rate received by the station, we can find the optimal set of distances where both APs can activate the network-coded mode of operation during which a soft handover process can take place seamlessly with least interruption over the service. It is not possible to find a closed-form expression for the optimal times to start coded packets transmission, or the optimal time $t^\star = d^\star/v$ to switch to the new AP.

Therefore, we can consider R_1 and R_2 using Shannon's capacity formula, this will provide more feasible way to find the optimal time to switch to the new AP, therefore, we can re-write the objective function as follows:

$$max \int_{t_1}^{t} log(1 + snr(x(t), y(t)))\,\mathrm{d}t + \int_{t+\tau}^{t_2} log(1 + snr(x(t), y(t)))\,\mathrm{d}t \tag{9.40}$$

Deriving the integral of the cost function with respect to the distance and plugging the distance-time-velocity relation into the result, the objective becomes as follows:

$$\begin{aligned} max \frac{vt_1}{2} ln(1 + \frac{1}{t_1^2 v^2}) + \frac{1}{2vt_1} ln(1 + t_1^2 v^2) - \frac{1}{2vt_1} \\ - \frac{vt}{2} ln(1 + \frac{1}{t^2 v^2}) - \frac{1}{2vt} ln(1 + t^2 v^2) + \frac{1}{2vt} \\ + \frac{v(t+\tau)}{2} ln(1 + \frac{1}{(t+\tau)^2 v^2}) + \frac{1}{2v(t+\tau)} ln(1 + (t+\tau)^2 v^2) - \frac{1}{2v(t+\tau)} \\ - \frac{vt_2}{2} ln(1 + \frac{1}{t_2^2 v^2}) - \frac{1}{2vt_2} ln(1 + t_2^2 v^2) + \frac{1}{2vt_2} \end{aligned} \tag{9.41}$$

The derivative of the integral form with respect to the distance, and in terms of time-velocity, is as follows:

$$\begin{aligned}
&\frac{v}{2}ln(1+\frac{1}{{t_1}^2v^2})-\frac{1}{2v{t_1}^2}ln(1+{t_1}^2v^2)+\frac{1}{2v{t_1}^2}\\
&\qquad-\frac{v}{2}ln(1+\frac{1}{t^2v^2})+\frac{1}{2vt^2}ln(1+t^2v^2)-\frac{1}{2vt^2}\\
&+\frac{v}{2}ln(1+\frac{1}{(t+\tau)^2v^2})-\frac{1}{2v(t+\tau)^2}ln(1+(t+\tau)^2v^2)+\frac{1}{2v(t+\tau)^2}\\
&\qquad-\frac{v}{2}ln(1+\frac{1}{{t_2}^2v^2})+\frac{1}{2v{t_2}^2}ln(1+{t_2}^2v^2)-\frac{1}{2v{t_2}^2}\\
&\qquad=0
\end{aligned} \tag{9.42}$$

Therefore, the optimal time to switch $t^\star$ is the numerical solution of the equation above. However, we can account for more delays via AP cooperation and coding across packets.

9.8 Simulation Results

The expected time to deliver one or N packets derived in the sequel of the previous sections can be adapted to different WiFi IEEE 802.11 systems. Table 2 provides a summary of the approved timing parameters in the specifications of the MAC-layer which decomposes the packets. The following simulations focus on the values of the IEEE 802.11g with legacy, using $T_p = 0.00144$ *sec*.

Figure 9.6 illustrates the time to deliver N packets by unicast, broadcast, and broadcast with ACK with respect to the probability of erasure. It is clear that the least delay incurred by using a broadcast mode under any number of coded or uncoded packets. Consequently, we can see in Figure 9.7 that the Throughput(packets/sec) is higher for the broadcast transmission mode than for the unicast while broadcast with ACK is a reliable alternative to the one without ACK. The previous results are emphasized through the relation between the packet error rate and the bit error rate, which makes it possible to derive the time to deliver N packets and the throughput with respect to the SNR and so with respect to the distance of the mobile station.

Figure 9.8 illustrates the time to deliver the N packets using different transmission modes with respect to the distance, and Figure 9.9 illustrates the Throughput using different transmission modes with respect to the distance from both AP1 and AP2. For

the sake of clarity, we suppose that the distance between AP1 and AP2 is 60 meters, i.e., AP1 is located at (0,0) and AP2 at (0,60). The negative distances correspond to the distances between the station and the access point at the other side of the coordinate. At a distance $d = 1$ meters for an indoor scenario, the SNR maintained is 64.124 dB, at a distance of $d = 32$ meters indoor or outdoor ($1000 > d > 100$) the maintained SNR is 42.11 dB. Therefore, we can chose $c = 100000$ which is an acceptable figure when a fading scenario is assumed.

We can see that the broadcast transmission always outperform the unicast one. However, we can also see the gain of introducing coding across the packets where the transmission of network-coded packets introduce a gain at earlier distances and with proper tuning to the system parameters, the throughput via network coding can approach the theoretical capacity of broadcast channels. Proper tuning means that we can limit the number of retransmissions of a lost packet, this is of practical relevance since the packets have a timeout that afterwards its worthless for the receiver.

To illustrate the optimal decision set of a network-coded handover corresponding to the optimal time or distance at which the mobile station can receive coded packets from both current and prospective AP, and during the handover process. Suppose that AP1 is located at $(0, 0)$ and AP2 at $(25, 0)$, and the path the mobile station is moving on is deterministic and given by a linear movement with $y(t) = x(t)$, the maximum of the sum rates can be received by the mobile at a specific (or several) points in time.

However, due to the fact that the optimal time is not only distance but also velocity dependent $t^\star = d/v$; if the mobile is moving with constant speed equals $v = 2$ meters/sec and the delays of switching to AP2 is $\tau = 10sec.$ $t^\star \in [t_1 : t_2]$ corresponds to the optimal time to switch $t^\star = 11.5$ sec which corresponds to time when the station is associated to AP2.

This means that the optimal distance at which its optimal to handoff to the new AP is equal to $d_2(t^\star) = t^\star v = 23$ meters; which is path and velocity dependent. This also defines a possible point in time where AP1 can start transmission of its coded packets, this may be possible at $d(t^\star) - d(\tau) = 3$ meters before the new association takes place which corresponds to $d_1(t) = 23 - 3 = 20$ meters.

If the path of travel of the mobile station is partially or totally unknown, location estimation techniques and some mobility models can be used instead, depending on the network topology under study.

