Departament de Teoria
del Senyal i Comunicacions

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Application of Network Coding in Satellite Broadcast and Multiple Access Channels

PhD Thesis Dissertation

By

Giuseppe Cocco

Submitted to the Universitat Politècnica de Catalunya (UPC)

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Barcelona, January 2013

Supervised by Dr. Christian Ibars Casas

Tutor: Prof. Miguel Ángel Lagunas Hernández

PhD program on Signal Theory and Communications

Abstract

The relaying and broadcasting capabilities of satellite platforms enable the deployment of mobile broadcast systems over wide geographical areas, opening large market possibilities for handheld, vehicular and fixed user terminals. Satellite broadcasting to mobiles, for instance, is of paramount importance for services like digital TV or machine-to-machine applications such as software and navigation maps updates, road safety or traffic congestion control. This, together with the availability of a return channel, gives rise to the possibility of creating an interactive multimedia satellite system, supporting services such as file downloading, messaging and on-demand video streaming and virtually available in any point on Earth. Geostationary satellites are particularly suited for such kinds of applications. Being a fixed point in the sky, geostationary satellites save the need to track the spacecraft movement with consequent savings in user terminal complexity and cost. The large radius of the satellite orbit, however, introduces considerable delays in the network, that prevents the availability of channel state information at the transmitter or the use of acknowledge-based error control systems in both the forward and the reverse link. The problem is aggravated by the large number of terminals usually served by satellite systems. The problem of packet loss in satellite networks is currently addressed through the deployment of terrestrial *gap fillers* and the use of large physical layer interleavers and packet-level error correcting codes in the forward link, while collision resolution methods based on retransmissions are envisaged for the return channel in case a random access is adopted. In this dissertation we consider several communication issues related to the forward and the reverse link of satellite networks. In particular we focus on three scenarios.

First we consider the problem of broadcast transmission in land mobile satellite networks

with focus on urban areas, challenging propagation environments characterized by high packet loss rates. We explore the possibility of adopting a cooperative approach in heterogeneous vehicular networks. After using the Max-flow Min-cut to derive a lower bound on the coverage that can be achieved through cooperation, we propose a practical cooperative scheme based on network coding, that allows to exploit the spatial diversity of the system in a decentralized way.

The second setup we consider is video streaming transmission, a type of traffic particularly sensible to packet losses and delay, which is endogenous in satellite networks. We consider real-time and non-real-time video streaming, and study different encoding schemes for which theoretical and numerical analysis are provided. One of the main results of tour study is the extension of the achievability of the ergodic channel capacity to the case in which the transmitter receives data over time rather than having it available since the beginning.

The third scenario we consider is satellite random multiple access. We develop a collision recovery scheme based on physical layer network coding over extended Galois fields, that shows promising performance in terms of packet recovery capability. Several practical issues related to the physical layer are also addressed which are fundamental in the perspective of a practical implementation of the system.

Resumen

Una de las características más importantes de las plataformas satelitales es su capacidad de retransmitir las señales recibidas a un gran número de terminales. Esto es fundamental en contextos como la transmisión a terminales móviles o la comunicación entre máquinas. Al mismo tiempo, la disponibilidad de un canal de retorno permite la creación de sistemas satelitales interactivos que, en principio, pueden alcanzar cualquier punto del planeta. Los satélites geoestacionarios son particularmente adecuados para cumplir esta tarea. Este tipo de satélites mantiene una posición fija respeto a la Tierra, ahorrando a los terminales terrestres la necesidad de seguir su movimiento en el cielo. Por otro lado, la grande distancia entre la Tierra y los satélites geoestacionarios introduce grandes retrasos en las comunicaciones que, también debido al gran número de terminales normalmente presentes en las redes satelitales, limitan el uso de técnicas de retransmisión basadas en acknowledgment en caso de pérdida de paquetes. Para solucionar el problema de la pérdida de paquetes las técnicas más utilizadas son el despliegue de repetidores terrestres llamados gap fillers, códigos de protección a nivel de paquete y mecanismos proactivos de resolución de colisiones en el canal de retorno. En ésta tesis se analizan y se estudian soluciones a problemas en la comunicación satelital tanto en el canal de bajada como en el de subida. En particular se consideran tres escenarios distintos. El primer escenario es la transmisión a poblaciones numerosas de terminales móviles en entornos urbanos, que están particularmente afectados por la pérdida de paquetes debido a la obstrucción de los edificios. La solución que consideramos consiste en la cooperación entre terminales. Una vez obtenida una medida de la ganancia que se puede alcanzar por medio de la cooperación en un modelo básico de red a través del teorema Max-flow Min-cut, proponemos un esquema de cooperación compatible con estándares de comunicación existentes. El segundo escenario

que consideramos es la transmisión de video, un tipo de tráfico particularmente sensible a la perdida de paquetes y retraso que es endógeno en los sistemas satelitales. Consideramos los casos de transmisión en tiempo real y tiempo diferido desde la perspectiva de la teoría de la información y estudiamos diferentes técnicas de codificación analítica y numéricamente. Uno de los resultados principales que hemos obtenido es la extensión de la alcanzabilidad de la capacidad ergodica del canal al caso en que el transmisor recibe los datos gradualmente en vez de tenerlos desde el principio de la transmisión. El tercer escenario que consideramos es al acceso múltiple aleatorio a satélite. Desarrollamos un esquema de recuperación de los paquetes perdidos basado en la codificación de red a nivel físico y extensiones de campos de Galois con resultados prometedores en términos de rendimiento. También estudiamos cuestiones relacionadas con la implementación práctica del esquema propuesto.

Resum

Una de les característiques mes importants de les plataformes de comunicacions per satèl.lit és la seva capacitat de retransmetre senyals rebuts a un gran número de terminals. Això es fonamental en contextes com la difusió a terminals mòbils o la comunicació entre màquines. Al mateix temps, la disponibilitat d'un canal de retorn permet la creació de sistemes de comunicacions per satèl.lit interactius que, en principi, poden arribar a qualsevol punt del planeta. Els satèl.lits Geoestacionaris son particularment adequats per a complir amb aquesta tasca. Aquest tipus de satèl.lits manté una posició fixa respecte a la Terra, estalviant als terminals terrestres la necessitat de seguir el seu moviment en el cel. Per altra banda, la gran distància que separa la Terra dels satèl.lits Geoestacionaris introdueix grans retrassos en les comunicacions que, afegit al gran número de terminals en servei, limita l'ús de tècniques de retransmissió basades en acknowledgments en cas de pèrdua de paquets. Per tal de sol.lucionar el problema de la pèrdua de paquets, les tècniques més utilitzades son el desplegament de repetidors terrestres, anomenats gap fillers, l'ús de codis de protecció a nivell de paquet i mecanismes proactius de resolució de col.lisions en el canal de retorn. En aquesta tesi s'analitzen i s'estudien sol.lucions a problemes en la comunicació per satèl.lit tant en el canal de baixada com el de pujada. En concret, es consideren tres escenaris diferents. El primer escenari es la transmissió a grans poblacions de terminals mòbils en enorns urbans, que es veuen particularment afectats per la pèrdua de paquets degut a l'obstrucció, per part dels edificis, de la línia de visió amb el satèl.lit. La sol.lució que considerem consisteix en la utilització de la cooperació entre terminals. Una vegada obtinguda una mesura del guany que es pot assolir mitjançant cooperació en un model bàsic de xarxa, a través del teorema Max-flow Min-cut, proposem un esquema de cooperació compatible amb estàndards de comunicació existents. El

segon escenari que considerem es la transmissió de vídeo, un tipus de tràfic particularment sensible a la pèrdua de paquets i retards endògens als sistemes de comunicació per satèl.lit. Considerem els casos de transmissió en temps real i en diferit, des de la perspectiva de teoria de la informació, i estudiem diferents tècniques de codificació analítica i numèrica. Un dels resultats principals obtinguts es l'extensió del límit assolible de la capacitat ergòdica del canal en cas que el transmissor rebi les dades de manera gradual, enlloc de rebre-les totes a l'inici de la transmissió. El tercer escenari que considerem es l'accés aleatori al satèl.lit. Desenvolupem un esquema de recuperació dels paquets perduts basat en la codificació de xarxa a nivell físic i en extensions a camps de Galois, amb resultats molt prometedors en termes de rendiment. També estudiem aspectes relacionats amb la implementació pràctica d'aquest esquema.

List of Publications

Journals

[J1] **G. Cocco**, D. Gündüz, C. Ibars, “Throughput and delay analysis in video streaming over block-fading channels", submitted to *IEEE Transactions on Wireless Communications.*

[J2] **G. Cocco**, N. Alagha, C. Ibars, S. Cioni, “A network-coded diversity protocol for collision recovery in slotted ALOHA networks", submitted to *Wiley’s International Journal of Satellite Communications and Networking*.

[J3] **G. Cocco**, C. Ibars, N. Alagha, “Coverage extension in heterogeneous satellite machine-to-machine networks", submitted to *Wiley’s Transactions on Emerging Telecommunications Technologies (ETT), special issue "Machine-to-Machine: An Emerging Communication Paradigm"*.

[J4] **G. Cocco**, D. Gündüz, C. Ibars, “Streaming transmitter over block-fading channels with delay constraint", second revision, *IEEE Transactions on Wireless Communications.*

[J5] **G. Cocco**, S. Pfletschinger, M. Navarro, C. Ibars, “Opportunistic adaptive transmission for network coding using nonbinary LDPC codes", *EURASIP Journal on Wireless Communications and Networking*, Vol. 2010, July 2010.

International Conferences

[C1] **G. Cocco**, D. Gündüz, C. Ibars, “Throughput and delay analysis in video streaming over block-fading channels", submitted to *IEEE International Conference on Communications (ICC 2013)*.

[C2] **G. Cocco**, C. Ibars, N. Alagha, “Cooperative coverage extension in heterogeneous machine-to-machine networks", in proceedings of *GLOBECOM ’12 Workshop: Second International Workshop on Machine-to-Machine Communications ’Key’ to the Future Internet of Things (GC’12 Workshop-IWM2M)*, 5-9 Dec. 2012, Houston, TX, U.S.A..

[C3] **G. Cocco**, C. Ibars, “On the feasibility of satellite M2M systems", in proceedings of *American Institute of Aeronautics and Astronautics (AIAA) International Communications Satellite Systems Conference (ICSSC 2012)*, 24-27 September 2012, Ottawa (Canada).

[C4] **G. Cocco**, N. Alagha, C. Ibars, S. Cioni, “Practical issues in multi-user physical layer network coding", in proceedings of *IEEE Advanced Satellite Mobile Systems Conference (ASMS 2012)*, 5-7 September 2012, Baiona (Spain).

[C5] **G. Cocco**, N. Alagha, C. Ibars, “Network-coded cooperative extension of link level FEC in DVB-SH", in proceedings of *American Institute of Aeronautics and Astronautics (AIAA) International Communications Satellite Systems Conference (ICSSC 2011)*, 28 November-1 December 2011, Nara (Japan).

[C6] **G. Cocco**, D. Gündüz, C. Ibars, “Real-time broadcasting over block-fading channels", in proceedings of *IEEE International Symposium on Wireless Communication Systems (ISWCS 2011)*, 6-9 November 2011, Aachen (Germany).

[C7] **G. Cocco**, D. Gündüz, C. Ibars, “Throughput analysis in asymmetric two-way relay channel with random access", in proceedings of *IEEE International Conference on Communications (ICC 2011)*, 5-9 June 2011, Kyoto (Japan).

[C8] **G. Cocco**, C. Ibars, D. Gündüz, O. d. R. Herrero, “Collision resolution in multiple access networks with physical-layer network coding and distributed fountain coding", in proceedings of *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011)*, 22-27 May 2011, Praga (Czech Republic).

[C9] **G. Cocco**, C. Ibars, D. Gündüz, O. d. R. Herrero, “Collision resolution in slotted ALOHA with multi user physical layer network coding", in proceedings of *IEEE Vehicular Technology Conference (VTC 2011-Spring)*, 15-18 May 2011, Budapest (Hungary).

[C10] **G. Cocco**, C. Ibars, O. d. R. Herrero, “Cooperative satellite to land mobile gap-filler-less interactive system architecture", in proceedings of *IEEE Advanced Satellite Mobile Systems Conference (ASMS 2010)*, 13-15 September 2010, Cagliari (Italy).

Contribution to Project Deliverables:

[T1] **G. Cocco**, technical contribution to deliverable D3.4 for the European FP7 project *EXALTED: Expanding LTE for Devices*, Dec. 2012.

[T2] **G. Cocco**, technical contribution to deliverable D4.5 for the Spanish national project *m:Via, autopistas de la Información*, 2009.

[T3] **G. Cocco**, technical contribution to deliverable D5.2 for the European FP7 project *NEWCOM++ - the FP7 Network of Excellence in Wireless COMmunications*, 2009.

Contents

List of Figures

List of Tables

List of Acronyms

ACK acknowledgement

ARQ Automatic Repeat Request

AWGN additive white Gaussian noise

CA collision avoidance

bpcu bits per channel use

CGC Complementary Ground Component

CRDSA Contention Resolution Diversity Slotted ALOHA

CSI channel state information

CSMA carrier sense multiple access

DSA Diversity Slotted ALOHA

DVB Digital Video Broadcasting

EGF extended Galois field

FER frame error rate

GEO geostationary

GOP group of pictures

IIC iterative interference cancelation

IFEC Inter-burst Forward Error Correction

i.i.d. independent and identically distributed

ISI inter-symbol interference

LL-FEC Link Layer-Forward Error Correction

LLR log-likelihood ratio

LMS lad mobile satellite

LNC linear network coding

M2M machine-to-machine

MAC medium access control

MI mutual information

MPE Multi Protocol Encapsulation

NACK negative acknowledgement

NC network coding

N-COCCO Network-COded Cooperative COverage

OFDM orthogonal frequency-division multiplexing

PLA physical layer abstraction

QoE quality of user experience

RBIR received bit mutual information rate

RCS Return Channel Satellite

RLNC random linear network coding

RTT round-trip time

S2 Satellite-Second Generation

SA Slotted ALOHA

SC Satellite Component

SH Satellite service to Handheld

SNR signal-to-noise ratio

VtoV vehicle-to-vehicle

Notation

x	Scalar notation
$\mathbf{x}$	Vector notation
$\mathbf{x}^H$	Hermitian transpose of $\mathbf{x}$
$\mathcal{X}$	Set notation
$\lvert x \rvert$	Absolute value of x
$\lvert \mathcal{X} \rvert$	Cardinality of the set $\mathcal{X}$
x^+	Maximum between 0 and x ($\max\{0, x\}$)
$\lfloor x \rfloor$	Floor function
$\lceil x \rceil$	Ceil function
$\ln(x)$	Natural logarithm of x
$\log_2(x)$	Logarithm in base 2 of x
$\arg\max_x f(x)$	Value of x that maximizes $f(x)$
$\arg\min_x f(x)$	Value of x that minimizes $f(x)$
$\frac{df(x)}{dx}$	Derivative of $f(x)$ with respect to x
$\triangleq$	Equal by definition

Chapter 1

Introduction

"Any sufficiently advanced technology is indistinguishable from magic."

Arthur C. Clarke

Contents

1.1 Motivation

The birth of satellite communications dates back to about 50 years ago. It was the 1962 when the first communication satellite, the *Telstar 1*, was launched, enabling radio communication between the two shores of the Atlantic Ocean. Since then, an uninterrupted scientific effort and the consequent technological advance lead to the modern high power and highly capable telecommunications satellites, making *satcom* the most mature of space applications. The evolving satellite technology allowed for a reduction in size, complexity and cost of the receiving terminals, opening up the possibility of a large interactive satellite network with many possible applications.

Satellite broadcasting and relaying capabilities enable mobile broadcast systems over wide geographical areas, which opens large market possibilities for handheld, vehicular and fixed user terminals. Satellite broadcasting to mobiles, for instance, is of paramount importance for services such as digital TV or machine-to-machine communication for software and navigation

maps update, road safety and traffic congestion control.

The availability of a return channel gives rise to the possibility of creating an interactive multimedia satellite system, supporting services such as file downloading, messaging and on-demand video streaming.

The geostationary (GEO) satellite orbit is highly suited for such applications, as it spares the need for satellite terminals to track the movement of the spacecraft, with important savings in terms of complexity and cost. The large radius of the GEO orbit (more than $40000\ \mathrm{km}$) has two main drawbacks. One is the large free space loss experienced by a signal traveling to or from the satellite, which limits the signal-to-noise ratio (SNR) margins in the link budget with respect to terrestrial systems. The second drawback of the GEO orbit is the large propagation delay (about $250\ \mathrm{msec}$) that limits the use of feedback in both the forward (satellite to satellite terminal) and the reverse (satellite terminal to satellite) link. The limited margin protection causes loss of service availability in environments where there is no direct line of sight to the satellite, such as urban areas. The large propagation delay on its turn, together with the large terminal population size usually served by a GEO satellite, limit the use of feedback, which is at the basis of error-control protocols such as Automatic Repeat Request (ARQ), in both the forward and the reverse link. Unlike in the forward link, in the reverse link, especially in the case of fixed terminals, packet losses are mainly due to collisions, that severely limit the access to satellite services in case a random access scheme is adopted.

Three scenarios of practical relevance are particularly vulnerable to these problems; these are broadcast transmission to mobiles in urban environments, real-time video streaming to a large population of terminals and satellite random access.

In absence of a line of sight between the terminal and the satellite, large physical layer interleavers and Link Layer-Forward Error Correction (LL-FEC) codes are employed in existing standards in order to exploit the time diversity of the system, while terrestrial gap-fillers are used for complementary coverage. The use of gap-fillers is a fixed solution which is not able to react quickly to changes in the propagation environment and is very costly in terms of investment, management, and bandwidth usage.

One of the main applications of broadcast satellite systems is the provision of digital video services. Video services, and particularly real-time streaming, have specific delay constraints which are not easily met in a satellite context. The study of the theoretical limits of video

streaming systems in the absence of feedback is a relatively new field, where even relatively simple channel models are not yet well understood in terms of achievable throughput, delay and the optimal coding techniques to be adopted.

Let us now consider the return link. Random multiple access systems are severely affected by the problem of collisions, that greatly limit the throughput of the system. Collision recovery techniques based on proactive retransmissions like Diversity Slotted ALOHA (DSA) have been proposed in order to increase the probability of successful detection. The limiting factor of this and other similar techniques is that they do not exploit the information that can be extracted from colliding signals.

The need for improvements and further understanding of these setups lead to the development of the present work. In this dissertation we study the application of *network coding* to counteract the above mentioned channel impairments in satellite systems. The idea of using network coding stems from the fact that it allows to efficiently exploit the diversity, either temporal or spatial, present in the system. Throughout the dissertation we adopt a cross-layer approach, putting emphasis on the interaction of the physical level with the packet level. Although the present work focuses on the satellite scenario, some of the proposed solutions and analysis can be easily extended to terrestrial networks.

1.2 Thesis Overview

In this section we outline the original contributions included in each of the chapters of the present dissertation and reference the papers in which they can be found.

In **Chapter 3** we address channel impairments in the forward link, and specifically we deal with the problem of limited coverage in urban environments for land mobile satellite (LMS) networks. We start by considering a mathematically tractable and yet practically interesting channel model. By applying the Max-flow Min-cut theorem we derive a lower bound on the maximum coverage that can be achieved through cooperation. Inspired by this result, we propose a practical scheme, keeping in mind the compatibility with the DVB-SH standard. In order to evaluate the performance of the proposed solution, we developed a simulator in Matlab/C++ based on *physical layer abstraction* and used it to compare the performance gain of our scheme

with a benchmark relaying scheme that does allow coding at packet level.

The technical contributions of the chapter can be found in the following pubblications:

- **G. Cocco**, C. Ibars, N. Alagha, "Coverage extension in heterogeneous satellite machine-to-machine networks", submitted to *Wiley's Transactions on Emerging Telecommunications Technologies (ETT), special issue "Machine-to-Machine: An Emerging Communication Paradigm"*.
- **G. Cocco**, C. Ibars, N. Alagha, "Cooperative coverage extension in heterogeneous machine-to-machine networks", in proceedings of *GLOBECOM '12 Workshop: Second International Workshop on Machine-to-Machine Communications 'Key' to the Future Internet of Things (GC'12 Workshop-IWM2M)*, 5-9 Dec. 2012, Houston, TX, U.S.A..
- **G. Cocco**, N. Alagha, C. Ibars, "Network-coded cooperative extension of link level FEC in DVB-SH", in proceedings of *American Institute of Aeronautics and Astronautics (AIAA) International Communications Satellite Systems Conference (ICSSC 2011)*, 28 November-1 December 2011, Nara (Japan).
- **G. Cocco**, C. Ibars, O. d. R. Herrero, "Cooperative satellite to land mobile gap-filler-less interactive system architecture", in proceedings of *IEEE Advanced Satellite Mobile Systems Conference (ASMS 2010)*, 13-15 September 2010, Cagliari (Italy).

Chapter 4 contributes to the information theoretical study of real-time streaming transmissions over fading channels with channel state information at the transmitter only. We introduce a new channel model and propose several transmission schemes, one of which is proved to be asymptotically optimal in terms of throughput in that it achieves the ergodic capacity. We also provide an upper bound on the achievable throughput for the proposed channel model and compare it numerically with the proposed schemes over a Rayleigh fading channel.

The following publications stemmed from the work in this chapter:

- **G. Cocco**, D. Gündüz, C. Ibars, "Streaming transmitter over block-fading channels with delay constraint", second revision, *IEEE Transactions on Wireless Communications*.
- **G. Cocco**, D. Gündüz, C. Ibars, "Real-time broadcasting over block-fading channels", in proceedings of *IEEE International Symposium on Wireless Communication Systems*

(ISWCS 2011), 6-9 November 2011, Aachen (Germany).

In **Chapter 5** we follow the same line of work of Chapter 4 and study throughput and delay in non-real-time streaming transmission over block fading channels. After introducing the channel model, we derive bounds on the throughput and the delay for this channel and propose different coding techniques based on time-sharing. For each of them we carry out an analytical study of the performance metrics. Finally, we compare numerically the performance of the proposed schemes over a Rayleigh fading channel.