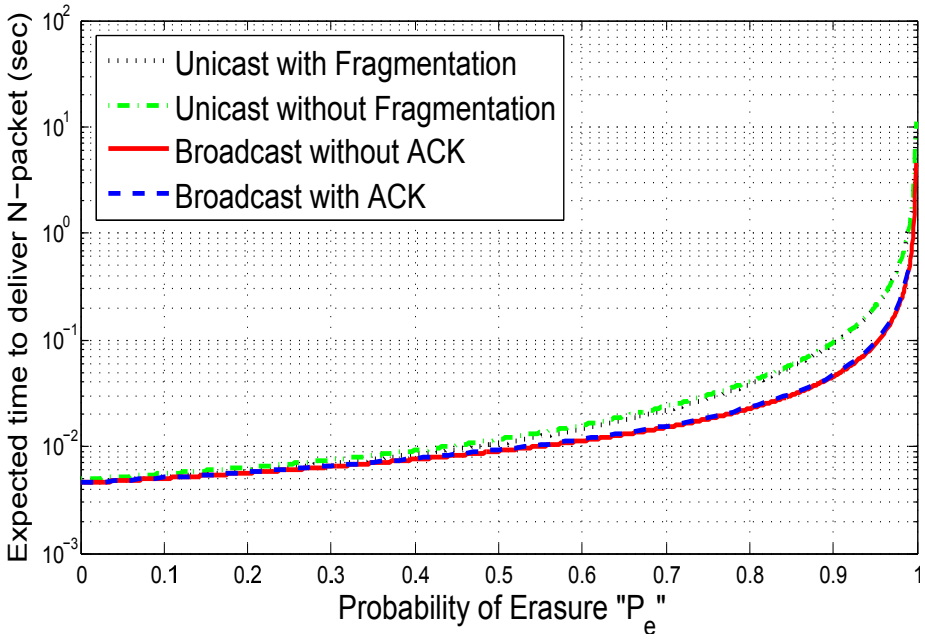

Figure 9.6: Time to deliver N packets(sec) vs. probability of erasure.

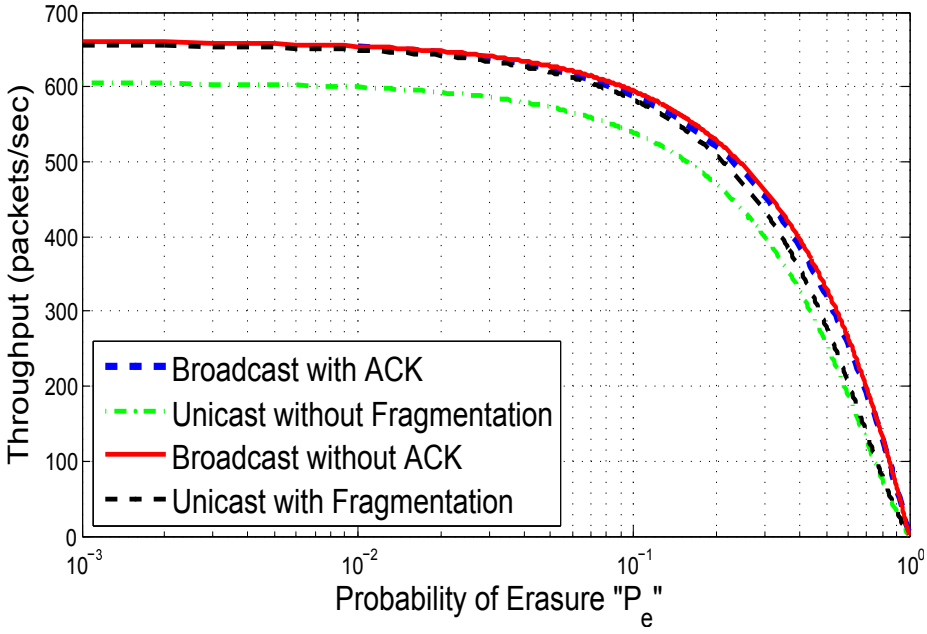

Figure 9.7: Throughput(packets/sec) vs. probability of erasure.

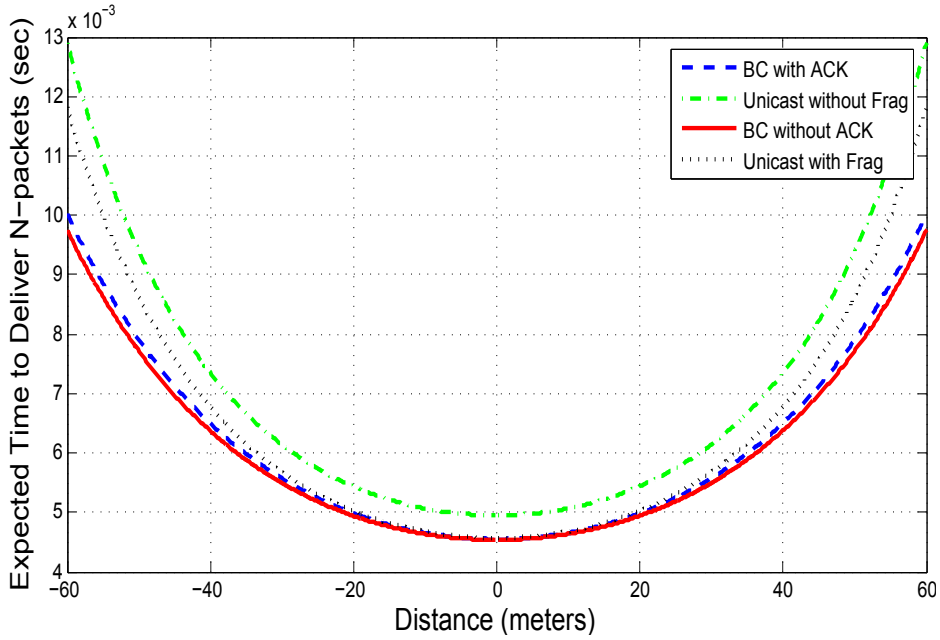

Figure 9.8: Time to deliver N packets(sec) vs. distance(meters).

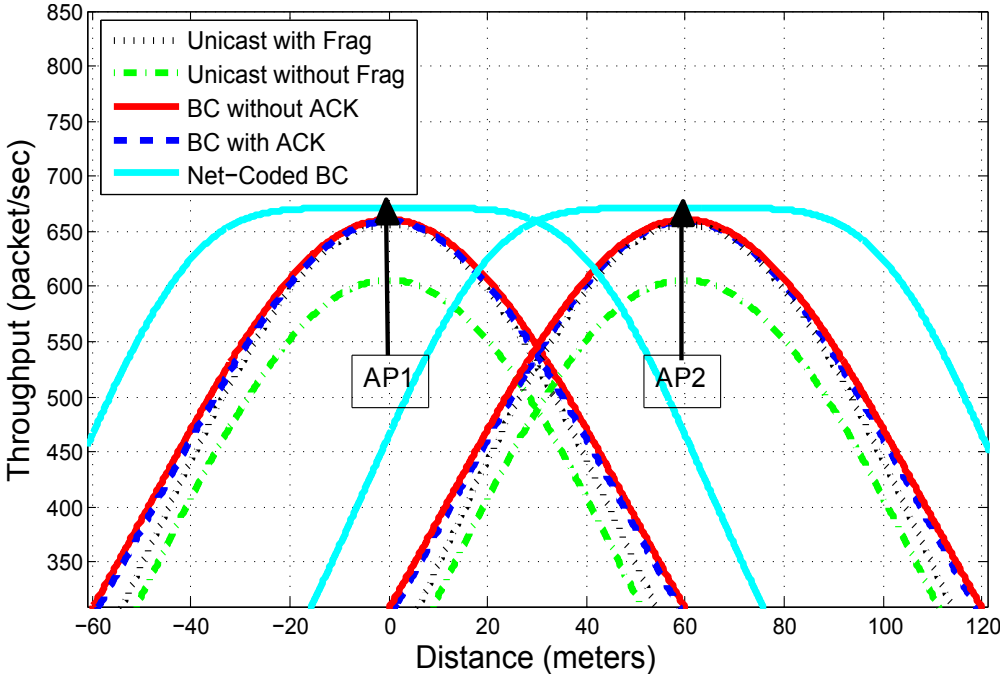

Figure 9.9: Throughput(packets/sec) vs. distance(meters).

Table 2: Timing parameters of the IEEE 802.11 MAC layer

Parameter	802.11b	802.11a	802.11g	802.11g+legacy
T_{slot}	$20\mu s$	$9\mu s$	$9\mu s$	$20\mu s$
$SIFS$	$10\mu s$	$16\mu s$	$10\mu s$	$10\mu s$
$DIFS$	$50\mu s$	$34\mu s$	$28\mu s$	$50\mu s$
ACK	$14Bytes$	$14Bytes$	$14Bytes$	$14Bytes$
CW_{min}	$31T_{slot}$	$15T_{slot}$	$15T_{slot}$	$15T_{slot}$
CW_{max}	$1023T_{slot}$	$1023T_{slot}$	$1023T_{slot}$	$1023T_{slot}$

9.9 Conclusion

In this Chapter, we propose a set of novel models of the DCF of the WiFi IEEE 802.11 with fixed time contention window for unicast transmission, with and without fragmentation, and for the uncoded and coded broadcast transmission, with and without acknowledgment. We analyze the delay over the unicast and broadcast transmission for a network topology that includes one AP and one station. We provide a closed-form expression for the expected time to deliver the N packets for the DCF mechanism, with unicast, the general broadcast, broadcast with ACK for the uncoded and coded transmission. We have shown that coding across packets in an acknowledged broadcast scenario encounters less delays, higher reliability, and higher throughout than for the uncoded broadcast or unicast cases. We propose a new protocol that utilizes network coding to broadcast coded packets to the station performing handover, this new proposed network-coded handover framework will immensely serve if implemented in the current standardized IEEE 802.11 systems. We build upon constraints that take into consideration the distance of the station and the degrees of freedom it owns to decode the received packets before it switches the connection to the new AP. This has been demonstrated by a novel mathematical formulation of a network-coded handover that optimally decides the optimal time when to switch to the new access point in order to maximize the sum rates received by both stations. Particularly, the optimal time to do a handover between two APs accounts for the delay penalty due to probing, association and authentication to the new AP. In addition, we propose a framework were the mobile can be associated to two APs simultaneously and so coding across

packets can be cooperated between the two APs to keep the stream of native (uncoded) packets and account for lost ones. Therefore, we provide a framework under which the QoS over delay sensitive streaming applications can be radically improved, providing a seamless handover. This in conjunction with the techniques proposed underlies a proposal on adding a network coding layer on top of the classical MAC layer which can improve the current standardized IEEE 802.11 and boost its performance from a delay and throughput perspective. This will be of particular relevance in delay sensitive applications if implemented in existing technologies. Worth to note that it is intuitive that the addition of network coding can boost the data rates, and minimize the delay. However, NC adds a decoding complexity if we consider a large set of users with high mobility. In such scenarios, a preservative approach can be thought of, in order to provide an optimal decision of when to switch on/off the network coding. This can be applied to mobile nodes based on their SNR levels, whether they are performing a handover or just losing the quality due to mobility or obstructions. Nevertheless, its instrumental to think of a NC layer as an enhancement layer that can be, adapted to different set of parameters, like the physical parameters of the communication channel, or the number of mobile stations associated to the APs. In addition, a NC layer can be adapted to solve a set of problems, like performing a seamless handover, coping with unbalanced data rate users demands, or solving the issues encountered by mobile stations experiencing poor coverage. In principle, a NC layer is worth to be implemented and verified in current technologies, with a centralized control mechanism via the APs on its adaptation, or with distributed control mechanisms via associated stations, i.e., like having on demand service. Future work will consider the multi-station modeling problem, the speed of the mobile station, and with mobility patterns or unknown paths of the mobile station.

Chapter 10
Concluding Remarks

In this thesis we addressed two fundamental issues in the current communication systems. In particular, the increased demand for higher data rates and stringent delay requirements. We utilized a set of tools to address each problem. The problem of increasing the data rates have been addressed from a physical layer approach to design, analyze and optimize the optimal power allocation and optimal precoding that can maximize the achievable data rates experienced by different types of multi-user multi-applications channels like the MAC Gaussian and MAC Poisson channels, that each of which models the medium without or with optical modulation techniques, respectively. The major contribution in this setup is a characterization of optimal designs in fixed point equations and an extension of the fundamental relation between the mutual information and the MMSE for multi-user channels. This was particularly the focus of Chapter 4, Chapter 6 and Chapter 7.

Moreover, the thesis contributed in novel formulations exploiting cooperative frameworks between cluster of base stations to provide a virtual MIMO setup that naturally increases the spectral efficiency of the network. This thesis also presents another framework with minimal cooperation, with a MAC setup that yet increases the spectral efficiency, taking into consideration the tradeoff between processing information, congestion, and imperfections in the delayed versions of the channel estimates, between the receiver and transmitter. Therefore, despite the fundamental contributions of the theoretical basis of this thesis, providing a characterization of the optimal precoder at the high-snr which is a novel solution of an NP-hard problem in the literature. However, we contribute from a practical perspective into the concept of timely designs through timely estimates, where we proposed a transmitter side channel estimation process that takes the pilot assisted channel estimates feedback from the receiver to auto-regressively oversee the future channel variation over time due to the channel distribution and correlation in a block by block basis, where typically, each OFDM

block is usually considered to be under a flat fading which is yet not necessarily real in practice, this was particularly the focus of Chapter 5, and Chapter 6.

Furthermore, the thesis tackles the problem of delay which is a fundamental issue in current systems especially with the increased demands of high data rates, and under the mobility of nodes. Therefore, we studied the network coding concept and how much it can offer from a data rate perspective and a delay perspective. We have addressed two scenarios in networks which are of particular importance: One is the coded transmission over time varying channels which are delay challenging, that is, we focus on the satellite channel as an example. The other is the coded transmission over fixed erasure channels, we focus on a handover scenario in WiFi networks. We had to do an in depth study of the existing standardization and the state of the art works. Therefore, we provide novel models underlying the time variation during the transmission of packets over channels with variable erasures due to fading as in real life scenarios. We approximated the delay and provided novel adaptive transmission scheme that outperform state of the art schemes. However, in the WiFi network-coded handover, we proposed novel models of the typical DCF, we find closed-form expressions of the delay, and compare it to the delay encountered using network-coded broadcast with ACK. We proposed a network coding protocol that can be added on top of the handover functionality of the WiFi IEEE 802.11 to provide higher performance and non-disruptive service of the mobile station performing handover. This was particularly the focus of Chapter 8 and Chapter 9.