The contributions of this chapter are contained in the following papers that are currently being evaluated for publication:

- **G. Cocco**, D. Gündüz, C. Ibars, "Throughput and delay analysis in video streaming over block-fading channels", submitted to *IEEE Transactions on Wireless Communications*.
- **G. Cocco**, D. Gündüz, C. Ibars, "Throughput and delay analysis in video streaming over block-fading channels", submitted to *IEEE International Conference on Communications (ICC 2013)*.

Finally, in **Chapter 6** we propose a collision resolution method for the return link based on physical layer network coding over an extended Galois field (EGF). The proposed scheme extracts information from the colliding signals and achieves important gains with respect to Slotted ALOHA systems as well as with respect to other collision resolution schemes.

The content of this chapter is contained in the following papers:

- **G. Cocco**, N. Alagha, C. Ibars, S. Cioni, "A network-coded diversity protocol for collision recovery in slotted ALOHA networks", submitted to *Wiley's International Journal of Satellite Communications and Networking*.
- **G. Cocco**, C. Ibars, "On the feasibility of satellite machine-to-machine (M2M) systems", in proceedings of *American Institute of Aeronautics and Astronautics (AIAA) International Communications Satellite Systems Conference (ICSSC 2012)*, 24-27 September 2012, Ottawa (Canada).

- **G. Cocco**, N. Alagha, C. Ibars, S. Cioni, "Practical issues in multi-user physical layer network coding", in proceedings of *IEEE Advanced Satellite Mobile Systems Conference (ASMS 2012)*, 5-7 September 2012, Baiona (Spain).

- **G. Cocco**, C. Ibars, D. Gündüz, O. d. R. Herrero, "Collision resolution in multiple access networks with physical-layer network coding and distributed fountain coding", in proceedings of *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011)*, 22-27 May 2011, Praga (Czech Republic).

- **G. Cocco**, C. Ibars, D. Gündüz, O. d. R. Herrero, "Collision resolution in slotted ALOHA with multi user physical layer network coding", in proceedings of *IEEE Vehicular Technology Conference (VTC 2011-Spring)*, 15-18 May 2011, Budapest (Hungary).

Apart from those presented in this dissertation, others results can be found in the following papers published during the author's PhD studies:

- **G. Cocco**, S. Pfletschinger, M. Navarro, C. Ibars, "Opportunistic adaptive transmission for network coding using nonbinary LDPC codes", *EURASIP Journal on Wireless Communications and Networking*, Vol. 2010, July 2010.

- **G. Cocco**, D. Gündüz, C. Ibars, "Throughput analysis in asymmetric two-way relay channel with random access", in proceedings of *IEEE International Conference on Communications (ICC 2011)*, 5-9 June 2011, Kyoto (Japan).

Chapter 2

Background and State of the Art

"De custu mundu intreu in d'ogni perra
de su duos connottu est su valore
sa linea immaginaria 'e s'Equatore
in duas partes dividit sa Terra
e deo tott'ai duas das cuncordo
dae su Polo Sud a Polo Nord."

Bernardo Zizi

Contents

2.1 Network Coding

A communication network can be modeled as a *graph* $G(\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the set of nodes (vertices) and $\mathcal{E}$ is the set of edges (communication links) between the nodes. In the classical networking approach a *path* on the graph is selected to deliver information from a source node to a destination (*terminal*) node in a unicast connection, while *trees* (and specifically *Steiner trees*) are used in the case of multicast connections. In the classical relaying approach a node takes a packet from one of its inputs, eventually replicates it and sends it out to a subset of its outputs. However, from an information-theoretical point of view there is no reason for limiting the operations of a node to simple replication and forward [1]. In the *network coding (NC)* approach nodes are allowed to perform coding operations across their inputs and forward functions of received packets. If the function is a linear combination we talk about *linear network coding (LNC)*, which is the most popular variant of network coding (NC) (even if not necessarily the optimal one [2]). NC can provide significant advantages in terms of throughput, delay and energy saving with respect to the classical approach. In the following we give three popular examples that provide an intuition on how such benefits are obtained [3].

Example 1: Throughput Gain

Consider the *butterfly network* depicted in Fig. 2.1. Each of the two source nodes S_1 and S_2

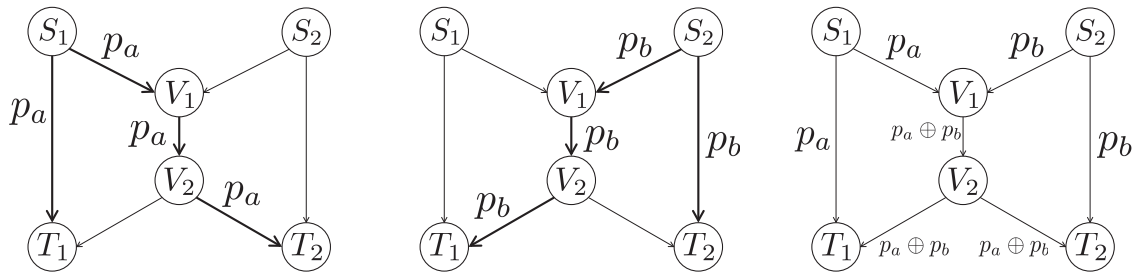


Figure 2.1: Throughput maximization.

has a packet to deliver to both terminal nodes T_1 and T_2, i.e., we have a multiple multicast connection. We assume for simplicity that only one packet per time unit can transit over each edge. In the classical approach each source transmits its packet over a Steiner tree (thick line paths in figures 2.1(a) and 2.1(b)) in order to deliver it to both terminals. It can be seen from

the picture that the two possible Steiner trees have a common edge $\{V_1, V_2\}$ (i.e., they are not *edge-disjoint*) which constitutes a bottleneck. A total of 4 time units is needed to forward the packet of each source to both the receivers. The total throughput of the network is, thus, 1 packet per unit of time. Now let us consider Fig. 2.1(c), where a network coded graph is illustrated. In this case node V_1 combines (symbol $\oplus$) the two packets received from S_1 and S_2 during the first transmission slot. In the second transmission slot V_1 forwards the combination to V_2 which, on its turn, delivers the new packet to both $T_1$and T_2 in the third transmission. Each of the terminals can obtain both packets by combining together those received during the first and the third transmission slot. In this case the network delivers 4 packets in 3 time slots, that is, it increases throughput of about 25% without increasing the usage of network resources.

Example 2: Delay Reduction

The directed graph in Fig. 2.2 describes a network with one source S and a set of three terminals $\mathcal{T} = \{T_1, T_2, T_3\}$. We want to count the number of time slots needed to transmit two distinct unit-time messages p_a and p_b from S to all nodes in $\mathcal{T}$, with S constrained to transmit both packets in the first time slot. As in the previous example, only one packet per time slot can be transmitted over a given edge and each time slot has a constant duration of 1 time unit. With the classical approach the fastest way to deliver the two packets is to transmit them over the two edge disjoint Steiner trees in figures 2.2(a) and 2.2(b), which requires three time slots. In Fig. 2.2(c) the same problem is solved with NC, which reduces the delay to 2 time slots. It can be proved that it is not possible to get a delay less than 3 without the use of NC.

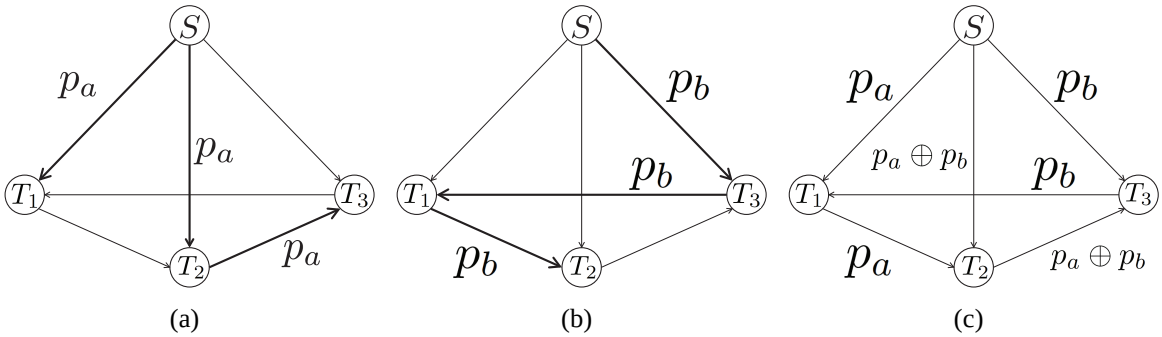
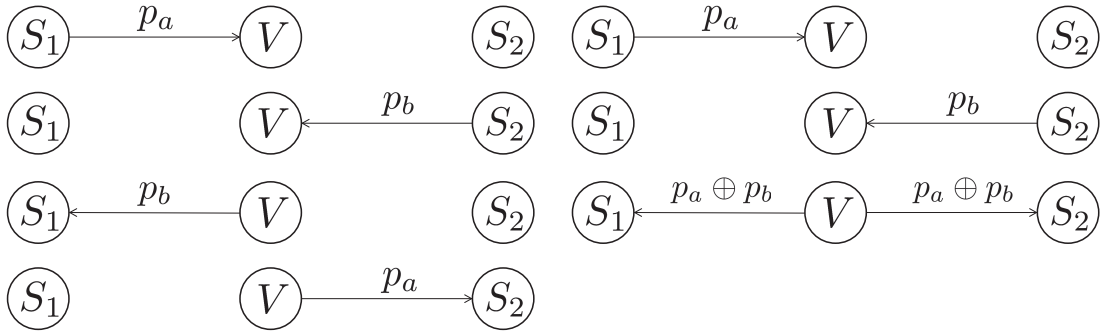


Figure 2.2: Delay minimization.

Example 3: Energy Saving

Fig. 2.3 illustrates a wireless *two-way relay channel* setup, in which the communication between the two nodes S_1 and S_2 is enabled by the relay node V. S_1 and S_2 each have a unit-time packet to deliver to the other node. Edges are assumed to have unit capacity. Assuming that one unit of energy is needed to transmit a packet, a total of 4 units are necessary when the classical approach is used (Fig. 2.3(a)), while just 3 energy units are needed with NC (Fig. 2.3(b)). Such energy saving is possible on one side thanks to the fact that the combined packet has exactly the same energy than a normal one, and on the other side thanks to the broadcast nature of the wireless channel.



(a) Four phases two-way relay channel. (b) Three phases two-way relay channel.

Figure 2.3: Energy saving.

2.1.1 Elements of Network Coding

Basics of Finite Fields

In the following we recall some basics of finite fields, which are at the basis of LNC.

A finite field is a closed set with respect to sum and multiplication with finitely many elements. Finite fields are often denoted as $GF(s^n)$, where s is a prime number, n is a positive integer and GF stands for *Galois field*. If $n = 1$ all operations (sum, subtraction, multiplication and division) in the field coincide with operations over natural numbers modulo s. If $n > 1$ the field is said to be an *extended Galois field (EGF)*. In an EGF each element can be represented as a polynomial of degree lower than n and coefficients in $GF(s)$. An element in an EGF can be represented using the coefficients of the corresponding polynomial representation. Thus, a

string of n bits can be interpreted as an element in $GF(2^n)$. Along the same line, a string of $N = n \cdot L$ bits, $L \in \mathbb{N}$, can be represented as a vector in an L-dimensional space over $GF(2^n)$ (see [4] for more details).

The sum operation in an EGF is done coefficient-wise. The sum of two elements in $GF(2^n)$ can be calculated as the bit-wise XOR of the two n-bits strings corresponding to the two elements to add.

The product in an EGF can be calculated through polynomial multiplication modulo an irreducible polynomial which characterizes the field. Subtraction and division are defined as the inverse operations of sum and product, respectively, and calculated accordingly.

Finally, let us consider a system of linear equations in $GF(2^n)$ with M variables and N equations, $N \geq M$, with an associated $N \times M$ coefficient matrix $\mathbf{A}$ having elements in $GF(2^n)$. The system admits a unique solution if and only if the associated coefficient matrix $\mathbf{A}$ has exactly M linearly independent columns (rows).

Network Coding

Let us consider a network $G(\mathcal{V}, \mathcal{E})$. Nodes in $\mathcal{V}$ exchange packets over edges in $\mathcal{E}$. Each packet is modeled as an element of $GF(q)$ and can be represented by a string of $\log_2(q)$ bits. For simplicity we will consider only unit capacity edges. The case of edges with different capacities can be taken into account by considering multiple parallel unit capacity edges between two nodes. Each edge e is indicated as (V, U) where V and U are the nodes connected by e. In the following we will consider only *linear network coding*. A linear network coding is indicated as $N(G, S, \mathcal{T})$ where G is the graph representing the network, S is the source node and $\mathcal{T}$ is the set of terminal nodes. We define the *network capacity* as the tightest upper bound on the maximum amount of information that can be transferred from S to $\mathcal{T}$ in a single communication slot. If $h(t)$ is the smallest amount of information that reaches one of the nodes in $\mathcal{T}$ up to time instant t, then the network capacity is defined as:

$$h^* = \limsup_{t \to +\infty} \frac{h(t)}{t}. \tag{2.1}$$

If the network that does not contain cycles then $h^* = h(1)$.

An upper bound on the rate at which messages can be delivered to a destination node in a network is provided by the Max-flow Min-cut theorem, which is given after the following

definition.

Definition. Consider a graph $G(\mathcal{V}, \mathcal{E})$ with a source node $S \in \mathcal{V}$ willing to transmit data to a terminal node $T \in \mathcal{V}$. A *cut* between S and T is defined as the set of edges whose removal disconnects S from T. The value of the cut is the sum of the capacities of the edges in the cut. The min-cut is the cut with the smallest value.

Max-flow Min-cut Theorem. *Consider a graph $G(\mathcal{V}, \mathcal{E})$ with unit capacity edges, a source vertex S, and a receiver vertex T. The maximum rate at which information can be sent from S to T is equal to the min-cut between S and T. Equivalently, there exist exactly h edge-disjoint paths between S and T [5].*

In case of directed graphs linear NC can achieve the network capacity of the Max-flow Min-cut theorem, while in the case of undirected graphs it constitutes an upper bound.

Linear Network Coding

Let $\mathbf{R} = \{p_1, p_2, ..., p_h\}$ be the set of h packets produced by the source node that must be delivered to all nodes in $\mathcal{T}$. We assume, without loss of generality that S has exactly h outgoing edges and each node in $\mathcal{T}$ has h incoming edges. If this is not verified for some $T \in \mathcal{T}$, a new terminal T' can be added and connected to T with h parallel edges. The generic packet $p_i \in \mathbf{R}$ is transmitted by the source S over its outgoing edge e_i.

We call p_e the packet transmitted over the network on the generic edge $e(v, u)$. Packet p_e can be expressed as:

$$p_e = \sum_{e' \in \mathcal{M}_e} \beta_{e',e} \cdot p_{e'} \tag{2.2}$$

where $\mathcal{M}_e$ is the set of incoming edges of node V and $\beta_{e',e} \in \mathbb{F}_q$. Eqn. (2.2) states that a node transmits on each of its outgoing edges a linear combination of the packets received on its incoming edges in previous time slots, using $\beta_{e',e}$ as coefficients. Thus each packet in the network can be expressed as a linear combination in $GF(q)$ of the source packets (Eqn. (2.2) is valid also for outgoing edges of S if we add h fictitious incoming edges to S and use $\beta_{e',e} = 1 \forall e', e$).

A *global encoding vector* is defined for each edge e: $\boldsymbol{\gamma}_e = \{\gamma_1^e \cdots \gamma_h^e\}$, whose elements are the encoding coefficients of source packets over edge e. In other words, the generic packet p_e transmitted over the edge e can be expressed as:

$$p_e = \sum_{i=1}^{h} p_i \cdot \gamma_i^e \tag{2.3}$$

where $p_i, i = 1, ..., h$ are the source packets. The global encoding vector on a given edge $e(v, u)$ can be expressed as a function of local coefficients (that form the *local encoding vector*) and the global encoding vectors of packets in $\mathcal{M}_e$ (the set of incoming edges of node V):

$$\boldsymbol{\gamma}(v, u) = \sum_{e' \in \mathcal{M}} \beta_{e',e_i} \cdot \boldsymbol{\gamma}_{e'} \tag{2.4}$$

In order to understand which are the characteristics that the coefficients must satisfy for a reliable multicast connection let us consider a generic terminal node T with its set of incoming edges $E_t = \{e_t^1, ..., e_t^h\}$. The *transfer matrix* $\mathbf{M}_t$ is a $h \times h$ matrix obtained stacking up the global encoding vectors of all edges in E_t, namely:

$$\mathbf{M}_t = \begin{bmatrix} \gamma_{e_t^1} \\ \gamma_{e_t^2} \\ \vdots \\ \gamma_{e_t^h} \end{bmatrix}.$$

Stacking up the source packets in vector $\mathbf{r}$:

$$\mathbf{r} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_h \end{bmatrix},$$

and packets received by T in vector $\mathbf{e}_t$:

$$\mathbf{e}_t = \begin{bmatrix} p_{e_t^1} \\ p_{e_t^2} \\ \vdots \\ p_{e_t^h} \end{bmatrix},$$

we can write:

$$\mathbf{e}_t = \mathbf{M}_t \cdot \mathbf{r}, \tag{2.5}$$

and finally:

$$\mathbf{r} = \mathbf{M}_t^{-1} \cdot \mathbf{e}_t. \tag{2.6}$$

Using Eqn. (2.6) the terminal T can decode the original messages transmitted by S, provided that matrix $\mathbf{M}_t^{-1}$ exists, or, equivalently, that $det\,(\mathbf{M}_t) \neq 0$. In this case we say that the network coding is *feasible*. We call a network where NC is applied a *coding network* and indicate it with $\mathbb{N}(G, S, \mathcal{T})$. More generally, for a unicast connection (i.e., $\mathcal{T} = t$) the following theorem holds:

Theorem 2.1: Let $\mathbb{N}(G, S, \mathcal{T})$ be a coding network, with $\mathcal{T} = T$, and h be the number of packets that need to be delivered from S to T. Then, the following three conditions are equivalent.

1. There exist a feasible network code for $\mathbb{N}(G, S, \mathcal{T})$ and h over $GF(q)$ for some finite value of q;
2. The determinant $det\,(\mathbf{M}_t)$ of the transfer matrix $\mathbf{M}_t$ is a multi-variate polynomial (in variables β) not identically equal to zero;
3. Every cut that separates S and T in $G(\mathcal{V}, \mathcal{E})$ includes at least h edges [3].

The following theorem is the extension of the previous one to multicast transmissions:

Theorem 2.2: Let $\mathbb{N}(G, S, \mathcal{T})$ be a multicast coding network and h be the number of packets that need to be delivered from S to all terminals in $\mathcal{T}$. Then, the following three conditions are equivalent.

1. There exist a feasible network code for $\mathbb{N}(G, S, \mathcal{T})$ and h over $GF(q)$ for some finite value of q;
2. The product $\prod_{T \in \mathcal{T}} det\,(\mathbf{M}_t)$ of the determinants of the transfer matrices is a multi-variate polynomial (in variables β) not identically equal to zero;
3. Every cut that separates S and T, $\forall T \in \mathcal{T}$, in $G(\mathcal{V}, \mathcal{E})$ includes at least h edges [3].

A criterion to determine whether a feasible network coding exists for a given multicast network $\mathbb{N}(G, S, \mathcal{T})$ has been given in [6]. It has been shown that a valid network code $\{\beta_{e',e}\}$ for $\mathbb{N}(G, S, \mathcal{T})$ exists on any field $GF(q)$ if q is larger than the number of terminals in the network.

2.1.2 Network Coding in Wireless Ad-Hoc Networks

Wireless ad-hoc networks are characterized by a changing topology that can make it difficult to determine a-priori the optimal network coding coefficients. A suitable approach to randomized settings which is totally distributed has been proposed in [7], where the network coding coefficients are randomly chosen by each node from a uniform distribution in $GF(q)$. A wireless network can be modeled as a directed hypergraph $\mathcal{H} = (\mathcal{N}, \mathcal{A})$, $\mathcal{N}$ being a set of nodes and $\mathcal{A}$ a set of hyperarcs. A hyperarc is a pair (i, J), where i is the *head* node of the hyperarc while J is the subset of $\mathcal{N}$ connected to the head through the hyperarc. A hyperarc (i, J) can be used to model a broadcast transmission from node i to nodes in J. As said before a network coding is feasible if $\prod_{T\in\mathcal{T}} det\left(\mathbf{M}_t\right) \neq 0$, where $det\left(\mathbf{M}_t\right)$ are multivariate polynomials in the variables $\beta_{e',e}$, e' and $e \in \mathcal{E}$. In [6] it has been shown that for a number of variables equal to $\eta \in GF(q)$, the probability to obtain a feasible network coding by randomly selecting the coefficients is upper bounded by $\left(1 - \frac{|\mathcal{T}|}{q}\right)^{\eta}$. If the field size q is sufficiently large and the random coding approach is used, a feasible network coding can be found with high probability.

In the work of Chou *et al.* [8] a practical method for the implementation of a network coded system is proposed. A single broadcast transmission is considered in which the source produces a string of bits. Bits are grouped into blocks of $\log_2(q)$ bits, each of which represents a symbol in $GF(q)$. Symbols are taken in groups of N. Each group of N symbols form a packet. h successive packets form a *generation*. Each node in the network transmits linear combinations of the received packets belonging to the same generation, using coefficients randomly chosen according to a uniform distribution in $GF(q)$. Packets produced combining other packets from generation k also belong to generation k. Note that, unlike the general NC approach previously described, in the method proposed in [8] a packet is not considered as a single symbol in a certain $GF(q')$, but as a sequence of symbols in a field $GF(q)$ with $q < q'$. Coefficients are also chosen from the same field $GF(q)$. However if we set $q' = q^N$ we can still see the packet as an N-dimensional vector in $GF(q')$. In order for the nodes to correctly

decode the source packets, the global encoding vector for each packet is attached to the packet itself. This determines a certain overhead which implies a negligible loss in spectral efficiency if the packets are long enough [8]. A fundamental advantage of random coding with respect to deterministic coding is that there is no need for the nodes to know the network topology. This makes random linear network coding particularly suited in settings where the network topology rapidly changes, which is the case for wireless ad-hoc network and particularly for Vehicular Ad-Hoc Networks (VANETS). In [9] and [10] NC was applied in the space segment of the European Telecommunications Standard Institute (ETSI) standard DVB-Satellite service to Handheld (SH) and DVB-Satellite-Second Generation (S2) in order to counteract channel impairments in mobile and fixed scenarios.