Therefore, we consider two main problems of study: First is the design of optimal transmit filters (precoders) that maximize the data rates in future telecommunication systems. We solve a set of scenarios underlining this framework with an information-theoretic estimation-theoretic approach. Second is the modeling and design of transmission schemes that minimize the delay in future telecommunication systems. We study the role of network coding in solving the problem, and we solve a set of scenarios underlying this framework using communication theoretic and probabilistic approaches.

The scenarios and the solutions provided in this thesis work build a novel framework for future communications systems with multi-layer paradigms - data rates and delay - which are also applicable to current communications systems. This demonstrates that exploiting the concepts of cooperation via transmit diversity and virtualizing the networks can move networks into upper bounds despite the fundamental limits of cooperation. Therefore, cooperation in small size clustered network frameworks from one side can boost network performance. However, from another side, demonstrates

that adaptivity and feedback can have a dramatic effect on the data rates when transmitter adapt to the channel experience. This introduce a significant effect on time variant channels with high error rates. This also build upon new schemes that can work in conjunction with coded networks to boost the network throughput. Moreover, the network coding layer on top of the multiple access mechanisms can be of particular relevance to reducing the delays in existing systems. This build the major contributions of the thesis work and provides a solid theoretical framework, as well as a practical framework through new schemes that can be introduced into current existing systems to be verified and validated.

This thesis through analytical work, design proposals, and future recommendations establishes the fact that future high rate delay tolerant telecommunications systems design should consider transmission and coding strategies that are physical layer aware and topology aware.

"Problems cannot be solved at the same level of awareness that created them"

Albert Einstein

The contributions of this thesis can be divided into six main parts:

The first part considers building connections between information-theoretic estimation-theoretic fundamental limits for multi-user channels. In particular, we devise a new relation between the mutual information and the minimum mean-squared error for multiple access channels. We devise optimal transmission strategies for the MAC Gaussian channel capitalizing on this relation. We provide a generalized closed-form expression for the optimal transmit filters (precoders) and optimal power allocation for the arbitrary inputs, we also specialize it for Gaussian inputs. We analyzed the characterization of the optimal transmission strategies, which are specialized for key asymptotic regimes. We provide an interpretation for the interference with respect to the channel, power, and input estimates of the main user and the interferer.

The second part introduces a multi-cell processing framework with data and CSI sharing among a cluster of base stations connected via a backhaul link. The scenario considers a network virtual MIMO system. We devise the optimal power allocation and optimal precoding for a cluster of two base stations which cooperate to jointly maximize the achievable rate for two users connecting to each BS. We provide a

generalized fixed point equation of the optimal precoder in the asymptotic regimes of the low- and high-snr. We introduce a new iterative approach that leads to a closed-form expression for the optimal precoding matrix in the high-snr regime which is known to be an NP-hard problem. We introduce MCP distributed algorithms; a power allocation algorithm for the uplink, and a precoding algorithm for the downlink.

The third part builds upon the previous contributions to introduce a multi-cell processing framework with minimal cooperation among the cooperating cluster of BSs. Through which the BSs only share their CSI. The scenario considered a MAC Gaussian channel in the presence of Rayleigh fading and pilot assisted channel estimation. We provide a generalized fixed point equation of the optimal precoder with respect to the estimated channel, power, and the MMSE. We provide the optimal power allocation with respect to the estimated channel and MMSE. We basically proposed a two way channel estimation process at the transmitter and receiver sides that allows timely designs. The designs introduced are optimal for multiple access (MAC) Gaussian coherent time-varying fading channels with arbitrary inputs and can be specialized to MIMO channels by decoding interference.

The forth part considers MAC channels with Poisson distribution. This scenario is relevant for channels with Poisson arrivals like in optical communications. we derive closed-form expressions for the capacity of the multiple access Poisson channel under the assumption of constant shot noise. We provide an empirical form of the k-users MAC Poisson channel capacity with average powers that are not necessarily equal. We devise optimal power allocation for this setup and compare it to the Gaussian setup, in respect to power allocation interpretation, and to the interplay between the mutual information and the MMSE.

The fifth part proposes a random linear network coding scheme for reliable communications over time varying channels. We consider small scale log-normal fading channel over Ka-band. We propose a novel model that exploits the channel delay profile and the dependency between channel states for satellite communications. We provide an approximation of the delay induced assuming finite number of time slots to transmit a given number of packets. We also propose a novel network-coding transmission strategy that can be employed to compensate for the lost degrees of freedom by tracking the packet erasures over time. This novel scheme is of practical relevance and can be implemented in existing systems validating the theoretical results which proved that this scheme outperforms schemes like selective repeat with ARQ.

The sixth part proposes a new network coding layer on top of the MAC layer of

WiFi IEEE 802.11 systems. We propose new models for the distributed coordination function. We analyze the delay over the unicast and broadcast transmission for a network topology with one mobile station. We propose a new network-coded protocol that utilizes network coding to broadcast coded packets to the station performing handover. This new proposed network-coded handover framework will immensely serve if implemented in the current standardized IEEE 802.11 systems. We build upon constraints that takes into consideration the distance of the station and the degrees of freedom it owns to decode the received packets before it switches the connection to the next AP. Therefore, we provide a framework under which the quality of service over delay sensitive streaming applications can be radically improved.

It is of particular relevance to consider the proposed schemes in this work to be evaluated in the current standardized system, and this will be the first recommendation to be considered for future consideration. The validation of the two way channel estimation, in conjunction with timely precoding and power allocation designs, the clustered cooperation under CSI and/or data sharing, the new adaptive network coding scheme for time variant channels over satellite or even other wireless medium, in addition to network coding handover over WiFi must be validated and verified in current systems to serve the enhancement of future communications systems.

10.1 Recommendations for Future Research

Based on the experience and knowledge acquired during this research work, several areas for future research can be suggested. These are:

- First: We have addressed the interplay between information-theoretic measures and estimation-theoretic measures in the MAC and MIMO Gaussian channels where we provided fundamental relation between the mutual information and the MMSE and we characterize the optimal precoding and optimal power allocation. However, it is of particular interest to test our designs in conjunction with error correcting codes like Turbo and LDPC codes. Moreover, it would be relevant to revisit the fundamental relations between information measures and estimation measures for other types of multi-user channels like the multi-user arbitrary channel. From another perspective, a future research work would consider the physical layer security of such type of channels where characterizations of optimal transmit strategies can be defined from a secrecy perspective.

- Second: One future research question would consider finding explicit expressions of the mutual information for different types of signaling and different types of Gaussian channel models driven by arbitrary inputs. Recent trials approaches a closed-form expression of the mutual information for BPSK input and their MMSE counter part. Therefore, it is of particular relevance to revisit the problem by an information-theoretic estimation-theoretic means or by geometric means.

- Third: The interplay between information-theoretic and estimation-theoretic measures is well established for Gaussian channels with solutions provided for the optimal power allocation like the waterfilling and mercury waterfilling, while we also provided in this thesis a quantified term for the mutual interference caused by the per user sub-channels and the interference from other users channels/sub-channels. However, an interpretation that joins channel coding as well as source coding - from a rate distortion perspective - will be of particular interest to investigate.

- Forth: We addressed network coding for time varying channels with a line network, and we consider the satellite channel as an example. However, its of particular interest to revisit the problem for scenarios of the wireless medium where muticast occurs and different propagation delays exist. However, we consider a log-normal fading distribution of the channel, therefore, its of interest to study the scenario with different types of fading distributions like Rayleigh, Ricean, etc. In addition, its of particular relevance to revisit the problem from an energy perspective, where we can consider scenarios of fixed power or adaptive power transmission policies.

- Fifth: We proposed a network-coded handover and an optimal time to switch from one AP to another utilizing the robustness of network coding. Our proposal of an enhanced soft handover using network coding provides a solution to different addressed issues like the decision on when to switch to the new AP based on a new design criterion. Despite that our solution gives intuitions on scenarios that can utilize network coding, but when network coding is required is another open question to validate, more specifically it is of interest to be able to define under which optimality conditions the network coding part can be switched on/off. However, this opens other questions of future research interest like, how many original packets to code over, and how much redundancy is required from network coding to assure that all information is recovered with high probability. Moreover, future work on the network-coded handover would consider verifying

the setups in an experimental framework, with multiple mobile stations, and with deterministic, random, predictive, and realistic mobility patterns.

- Sixth: One of the main tools we used in this thesis is the interplay between information theory and estimation theory at the physical layer. However, finding connections between information theory and estimation theory on a network level is of particular importance. On the network level, extracting such connections will enforce a physical layer view into the network taking into consideration that the noiseless assumption is not anymore interesting. It is also of interest to incorporate the effect of the existence of a network topology exploiting a structure that contains the channel, the precoding matrix, as well as the topology. The derivations can be analyzed for scenarios where the nodes are mobile with a sample of known topologies. This may also be applied to vehicular networks with mobile nodes, where we can build optimal designs based on the relations derived, that are topology-aware and physical layer aware.

- Finally, the previous recommendations for future research work shed light on future options within the scope of the contributions of this PhD thesis. However, a lot of problems are still wide open for future research in the practical as well as the theoretical levels which can use similar tools. Compressed sensing and cognitive radios are examples on applications that can be inspired through the frameworks presented.

References

[1] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Transactions on Communications*, vol. 44, no. 10, pp. 1261–1271, 1996.

[2] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 399–431, 1999.

[3] C.-C. JayKuo and S. H. Tsai and L. Tadjpour and Yu-Hao Chang, *Precoding Techniques for Digital Communication Systems.* Springer, 2008.

[4] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.

[5] V. Tarokh, A. Naguib, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communication: performance criteria in the presence of channel estimation errors, mobility, and multiple paths," *IEEE Transactions on Communications*, vol. 47, no. 2, pp. 199–207, Feb. 1999.

[6] A. Naguib, V. Tarokh, N. Seshadri, and A. Calderbank, "A space-time coding modem for high-data-rate wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1459–1478, 1998.

[7] V. Tarokh, A. Naguib, N. Seshadri, and A. Calderbank, "Combined array processing and space-time coding," *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1121–1128, 1999.

[8] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi element antennas," *Bell Labs. Technical Journal*, vol. 1, no. 2, 1996.

[9] B. Hochwald and T. Marzetta, "Unitary space-time modulation for multiple-antenna communications in rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 543–564, 2000.

[10] W. Keusgen, "On limits of wireless communications when using multiple dual-polarized antennas," in *10th International Conference on Telecommunications (ICT)*, vol. 1, 2003, pp. 204–210.

[11] G. Raleigh and J. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Transactions on Communications*, vol. 46, no. 3, pp. 357–366, 1998.

[12] G. Raleigh and V. K. Jones, "Multivariate modulation and coding for wireless communication," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 5, pp. 851–866, 1999.

[13] L. H. Brandenburg and A. D. Wyner, "Capacity of the Gaussian channel with memory: The multivariate case," *Bell Labs. Technical Journal*, vol. 53, 1974.

[14] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input-multiple-output (mimo) transmission systems," *IEEE Transactions on Communications*, vol. 42, no. 12, pp. 3221–3231, 1994.

[15] J. Salz, "Digital transmission over cross-coupled linear channels," *AT&T Technical Journal*, vol. 64, pp. 1147–1159, July 1985.

[16] H. Malvar and D. Staelin, "Optimal pre- and postfilters for multichannel signal processing," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 2, pp. 287–289, 1988.

[17] D. P. Palomar and J. M. Cioffi and M. A. Lagunas, "Joint tx-rx beamforming design for multicarrier mimo channels: A unified framework for convex optimization," *IEEE Transactions on Signal Processing*, vol. 51, no. 9, pp. 2381–2401, September 2003.

[18] A. Scaglione, P. Stoica, S. Barbarossa, G. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Transactions on Signal Processing*, vol. 50, no. 5, pp. 1051–1064, 2002.

[19] R. F. Fischer, C. Stierstorfer, and C. Windpassinger, "Precoding and signal shaping for transmission over mimo channels," *Canadian Workshop on Information Theory*, May 2003.

[20] Minqi Shen and A. Host-Madsen and J. Vidal, "An improved interference alignment scheme for frequency selective channels," in *IEEE International Symposium on Information Theory (ISIT)*, July 2008, pp. 559–563.

[21] Hakjea Sung and Seok-Hwan Park and Kyoung-Jae Lee and Inkyu Lee, "A Two-Stage Precoding Method Based on Interference Alignment for Interference Channel Systems," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Nov. 2009, pp. 1–6.

[22] K. Gomadam, V. Cadambe, and S. Jafar, "Approaching the capacity of wireless networks through distributed interference alignment," in *IEEE Global Telecommunications Conference (GLOBECOM)*, 2008, pp. 1–6.

[23] M. L. Honig, P. Crespo, and K. Steiglitz, "Suppression of near- and far-end crosstalk by linear pre- and post-filtering," *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 3, pp. 614–629, 1992.

[24] W. Yu, W. Rhee, S. Boyd, and J. Cioffi, "Iterative water-filling for gaussian vector multiple-access channels," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 145–152, 2004.

[25] D. N. Tse and S. V. Hanly, "Multiaccess fading channels. i. polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2796–2815, September 2006.

[26] S. Hanly and D. Tse, "Multi-access fading channels: delay-limited capacities," *IEEE International Symposium on Information Theory (ISIT)*, 1998.

[27] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels .i. ergodic capacity," *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 1083–1102, 2001.

[28] ——, "Capacity and optimal resource allocation for fading broadcast channels .ii. outage capacity," *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 1103–1127, 2001.