A practical architecture for the implementation of NC in the current network stack called COPE has been proposed in [11]. The architecture in [11] exploits the broadcast nature of the wireless medium by allowing all nodes to store overheard packets (promiscuous mode). Packets are added together in $GF(2)$. In COPE nodes locally exchange reception reports in which they communicate which packets have been stored. Based on this knowledge each of the nodes chooses which packets to combine in order to limit the number of transmissions, thus increasing network throughput and reducing delay. The problem of choosing the correct packets so as to minimize the number of transmission is referred to as *Index Coding* [12][13]. Results in [11] show that, at moderate traffic loads, COPE can achieve an increase in throughput of 3x-4x on average with respect to classical routing transmission. When the network is lightly loaded the performance is similar to the case of uncoded network due to scarce coding opportunity, while at high loads COPE suffers from reception reports losses, which determines a decrease in network throughput with respect to the moderate-load case.

2.1.3 Coding Gain

Let us consider a multicast network $\mathbb{N}(G, s, \mathcal{T})$. We call h^*_{NC} the capacity of the network coded network and h^*_{UN} the network capacity of the uncoded one. Let us define the *coding advantage* $\mathfrak{A}$ as:

$$\mathfrak{A} = \frac{h^*_{NC}}{h^*_{UN}} \tag{2.7}$$

$\mathfrak{A}$ catches the advantage of network coding with respect to the classical approach. In [14] it was shown that for a directed multicast network (a network where each edge has a fixed verse)

the coding advantage can be as large as $\Omega\left(\frac{\log|\mathcal{T}|}{\log(\log|\mathcal{T}|)}\right)$, where $|T|$ is the number of receivers. In case of undirected networks the coding advantage is upper bounded by two. If we consider multiple unicast connections, then the maximum $\mathfrak{A}$ grows as the number of unicast pairs, while there is no coding gain in case of single unicast connections. More references on network coding as well as theorems' proofs can be found in [15].

2.1.4 Physical Layer Network Coding

Physical layer network coding (PHY NC) is a technique that allows extracting of information from colliding signals, i.e., signals that interfere each other. PHY NC was originally proposed to save bandwidth in two-way relay communications [16] by decoding the collision of two signals under the hypothesis of symbol, frequency and phase synchronism. Several studies have been published on PHY NC about synchronization issues, gain analysis and ad-hoc modulation techniques for the case of two colliding signals [17], [18], [19]. In [20] PHY NC was applied in the satellite context for pairwise node communication. In [21] and [22] PHY NC was used to determine the identity of transmitting nodes in case of acknowledgement (ACK) collision in multicast networks by using energy detection and ad-hoc coding schemes for the case of phase synchronous transmission. In [23] an overview of the state of the art on PHY NC has been presented from an information theoretical point of view.

Among the different variants of PHY network coding for two-way communication we recall the most popular ones. One is the amplify-and-forward (AF) approach, in which the relay node amplifies and broadcasts the collided signal to the two terminals. The other approach is the decode-and forward (also known as denoise-and-forward), in which the relay decodes the sum of the interfering signals.

2.2 Elements of Information Theory

2.2.1 Stochastic Convergence

Weak Law of Large Numbers

Let $X_1, \ldots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables distributed according to the probability density function (pdf) $f_X(x)$ with mean $\overline{X}$ and

such that

$$\int_{-\infty}^{+\infty} |x| f_X(x) dx < +\infty.$$

Be S_n the *sample mean* of a realization of the sequence, defined as:

$$S_n = \frac{1}{n} \sum_{i=1}^{n} x_i. \tag{2.8}$$

The *weak law of large numbers* states that the sequence $X_1, \ldots, X_n$ converges *in probability* to $\overline{X}$, that is, for any $\epsilon > 0$:

$$\lim_{n \to \infty} Pr\left\{ \left| S_n - \overline{X} \right| > \epsilon \right\} = 0. \tag{2.9}$$

Central Limit Theorem

Let $X_1, \ldots, X_n$ be a sequence of i.i.d. random variables distributed according to the probability density function $f_X(x)$ having mean $\overline{X}$ and variance $\sigma^2 < \infty$, and be S_n the sample mean of the sequence. The *central limit theorem* states that the sequence of random variables $\sqrt{n}\left(S_n - \overline{X}\right)$ converges in distribution to a Gaussian random variable with zero mean and variance σ^2. Moreover, if $\sigma > 0$, the convergence is pointwise, that is:

$$\lim_{n \to \infty} \sup_{z \in \mathbb{R}} \left| Pr\left\{ \sqrt{n}\left(S_n - \overline{X}\right) \leq z \right\} - \Phi(z) \right| = 0, \tag{2.10}$$

where

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx,$$

and $\mathbb{R}$ is the set of real numbers.

2.2.2 Some Definitions

Differential Entropy

Consider a continuous random variable X having probability density function $f_X(x)$. The *differential entropy* of the variable X is defined as

$$h(X) \triangleq \int_{-\infty}^{\infty} f_X(x) \log_2 \left[\frac{1}{f_X(x)} \right] dx = E[-\log_2(f_X(x))], \tag{2.11}$$

if the integral exists, and is expressed in bits.

The differential entropy of Gaussian variables is important in the study of several communication channels of practical interest. Its expression is derived in the following.

The differential entropy of the random variable $X \sim \mathcal{N}(0, \sigma^2)$ is

$$\begin{aligned} h(X) &= -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \log_2 \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \right) dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \left[\frac{1}{2} \log_2(2\pi\sigma^2) + \frac{x^2}{2\sigma^2} \log_2(e) \right] dx \\ &= \frac{1}{2} \log_2(2\pi\sigma^2) + \frac{1}{2} \log_2(e) \\ &= \frac{1}{2} \log_2(2\pi e\sigma^2) \qquad \text{bits.} \end{aligned} \tag{2.12}$$

Given two random variables X and Y distributed according to $f_X(x)$ and $f_Y(y)$, respectively, and having joint pdf $f_{XY}(x, y) = f_{X|Y}(x|y) f_Y(y)$, the *conditional differential entropy* is defined as

$$h(X|Y) \triangleq \int_{-\infty}^{\infty} f_{XY}(x, y) \log_2 \left[\frac{1}{f_{X|Y}(x|y)} \right] dxdy. \tag{2.13}$$

Mutual Information

The mutual information $I(X;Y)$ between the random variables X and Y having joint pdf $f_{XY}(x, y)$ and marginal pdf's $f_X(x)$ and $f_Y(y)$, respectively, is defined as

$$I(X;Y) \triangleq \int_{-\infty}^{+\infty} f_{XY}(x, y) \log_2 \left[\frac{f_{XY}(x, y)}{f_X(x) f_Y(y)} \right] dxdy. \tag{2.14}$$

The mutual information can also be expressed as a function of the differential entropy:

$$I(X;Y) = h(X) - h(X|Y) = h(Y) - h(Y|X) = h(X) + h(Y) - h(X, Y). \tag{2.15}$$

Joint Typicality

Given a sequence of i.i.d. random variables $X_1, \ldots, X_n$ distributed according to the probability density function $f_X(x)$, for a given $\epsilon > 0$ and for any n the *typical set* $A_\epsilon^{(n)}$ is defined as

$$A_\epsilon^{(n)} \triangleq \left\{ (x_1, \ldots, x_n) \in S^n : \left| -\frac{\log_2[f_X(x_1, \ldots, x_n)]}{n} - h(X) \right| \leq \epsilon \right\}, \tag{2.16}$$

where S^n is the set of all possible length-n sequences. Let us now define the *volume* $Vol(A)$ of a set $A \in \mathbb{R}^n$ as:

$$Vol(A) = \int_A dx_1 \ldots dx_n. \tag{2.17}$$

The following is true for the joint typical set $A_\epsilon^{(n)}$:

1. $Pr\{(x_1, \ldots, x_n) \in A_\epsilon^{(n)}\} > 1 - \epsilon$ for n sufficiently large.

2. $Vol(A_\epsilon^{(n)}) \leq 2^{n(h(X)+\epsilon)}$ for all n.

3. $Vol(A_\epsilon^{(n)}) \geq (1-\epsilon)2^{n(h(X)-\epsilon)}$ for n sufficiently large.

The above mentioned properties follow from the weak law of large numbers, and particularly from the fact that:

$$-\frac{1}{n}\log_2[f(x_1, \ldots, x_n)] \to E[-\log_2(f_X(x))] = h(X) \text{ in probability.} \tag{2.18}$$

Channel Capacity

A time-discrete communication channel with continuous input X and continuous output Y is characterized by the conditional pdf $f_{Y|X}(y|x)$ that describes the statistical relationship between the input and the output of the channel. The *channel capacity* C is defined as the maximum of the mutual information between the input and the output of the channel taken over all input distributions, i.e.:

$$C \triangleq \max_{f_X(x)} I(X;Y). \tag{2.19}$$

The channel capacity is the maximum rate of information, measured in bits per channel use (bpcu), that can be transferred from the input to the output of the channel with arbitrarily small probability of error.

2.2.3 The Gaussian Channel

A *Gaussian channel* is a communication channel for which the following holds:

$$Y = X + Z, \tag{2.20}$$

where X and Y are the input and the output of the channel, respectively, while $Z \sim \mathcal{N}(0, N)$. Most practical systems have a constraint on the average power transmitted. This can be accounted for in the Gaussian channel model by imposing the constraint on the input $\sum_{i=1}^{n} |x_i|^2 \leq nP$, P being the maximum average input power. For this channel model we have that

$$\underset{f_X(x):\sum_{i=1}^{n}|x_i|^2 \leq nP}{\arg\max} I(X;Y) = \frac{1}{\sqrt{2\pi P}} e^{-\frac{x^2}{2P}}, \tag{2.21}$$

that is, the mutual information between the input and the output of a Gaussian channel with power constraint P is maximized by using a Gaussian input alphabet with variance P. The capacity of the Gaussian channel with power constraint can be derived from Eqn. (2.12) and Eqn. (2.15):

$$C \triangleq \max_{f_X(x):\sum_{i=1}^{n}|x_i|^2 \leq nP} I(X;Y) = \frac{1}{2}\log_2\left(1+\frac{P}{N}\right) \text{ bpcu.} \tag{2.22}$$

Channel Coding Theorem

Definition: an (M, n) channel code for the Gaussian channel with power constraint P consists of:

1. A set of indexes $\{1, \ldots, M\}$ representing the information messages.
2. An encoding function $x : \{1, \ldots, M\} \rightarrow \mathcal{X}^n$ that associates a codeword $x^n(w)$ to the index w, $w \in \{1, \ldots, M\}$, such that
$$\sum_{i=1}^{n} |x_i|^2 \leq nP.$$
3. A decoding function that associates the received sequence Y^n to an index $\hat{w}$, i.e.:
$$g : \mathcal{Y}^n \rightarrow \{1, \ldots, M\}.$$

An information rate R can be adopted over a Gaussian channel with an arbitrarily small probability of error (i.e., R is *achievable*) provided that $C \geq R$. R can be achieved as follows:

- A $(2^{nR}, n)$ channel code according to a Gaussian distribution with variance $P - \epsilon$ is generated.
- The code is revealed to both the transmitter and the receiver.
- The transmitter selects the codeword $X^n(w)$ relative to the message index w to be transmitted.
- The receiver selects the codeword that is jointly typical with the received sequence Y^n. If more than a codeword is jointly typical with Y^n or if there is no such codeword, an error is declared.

Further details and proofs can be found in [24].

2.3 DVB Standards for Satellite Communications

2.3.1 Forward link: Introduction to ETSI DVB-SH Standard

The DVB-SH standard provides broadcast services over a hybrid satellite and terrestrial infrastructure operating at frequencies below 3 GHz to a variety of portable, mobile and fixed terminals. Target terminals have compact antennas with limited directivity and include handheld, defined as light-weight and battery-powered apparatus (e.g., mobile phones), vehicle-mounted, nomadic (e.g. laptops, palmtops, etc.) and stationary terminals.

The broadcast services provided by the DVB-SH include streaming services such as television, radio programs as well as download services. The DVB-SH system coverage is obtained by combining a Satellite Component (SC) and, where necessary, a Complementary Ground Component (CGC) to ensure service continuity in areas where the satellite alone can not provide the required QoS. The SC ensures wide area coverage while the CGC provides cellular-type coverage. All types of environment (outdoor, indoor, urban, suburban and rural) can be served.

The CGC consists of terrestrial repeaters fed by a satellite (DVB-S/S2) or terrestrial (fibre, xDSL, etc.) broadcast distribution network. Three kinds of terrestrial repeaters (TR) are envisaged: TR(a) are broadcast infrastructure transmitters which complement reception in areas where satellite reception is difficult, particularly in urban areas; they may be collocated with mobile cellular base stations or standalone. Local content insertion at that level is possible, relying on adequate radio frequency planning and/or waveform optimizations. TR(b) are personal gap-fillers of limited coverage providing local re-transmission, on-frequency and/or with frequency conversion; a typical application is indoor coverage provision, locally repeating the satellite signal available outdoor. No local content insertion is foreseen. TR(c) are mobile broadcast infrastructure transmitters creating a “moving complementary infrastructure” on board moving platforms (cars, trains, bus). Depending on waveform configuration and radio frequency planning, local content insertion may be possible [25].

2.3.2 Return link: Introduction to ETSI DVB-RCS Standard

DVB-Return Channel Satellite (RCS) is a technical standard that defines a complete air interface specification for two-way satellite broadband very small aperture terminal (VSAT) sys-

tems. Low cost VSAT equipment can provide dynamic, demand-assigned transmission capacity to residential and commercial/institutional users. DVB-RCS provides users with the equivalent of an ADSL or cable Internet connection, without the need for local terrestrial infrastructure. Depending on the satellite link budget and other system design parameters DVB-RCS implementations can dynamically provide up to 20 Mbit/s to each terminal on the forward link, and up to 5 Mbit/s or more from each terminal on the return link.

The core of DVB-RCS is a multi-frequency Time Division Multiple Access (MF-TDMA) transmission scheme for the return link, in which a frame may include several carriers. The demand-assignment scheme uses several capacity allocation mechanisms that allow optimization for different classes of applications, so that voice, video streaming, file transfers and web browsing can all be handled efficiently. DVB-RCS supports several access schemes aiming at increasing the efficiency with respect to traditional demand-assigned satellite systems. These access schemes are combined with a flexible transmission scheme that includes state-of-the-art turbo coding, several burst size options and efficient IP encapsulation options [26].

Chapter 3

Cooperative Coverage Extension in Land Mobile Satellite Networks

"[...] e su tempus presente mi anto 'e tennere
cun su passadu e su chi devet bennere."

Mario Masala

Contents

3.1 Introduction

In the last decade several proprietary solutions as well as open standards such as the DVB-SH [25] have been developed to enable data broadcasting via satellite to mobile users. Today there exist lad mobile satellite (LMS) solutions already implemented for maritime and aeronautical communications. Satellite broadcast and relaying capabilities give rise to the possibility of creating mobile broadcast systems over wide geographical areas, which opens large market possibilities for both handheld and vehicular user terminals. Mobile broadcasting is of paramount importance for services such as digital TV or M2M communication, a new communication paradigm which will bring about a tremendous increase in the number of deployed wireless terminals [27].

Coverage, intended as the possibility for all nodes to correctly receive the data transmitted by a central node (e.g., a satellite or a base station), is a main issue for networks with a large number of terminals. As an example, in M2M networks reliable broadcast transmission is of primary importance for terminal software and firmware update, in which all terminals need to correctly receive all the data or, for instance, navigation maps update in vehicle-mounted positioning systems. Protocols such as ARQ, although very effective in point-to-point communication (see, e.g., [24, section 7.1.5]), may not be applicable in multicast contexts, as there may be many retransmission requests by the terminals in case packets are lost, which would saturate the return channel and overwhelm the source [28]. A cooperative approach may be applied in heterogeneous networks [27], in which terminals are equipped with both a long range and a short range communication interface.

A lot of work has been done on the use of cooperation in multicast and broadcast communications in both terrestrial [29][30] and satellite networks [31][32][33]. Many of the proposed solutions [28][34][8] are based on network coding [1], that can achieve the Max-flow Min-

cut capacity in ad-hoc networks. Cooperative content dissemination from road side units to vehicular networks based on rateless codes has also been studied in [35] and [36].

In the present chapter we consider the application of network coding for cooperative coverage extension in satellite broadcast channels. Our main contributions are the following:

1) We carry out an analytical study on the benefits and limits of a cooperative approach in providing missing coverage in single multicast networks. We consider a mathematically tractable and yet practically interesting network model, in which fading and shadowing in the communication channels as well as the medium access mechanism of the ad-hoc network are taken into account. By applying the Max-flow Min-cut theorem we derive an analytical lower bound on the coverage as a function of both the transmission rate at physical level and the rate of innovative packets per unit-time at link level. Our results show a tradeoff between the coverage and the rate at which the information can be injected in the network, and at the same time quantify the gain deriving from cooperation, giving hints on how to tune important parameters such as the medium access probability.

2) We propose a practical scheme for cooperative coverage extension in heterogeneous satellite LMS systems. We adopt the Digital Video Broadcasting - Satellite services to Handheld (DVB-SH) [25] as a reference. The importance of coverage extension in LMS systems stems from the fact that only terminals with an adequate channel quality are able to access satellite services. Poor channel conditions frequently occur due to the shadowing effect of surrounding environment especially in case of low satellite elevation angles. In order to counteract channel impairments, terrestrial repeaters, called *gap-fillers*, and the LL-FEC [25], are envisaged in DVB-SH. The gap-filler solution has two main shortcomings: i) it is a fixed solution which is not able to react quickly to changes in the propagation environment, which may create new dead spots; ii) it is very costly in terms of investment, management, and bandwidth usage. A hybrid satellite-terrestrial networking approach has several advantages with respect to the fixed gap-filler solution as we will argue later. We focus on vehicular terminals and adopt the IEEE 802.11p as reference standard for node-to-node communication in the capillary network. In the proposed scheme the space segment of DVB-SH is not modified. Network coding is merged with the DVB-SH LL-FEC in the terrestrial segment. The proposed ad-hoc cooperative approach based on Network Coding aims at improving coverage while reducing the number of fixed gap-fillers. To the best of our knowledge, the proposed approach was not

applied before to enhance coverage in multicast satellite transmission.

3) We compare the performance of the proposed scheme with our analytical results in order to study the gap deriving from a practical implementation. Furthermore, we evaluate the performance of the proposed system in a challenging propagation environment, such as the urban LMS scenario with a low satellite elevation angle, through extensive simulations in which state-of-the-art satellite channel models are used. The simulator interface between the physical channel and the protocol stack at higher layers leverages on Physical Layer Abstraction (PLA) [37] [38]. More specifically, we take the *received bit mutual information rate (RBIR)* approach, which has been adopted in the standardization process of the IEEE 802.16 standard [39]. The PLA allows to accurately predict the link level performance of a communication system in a computationally simple way. Our simulations show how the proposed scheme enhances the coverage with respect to a system where simple relaying is used and can help to decrease the system outage probability in areas with low gap-filler density.

The rest of the chapter is organized as follows. In Section 3.2 the system model is described. In Section 3.3 the non cooperative case is studied, while in Section 3.4 the proposed cooperative scheme is presented, and analytical expressions for the coverage are derived. In Section 3.5 we introduce the DVB-SH standard, which is used as reference standard in Section 3.6, where a practical scheme called Network-COded Cooperative COverage (N-COCCO) is described, which implements the cooperative approach described in Section 3.4. In Section 3.8 we describe the adopted simulation approach and the simulation setup, while in Section 3.9 we present the numerical results. Finally, Section 3.10 contains the conclusions.

3.2 System Model

A network is considered in which a source (S), that may represent either a satellite or a base station, has a set of K source messages $\mathbf{w}_1, \ldots, \mathbf{w}_K$ of k bits each, to broadcast to a population of terminal nodes, each of which has both long range an short range communication capabilities. No feedback is assumed from the terminals to the source and no channel state information channel state information (CSI) is assumed at S, which implies a non-zero probability of packet loss. S protects each message using a channel code, in order to decrease the

probability of packet loss on the channel. A second level of protection is also applied by S at packet level in order to compensate for packet losses. The encoding at packet level takes place before the channel encoding. $N \geq K$ coded packets are created by S applying a random linear network code (RLNC) to the K source messages. We define $R = K/N$ as the rate of the NC encoder at S. Network coding operates in a finite field of size q ($GF(q)$), so that each message is treated as a vector of $k/\log_2(q)$ symbols. Source messages are linearly combined to produce encoded packets. An encoded packet $\mathbf{x}$ is generated as follows:

$$\mathbf{x} = \sum_{i=1}^{K} \varrho_i \mathbf{w}_i,$$

where ϱ_i, $i = 1, \ldots, K$ are random coefficients drawn according to a uniform distribution in $GF(q)$. The coefficients ϱ_i, $i = 1, \ldots, K$, are appended to each message $\mathbf{x}$ before its transmission. The set of appended coefficients represents the coordinates of the encoded message $\mathbf{x}$ in $GF(q)$ with respect to the basis $\{\mathbf{w}_i\}$, $i = 1, \ldots, K$, and is called *global encoding vector*.

The encoding at the physical layer is applied to network-encoded packets, each consisting of of k bits. The transmitter encodes a packet using a Gaussian codebook of size 2^{nr}, with $r = \frac{k}{n}$ bits per channel use (bpcu), associating a codeword $\mathbf{c}_m$ of n i.i.d. symbols drawn according to a Gaussian distribution to each $\mathbf{x}_m$, $m = 1 \ldots, N$ [24]. The time needed for S to transmit a packet is called *transmission slot (TS)*.

Terminals cooperate in order to recover the packets that are lost in the link from the transmitter. We assume terminals with high mobility, which is the case of, e.g., vehicular networks. Thus, nodes have little time to set up connections with each other. For this, and in order to exploit the broadcast nature of the wireless medium, nodes act in *promiscuous mode*, broadcasting packets to all terminals within reach. Similarly as in the broadcast mode of IEEE 802.11 standards, no request to send (RTS)/clear to send (CTS) mechanism is assumed [40]. No CSI is assumed at the transmitter in the terminal to terminal communication, so that there is always a non zero probability of packet loss. Like the source, also each terminal uses two levels of encoding, which are described in the following.

Let L be the number of packets correctly decoded at the physical level by a terminal. The terminal selects the $L' \leq L$ packets which constitute the largest set of linearly independent packets with respect to the basis $\mathbf{w}_i$, $i = 1, \ldots, K$. Without loss of generality we assume that such set be $\mathbf{x}_1, \ldots, \mathbf{x}_{L'}$. Linear independence can be verified through the global encoding vectors of the packets. The L' selected packets are re-encoded together using random linear

network coding (RLNC), and then re-encoded at the physical layer. RLNC encoding at the terminals works as follows. Given the set of received packets $\mathbf{x}_1, \ldots, \mathbf{x}_{L'}$, the message $\mathbf{y} = \sum_{m=1}^{L'} \sigma_m \mathbf{x}_m$ is generated, σ_m, $m = 1, \ldots, L'$, being coefficients drawn at random according to a uniform distribution in $GF(q)$. Each time a new encoded message is created, it is appended its global encoding vector. The overhead this incurs is negligible if messages are sufficiently long [41]. The new global encoding vector $\boldsymbol{\eta}$ can be easily calculated by the transmitting node as follows:

$$\boldsymbol{\eta} = \boldsymbol{\sigma}\boldsymbol{\Psi},$$

where $\boldsymbol{\sigma} = [\sigma_1 \ \cdots \ \sigma_{L'}]$ is the *local encoding vector*, i.e., the vector of random coefficients chosen by the transmitting node, while $\boldsymbol{\Psi}$ is an $L' \times K$ matrix that has the global encoding vector of $\mathbf{x}_m$, $m = 1, \ldots, L'$, as row m. We assume that the transmission of a message by a terminal is completed within one TS. The physical layer encoding at a mobile node takes place in the same way as at the source, and using the same average transmission rate r.

3.2.1 Source-to-Node Channel Model

The channel from the source S to a generic terminal N_i (S-N channel) is affected by both Rayleigh fading and log-normal shadowing. The power of the signal received at the terminal is modeled as the product of a unit-mean exponential random variable γ and a log-normal random variable Γ_S which accounts for large scale fading. This model has been largely used to model propagation in urban scenarios [42] and, with some modifications, in LMS systems [43]. The fading coefficient γ takes into account the fast channel variations due to the terminal motion and is assumed to remain constant within a TS, while changing in an i.i.d. fashion at the end of each channel block. The shadowing coefficient Γ_S includes the transmitted power at S and accounts for the obstruction of buildings in the line of sight and changes much slowly with respect to γ. For mathematical tractability we assume that Γ_S remains constant for N channel blocks, i.e., until all encoded packets relative to the K source messages have been transmitted by S. We call the time needed to transmit N messages a *generation period (GP)*. The fading and shadowing processes of two different nodes are assumed to be independent. We further assume that shadowing and fading statistics are the same for all nodes, which is the case if nodes are located at approximately the same distance from S.

A message is lost in the S-N channel if the instantaneous channel capacity is lower than

the transmission rate at the physical layer r. Thus the packet loss probability in the S-N channel for a generic node is:

$$P_{SN} = Pr\left\{\log_2(1+\gamma\Gamma_S) < r\right\}, \tag{3.1}$$

where $\gamma \sim exp(1)$ while $\Gamma_S = e^{\frac{X}{10}}$ with $X \sim \mathcal{N}(\mu, \sigma^2)$. Γ_S is constant within a GP, while γ changes independently at the end of each channel block. Fixing the value of Γ_S, the packet loss probability P_{SN} in the S-N link is:

$$P_{SN} = 1 - e^{\frac{1-2^r}{\Gamma_S}}. \tag{3.2}$$

In the rest of the chapter we will use the expressions "packet loss rate" and "probability of packet loss" interchangeably. Due to shadowing, Γ_S changes randomly and independently at each generation period and, within a generation, from one node to the other. Thus the packet loss rate P_{SN} is also a random variable that remains constant within a generation and changes in an i.i.d. fashion across generations and terminals.

3.2.2 Node-to-Node Channel Model

We model the channels between the transmitting terminal and each of the receiving terminals (N-N channel) as independent block fading channels, i.e., the fading coefficient of each channel changes in an i.i.d. fashion at the end of each channel block. The probability of packet loss in the N-N channel P_{NN} is:

$$P_{NN} = Pr\left\{\log_2(1+\gamma\Gamma_N) < r\right\} = 1 - e^{\frac{1-2^r}{\Gamma_N}}, \tag{3.3}$$

where Γ_N accounts for path loss and transmitted power, and is assumed to remain constant for a whole generation period and across terminals. In order not to saturate the terrestrial channel, we assume that a node can transmit at most one packet within one TS.

3.3 Non-Cooperative Coverage

Let us consider a network with a source S and M terminals. We define the *coverage* (Ω) as the probability that all M terminals correctly decode the whole set of K source messages.

Assuming that K is large enough, and using the results in [28], the probability that node N_i can decode all the K source messages of a given generation in case of no cooperation is:

$$Pr\left\{P_{SNi} < 1 - R\right\} = F_{P_{SNi}}\left(1 - R\right), \tag{3.4}$$

$F_{P_{SN}}$ being the cumulative density function (cdf) of P_{SN} and $R = K/N$ being the rate of the NC encoder at S. We recall that, due to the shadowing, the packet loss rate P_{SN} is a random variable which changes in an i.i.d. fashion across generations and terminals. Plugging Eqn. (3.2) into Eqn. (3.4) we find:

$$Pr\left\{1 - e^{\frac{1-2^r}{\Gamma_S}} < 1 - R\right\}. \tag{3.5}$$

The coverage is the probability that each of the nodes decodes all source messages, that is:

$$\Omega = Pr\left\{P_{SN1} < 1 - R, \ldots, P_{SNM} < 1 - R\right\}, \tag{3.6}$$

where P_{SNi} is the packet loss rate in the S-N link of node N_i, $i = 1, \ldots, M$. Under the assumption of i.i.d. channels we have $F_{P_{SNi}} = F_{P_{SN}}$, $\forall i \in \{1, \ldots, M\}$. Thus, (3.6) can be written as:

$$\Omega = \left(Pr\left\{P_{SN} < 1 - R\right\}\right)^M = F_{P_{SN}}^M(1 - R). \tag{3.7}$$

$F_{P_{SN}}(y)$, which is derived in the Appendix of this chapter, is the cdf of P_{SN}, and takes the form:

$$F_{P_{SN}}(y) = \frac{1}{2} - \frac{1}{2}\mathrm{erf}\left(\frac{10\ln\left[\frac{1-2^r}{\ln(1-y)}\right] - \mu}{2\sigma^2}\right), \tag{3.8}$$

for $y \in (0, 1)$, where $\mathrm{erf}(x)$ is the *error function*, defined as $\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$. Finally, plugging Eqn. (3.8) into Eqn. (3.7), we find the coverage in the non cooperative case:

$$\Omega = \frac{1}{2^M}\left[1 - \mathrm{erf}\left(\frac{10\ln\left[\frac{1-2^r}{\ln(R)}\right] - \mu}{2\sigma^2}\right)\right]^M, \tag{3.9}$$

for $R \in (0, 1)$. Note that, fixing R and M, the expression in Eqn. (3.9) goes to 0 as the rate at physical level r goes to infinity. This confirms the intuition according to which in the non-cooperative case the coverage decreases as the transmission rate increases. As said previously, this result holds for any value of q as long as K is large enough. Thus, Eqn. (3.9) can also be interpreted as the coverage in a network of M nodes in presence of fading and shadowing that can be achieved for a rate couple (r, R) by a fountain code such as, e.g., a Raptor code.

3.4 Cooperative Coverage

The wireless network is modeled as a directed hypergraph $\mathcal{H} = (\mathcal{N}, \mathcal{A})$, $\mathcal{N}$ being a set of nodes and $\mathcal{A}$ a set of hyperarcs. A hyperarc is a pair (i, J), where i is the *head* node of the hyperarc while J is the subset of $\mathcal{N}$ connected to the head through the hyperarc. J is also called *tail*. A hyperarc (i, J) can be used to model a broadcast transmission from node i to nodes in J. Packet losses can also be taken into account. We want to study the relationship between the coverage and the rate at which the information is transferred to mobile terminals, which depends on both the rate at physical level r and the rate at which new messages are injected in the network, which is the rate at packet level R. In [28] (Theorem 2) it is shown that, if K is large, random linear network coding achieves the network capacity in wireless multicast and unicast connections, even in case of lossy links, if the number of innovative packets transmitted by the source per unit of time is lower than or equal to the flow across the minimum flow cut between the source and each of the sink nodes, i.e.:

$$R \leq \min_{Q \in \mathcal{Q}(S,t)} \left\{ \sum_{(i,J) \in \Gamma_+(Q)} \sum_{T \nsubseteq Q} z_{iJT} \right\} \tag{3.10}$$

where z_{iJT} is the average injection rate of packets in the arcs departing from i to the tail subset $T \subset J$, $\mathcal{Q}(S, t)$ is the set of all cuts between S and t, and $\Gamma_+(Q)$ denotes the set of forward hyperarcs of the cut Q, i.e.:

$$\Gamma_+(Q) = \{(i, J) \in \mathcal{A} | i \in Q, J \setminus Q \neq 0\}. \tag{3.11}$$

In other words, $\Gamma_+(Q)$ denotes the set of arcs of Q for which the head node is on the same side as the source, while at least one of the tail nodes of the relative hyperarc belongs to the other side of the cut. The rate z_{iJT} is defined as:

$$z_{iJT} = \lim_{\tau \to \infty} \frac{A_{iJT}(\tau)}{\tau}, \tag{3.12}$$

where $A_{iJT}(\tau)$ is the counting process of the packets sent by i that arrive in $T \subset J$ in the temporal interval $[0, \tau)$. The existence of an average rate is a necessary condition for the applicability of the results in [28].

In the following we derive z_{iJT} for the considered network setup as a function of both physical layer and medium access control (MAC) layer parameters such as transmission rate, transmission power and medium access probability.

3.4.1 Multiple Access

Let us consider a network with M nodes. We assume that all nodes have independent S-N and N-N channels. We further assume that channel statistics are the same for all terminals, which is the case if the distances from node N_i to node N_j change little $\forall i, j \in \{1, \ldots, M\}$, $i \neq j$ and with respect to each node's distance to the source.

In our setup the terminals are set in *promiscuous mode* so that each node can receive the broadcast transmissions of any other node [40]. The terminals share the wireless medium, i.e., they transmit in the same frequency band. We assume that a carrier sense multiple access (CSMA)/collision avoidance (CA) protocol is adopted by the nodes and that all nodes hear each other, so that the medium is shared among the terminals willing to transmit but no collision happens.

We now derive the communication rate z_{iJT}. We start by finding out the communication rate z_{ij} between a transmitting node N_i and a receiving node N_j. By the symmetry of the problem all links have the same average rate. Consider the generic node N_i. The average transmission rate from node N_i to node N_j is:

$$\begin{aligned} z_{i,j} &= p_a \cdot Pr\{\text{No one else transmits}\}(1 - P_{NN}) \\ &= p_a \cdot [Pr\{\text{No one else tries to transmit}\} \\ &+ Pr\{N_i \text{ wins contention}\}](1 - P_{NN}), \end{aligned} \tag{3.13}$$

where p_a is the probability that a node tries to contend for the channel, and is fixed for all nodes. The first term in the sum of Eqn. (3.13) is:

$$Pr\{\text{No one else tries to transmit}\} = (1 - p_a)^{M-1}. \tag{3.14}$$

The second term in the sum of Eqn. (3.13) is the probability that one or more other nodes try to access the channel, but N_i transmits first. To calculate this probability, we note that, if k other nodes try to access the channel (for a total of $k + 1$ nodes trying to access the channel), the probability for each of them to occupy the channel before the others is $1/(k + 1)$, which

follows by the symmetry of the problem. Thus we can write:

$$
\begin{aligned}
Pr\left\{N_i \text{ wins contention}\right\} &= \sum_{k=1}^{M-1} \binom{M-1}{k} \frac{p_a^k (1-p_a)^{M-1-k}}{k+1} \\
&= \frac{1}{Mp_a} \sum_{k=1}^{M-1} \binom{M}{k+1} p_a^{k+1} (1-p_a)^{M-1-k} \\
&= \frac{1}{Mp_a} \sum_{k=2}^{M} \binom{M}{k} p_a^{k} (1-p_a)^{M-k} \\
&= \frac{1}{Mp_a} \left[1 - \binom{M}{0} (1-p_a)^M - \binom{M}{1} p_a (1-p_a)^{M-1} \right] \\
&= \frac{1}{Mp_a} \left[1 - (1-p_a)^M - Mp_a (1-p_a)^{M-1} \right]. \qquad (3.15)
\end{aligned}
$$

Plugging (3.14) and (3.15) into (3.13) we obtain:

$$z_{i,j} = \frac{1-(1-p_a)^M}{M}(1-P_{NN}). \qquad (3.16)$$

Using the definition given in Eqn. (3.12) together with Eqn. (3.16), we finally find

$$z_{iJT} = \frac{1-(1-p_a)^M}{M} \left[1 - (P_{NN})^{|T|} \right], \qquad (3.17)$$

where $|T|$ is the cardinality of T, and the term $\left[1-(P_{NN})^{|T|}\right]$ is the probability that at least one of the $|T|$ nodes whose S-link belongs to the cut receives correctly a transmission from a node that is in the other side of the cut. In other words, Eqn. (3.17) is the rate at which packet are received by the set T considered as a single node, that is, the counting process $A_{iJT}(\tau)$ increases of one unit when at least one of the terminals in T receives one packet, independently from how many terminals receive it.

3.4.2 Coverage Analysis

In the following we derive the conditions that maximize the coverage as a function of relevant network parameters by applying the Max-flow Min-cut theorem. We recall that such maximum coverage can be attained by using the random coding scheme described in Section 3.2.

Let us consider Eqn. (3.10). For each of the M nodes we must consider all the possible cuts of the network such that the considered node and the satellite are on different sides of the cut. Let us fix a receiving node N_t. We recall that a cut is a set of edges that, if removed from a graph, separates the source from the destination. Fig. 3.1 gives an example of a network with four nodes where the cut Q_{SN_4} (i.e., the cut such that N_4 and S are on the same side) is put into

evidence. In the example, the destination node is $N_t = N_1$. The dashed black lines represent the edges which are to be removed to get the cut. Note that the set of nodes that receive from S (only node N_4 in the figure) are isolated by the cut from the nodes with satellite cut (nodes N_1, N_2 and N_3 in Fig. 3.1). We define an *S-edge* as an edge of the kind $(S, N_j), j \neq t$. We further define a *T-edge* as one of the kind: $(N_j, N_t), j \neq t$. First of all, note that in each possible cut

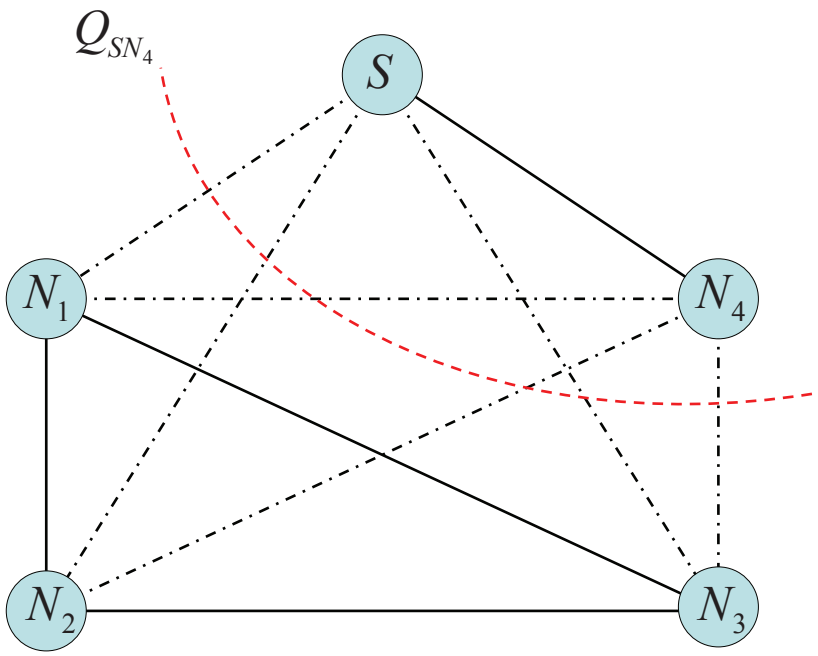


Figure 3.1: Graph model for a network with four terminals. There are $2^{M-1} = 8$ possible cuts for each of the M nodes. The set of nodes that receive from S (only node N_4 in the figure) are isolated by the cut from the nodes with satellite cut.

of N_t the arc joining the node with the source is always present. For the particular network topology considered, the rest of the cuts consist in removing, for each of the $M-1$ remaining nodes, either the S-link or the T-link between the considered node and N_t. The number of possible cuts is thus equal to 2^{M-1}. Two distinct cuts differ in either the number n_s of S-edges which are included in the cut or the identity of the nodes for which the S-edge is part of the cut. For each $N_t \in \mathcal{N}$ and for each cut so that $n_s \in \{1, \cdots, M-1\}$ S-links are present, the average message rate R at the source must be lower than or equal to the capacity of the cut, i.e.:

$$R \leq 1 - \prod_{j \in Q_{n_s}} Y_j + (M - n_s) \frac{1 - (1 - p_a)^M}{M} \left[1 - (P_{NN})^{n_s}\right],$$

that can be rewritten as

$$\alpha(n_s) - \prod_{j \in Q_{n_s}} Y_j \geq 0, \tag{3.18}$$

where Q_{n_s} is one of the cuts with n_s satellite links relative to the node N_t and we defined:

$$\alpha(n_s) = 1 - R + (M - n_s)\frac{1-(1-p_a)^M}{M}\left[1-(P_{NN})^{n_s}\right].$$

The condition in Eqn. (3.18) must hold for any number n_s of S-edges. This is equivalent to imposing a new condition which is the intersection of all the conditions of the kind of Eqn. (3.18), i.e.:

$$\bigcap_{Q_{n_s} \in \mathcal{S}(n_s, \overline{N}_t)} \left[\prod_{j \in Q_{n_s}} Y_j \leq \alpha(n_s) \right], \tag{3.19}$$

where $\mathcal{S}(n_s, \overline{N}_t)$ is the set of subsets of $\mathcal{N} \backslash N_t$ with n_s elements. The number of elements in $\mathcal{S}(n_s, \overline{N}_t)$ is $\binom{M-1}{n_s}$, as each of them is obtained by choosing n_s elements from a set with cardinality $M-1$. As previously said, for a given N_t to decode all messages, the condition on the flow must be satisfied for all cuts, which is equivalent to imposing the condition (3.19) for all n_s. Finally, in order for all nodes to decode all source messages the condition on the minimum flow cut must hold $\forall t \in \mathcal{N}$. Imposing this, we obtain the expression for the coverage that is reported in Eqn. (3.20) at the bottom of the page.