[29] P.S. Chow and J.M. Cioffi and J. A C Bingham, "A practical discrete multitone transceiver loading algorithm for data transmission over spectrally shaped channels," *IEEE Transactions on Communications*, vol. 43, no. 234, pp. 773–775, 1995.

[30] D. Guo and S. Shamai and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, pp. 1261–1282, April 2005.

[31] D. P. Palomar and S. Verdu, "Gradient of mutual information in linear vector gaussian channels," *IEEE Transactions on Information Theory*, vol. 52, No.1, pp.141-154, January 2006.

[32] ——, "Representation of mutual information via input estimates," *IEEE Transactions on Information Theory*, vol. 53, no. 2, pp. 453–470, 2007.

[33] S. Verdu and D. Guo, "A simple proof of the entropy-power inequality," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2165–2166, 2006.

[34] A. Tulino and S. Verdu, "Monotonic decrease of the non-gaussianness of the sum of independent random variables: A simple proof," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4295–4297, 2006.

[35] A. Lozano, A. Tulino, and S. Verdu, "Optimum power allocation for parallel gaussian channels with arbitrary input distributions," *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 3033–3051, 2006.

[36] F. Perez-cruz, M. R. D. Rodrigues, and S. Verdu, "Generalized mercury/waterfilling for multiple-input multiple-output channels," *45th Allerton Conference on Communication, Control, and Computing, Monticello, Illinois, U.S.A.*, September 2007.

[37] F. Perez-Cruz, M. R. D. Rodrigues, and S. Verdu, "Mimo gaussian channels with arbitrary inputs: optimal precoding and power allocation," *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1070–1084, March 2010.

[38] ——, "Optimal precoding for digital subscriber lines," *IEEE International Conference on Communications (ICC), Beijing, China*, pp. 1200–1204, May 2008.

[39] M. R. D. Rodrigues, F. Perez-Cruz, and S. Verdu, "Multiple-input multiple-output gaussian channels: Optimal covariance for non-gaussian inputs," *IEEE Information Theory Workshop (ITW), Porto, Portugal*, pp. 445–449, May 2008.

[40] J. A. C. Bingham, "Multicarrier modulation for data transmission: an idea whose time has come," *IEEE Communications Magazine*, vol. 28, no. 5, pp. 5–14, 1990.

[41] I. Telatar, "Capacity of multi-antenna gaussian channels," *Bell Labs Technical Memorandum*, June 1995.

[42] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1986–1992, 1997.

[43] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. 2nd ed. New York: Wiley, 2006.

[44] A. Lozano, A. Tulino, and S. Verdu, "Mercury/waterfilling: Optimum power allocation with arbitrary input constellations," *International Symposium on Information Theory (ISIT)*, pp. 1773–1777, 2005.

[45] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, 1968.

[46] K.-H. Lee and D. Petersen, "Optimal linear coding for vector channels," *IEEE Transactions on Communications*, vol. 24, no. 12, pp. 1283–1290, 1976.

[47] M. Payaro and D. P. Palomar, "A note on gradient of mutual information in linear vector gaussian channels," 2008.

[48] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.

[49] R. Ahlswede, N. Cai, S.-Y. Li, and R. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, 2000.

[50] Y. Chen, S. Kishore, and J. Li, "Wireless diversity through network coding," in *IEEE Wireless Communications and Networking Conference (WCNC)*, vol. 3, April 2006, pp. 1681–1686.

[51] S.-Y. Li, R. Yeung, and N. Cai, "Linear network coding," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 371–381, 2003.

[52] R. Koetter and M. Medard, "An algebraic approach to network coding," *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 782–795, 2003.

[53] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," *Proceedings of Allerton Conference on Communication, Control, and Computing*, Oct. 2003.

[54] J.-S. Park, M. Gerla, D. Lun, Y. Yi, and M. Medard, "Codecast: a network-coding-based ad hoc multicast protocol," *IEEE Wireless Communications*, vol. 13, no. 5, pp. 76–81, 2006.

[55] D. Lun, M. Medard, and R. Koetter, "Network coding for efficient wireless unicast," in *International Zurich Seminar on Communications*, 2006.

[56] C. Gkantsidis and P. Rodriguez, "Network coding for large scale content distribution," in *IEEE International Conference on Computer Communications (INFOCOM)*, vol. 4, Mar. 2005, pp. 2235–2245.

[57] A. Dimakis, P. Godfrey, Y. Wu, M. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4539–4551, 2010.

[58] S. Deb, M. Medard, and C. Choute, "On random network coding based information dissemination," in *International Symposium on Information Theory (ISIT)*, Sept. 2005, pp. 278–282.

[59] M. Chuah and P. Yang, "Impact of selective dropping attacks on network coding performance in dtns and a potential mitigation scheme," in *Proceedings of 18th Internatonal Conference on Computer Communications and Networks (ICCCN)*, Aug. 2009, pp. 1–6.

[60] M. Radenkovic and S. Zakhary, "Flexible and dynamic network coding for adaptive data transmission in dtns," in *8th International Wireless Communications and Mobile Computing Conference (IWCMC)*, Aug. 2012, pp. 567–573.

[61] D. Guo, S. Verdu, and S. Shamai, "Mutual information and conditional mean estimation in poisson channels," in *IEEE Information Theory Workshop (ITW)*, Oct. 2004, pp. 265–270.

[62] S. A. M. Ghanem and M. Ara, "The mac poisson channel: Capacity and optimal power allocation," *IAENG Transactions on Engineering Technologies, Lecture Notes in Electrical Engineering, ©Springer*, vol. 170, pp. 45–60, 2013.

[63] V. Anantharam and S. Verdu, "Bits through queues," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 4–18, 1996.

[64] B. Dunn, M. Bloch, and J. Laneman, "Secure bits through queues," in *IEEE Information Theory Workshop on Networking and Information Theory (ITW)*, June 2009, pp. 37–41.

[65] G. Bianchi, "Performance analysis of the ieee 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 3, pp. 535 –547, Mar. 2000.

[66] A. Eryilmaz, A. Ozdaglar, and M. Medard, "On delay performance gains from network coding," in *40th Annual Conference on Information Sciences and Systems*, Mar. 2006, pp. 864–870.

[67] E. Ahmed, A. Eryilmaz, M. Medard, and A. E. Ozdaglar, "On the scaling law of network coding gains in wireless networks," in *IEEE Military Communications Conference (MILCOM)*, 2007.

[68] M. Ghaderi, D. Towsley, and J. Kurose, "Network coding performance for reliable multicast," in *IEEE Military Communications Conference (MILCOM)*, Oct. 2007, pp. 1–7.

[69] C. Fragouli, D. Lun, M. Medard, and P. Pakzad, "On feedback for network coding," in *41st Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2007, pp. 248–252.

[70] J. Sundararajan, D. Shah, and M. Medard, "Arq for network coding," in *IEEE International Symposium on Information Theory (ISIT)*, July 2008, pp. 1651–1655.