3.4.3 Lower Bound on Achievable Coverage

Although Eqn. (3.20) might be used to evaluate Ω numerically, a closed-form expression would give more insight into the impact of cooperation on the considered setup. Finding a simple closed form expression for Eqn. (3.20) is a challenging task. Thus in the following we derive a lower bound Ω_{LB} on Ω. Ω can be lower bounded by substituting in Eqn. (3.20) the packet loss

$$\Omega = Pr\left\{ \bigcap_{N_t \in \mathcal{N}} \bigcap_{n_s \in \{1,\ldots,M-1\}} \bigcap_{Q_{n_s} \in \mathcal{S}(n_s, \overline{N}_t)} \left[\prod_{j \in Q_{n_s}} Y_j < 1 - R + (M - n_s)\frac{1-(1-p_a)^M}{M}\left[1-(P_{NN})^{n_s}\right] \right] \right\}. \tag{3.20}$$

rate Y_j for each cut with the largest packet loss rate among all the S-links in the network, i.e.:

$$\begin{aligned}
\Omega &= Pr\left\{\bigcap_{N_t\in\mathcal{N}}\bigcap_{n_s\in\{1,\dots,M\}}\bigcap_{Q_{n_s}\in\mathcal{S}(n_s,\overline{N}_t)}\left[\prod_{j\in Q_{n_s}} Y_j < \alpha(n_s)\right]\right\} \\
&\geq Pr\left\{\bigcap_{N_t\in\mathcal{N}}\bigcap_{n_s\in\{1,\dots,M\}}\left[\prod_{j=1}^{n_s} Y_{(j)} < \alpha(n_s)\right]\right\} \\
&\geq Pr\left\{\bigcap_{N_t\in\mathcal{N}}\bigcap_{n_s\in\{1,\dots,M\}}\left[Y_{(1)}^{n_s} < \alpha(n_s)\right]\right\} \\
&= Pr\left\{\bigcap_{N_t\in\mathcal{N}}\bigcap_{n_s\in\{1,\dots,M\}}\left[Y_{(1)} < \sqrt[n_s]{\alpha(n_s)}\right]\right\} \\
&= Pr\left\{Y_{(1)} < \min_{n_s\in\{1,\dots,M\}}\sqrt[n_s]{\alpha(n_s)}\right\} \\
&= F_Y^M(\beta), \qquad (3.21)
\end{aligned}$$

where $Y_{(i)}$ is the i-th largest packet loss rate across all S-edges of the network, i.e., $Y_{(i)} \geq Y_{(j)}$ if $i < j, \forall i, j \in \mathcal{N}$, and we defined

$$\beta = \min_{n_s\in\{1,\dots,M\}}\sqrt[n_s]{\alpha(n_s)}.$$

The first inequality in Eqn. (3.21) derives from the fact that:

$$\prod_{j\in S} Y_j \leq \prod_{j=1}^{n_s} Y_{(j)}, \text{ for } S \in \mathcal{S}(n_s,\overline{t}), \forall\, n_s, t, \qquad (3.22)$$

i.e., we substitute the product of n_s random variables, chosen within a set of M variables, with the product of the n_s largest variables of the same set. The second inequality in Eqn. (3.21) follows from the fact that

$$\prod_{j=1}^{n_s} Y_{(j)} \leq Y_{(1)}^{n_s}, \forall\, n_s, t.$$

By plugging Eqn. (3.8) into Eqn. (3.21) we finally find:

$$\Omega_{LB} = \frac{1}{2^M}\left[1 - \text{erf}\left(\frac{10\ln\left[\frac{1-2^r}{\ln(1-\beta)}\right] - \mu}{2\sigma^2}\right)\right]^M. \qquad (3.23)$$

Example: A Two-nodes Network In order to clarify the concepts just described, in the following we consider the case of a network with only two nodes, such as the one depicted in Fig. 3.2. We start by deriving the communication rates over the terrestrial edge. In each slot node N_i tries to access the channel with probability p_{ai}. In case only node N_i tries to access the channel, the transmission will be successful with probability $1 - P_{LT}$, where P_{LT}

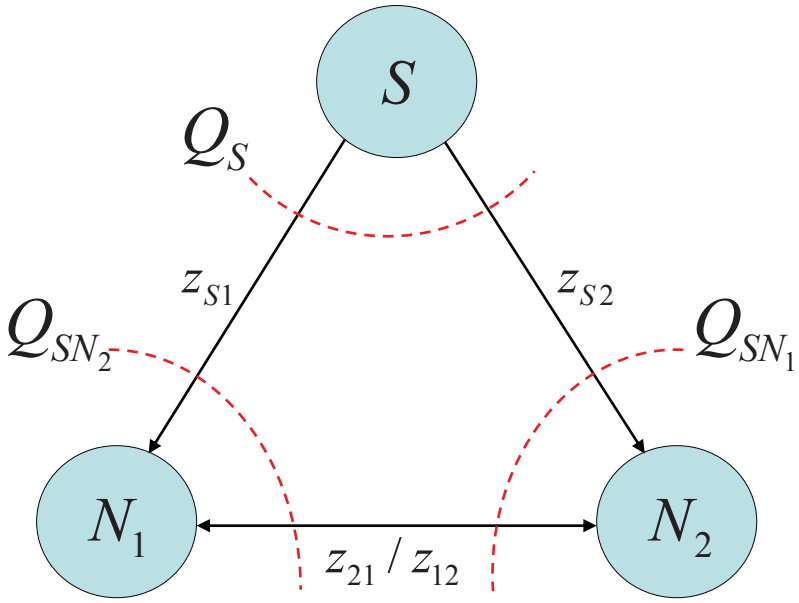


Figure 3.2: Graph model for a network with two nodes. Q_S, Q_{SN_1} and Q_{SN_2} are the three cuts of the network. Q_S is the cut in which the satellite and the nodes lie in different sides, Q_{SN_1} is the cut in which node N_1 is on the satellite side and Q_{SN_2} is the cut in which node N_2 is on the satellite side. z_{ij} is the average injection rate in the edge (i, j).

is the packet loss probability in the link between the two nodes. In case both nodes try to access the channel in the same slot, the CSMA/CA mechanism determines which of the two nodes transmits. By the symmetry of the problem, in case of contention each of the two nodes occupies the channel with probability $1/2$ and the transmission is successfully received by the other node with probability $1 - P_{LT}$. According to Eqn. (3.15), the average rate on the edge (N_1, N_2) can be written as:

$$z_{1,2} = p_{a1}\left[(1-p_{a2})(1-P_{LT}) + \frac{p_{a2}}{2}(1-P_{LT})\right] = p_{a1}\left(1-\frac{p_{a2}}{2}\right)(1-P_{LT}),$$

while

$$z_{2,1} = p_{a2}\left(1-\frac{p_{a1}}{2}\right)(1-P_{LT}).$$

With reference to Fig. 3.2, the cuts in the network graph are: Q_S in which the satellite and the nodes lie in different sides of the cut, Q_{SN_1}, in which node N_1 is on the satellite side and Q_{SN_2},

in which node N_2 is on the satellite side. The conditions on the flows across the three cuts are:

$$\begin{aligned} &Q_S : 1 - P_{LS1} \cdot P_{LS2} \geq R \\ &Q_{SN_1} : 1 - P_{LS2} + p_{a2}(1 - p_{a1})(1 - P_{LT}) \geq R \\ &Q_{SN_2} : 1 - P_{LS1} + p_{a1}(1 - p_{a2})(1 - P_{LT}) \geq R. \end{aligned} \tag{3.24}$$

Hence the maximum achievable rate R^* is:

$$R^* = \min\left\{1 - P_{LS1} \cdot P_{LS2}, 1 - P_{LS2} + p_{a2}(1 - p_{a1})(1 - P_{LT}), 1 - P_{LS1} + p_{a1}(1 - p_{a2})(1 - P_{LT})\right\}. \tag{3.25}$$

Note that, in Eqn. (3.25), P_{LS1} and P_{LS2} are i.i.d. random variables, and thus also R^* is a random variable. As the pair (r, R) is fixed, there is a nonzero probability that $R > R^*$, i.e., the packet injection rate at the satellite is not supported, which implies that either one or both the terminals are not able to recover all source packets. By definition of coverage we have:

$$\Omega = Pr\{R^* \geq R\}. \tag{3.26}$$

Imposing $p_{a1} = p_{a2} = p_a$ we have $z_{1,2} = z_{2,1}$. According to the notation defined in previous subsection, we define $Y_{(1)} = \max\{P_{LS1}, P_{LS2}\}$, $Y_{(2)} = \min\{P_{LS1}, P_{LS2}\}$,

$$\begin{aligned} \alpha(1) &= 1 - R + p_{a1}\left(1 - \frac{p_{a2}}{2}\right)(1 - P_{LT}) \\ &= 1 - R + \frac{1 - (1 - p_a)^2}{2}(1 - P_{LT}), \end{aligned}$$

and

$$\alpha(2) = 1 - R.$$

Finally, applying Eqn. (3.21) we derive the following lower bound on Ω for a network with 2 nodes:

$$\Omega \geq F_Y^2\left(\min\left\{\alpha(1), \sqrt{\alpha(2)}\right\}\right). \tag{3.27}$$

3.5 Cooperative Coverage Extension in DVB-SH

In this section we propose a practical scheme that implements the cooperative approach described in the previous section in heterogeneous satellite vehicular networks.

3.5.1 Space Segment

Satellite Channel

The considered setup is an LMS system with a GEO satellite in L band (or low S band) broadcasting a DVB-SH-B (time-division multiplexing (TDM) waveform from satellite and orthogonal frequency-division multiplexing (OFDM) from the gap-fillers) signal to a population of mobile terminals. Propagation conditions change due mainly to the shadowing effect of building and trees and are classified in urban, suburban and rural. The main cause of channel impairment in urban and suburban environments is the long-lasting shadowing of the buildings that causes an intermittent satellite connectivity, while in the rural propagation scenarios the main source of impairment is tree shadowing. Signal reception in LMS systems is limited by three phenomena:

- Path loss at large scale,
- Shadowing at mid-scale,
- Multipath fading at small scale.

We adopt the Perez-Fontan land mobile satellite (LMS) channel model [44], based on a three-state Markov chain in which the possible states represent line of sight reception, moderate shadowing reception and deep shadowing reception. In each of the states the signal amplitude is modeled as a Loo process (sum of a Log-normal and a Rayleigh random variables) with different parameters. In Fig. 3.3 the channel amplitude of a 10 seconds simulation according to the Perez-Fontan channel model in urban scenario is shown.

Channel Impairment Countermeasures in DVB-SH

In this section we recall the channel impairment countermeasures foreseen in the DVB-SH standard.

Physical Layer The physical layer error protection scheme of the DVB-SH standard is shown in Fig. 3.4. The main blocks of the scheme are the duobinary turbo code with different rates/word lengths, the bit interleaver, which works at bit level within a turbo codeword, and the time interleaver, the depth of which which spans more than one codeword and uses interleaving

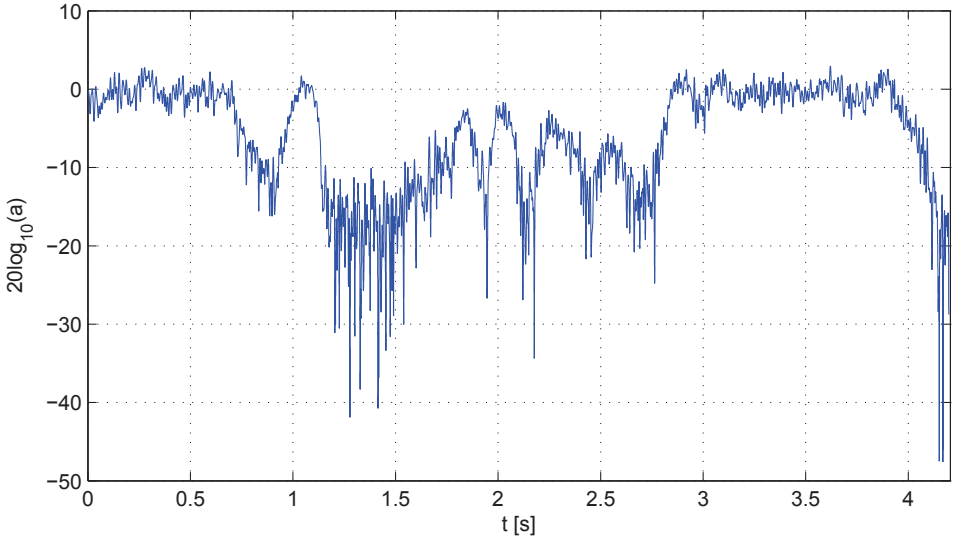


Figure 3.3: Channel realization in urban scenario for a satellite elevation angle 40^o and a node speed of 50 kmph. Building obstruction can lead to long-lasting deep fading reception conditions. An attenuation of 20 dB in the received signal strength lasting about 300 msec can be observed starting from $t = 1.2$ seconds.

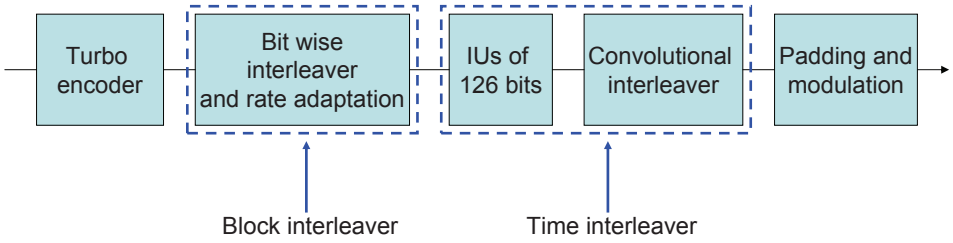


Figure 3.4: Physical layer error protection in DVB-SH-B. The block interleaver works bit wise, while the time interleaver works on blocks of 126 bits, called interleaving units (IU).

blocks of 126 bits each. This last element is particularly important to counteract long blockage periods, as it can span time intervals of up to about 10 seconds. The drawbacks in using a long time interleaver are the large decoding delay and the memory requirements at mobile terminals, which can be met only by high class nodes.

MPE-IFEC in DVB-SH The Multi Protocol Encapsulation (MPE)-Inter-burst Forward Error Correction (IFEC) is a process section between the IP and the transport layers introduced in DVB-SH in order to counteract the disturbances in reception and transmission. This can be achieved by applying FEC over multiple *datagram bursts*, i.e., groups of datagrams. The long high-layer interleaver used in IFEC allows for significant performance enhancements with respect to FEC [25], as it can better counteract long lasting shadowing which is typical of the

LMS channel.

The encoding is made over several datagram bursts. Let us consider a datagram burst entering the MPE-IFEC process. The burst is reshaped in a matrix of T by C bytes called Application Data Sub-Table (ADST) illustrated in Fig. 3.5 [25]. The columns of the ADST

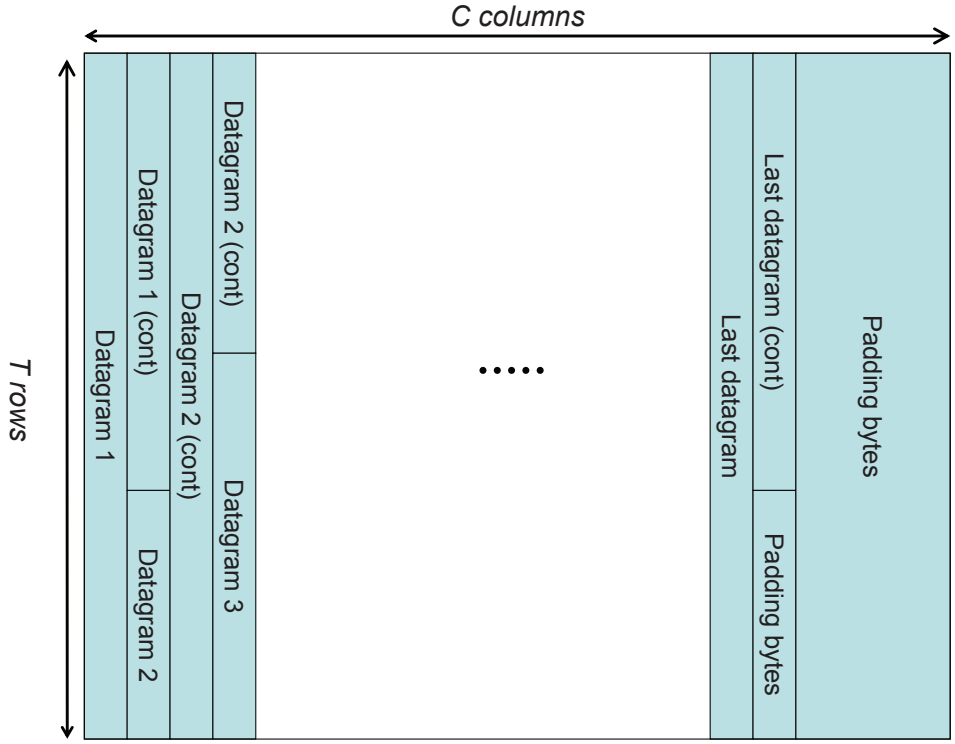


Figure 3.5: ADST reshaping of datagram bursts.

are then distributed in a round robin fashion among B matrices called Application Data Tables (ADT). An ADT is a T by K matrix. The FEC, always systematic, is applied on the ADT producing a T by N_r parity matrix, called IFEC Data Table (iFDT). An ADT is filled up and the encoding takes place every EP bursts, EP being the Encoding Period, which determines the number of datagram bursts over which the parity is calculated. The ADT and the iFDT together form an *encoding matrix*. It takes $B \times EP$ bursts to fill up a single ADT. Once an ADT is full (this happens to B ADT at the same time) the iFDT is calculated. As soon as the B iFDTs are calculated an *IFEC burst* is generated by taking groups of columns from S different iFDTs. An IFEC burst is made up of several IFEC sections. Each section is comprised of a header, a payload containing g columns from the same iFDT and a cyclic redundancy check (CRC). The k-th IFEC burst is merged with the $(k - D)$-th datagram burst (and eventual MPE-FEC redundancy) to form a *time-slice burst*. The time slice burst is then multiplexed on MPEG2-TS

frames and passed down to lower layers.

Depending on the FEC technique applied, different values of EP, B and S are adopted. In case a Raptor code is used EP is generally greater than 1, while $B = S = 1$. This is because Raptor codes, unlike other FEC codes such as Reed-Solomon codes [45], are capable of handling large source matrices (i.e., ADT), that can span several datagram bursts.

Raptor Codes in DVB-SH The Raptor code adopted for the DVB-SH is the same as in the 3GPP standard, which has also been adopted in the DVB-Handheld (DVB-H) standard [25]. Its description can be found in [46]. A source block in [46] corresponds to an ADT and a source symbol is a column of the ADT. Thus a source block has K symbols of T bytes each. The Raptor encoder is applied independently to each source block, each of which is identified by a Source Block Number (SBN). The encoder produces K systematic symbols (the ADT matrix) and N_r *repair* (parity) symbols. Systematic and repair symbols are called *encoding symbols*. Each symbol is identified by an Encoding Symbol Identifier (ESI). Values from 0 to $K - 1$ are assigned to the systematic symbols, while values from K to $N_r + K - 1$ identify repair symbols. The encoding procedure consists of two parts. In the first part L *intermediate symbols* are produced starting from the K source symbols, while in the second part $K + N_r$ encoding symbols are generated starting from the L intermediate symbols.

The intermediate symbols from 0 to $K - 1$ are systematic (i.e., are the source symbols). The S intermediate symbols from K to $K + S - 1$ are generated using an LDPC encoder while the last H symbols from $K + S$ to L are called *Half Symbols* and are generated using a binary reflected Gray encoder [25].

The encoding symbols are generated applying a Luby Transform (LT) encoder to the L intermediate symbols. The LT encoder operates a bit-wise XOR of intermediate symbols chosen according to a certain degree distribution. Each of the encoding symbols is transmitted together with its ESI and a triple (d, a, b) where d is the symbol degree and a and b are integers from the sets $1, \ldots, L'' - 1$ and $0, \ldots, L'' - 1$ respectively, $L'' - 1$ being the smallest prime integer greater than or equal to L. At the end of the encoding process, K systematic symbols plus N_r parity symbols are produced. The parity symbols are linear combinations of systematic symbols in $GF(2)$. The encoding symbol triple together with the ESI and the value K allows the decoder to determine which intermediate symbols (and thus which source symbols) were

combined to form each of the encoding symbols.

3.5.2 Ground Segment

Terminal Types

We consider high class terminals as defined in [47]. High class terminals are not energy constrained and have relatively good computation capabilities and memory [47]. This is the case of vehicular terminals, which are powered by rechargeable batteries and can host computation units of high speed, thanks to the relative low impact in terms of cost, space and weight. We assume that each terminal has both satellite and ad-hoc networking capabilities, which give rise to the possibility of implementing a vehicular capillary network.

Terrestrial Channel

A possible classification of the vehicle-to-vehicle (VtoV) communication channel is based on roadside environments such as buildings, trees and bridges. In this case the propagation scenarios are categorized as urban canyon, suburban street and expressway. Signal propagation in VtoV systems has been studied in [48], [49] and [50].

In [51] a measurements campaign made on the 5.9 GHz frequency is presented. The measurements presented in the paper have been made using a Dedicated Short Range Communication (DSRC)/IEEE 802.11p prototype radio. In the paper a dual slope model for the path loss in urban VtoV scenarios is derived based on real measurements. We adopt the model of [51] for the path loss together with the TU6 multi-tap channel model. An OFDM signal with 52 carriers (48 information carriers) and a rate $1/2$ convolutional encoder are assumed. All physical layer parameters are taken from the 802.11p standard.

As usual practice in VtoV simulations, we assume a finite communication range which is fixed for all vehicles and set to 300 meters.

Furthermore, we take into account collisions as they constitute an important throughput-limiting and delay-increasing factors in ad-hoc wireless networks [52].

3.6 Network-Coded Cooperation for DVB-SH

In the following we describe our proposed cooperative scheme for coverage enhancement in the forward link. We call such cooperation scheme Network-COded Cooperative COverage (N-COCCO). Let us consider a satellite broadcasting a DVB-SH-B signal with MPE-IFEC protection to a population of vehicular terminals with both DVB-SH-B and IEEE 802.11p radio interfaces. During a time window $(0, t)$ the satellite transmits $K + N_r$ IFEC symbols obtained from an ADT. Terrestrial and satellite communications take place in orthogonal frequency bands. Due to long-lasting shadowing caused by urban propagation conditions, it can happen that a user decodes a number of symbols equal to $M < K$ during the interval $(0, t)$. In this case the user cannot decode the entire source data block. In order to enhance satellite coverage each node re-encodes the received packets (either received directly from the satellite or from other terminals) and broadcasts them to nodes within its transmission range. In the following sections we describe the encoding procedure at land mobile nodes.

3.6.1 Encoding at Land Mobile Nodes

Let us assume that a node is able to decode some of the encoding symbols directly from the satellite. Each symbol carries an ESI and a triple (d, a, b). As described in Section 3.5.1 the node can use this information to find out which of the source symbols were combined together to form that encoding symbol. We propose to apply a network encoding scheme at land mobile nodes using the source symbols of iFEC as source symbols of the network code. In other words, nodes exchange linear combinations of encoding symbols in some finite field, with the aim of recovering all the source symbols.

3.6.2 Terrestrial Channel Usage

Each received encoding symbol is interpreted by a node as a linear combination of source symbols with coefficients 0 or 1 in $GF(q)$. The node, then, applies the network encoding procedure described in Section 3.2. The encoding vector of the received encoding symbol can be derived from symbol's ESI and triple (d, a, b).

The probability to access the channel in each slot is determined by the parameter *cooperation level*, fixed for all nodes, which we indicate with ζ, $0 \leq \zeta \leq 2$. If $\zeta \leq 1$, in each slot,

if a node stored a number of linearly independent packets which is larger than the number of transmitted packets in the current generation, it creates a linear combination of all the stored packets as described in Section 3.4 and tries to access the channel with probability ζ. If $\zeta > 1$ two cases must be considered. In case the number of transmissions made by the node is lower than the number of linearly independent packets received, the node tries to access the channel with probability $p_a = 1$. If the node has a number of stored packets which is lower than or equal to the number of those transmitted, instead, it tries to access the channel with probability $p_a = \zeta - 1$.

When a node receives a packet from another node, it checks whether the packet is linearly independent with the stored packets and, if this is the case, the new packet is stored. If the received packet is not linearly independent with the stored ones, it is discarded.

Another possible relaying choice is to have the nodes simply forwarding the received symbols without combining them. We call this scheme *simple relaying* (SR) and use it as a benchmark. SR is described more in detail in Section 3.8.

3.6.3 Implementation Aspects

According to the DVB-SH standard we consider a source symbol size of 1024 Bytes each. Each source symbol is divided into n_{ss} subsymbols, each of which containing $\frac{1024}{n_{ss}}$ bytes. Each of these subsymbols is multiplied by a randomly chosen coefficient in a field with $q = \frac{1024}{n_{ss}} 2^3$ elements. The coefficient is the same for all subsymbols within a symbol. In this way the complexity of the network encoder/decoder can be kept at a reasonable level [8]. A field size of 2^8 or 2^{16} (one or two bytes) may constitute a valid choice. The NC is applied as in [8], adding the encoding vector at the end of each packet. Thus, for a K symbols generation, a header with $K \times q$ bits is appended to each symbol. The loss in spectral efficiency is then $(Kq)/8192$. Assuming coefficients of 1 byte are used, the loss becomes $K/1024$. In order to keep the loss at a reasonable value we should limit the size of the generation. For instance, if generations of $K = 100$ symbols are used, the loss is below 10%. The adoption of small generation sizes has the drawback that the code efficiency is reduced. For example, it is known that the efficiency of the Raptor code increases with the source block. There is, however, advantages in using small blocks. Actually, if a short interleaver is used together with blocks of small size, the data is readily available to the upper layer sooner than in the case of large blocks, i.e., the

delay decreases. In Section 3.9 we show the gap between the asymptotic results obtained in Section 3.4 and the simulation results obtained in the same setup but with the 3GPP Raptor code, having finite block-length.