[71] D. Lucani, M. Stojanovic, and M. Medard, "Random linear network coding for time division duplexing: When to stop talking and start listening," in *IEEE International Conference on Computer Communications (INFOCOM)*, April 2009, pp. 1800–1808.

[72] A. Moreira and D. Lucani, "On coding for asymmetric wireless interfaces," in *International Symposium on Network Coding (NetCod)*, June 2012, pp. 149–154.

[73] S. Katti, S. Gollakota, and D. Katabi, "Embracing wireless interference: analog network coding," *SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 4, Aug. 2007.

[74] S. Gollakota and D. Katabi, "Zigzag decoding: Combating hidden terminals in wireless networks," in *Proceedings of the ACM SIGCOMM Conference on Data communication*, New York, NY, USA, 2008.

[75] S. Katti, D. Katabi, W. Hu, H. Rahul, and M. Medard, "The Importance of Being Opportunistic: Practical Network Coding for Wireless Environments," in *Allerton Annual Conference on Communication, Control and Computing*, 2005.

[76] M. Xiao and M. Skoglund, "On network coding with finite channel state information," in *8th International Symposium on Wireless Communication Systems (ISWCS)*, Nov. 2011.

[77] B. Shrader, A. Babikyan, N. Jones, T. Shake, and A. Worthen, "Rate control for network-coded multipath relaying with time-varying connectivity," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 5, pp. 1106–1117, April 2011.

[78] D. Lucani and J. Kliewer, "On the delay and energy performance in coded two-hop line networks with bursty erasures," in *8th International Symposium on Wireless Communication Systems (ISWCS)*, Nov. 2011.

[79] F. Cali, M. Conti, and E. Gregori, "Dynamic tuning of the ieee 802.11 protocol to achieve a theoretical throughput limit," *IEEE/ACM Transactions on Networking*, vol. 8, no. 6, pp. 785–799, Dec. 2000.

[80] C. Hu, H. Kim, J. C. Hou, C. Hu, H. Kim, and J. C. Hou, "An analysis of the binary exponential backoff algorithm in distributed mac protocols," 2005.

[81] R. Oliveira, L. Bernardo, and P. Pinto, "Performance analysis of the ieee 802.11 distributed coordination function with unicast and broadcast traffic," in *IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept. 2006.

[82] B. Park, I. Huh, and H. A. Latchman, "Performance improvement of tcp with an efficient contention window control mechanism (ecwc) in ieee 802.11 based multi-hop wireless networks," in *ADHOC-NOW*, 2006.

[83] C. Shi, X. Dai, Q. Zhou, and J. Nie, "A novel fixed contention window backoff algorithm for ieee 802.11 wlan," in *International Conference on Multimedia Information Networking and Security (MINES)*, Nov. 2010.

[84] A. Mishra, M. Shin, and W. Arbaugh, "An empirical analysis of the ieee 802.11 mac layer handoff process," *SIGCOMM Comput. Commun. Rev*, vol. 33, pp. 93–102, 2003.

[85] S. Shin, A. G. Forte, A. S. Rawat, and H. Schulzrinne, "Reducing mac layer handoff latency in ieee 802.11 wireless lans," in *Proceedings of the second international workshop on Mobility management & wireless access protocols (MobiWac)*, 2004.

[86] P. Elias, A. Feinstein, and C. Shannon, "A note on the maximum flow through a network," *IRE Transactions on Information Theory*, vol. 2, no. 4, pp. 117–119, 1956.

[87] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Medard, and J. Crowcroft, "Xors in the air: Practical wireless network coding," *IEEE/ACM Transactions on Networking*, vol. 16, no. 3, pp. 497–510, 2008.

[88] D. Lun, M. Medard, R. Koetter, and M. Effros, "Further results on coding for reliable communication over packet networks," in *International Symposium on Information Theory (ISIT)*, 2005, pp. 1848–1852.

[89] S. Chachulski, M. Jennings, S. Katti, and D. Katabi, "Trading structure for randomness in wireless opportunistic routing," *SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 4, pp. 169–180, Oct. 2007.

[90] C. Fragouli, J. Widmer, and J.-Y. L. Boudec, "Efficient broadcasting using network coding," *IEEE/ACM Transactions on Networking*, vol. 16, no. 2, pp. 450–463, 2008.

[91] D. Johnson, C. Perkins, and J. Arkko, "Ip mobility support for ipv6," *IETF RFC 3775*, June 2004.

[92] R. Koodli, "Mobile ipv6 fast handovers," *IETF RFC 5568*, July 2009.

[93] H. Soliman, C. Castelluccia, K. ElMalki, and L. Bellier, "Hierarchical mobile ipv6 (hmipv6) mobility management," *IETF RFC 5568*, Oct. 2008.

[94] K.-W. Lee, W.-K. Seo, D.-W. Kum, and Y.-Z. Cho, "Global mobility management scheme with interworking between pmipv6 and mipv6," in *IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WIMOB)*, 2008.

[95] I. . WG, "IEEE Draft Standard for Local and Metropolitan Area Networks: Media Independent Handover Services - Amendment for Security Extensions to Media Independent Handover Services and Protocol," *IEEE P802.21a/D04*, pp. 1–85, June 2011.

[96] M. Chuah, P. Yang, and Y. Xi, "How mobility models affect the design of network coding schemes for disruption tolerant networks," in *29th IEEE International Conference on Distributed Computing Systems Workshops, ICDCS Workshops*, 2009.

[97] J. Jin, B. Li, and T. Kong, "Is random network coding helpful in wimax?" in *IEEE Conference on Computer Communications (INFOCOM)*, 2008.

[98] A. Yazdi, S. Sorour, S. Valaee, and R. Kim, "Optimum network coding for delay sensitive applications in wimax unicast," in *IEEE International Conference on Computer Communications (INFOCOM)*, 2009.

[99] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*. The University of Michigan: Athena Scientific, 2002.

[100] J. Cioffi, S. Jagannathan, M. Mohseni, and G. Ginis, "Cupon: the copper alternative to pon 100 gb/s dsl networks [accepted from open call]," *IEEE Communications Magazine*, vol. 45, no. 6, pp. 132–139, 2007.

[101] G. Kramer, "Interference channels handout," *IEEE Winter school of Information Theory, UPF and CTTC, Barcelona*, 2011.

[102] J. Andrews, "Interference cancellation for cellular systems: a contemporary overview," *Wireless Communications, IEEE*, vol. 12, no. 2, pp. 19 –29, 2005.

[103] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the gaussian multiple-access channel," *Information Theory, IEEE Transactions on*, vol. 42, no. 2, pp. 364–375, 1996.

[104] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[105] S. P. Weber, J. Andrews, X. Yang, and G. de Veciana, "Transmission capacity of wireless ad hoc networks with successive interference cancellation," *Information Theory, IEEE Transactions on*, vol. 53, no. 8, pp. 2799–2814, 2007.