3.7 Interaction of Physical Layer and Upper Layers

In order to evaluate numerically the performance of the proposed methods at system level, the simulator must be capable of taking into account the channel impairments of the physical layer. In order to do this, physical layer simulations should be run for each of the nodes, taking into account the channel characteristics and the error correction capabilities of the considered PHY layer standard as done in [53]. Such approach is, however, extremely time consuming, which makes it unfit for a system level simulation. A valid alternative is given by the physical layer abstraction (PLA) [37] [38], which significantly decreases the required computational resources while being able to take into account important physical layer elements such as the effect of coding and modulation, interleaver and interference.

3.7.1 Physical-Layer Abstraction

The use of PLA allows to take into account the effects of physical layer at system level in a computationally efficient way. This is particularly useful in case of time-selective channels, in which the channel gain changes within the duration of a codeword. The PLA has been widely studied in the last decade achieving a growing accuracy for a wide range of transmission setups. A method used in the past to obtain instantaneous link performances is the average signal to interference plus noise ratio (SINR) mapping, which consists in calculating the arithmetic (or geometric) average SINR experienced by the channel symbols of a codeword and map such average on the frame error rate (FER) curve in additive white Gaussian noise (AWGN) for the considered code and modulation. This technique shows in general an optimistic behavior of the channel, as high SINR experienced in some parts of the codeword would distort the results.

In more recent types of PLA the instantaneous symbol SINR vector is compressed in a single SINR value, the effective SINR ($SINR_{eff}$). Such approach is called *effective SINR mapping* (ESM). Several ESM PHY abstraction methods have been proposed in the literature based on mean instantaneous capacity, exponential-effective SINR mapping and Mutual In-

formation Effective SINR Mapping (MIESM). A more detailed description as well as more references for these methods can be found in [39]. The $SINR_{eff}$ in the ESM methods is obtained as follows:

$$SINR_{eff} = \Phi^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\Phi(SINR_i)\right), \tag{3.28}$$

where $\Phi(x)$ is an invertible function that depends on the specific ESM method and n is the codeword length. In MIESM such function can be related to the mutual information per received coded bit. This approach is referred to as RBIR. The function $\Phi(x)$ is provided by the so called *modulation-channel model* through a function obtained by normalizing the modulation-constrained symbol mutual information (SI) vs SNR function. Once $SINR_{eff}$ is obtained, it is used to determine the FER using curves for the considered channel code in AWGN. Note that $SINR_{eff}$ is referred to the coded symbol, which means that modulation order and coding rate must be taken into account before using it in the FER curves. If, for instance, E_b/N_o FER curves are used, the $SINR_{eff}$ must be multiplied by a factor $\frac{1}{\log_2(M_o)r}$, M_o and r being the modulation order and the coding rate, respectively.

3.7.2 Simulator Validation

We implemented the RBIR approach in our simulator and validated it by comparing the FER curve obtained with our simulator with the FER curve obtained from the physical layer simulation (which includes the whole transmission chain) in the same scenario. An intermediate tree shadowing (ITS) scenario with DVB-SH physical layer has been considered. The simulation parameters are reported in Table 3.1. Fig. 3.6 shows the comparison of the FER curves obtained with the two methods for a range of E_s/N_o. We see how they almost coincide, which validates our implementation of the RBIR approach.

3.8 Simulation Setup

We evaluated the performance of the proposed scheme by developing a simulator that models a satellite to land mobile broadcast transmission over DVB-SH-B. 150 nodes were randomly placed on a Manhattan grid of one square kilometer with 10 intersecting roads. The distance between two parallel roads is 110 m. Each node moves at a speed of 50 km/h along one of the

Table 3.1: Physical layer abstraction validation scenario.

Environment	ITS
Carrier frequency	2.2 GHz
Terminal speed	50 km/h
Elevation angle	40°
Time Interleaver Depth	200 ms
Modulation	QPSK

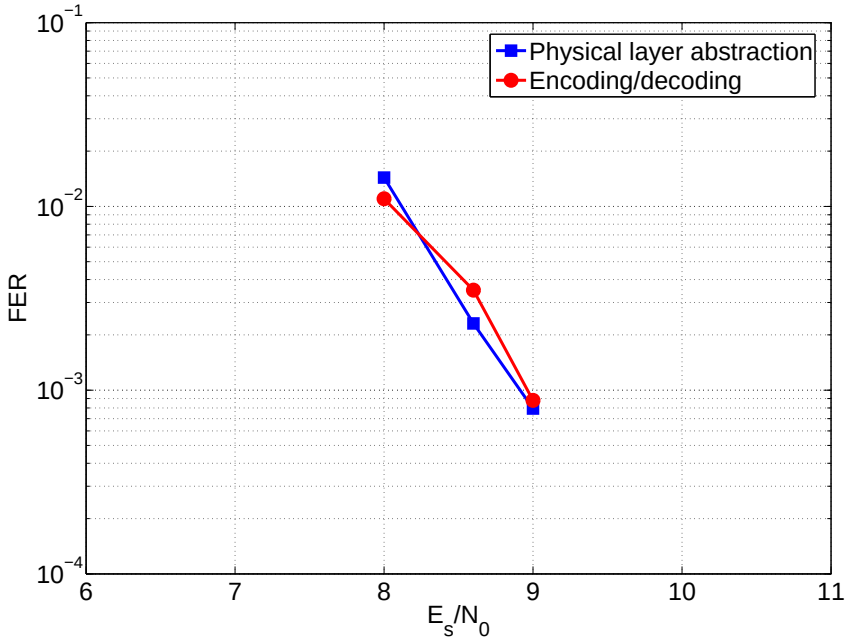


Figure 3.6: Validation of the simulator implementing RBIR physical layer abstraction.

roads, keeping a constant direction of motion during the whole simulation. The direction of motion is chosen at random for each node. When a node reaches the border of the map it enters back into the map from the opposite side, as is common practice in this kind of simulations. Nodes can communicate with each other and have network coding capabilities. Communication can take place between two nodes only if they are within a radius of 300 m. A combination of the path loss model derived in [51] and the TU6 multi-tap propagation model is used. The coding and modulation considered are the ones of 802.11p, namely OFDM modulation and rate $1/2$ convolutional code. The correctness of the reception is evaluated through PLA. One

IFEC block of $K = 150$ IFEC symbols is transmitted at each trial. In the following we will use interchangeably the terms "IFEC block" and "generation". Each block contains K source symbols of 1024 bytes each. The total number of coded symbols transmitted for a single generation is $\lceil K/R_{IFEC} \rceil$, where R_{IFEC} is the rate of the Raptor encoder and $\lceil x \rceil$ is the smallest integer larger than or equal to x. We use the 3GPP Raptor encoder described in [46]. Each IFEC symbol is encapsulated within an MPEG2 TS packet and sent to the channel encoder. The channel encoder is the 3GPP2 turbo encoder specified in [54]. Each source message of the channel encoder has a fixed length of 12288 bits, which means that about one and a half IFEC symbols fit within one Turbo codeword. Once encoded with a rate R_{turbo} the IFEC symbols are first interleaved with the bit interleaver and successively with the time interleaver, which provides time diversity to the signal. In the simulator we implemented two of the time interleavers described in [47], namely the *short uniform* interleaver and the *long uniform* interleaver. The former has a depth on the order of 200 milliseconds while the latter has a depth on the order of 10 seconds. After time interleaving, the bits are QPSK modulated and transmitted with a roll off factor equal to 0.35. For each of the mobile nodes we generate a channel series using a generator implementing the three state Perez-Fontan LMS channel model. The correctness of the reception of each turbo codeword is evaluated using PLA as described in Section 3.7.1, taking into account data rate, channel interleaver, channel code rate, and other relevant parameters. In the setup in which the gap-filler is present, an OFDM modulation with 6048 carriers and a guard interval GI of 224 microseconds is used by the gap-filler. All parameters conform to the DVB-SH-B standard. The propagation model from the gap-filler to each of the nodes is a combination of the modified COST 231 Hata path loss model with the classical TU6 channel model, as suggested in [25]. The signals from the satellite and the gap-filler are combined at the physical level by the receiver using maximal ratio combining. The same channel code and interleaver are used in both the satellite and the gap-filler. The gap-filler is located at a distance $d_{gap\_fill} < GI \cdot c$ from the center of the map, where $c = 3 \cdot 10^8$ m/sec is the speed of light.

The link budget adopted is the one in [25], Table 11.28. The table 3.2 below summarizes the main simulation parameters.

Depending on the sequence of correctly decoded codewords, the decoded IFEC symbols can be determined. Nodes exchange IFEC messages using DSRC/IEEE 802.11p interfaces. The transmission rate in the ground segment is set high enough so that an IFEC symbol can be transmitted before the next one is received on the satellite channel. The MAC mechanism

Table 3.2: Simulation parameters.

Environment	Urban
Satellite carrier frequency	2.2 GHz
Satellite SNR (LOS)	12 dB
Time interleaver depth	200 ms - 10 s
Modulation	QPSK
Roll-off factor	0.35
Bandwidth	5 MHz
LL-FEC symbol size	1024 bytes
Size of LL-FEC blok (K)	150 ($\sim$ 150 kB)
Rate Turbo Code	$1/2$
Rate Raptor Code	$1/4$
Gap filler distance (d_{gap_fill})	3 km
Gap-filler carrier frequency	2.12 GHz
EIRP gap filler	25 dBW
Number of gap-filler OFDM carriers	6048
Subcarrier spacing gap-filler	0.69754 kHz
Scenario surface	1 sq. km
Number of terminals	150
Terminal type	Vehicular
Terminal speed	50 km/h
V2V carrier frequency	5.9 GHz
V2V transmission power	20 dBm
Number of IEEE 802.11p OFDM carriers	52
Subcarrier spacing IEEE 802.11p	0.15625 MHz
Conv. code rate IEEE 802.11p	$1/2$

in the terrestrial segment is CSMA as in 802.11p. Nodes are set in *promiscuous mode* so that each node can receive the transmissions of any other node.

We compare two different relay methods. One is the N-COCCO scheme described in

Section 3.6, which is based on network coding. The other relay scheme is the simple relaying (SR) scheme, also introduced in Section 3.6. Unlike in the N-COCCO scheme, in SR nodes do not combine IFEC symbols, they just transmit the oldest non transmitted packet. In SR, if all the received packets have already been transmitted, then, if $\zeta > 1$, a node transmits (with probability $1 - \zeta$) a randomly chosen packet.

The amount of received data is measured at the interface between the IFEC and the upper layers, as indicated in Fig. 3.7, considering the IFEC block as a fundamental data unit. The reason for this choice is that data coming from the upper layers are reshaped in the ADST's. Thus, receiving one or more IFEC symbols, even if systematic, may not be useful, as they are part of a larger bunch of data like, e.g., firmware or road map updates, or may be parts of incomplete IP datagrams. Thus when we refer to *decoded data* we mean decoded IFEC blocks. The decoding is possible if and only if a number of linearly independent IFEC coded symbols equal to the number of IFEC source symbols is correctly received. The decoding is possible since each of the IFEC coded symbols, as well as each of the NC packets, embeds information about which source symbols were combined to form it, and thus common matrix manipulation techniques can be used to retrieve the source symbols.

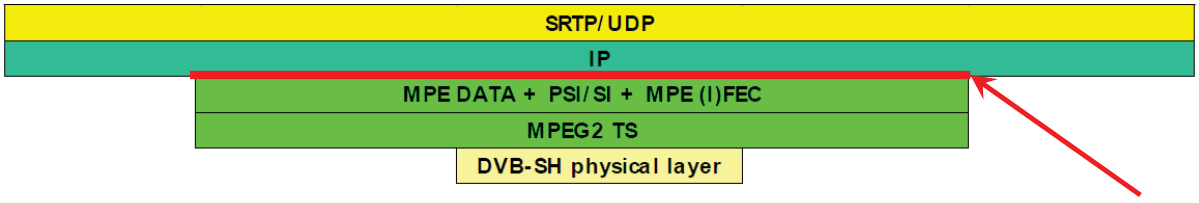


Figure 3.7: The amount of received data is measured at the interface between the IFEC and the upper layers.

3.9 Numerical Results

Fig. 3.8 shows the coverage Ω, obtained by evaluating numerically Eqn. (3.20), plotted against the rate at physical level r for a fixed message rate R and different network sizes. The relative lower bounds and the coverage curve in case of no cooperation are also shown. In the simulation we set $R = 2/3$, $p_a = 0.2$, $\Gamma_N = 10$ dB in the N-N channel, $\mu = 3$ and $\sigma = 1$ in the S-N channel. It is interesting to note how increasing the number of nodes also increases the achievable rate r for a given Ω. In other words, the higher the number of nodes, the higher the

probability that all the information broadcasted by S reaches the network, i.e., it has not been lost. Once the information has reached the network, it can be efficiently distributed among the terminals thanks to the properties of random linear network coding. An important gain in transmission rate can be observed, with an increase of about 0.4 bpcu when passing from no cooperation to cooperation in a network with 2 nodes, and about 1 bpcu in case of a network with 4 nodes. The lower bound is fairly tight for $M = 2$ and $M = 4$. An important point is that this result is achieved without any feedback to the source or any packet request among nodes, as the decision on whether to encode and transmit or not is taken autonomously by each terminal depending on the probability of media contention p_a. In Fig. 3.8 the curves obtained for the same setup but with a finite block-length Raptor code are also shown. The Raptor encoder is the one indicated in Section 3.8. A block length of $K = 150$ source symbols was chosen. We see that, although an important gain in terms of physical layer rate is achieved thanks to cooperation and such gain increases with the number of terminals as in the asymptotic case, a gap between theoretical and numerical results is present. This is due to the finite and relatively small block length. Such gap can be reduced by applying NC directly in the space segment [55]. However, such approach has the drawback that the decoder complexity is higher also in case no cooperation is used, which is not the case when a Raptor code is adopted. Moreover, it would imply a modification in the satellite segment, which, in, our proposed scheme, remains unaltered. In Fig. 3.9 the coverage is plotted against the per-node probability of transmission attempt p_a for $M = 4$, $\Gamma_N = 10$ dB, $r = 1$ bpcu and $R = 2/3$. It is interesting to note that relatively small values of p_a (lower than 0.15 for the asymptotic case) are sufficient to achieve full coverage for values of r and R which are of practical interest. We further observe that the lower bound tightly approximates the simulated theoretical curve. In the figure we also plotted the curve for the case of a practical cooperative scheme using the 3GPP Raptor code with source block length $K = 150$. As in Fig. 3.8, the loss with respect to the theoretical curve is due to the finite block length. The coverage for the non cooperative case in the setup considered in Fig. 3.9 is 0, coherently with Fig. 3.8.

In the rest of the section we compare the performance of the three practical schemes described in previous sections. One is the N-COCCO system described in Section 3.6, one is the SR system described in Section 3.8 while the last system we consider is a non cooperative system in which the nodes can receive only from the satellite. We consider as performance metric the average percentage of nodes that receive all the transmitted data. The metric is evaluated for

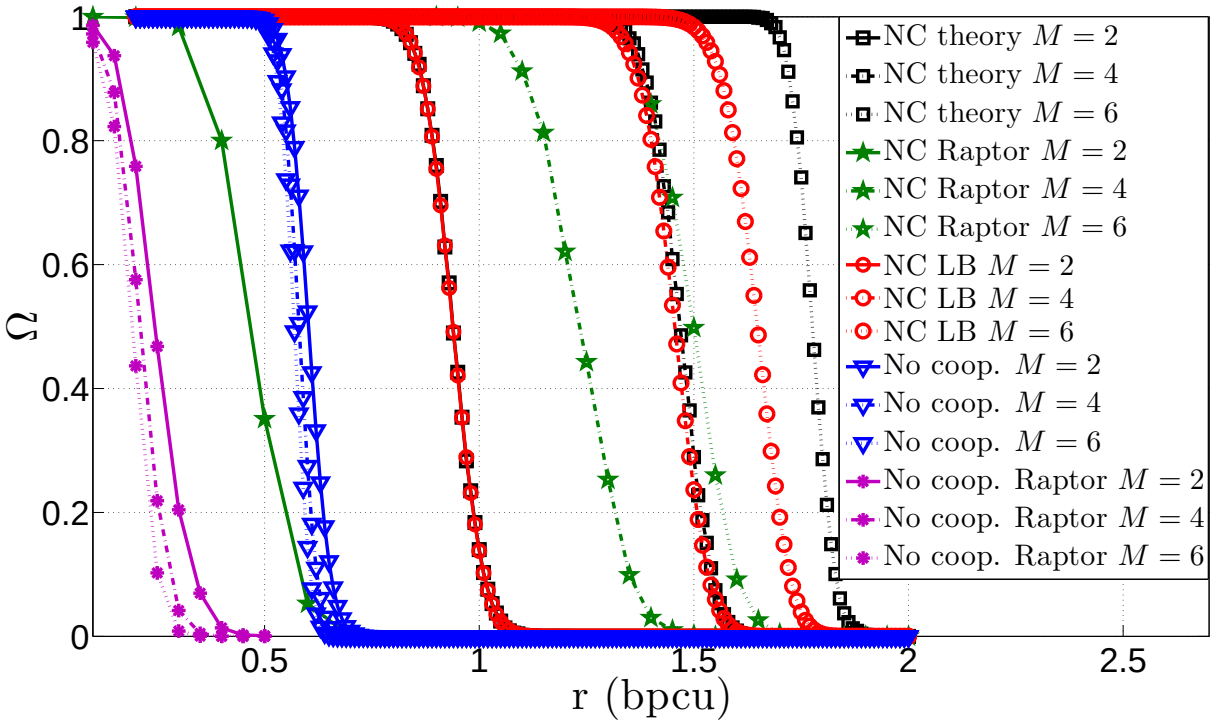


Figure 3.8: Coverage Ω plotted against rate at physical layer r in the cooperative case for different values of M. The lower bound and the non cooperative case are also shown. In the simulation we set $R = 2/3$ messages/slot, $p_a = 0.2$, $\Gamma_N = 10$ dB in the N-N channels, $\mu = 3$ and $\sigma = 1$ in the S-N channel.

different values of the cooperation level ζ in the range $[0, 2]$. Note that the system with satellite only reception corresponds to a cooperative system with $\zeta = 0$. Considering different values of ζ we can evaluate the performance gain of the cooperative methods with respect to the non cooperative system as a function of the terrestrial channel utilization. Fig. 3.10 shows the average percentage of nodes that receive all data plotted against ζ. In the simulations we set the rate at physical level to $1/2$ while the rate of the Raptor encoder has been set to $R_{IFEC} = 1/4$. The short interleaver has been used. We also evaluated the case of long interleaver with and without gap-filler and with no IFEC protection (which corresponds to a Raptor rate of $R_{IFEC} = 1$). We did not consider the case of long interleaver with forward error correction because, according in the DVB-SH-B standard, the IFEC protection is meant to be applied only in combination with the short interleaver. In case the long interleaver is used together with a gap filler (not shown in the figure) 100% of the nodes are covered. The N-COCCO scheme achieves the best performance among all others setups, with a gain of about 25% with respect to the non cooperative scheme and a gain of about 29% with respect to the SR scheme in case no gap-fillers are user (w.o. gf). A notable fact that emerges from the plots is that full coverage is achieved

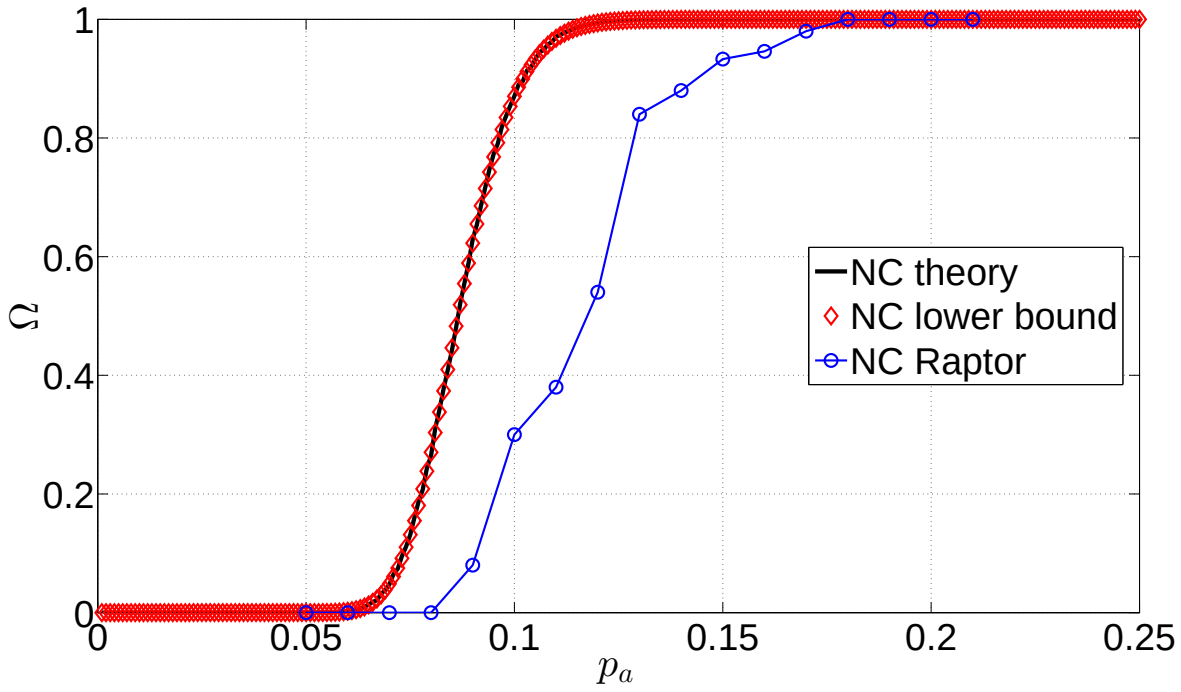


Figure 3.9: Coverage Ω plotted against the probability of media contention p_a in the cooperative case for a network with $M = 4$ and $\Gamma_N = 10$ dB. The lower bound Ω_{LB} curve and the curve of a practical scheme with finite block length Raptor code are also shown. In the simulation we set $R = 2/3$ messages/slot, $r = 1$ bpcu, $\mu = 3$ and $\sigma = 1$ in the S-N channel.

by the N-COCCO scheme with little use of the terrestrial channel for $\zeta = 0.05$ either in case a gap filler is used or not. The performance of the scheme worsens as ζ approaches 2. This is due to the fact that the terrestrial channel load increases with ζ, determining an increase in the number of collisions and thus diminishing the spectral efficiency of the capillary network. We further notice that this is similar to what shown in Fig. 3.9, in that the maximum advantage of the network-coded cooperative scheme is achieved for small values (smaller than 0.15) of the channel access probability p_a. From Fig. 3.10 we also notice that the N-COCCO scheme with short interleaver achieves a higher percentage of covered nodes with respect to the non cooperative configuration with long interleaver. On the one hand this result suggests that a short interleaver can be used instead of a long one, with a huge memory saving in the physical layer architecture of the receiver. Of course this comes at the expense of larger memory resources at higher levels (IFEC), which are likely to have, however, an overall cost which is lower than the memory at lower levels. On the other hand, for a fair comparison we must take into account that the long interleaver scheme does not use IFEC protection, which implies a gain in

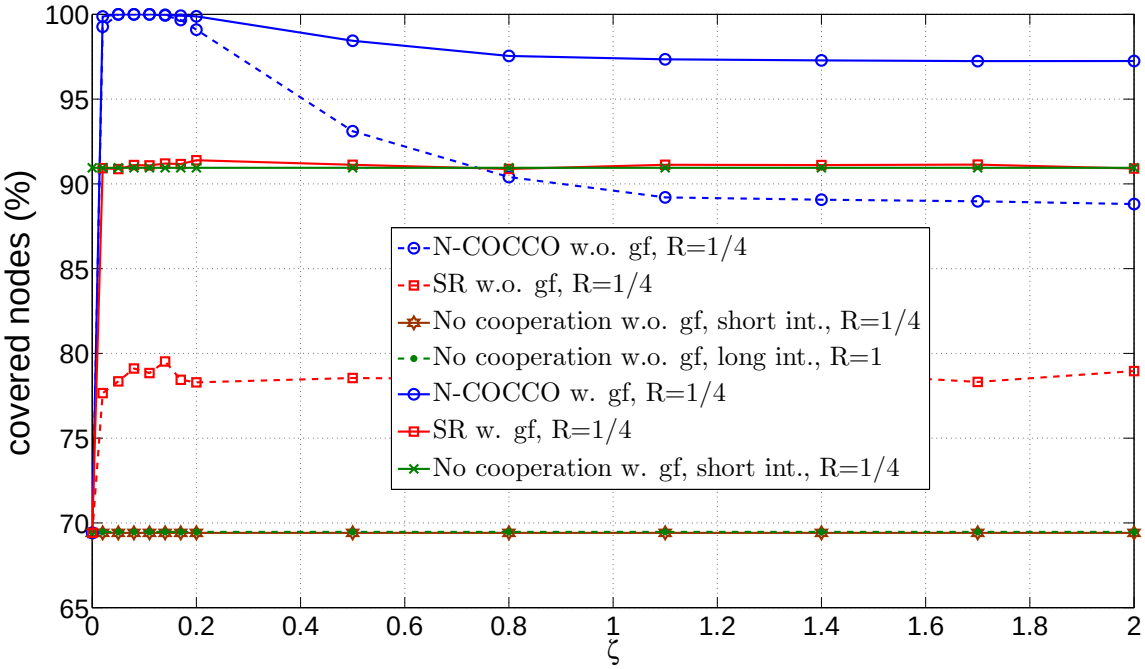


Figure 3.10: Average percentage of nodes that decode all data plotted against the cooperation level ζ for N-COCCO, SR and the non cooperative scheme. The rate couple $(r, R) = (1/2, 1/4)$ has been set in the simulation and the DVB-SH short interleaver has been considered. The non cooperative case with long interleaver and $R = 1$ is also shown for comparison.

terms of spectral efficiency of $1/R_{IFEC} = 4$, i.e., there is a tradeoff between complexity and transmission rate.

The gain in terms of percentage of covered nodes of the cooperative schemes with respect to the non cooperative case derives from the use of the terrestrial channel bandwidth. In order to evaluate which between the N-COCCO scheme and SR scheme uses such resources more efficiently, we plot the average number of retransmissions needed to decode a message (IFEC symbol) against ζ in Fig. 3.11. From the figure we see how the N-COCCO needs on average less transmissions per decoded message with respect to the SR scheme in the whole range of ζ considered. From Fig. 3.11 and Fig. 3.10 emerges that the N-COCCO scheme achieves a larger gain in terms of percentage of covered nodes with respect to the SR scheme using the resources of the terrestrial channel in a more efficient way.

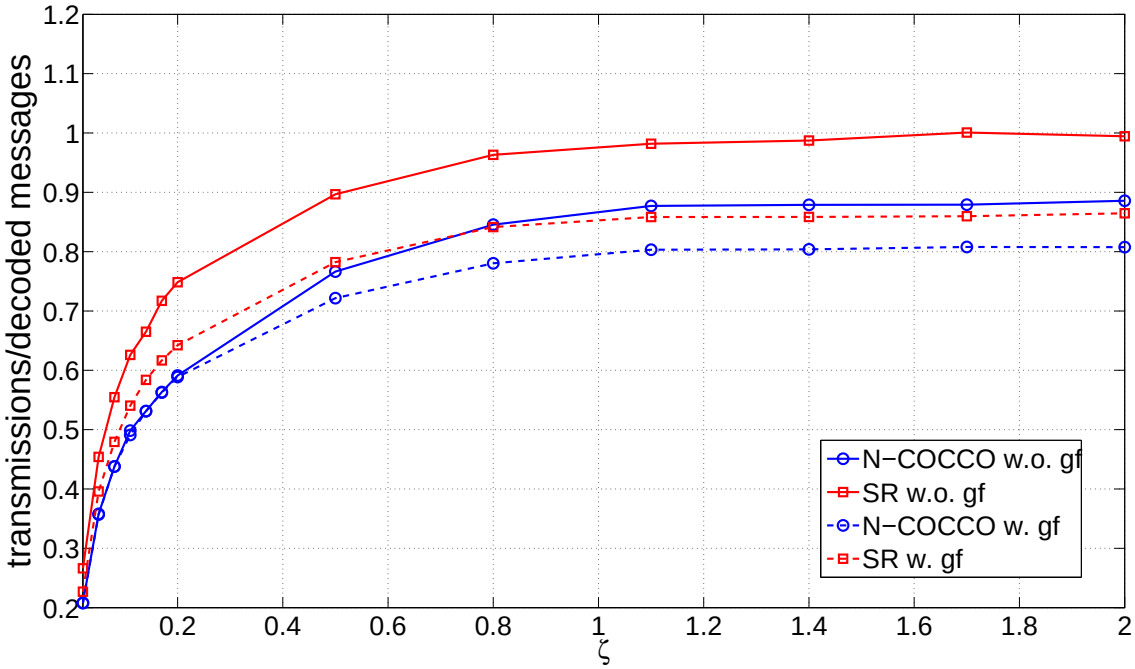


Figure 3.11: Average number of transmissions needed to decode one IFEC symbol plotted against the cooperation level for the N-COCCO and the SR schemes. The N-COCCO scheme makes a much more efficient use of terrestrial channel resources.

3.10 Conclusions

We investigated the performance of a cooperative approach in providing missing coverage for heterogeneous LMS networks. We carried out an analytical study considering a mathematically tractable and yet practically interesting network model, in which fading and shadowing effects in the communication channels as well as the medium access mechanism of the cooperating nodes have been taken into account. By applying the Max-flow Min-cut theorem we derived an analytical lower bound on the coverage as a function of both the information rate at physical layer and the rate of innovative packets injected in the network per unit-time. Our results give a tradeoff between the coverage and the rate at which the information can be injected in the network, and at the same time quantify the gain derived from node cooperation through the short range interface. We showed that the diversity gain grows with the number of terminals, which indicates that important gains in terms of transmission rate at the source source node (e.g., a satellite or a base station) can be achieved through cooperation.

Based on the considered theoretical model we proposed a practical cooperative scheme

which leverages on network coding for enhancing coverage in heterogeneous satellite vehicular LMS systems over DVB-SH. Our numerical results, based on physical layer abstraction, showed that a cooperative relaying system based on network coding can bring important benefits in terms of both coverage and terminal complexity with respect to a system in which nodes receive from satellite only, as well as with respect to a relaying scheme in which network coding is not used.

Appendix 3

Derivation of the cdf of the packet loss rate for the case of Rayleigh fading with log-normal shadowing

The log-normal variable Γ_S can be written as: $\Gamma_S = e^{\frac{X}{10}}$, where $X \sim \mathcal{N}(\mu, \sigma^2)$. Fixing the variable X the packet loss rate Y is:

$$Y = 1 - e^{(1-2^r)\cdot e^{-\frac{X}{10}}}.$$

The cdf of Y can be derived as follows:

$$\begin{aligned}
F_Y(y) &= Pr\{Y < y\} \\
&= Pr\left\{1 - e^{(1-2^r)\cdot e^{-\frac{X}{10}}} < y\right\} \\
&= Pr\left\{\ln(1-y) < (1-2^r)\cdot e^{-\frac{X}{10}}\right\} \\
&= Pr\left\{X > 10\ln\left[\frac{1-2^r}{\ln(1-y)}\right]\right\} \\
&= 1 - F_X\left(10\ln\left[\frac{1-2^r}{\ln(1-y)}\right]\right) \\
&= \frac{1}{2} - \frac{1}{2}\mathrm{erf}\left(\frac{10\ln\left[\frac{1-2^r}{\ln(1-y)}\right] - \mu}{2\sigma^2}\right),
\end{aligned}$$

for $y \in (0, 1)$.

Chapter 4

Streaming Transmitter over Block Fading Channels with Delay Constraint

"We want information, information, information!"

"Who are you?"

"The new number 2."

"Who is number 1?"

"You are number 6."

"I am not a number! I am a free man!"

"Ha ha ha ha ha ha!"

Patrick McGoohan

Contents

4.1 Introduction

In the previous chapter we took a cooperative approach to the problem of missing coverage in satellite broadcast systems. We intentionally did not modify the space segment, in order to enable compatibility with the DVB-SH standard. However, from a long term perspective it is reasonable to assume that some modifications to existing standard will be possible in the future. Thus it is important to understand the fundamental limits and to look for the best encoding scheme at both physical and packet level to be applied at the satellite (or gateway station) in the broadcast scenario. This is particularly true in case of video streaming transmission, due to the delay constraint that characterizes this type of traffic and to the difficulty to meet such constraint in networks with many users and large delays. The present chapter has been developed along this line taking an information theoretical approach to the problem.

We start by defining the *streaming transmitter* setup. In a streaming transmitter data becomes available at the transmitter over time rather than being available at the beginning of transmission. Consider, for example, digital TV satellite broadcasting. The satellite receives video packets from a gateway on Earth at a fixed data rate and has to forward the received packets to the users within a certain deadline. Hence, the transmission of the first packet starts before the following packets arrive at the transmitter. We consider streaming transmission over a block fading channel with CSI available only at the receiver. This assumption results from practical constraints when the receiver belongs to a large population of terminals receiving a broadcast transmission, or when the transmission delay is significantly larger than the channel coherence time [56]. Transmission rate can be adjusted to the channel state through adaptive coding and modulation (ACM) driven by a feedback channel. However, in real-time broadcast systems with large delays and many receivers, such as satellite systems, this is not practical. For instance, according to [57, Section 4.5.2.1] in real-time video transmission the ACM bit-rate control-loop may drive the source bit-rate (e.g., variable bit rate video encoder), but this may lead to a large delay (hundreds of milliseconds) in executing rate variation commands. In such cases the total control loop delay is too large to allow real time compensation of fading.

The data that arrives at the transmitter over a channel block can be modeled as an independent message whose rate is fixed by the quality of the gateway-satellite link and the video encoding scheme used for recording the event. We assume that the transmitter cannot modify the contents of the packets to change the data rate. This follows from the fact that the satellite

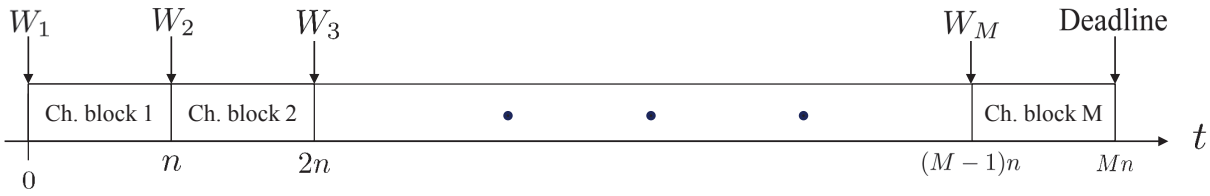


Figure 4.1: The transmitter receives message W_i of rate r at the beginning of channel block i. All the M messages need to be transmitted to the receiver by the end of channel block M.

transmitter is oblivious to the underlying video coding scheme adopted by the source, and considers the accumulated data over each channel block as a single data packet that can be either transmitted or dropped.

We further impose a delay constraint on the transmission such that the receiver buffers the received messages for M channel blocks before displaying the content, which is typical of multimedia streaming applications (see Fig. 4.1). As the messages arrive at the transmitter gradually over M channel blocks, the last message sees only a single channel realization, while the first message can be transmitted over the whole span of M channel blocks. For a finite number M of messages and M channel blocks, it is not possible to average out the effect of fading in the absence of CSI at the transmitter, and there is always a non-zero outage probability [58]. Hence, the performance measure we study is the throughput, that is, the average decoded data rate by the user.

Communication over fading channels has been extensively studied [59]. The capacity of a fading channel depends on the available information about the channel behavior [60]. When both the transmitter and the receiver have CSI, the capacity is achieved though waterfilling [61]. This is called the ergodic capacity as the capacity is averaged over the fading distribution. In the case of a fast fading channel without CSI at the transmitter ergodic capacity is achieved with constant power transmission [59]. However, when there is a delay requirement on the transmission as in our model, and the delay constraint is short compared to the channel coherence time, we have a slow fading channel. In a slow-fading channel, if only the receiver can track the channel realization, outage becomes unavoidable [59]. An alternative performance measure in this case is the ϵ-outage capacity [62]. In general it is hard to characterize the outage capacity exactly; hence, many works have focused on either the high SNR [63] or low SNR [64] asymptotic regimes. Another approach, which is also adopted in this chapter, is to study the average transmission rate as in [65] and [66]. Outages may occur even if the transmitter has

access to CSI when the system is required to sustain a constant transmission rate at all channel states, called the delay-limited capacity [67], [68]. Due to the constant rate of the arriving messages at all channel blocks, our problem is similar to the delay-limited capacity concept. However, here we neither assume CSI at the transmitter nor require all arriving messages to be transmitted. Our work also differs from the average rate optimization in [65] since the transmitter in [65] can adapt the transmission rate based on the channel characteristics and the delay constraint, whereas in our model the message rate is fixed by the underlying application. The only degree-of-freedom the transmitter has in our setting is the multiple channel blocks it can use for transmitting the messages while being constrained by the causal arrival of the messages and the total delay constraint of M channel blocks.

Streaming transmission has received significant attention recently especially with the increasing demand for multimedia streaming applications [69]. Most of the work in this area focus on practical code construction [70], [71], [72]. The diversity-multiplexing tradeoff in a streaming transmission system with a maximum delay constraint for each message is studied in [73]. Unlike in [73], we assume that *the whole* set of messages has a common deadline; hence, in our setting the degree-of-freedom available to the first message is higher than the one available to the last.

In this chapter we extend our work in [74] by presenting analytical results and introducing more effective transmission schemes. We first study joint encoding (JE) which encodes all the available messages into a single codeword at each channel block. We also study time-sharing (TS) and superposition (ST) schemes. The main contributions of the present chapter can be summarized as follows:

1. We introduce a channel model for streaming transmission over block fading channels with a common decoding deadline to study real-time multimedia streaming in networks with large delays, such as digital satellite broadcasting systems.

2. We study the *joint encoding (JE)* scheme that encodes all available messages into a single channel codeword. We show that the JE scheme has a threshold behavior that depends on the average channel quality.

3. We introduce an informed transmitter upper bound on the performance assuming the availability of perfect CSI at the transmitter.

4. We propose the *adaptive JE (aJE)* scheme and show that it performs very close to the informed transmitter upper bound for a finite number of messages, and approaches the ergodic capacity as the number of channel blocks goes to infinity.

5. We propose a *generalized TS (gTS)* scheme in which each message is transmitted over a window of B channel blocks through time-sharing. We show that optimizing the window size B significantly improves the throughput in the high SNR regime.

6. We show that the gTS and the ST schemes provide gradual performance improvement with increasing average SNR. This shows the advantage of these practically simple schemes when broadcasting to multiple users with a wide range of SNR values, or in a point-to-point system with inaccurate CSI.

We support our analytical results with extensive numerical simulations. The rest of the chapter is organized as follows. In Section 4.2 we describe the system model. In Section 4.3 we describe the proposed transmission schemes in detail. In Section 4.4 we provide an informed transmitter upper bound on the throughput, while Section 4.5 is devoted to the numerical results. Finally, Section 4.6 contains the conclusions.

4.2 System Model

We consider streaming transmission over a block fading channel. The channel is constant for a block of n channel uses and changes in an independent and identically distributed (i.i.d.) manner from one block to the next. We assume that the transmitter accumulates the data that arrives at a fixed rate during a channel block, and considers the accumulated data as a single message to be transmitted during the following channel blocks. We consider streaming of M messages over M channel blocks, such that message W_t becomes available at the beginning of channel block t, for $t = 1, \dots, M$ (see Fig. 4.1). Each message W_t has rate r bits per channel use (bpcu), i.e., W_t is chosen randomly with uniform distribution from the set $\mathcal{W}_t = \{1, \dots, 2^{nr}\}$, where n is the number of channel uses per channel block. Following a typical assumption in the literature (see, e.g., [65]), we assume that n, though still large (as to give rise to the notion of reliable communication [75]), is much shorter than the dynamics of the slow

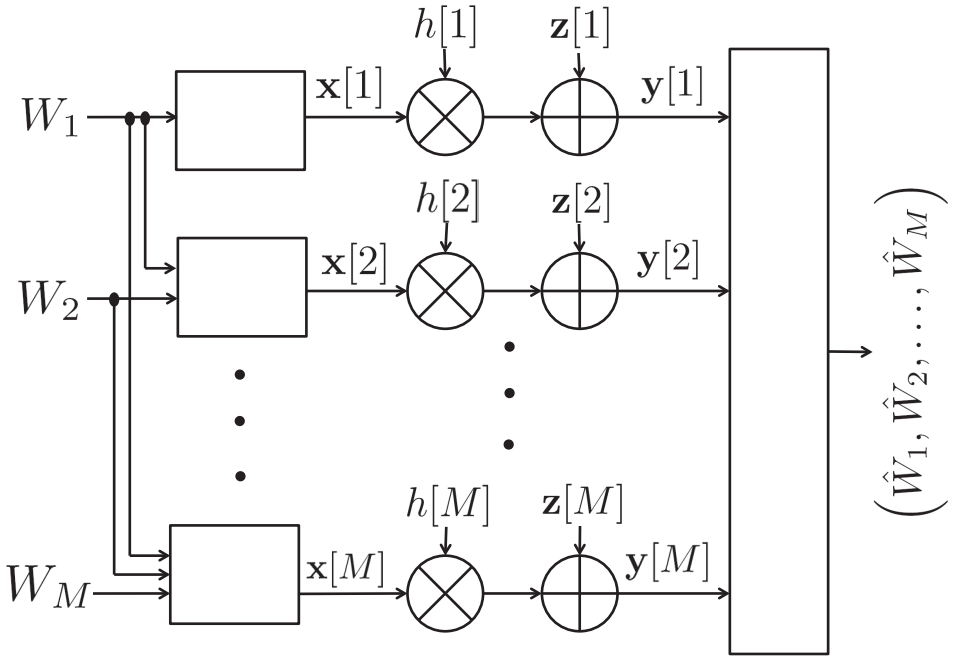


Figure 4.2: Equivalent channel model for the sequential transmission of M messages over M channel blocks to a single receiver.

fading process. The channel in block t is given by

$$\mathbf{y}[t] = h[t]\mathbf{x}[t] + \mathbf{z}[t], \tag{4.1}$$

where $h[t] \in \mathbb{C}$ is the channel state, $\mathbf{x}[t] \in \mathbb{C}^n$ is the channel input, $\mathbf{z}[t] \in \mathbb{C}^n$ is the i.i.d. unit-variance Gaussian noise, and $\mathbf{y}[t] \in \mathbb{C}^n$ is the channel output. The instantaneous channel gains are known only at the receiver, while the transmitter only has knowledge of the statistics of the process $h[t]$. We have a short-term average power constraint of P, i.e., $E[\mathbf{x}[t]\mathbf{x}[t]^\dagger] \leq nP$ for $t = 1, \ldots, M$, where $\mathbf{x}[t]^\dagger$ represents the Hermitian transpose of $\mathbf{x}[t]$ and $E[x]$ is the mean value of x. As we assume a unitary noise power, in the following we will use interchangeably the quantities P and SNR. The short-term power constraint models the restriction on the maximum power radiated by the transmitter which is present in many practical systems

The channel from the transmitter to the receiver can be seen as an orthogonal multiple access channel with a special message hierarchy [76], in which the encoder at each channel block acts as a separate virtual transmitter (see Fig. 4.2). The receiver tries to decode as many of the messages as possible, and the performance measure is the throughput. We denote the instantaneous channel capacity over channel block t by $C_t \triangleq \log_2(1 + \phi[t]P)$, where $\phi[t]$ is a random variable distributed according to a generic probability density function (pdf) $f_\Phi(\phi)$.