[106] J. Blomer and N. Jindal, "Transmission capacity of wireless ad hoc networks: Successive interference cancellation vs. joint detection," in *Communications, 2009. ICC '09. IEEE International Conference on*, 2009, pp. 1–5.

[107] M. Lamarca, "Linear precoding for mutual information maximization in mimo systems," in *International Symposium on Wireless Communication Systems (ISWCS)*, 2009.

[108] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1319–1343, 2002.

[109] A. Lozano, R. J. Heath, and J. Andrews, "Fundamental limits of cooperation," *IEEE Transactions on Information Theory*, vol. 59, no. 99, 2013.

[110] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu, "Multi-cell mimo cooperative networks:a new look at interference," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, Dec. 2010.

[111] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of gaussian mimo broadcast channels," in *IEEE Transactions on Information Theory*, vol. 49, no. 10, Oct. 2003.

[112] H. Sato, "An outer bound on the capacity region of broadcast channels," in *IEEE Transactions on Information Theory*, vol. IT-24, no. 3, May 1978.

[113] S. A. M. Ghanem, "Mac gaussian channels with arbitrary inputs: Optimal precoding and power allocation," in *IEEE International Conference on Wireless Communications and Signal Processing (WCSP), Huangshan, China*, October 2012.

[114] M. Payaro and D. P. Palomar, "On optimal precoding in linear vector gaussian channels with arbitrary input distribution," in *IEEE International Symposium on Infromation Theory (ISIT)*, June 2009, pp. 1085–1089.

[115] J. H. Wilkinson, "Elementary proof of the wielandt-hoffman theorem and its generalization," *Technical Report*, no. CS 150, January 1970.

[116] A. Hjorungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," in *IEEE Transactions on Signal Porcessing*, vol. 55, no. 6, pp. 2740 - 2746, May 2007.

[117] H. Zhang and H. Dai, "Cochannel interference mitigation and cooperative processing in downlink multicell multiuser mimo networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2004, no. 2, pp. 222–235, December 2004.

[118] S. A. M. Ghanem, "Optimal power allocation and optimal precoding with multi-cell processing," *IEEE 77th Vehicular Technology Conference: VTC-Spring, Dresden, Germany*, June 2013.

[119] P. Choi, *Channel Prediction and Adaptation over Satellite Channels with Weather-Induced Impairment, M.S. Thesis, 2000.* M.I.T, Cambridge, May 2000.

[120] Lifeng Lai and Hesham El Gamal, "The Water-Filling Game in Fading Multiple-Access Channels," *IEEE Transactions on Information Theory*, vol. 54, No.5, May 2008.

[121] M. Ara, H. Reboredo, S. A. M. Ghanem, and M. R. D. Rodrigues, "A zero-sum power allocation game in the parallel gaussian wiretap channel with an unfriendly jammer," in *IEEE International Conference on Communication Systems (ICCS), Singapore*, November 2012.

[122] C. E. Shannon, "A mathematical theory of communication," *Bell Systems Technical Journal*, vol. 27, pp. 379–423, 1948.

[123] R. Essiambre, G. Kramer, P. Winzer, G. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *Journal of Lightwave Technology*, vol. 28, no. 4, pp. 662–701, 2010.

[124] Y. M. Kabanov, "The capacity of a channel of the poisson type," *Theory of Probability and Its Applications*, vol. 23, no. 1, 1978.

[125] S. A. M. Ghanem and M. Ara, "The Poisson Optical Communication Channels: Capacity and Optimal Power Allocation," *IAENG International Journal of Computer Science (IJCS)*, no. 1, pp. 102–108, 2012.

[126] D. H. Johnson and I. N. Goodman, "Inferring the capacity of the vector Poisson channel with a Bernoulli model," *Network-computation in Neural Systems*, no. 1, 2008.

[127] A. Lapidoth and S. Shamai, "The Poisson Multiple Access Channel," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 488–501, 1998.

[128] L. Lai, Y. Liang, and S. Shamai, "On the capacity region of the poisson interference channels," in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2010, pp. 330–334.

[129] H. H. Han, "Capacity of Multimode Direct Detection Optical Communication Channel," *TDA Progress Report*, pp. 42–63, 1981.

[130] M. M. Alem-Karladani, L. Sepahi, M. Jazayerifar, and K. Kalbasi, "Optimum power allocation in parallel poisson optical channel," in *10th International Conference on Telecommunications (ConTEL), Zagreb*, 2009, pp. 285–288.

[131] A. Segall and T. Kailath, "The modeling of randomly modulated jump processes," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 135–143, 1975.

[132] M. H. A. Davis, "Capacity and cutoff rate for poisson-type channels," *IEEE Transactions on Information Theory*, vol. 26, no. 6, pp. 710–715, 1980.

[133] M. Karaliopoulos and F.-N. Pavlidou, "Modelling the land mobile satellite channel: a review," *Electronics Communication Engineering Journal*, vol. 11, no. 5, pp. 235–248, Oct. 1999.

[134] M. Stojanovic and V. Chan, "Adaptive power and rate control for satellite communications in ka band," in *IEEE International Conference on Communications (ICC)*, vol. 5, 2002, pp. 2967–2972.

[135] B. Gremont, M. Filip, P. Gallois, and S. Bate, "Comparative analysis and performance of two predictive fade detection schemes for ka-band fade countermeasures," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 2, pp. 180 –192, February 1999.

[136] K. Baddour and N. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Transactions on Wireless Communications*, vol. 4, no. 4, pp. 1650 – 1662, july 2005.

[137] M. Safari and M. Uysal, "Cooperative diversity over log-normal fading channels: performance analysis and optimization," *IEEE Transactions on Wireless Communications*, vol. 7, no. 5, pp. 1963–1972, May 2008.

[138] N. Mehta, A. Molisch, J. Wu, and J. Zhang, "Approximating the sum of correlated lognormal or, lognormal-rice random variables," in *IEEE International Conference on Communications (ICC)*, vol. 4, June 2006, pp. 1605–1610.

[139] V.-P. Kaasila and A. Mammela, "Bit error probability of a matched filter in a rayleigh fading multipath channel," *IEEE Transactions on Communications*, vol. 42, no. 234, pp. 826–828, feb/mar/apr 1994.

[140] E. Malkamaki and H. Leib, "Evaluating the performance of convolutional codes over block fading channels," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1643–1646, July 1999.

[141] P. Frenger, P. Orten, and T. Ottosson, "Convolutional codes with optimum distance spectrum," *IEEE Communications Letters*, vol. 3, no. 11, pp. 317 –319, November 1999.

[142] L. Perez, J. Seghers, and D. J. Costello, "A distance spectrum interpretation of turbo codes," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 1698–1709, Nov. 1996.

[143] P. J. Lee, *A very efficient transfer function bounding technique on bit error rate for Viterbi decoder, rate 1/N convolutional codes.* TDA Progress Report, 1984.

[144] A. van der Poel, *A computer search for good convolutional codes.* TH-Report 74-E-50, 1974.

[145] *IEEE Standard 802.11.* IEEE Computer Society, 2007.