Note that C_t is also a random variable. We define $\overline{C} \triangleq E[\log_2(1+\phi P)]$, where the expectation is taken over $f_\Phi(\phi)$. $\overline{C}$ is the ergodic capacity of this channel when there is no delay constraint on the transmission.

4.3 Transmission Schemes

The most straightforward transmission scheme is to send each message only within the channel block following its arrival. This is called memoryless transmission (MT). Due to the i.i.d. nature of the channel, successful decoding probability is constant over messages. Denoting this probability by $p \triangleq Pr\{C_t \geq r\}$, the probability that exactly m messages are decoded is

$$\eta(m) \triangleq \binom{M}{m} p^m (1-p)^{M-m}. \tag{4.2}$$

The throughput of the MT scheme $\overline{r}_{MT}$ is found by evaluating $\frac{r}{M}\sum_{m=1}^{M} m\eta(m)$. The MT scheme treats all messages equally. However, depending on the average channel conditions, it might be more beneficial to allocate more resources to some of the messages in order to increase the throughput. In the following, we will consider three transmission schemes based on the type of resource allocation used. We will find the throughput for these schemes and compare them with an upper bound that will be introduced in Section 4.4.

4.3.1 Joint Encoding (JE) Transmission

In the *joint encoding (JE)* scheme we generate a single multiple-index codebook for each channel block. For channel block t, we generate a t dimensional codebook of size $s_1 \times \cdots \times s_t$, $s_i = 2^{nr},\ \forall i \in \{1,\ldots,t\}$, with Gaussian distribution, and index the codewords as $\mathbf{x}_t(W_1,\ldots,W_t)$ where $W_i \in \mathcal{W} = \{1,\ldots,2^{nr}\}$ for $i = 1,\ldots,t$. The receiver uses joint typicality decoder and tries to estimate as many messages as possible at the end of block M. Such encoding scheme has been studied in terms of the diversity-multiplexing tradeoff (DMT) in [73] for transmission over block fading channels with a per-message delay constraint. Unlike in [73], here we study the JE scheme for the case of common deadline considering throughput as our performance metric. With high probability, the decoder will be able to decode the first m messages correctly if [76]:

$$(m-j+1)r \leq \sum_{t=j}^{m} C_t, \quad \forall\ j = 1,2,\ldots,m. \tag{4.3}$$

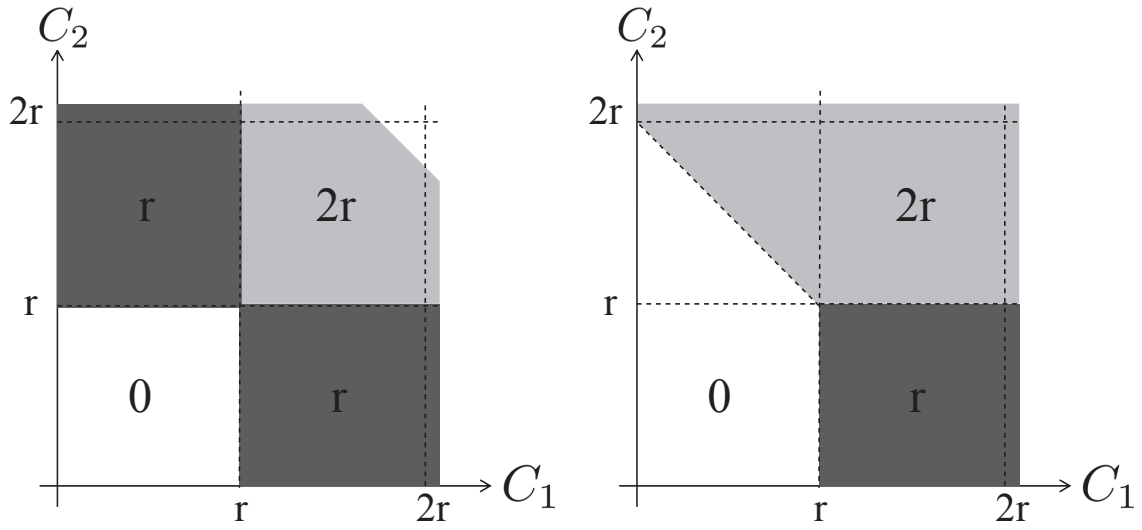


Figure 4.3: Total decoded rate regions illustrated on the (C_1, C_2) plane with $M = 2$ messages for MT (on the left) and JE (on the right) schemes.

As a comparison, we illustrate the achievable rate regions for MT and JE schemes for $M = 2$ in Fig. 4.3. In the case of MT, a total rate of $2r$ can be decoded successfully if both C_1 and C_2 are above r. We achieve a total rate of r if only C_1 or C_2 is above r. On the other hand, in the case of joint encoding, we tradeoff a part of the region of rate r for rate $2r$.

Using the conditions in Eqn. (4.3) we define functions $g^m(r)$, for $m = 0, 1, \ldots, M$, as follows:

$$g^m(r) = \begin{cases} 1, & \text{if } (m-j+1)r \leq \sum_{t=j}^{m} C_t, j = 1, \ldots, m, \\ 0, & \text{otherwise.} \end{cases}$$

Then the probability of decoding exactly m messages can be written as,

$$\eta(m) = Pr\left\{g^m(r) = 1 \text{ and } g^{m+1}(r) = 0\right\}. \tag{4.4}$$

After some manipulation, it is possible to prove that exactly m messages, $m = 0, 1, \ldots, M$, can be decoded if:

$$C_{m-i+1} + \cdots + C_m \geq ir, \;\; i = 1, \ldots, m, \tag{4.5}$$

$$C_{m+1} + \cdots + C_{m+i} < ir, \;\; i = 1, \ldots, M-m. \tag{4.6}$$

Then $\eta(m)$ can be calculated as in Eqn. (4.7) at the bottom of the next page, where we have defined $x^+ = \max\{0, x\}$, and $f_{C_1 \cdots C_m}(c_1, \ldots, c_m)$ as the joint pdf of $C_1, \ldots, C_m$, which is equal to the product of the marginal pdf's due to independence. The probability in Eqn.

(4.7) cannot be easily evaluated for a generic M. However, we provide a much simpler way to calculate the throughput $\overline{r}_{JE}$. This simplification is valid not only for i.i.d. but also for *conditionally i.i.d.* channels. Random variables $\{C_1, \cdots, C_M\}$ are said to be conditionally i.i.d. given a random variable U if the joint distribution is of the form

$$f_{C_1,\cdots,C_M,U}(c_1,\ldots,c_M,u) = f_{C_1|U}(c_1|u) \times \cdots \times f_{C_M|U}(c_M|u) f_U(u), \tag{4.8}$$

where

$$f_{C_i|U}(c_i|u) = f_{C_j|U}(c_j|u),\ \forall i,j \in \{1,\ldots M\}. \tag{4.9}$$

Note that i.i.d. channels is a particular case of conditionally i.i.d. channels where U is a constant.

Theorem 4.1: The throughput for the JE scheme in the case of conditionally i.i.d. channel capacities is given by:

$$\overline{r}_{JE} = \frac{r}{M} \sum_{m=1}^{M} Pr\{C_1 + \cdots + C_m \geq mr\}. \tag{4.10}$$

Proof: See Appendix 4.

In general it is still difficult to find an exact expression for $\overline{r}_{JE}$, but Theorem 4.1 simplifies the numerical analysis significantly. Moreover, it is possible to show that $\overline{r}_{JE}$ approaches r for large M if $\overline{C} > r$. To prove this, we rewrite Eqn. (4.10) as:

$$\overline{r}_{JE} = r - \frac{r}{M} \sum_{m=1}^{M} a_m, \tag{4.11}$$

where we have defined

$$a_m \triangleq Pr\left\{\frac{C_1 + \cdots + C_m}{m} < r\right\}. \tag{4.12}$$

It is sufficient to prove that, if $\overline{C} > r$, then $\lim_{M\to\infty} \sum_{m=1}^{M} a_m = c$, for some $0 < c < \infty$. We start by noting that $\lim_{m\to+\infty} a_m = 0$, since, by the law of large numbers, as m goes to infinity

---

$$\eta(m) = \int_r^{\infty} \int_{(2r-x_m)^+}^{\infty} \cdots \int_{(mr-x_m-\cdots-x_2)^+}^{\infty} f_{C_1\cdots C_m}(x_1,\ldots,x_m)dx_1\cdots dx_m$$
$$\times \int_0^r \int_0^{2r-x_{m+1}} \cdots \int_0^{(M-m)r-x_{m+1}-\cdots-x_{M-1}} f_{C_{m+1}\cdots C_M}(x_{m+1},\ldots,x_M)dx_{m+1}\cdots dx_M \tag{4.7}$$

$\frac{C_1+\cdots+C_m}{m}$ converges almost surely to $\overline{C} > r$. To prove the convergence of the series sum we show that

$$\lim_{m\to+\infty} \frac{a_{m+1}}{a_m} = \lambda, \tag{4.13}$$

We define

$$l_m \triangleq \sqrt{m}\left(\overline{C} - \frac{C_1+\cdots+C_m}{m}\right), \quad m = 1, 2, \ldots, M, \tag{4.14}$$

where each l_m is a random variable with zero mean and variance σ_c^2, which corresponds to the variance of the channel. From the central limit theorem we can write:

$$\lim_{m\to+\infty} \frac{a_{m+1}}{a_m} = \lim_{m\to+\infty} \frac{Pr\left\{l_{m+1} > \frac{\overline{C}-r}{1/\sqrt{m+1}}\right\}}{Pr\left\{l_m > \frac{\overline{C}-r}{1/\sqrt{m}}\right\}} \tag{4.15}$$

$$= \lim_{m\to+\infty} \frac{Q\left(\frac{\overline{C}-r}{\sigma_c/\sqrt{m+1}}\right)}{Q\left(\frac{\overline{C}-r}{\sigma_c/\sqrt{m}}\right)} \tag{4.16}$$

$$\leq \lim_{m\to+\infty} \frac{\frac{\sigma_c/\sqrt{m+1}}{(\overline{C}-r)\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\overline{C}-r}{\sigma_c/\sqrt{m+1}}\right)^2}}{\frac{\frac{\overline{C}-r}{\sigma_c/\sqrt{m}}}{1+\left(\frac{\overline{C}-r}{\sigma_c/\sqrt{m}}\right)^2}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\overline{C}-r}{\sigma_c/\sqrt{m}}\right)^2}} \tag{4.17}$$

$$= \lim_{m\to+\infty} \frac{\sigma_c^2 + m(\overline{C}-r)^2}{\sqrt{m(m+1)}(\overline{C}-r)^2} e^{-\frac{(\overline{C}-r)^2}{2}\left[\frac{m+1}{\sigma_c^2}-\frac{m}{\sigma_c^2}\right]} \tag{4.18}$$

$$= e^{-\frac{(\overline{C}-r)^2}{2\sigma_c^2}} < 1, \tag{4.19}$$

with $0 < \lambda < 1$. where inequality (4.17) follows from the bounds on the Q-function:

$$\frac{x}{(1+x^2)\sqrt{2\pi}} e^{-\frac{x^2}{2}} < Q(x) < \frac{1}{x\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \text{for } x > 0. \tag{4.20}$$

Similarly, we prove that if $\overline{C} < r$, the average rate tends to zero asymptotically with M. To see this, we consider the series in Eqn. (4.10) defining $b_m = Pr\{C_1 + \cdots + C_m \geq mr\}$. We want to prove that $\sum_{m=1}^{M} b_m$ converges to a constant, i.e., $M^{-1}\sum_{m=1}^{M} b_m$ converges to zero. We first notice that $\lim_{m\to+\infty} b_m = 0$ by the law of large numbers. One can also show that $\lim_{m\to+\infty} \frac{b_{m+1}}{b_m} < 1$; and hence, $\overline{r}_{JE}$ goes to zero with M. Overall we see that the average rate of the JE scheme shows a threshold behavior, i.e., we have:

$$\lim_{M\to\infty} \overline{r}_{JE} = \begin{cases} r, & \text{if } \overline{C} > r \\ 0, & \text{if } \overline{C} < r. \end{cases} \tag{4.21}$$

Adaptive Joint Encoding (aJE) Transmission Eqn. (4.21) indicates a phase transition such that $\overline{r}_{JE}$ is zero even for large M if $r > \overline{C}$, and the transmission rate cannot be modified. However, the transmitter may choose to transmit only a fraction $\alpha = \frac{M'}{M} < 1$ of the messages, allocating the extra $M - M'$ channel blocks to the M' messages, effectively controlling the transmission rate. In other words, the M' messages are encoded and transmitted in M' channel blocks as described above, while each of the remaining $M - M'$ blocks is divided into M' equal parts, and the encoding process used for the first M' blocks is repeated, using independent codewords, across the M' parts of each block. For instance, if $M = 3$ and $M' = 2$, $\mathbf{x}_1(W_1)$ and $\mathbf{x}_2(W_1, W_2)$ are transmitted in the first and second channel blocks, respectively. The third channel block is divided into $M' = 2$ equal parts and the independent codewords $\mathbf{x}_{31}(W_1)$ and $\mathbf{x}_{32}(W_1, W_2)$ are transmitted in the first and in the second half of the block, respectively. We call this variant of the JE scheme *adaptive JE (aJE)* scheme. The conditions for decoding exactly m messages, $m = 0, 1, \ldots, M'$, in aJE can be obtained from those given in Eqn. (4.5) and Eqn. (4.6) by replacing C_i with $C_i^* = C_i + \frac{1}{M'} \sum_{j=M'+1}^{M} C_j$, $i \in \{1, \ldots, M'\}$. Note that the random variables C_i^*, $i \in \{1, \ldots, M'\}$, are conditionally i.i.d., i.e., they are i.i.d. once the variable $U = \frac{1}{M'} \sum_{j=M'+1}^{M} C_j$ is fixed. This implies that Theorem 4.1 holds.

In the Appendix of this chapter we prove that, by choosing α appropriately, we can have

$$\lim_{M \to \infty} \overline{r}_{aJE} = \min\{r, \overline{C}\}. \tag{4.22}$$

Eqn. (4.22) suggests that the average transmission rate can be adapted at the message level while keeping a fixed rate at the physical layer. Rate adaptation at packet level can be implemented through rateless erasure codes. A rateless code can recover all the original packets as long as the rate at packet level is below the probability of correct decoding p, which depends on the average channel SNR and on the code rate at physical level r. The average throughput achieved by a rateless code when $M \to \infty$ is

$$\lim_{M \to \infty} \overline{r}_{RC} = r \cdot p(SNR, r). \tag{4.23}$$

Unlike the aJE scheme, rateless codes perform the encoding at the physical level and at the packet level independently, while in aJE the two encoding processes are performed jointly, which translates into a more efficient use of the channel.

In Fig. 4.4 we compare the asymptotic average throughput of the aJE scheme and of a rateless code in case of Rayleigh fading and for different values of r. We can see how the aJE scheme outperforms the RC for all considered values of r and in the whole SNR range.

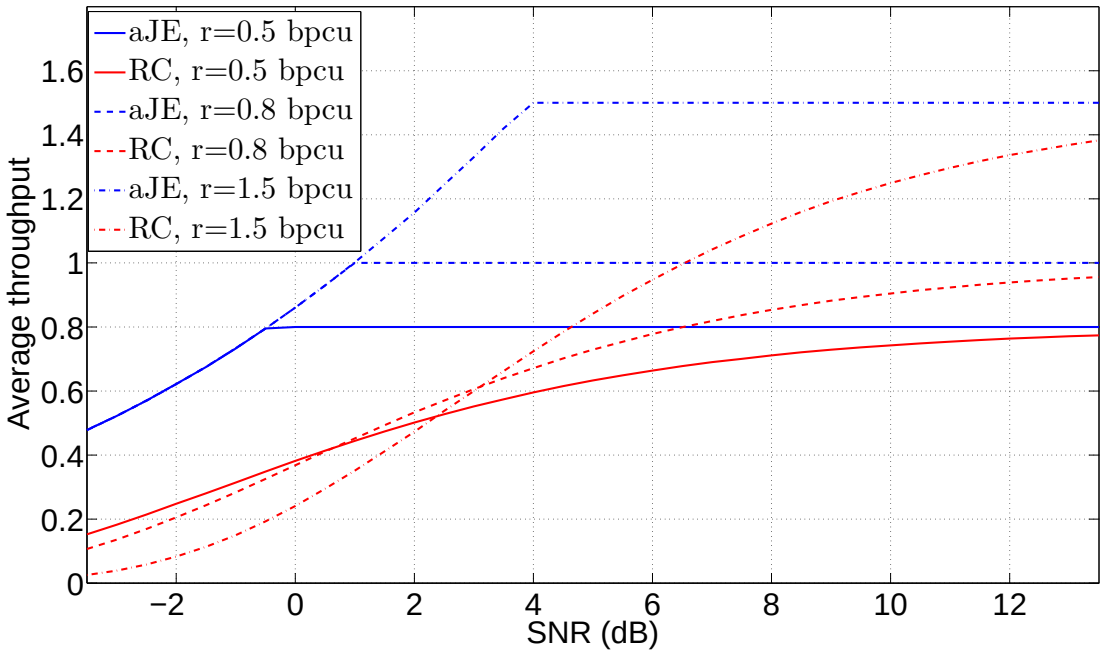


Figure 4.4: Average throughput of the aJE scheme and of a rateless code in case of Rayleigh fading for $M \to \infty$.

We will see in Section 4.4 that the maximum average throughput cannot be above the average throughput of the aJE scheme given in Eqn. (4.22); hence, as the number of messages and the channel blocks go to infinity, the aJE scheme achieves the optimal performance. We will show in Section 4.5 through numerical analysis that the near optimality of the aJE scheme is valid even for finite M. However, we also note the threshold behavior of the performance of aJE; that is, when there are multiple users or inaccuracy in the channel statistics information at the transmitter, aJE performs very poorly for users whose average received SNR is below the target value. In the following we propose alternative transmission schemes providing gradual performance change with the SNR.

4.3.2 Time-Sharing (TS) Transmission

One of the resources that the encoder can allocate among different messages is the total number of channel uses within each channel block. While the whole first channel block has to be dedicated to message W_1 (the only available message), the second channel block can be divided among the messages W_1 and W_2, and so on so forth. Assume that the encoder divides the channel block t into t portions $\alpha_{1t}, \ldots, \alpha_{tt}$ such that $\alpha_{it} \geq 0$ and $\sum_{i=1}^{t} \alpha_{it} = 1$. In channel

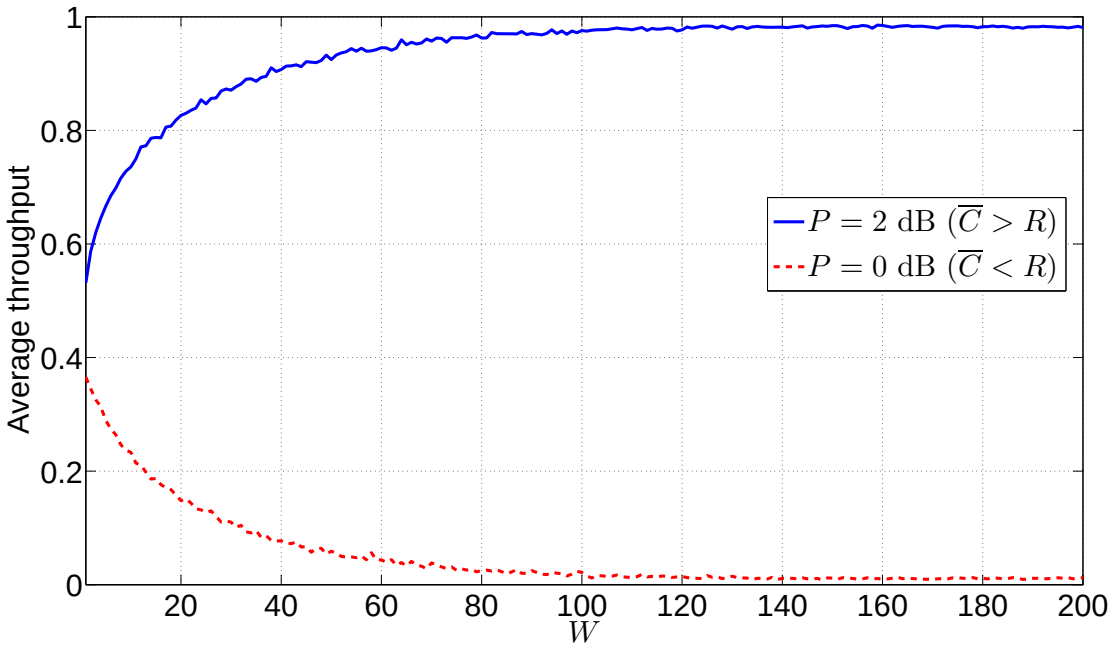


Figure 4.5: Average throughput for the gTS scheme plotted against the window size B for $M = 10^4$ messages and $r = 1$ bpcu for two different average SNR values.

block t, $\alpha_{it}n$ channel uses are allocated to message W_i. A constant power P is used throughout the block. Then the total amount of received mutual information (MI) relative to message W_i is $I_i^{tot} \triangleq \sum_{t=i}^{M} \alpha_{it} C_t$. Letting $\alpha_{it} = 1$ if $t = i$ and $\alpha_{it} = 0$ otherwise, we obtain the MT scheme.

For simplicity, in the *time sharing (TS)* scheme we assume equal time allocation among all the available messages; that is, for $i = 1, \ldots, M$, we have $\alpha_{it} = \frac{1}{t}$ for $t = i, i+1, \ldots, M$, and $\alpha_{it} = 0$ for $t = 1, \ldots, i$. The messages that arrive earlier are allocated more resources; and hence, are more likely to be decoded. We have $I_i^{tot} > I_j^{tot}$ for $1 \leq i < j \leq M$. Hence,

$$\overline{r}_{TS} = \frac{r}{M} \sum_{m=1}^{M} Pr\left\{ \frac{C_m}{m} + \frac{C_{m+1}}{m+1} + \cdots + \frac{C_M}{M} \geq r \right\}. \tag{4.24}$$

Generalized Time-Sharing (gTS) Transmission Note that, in TS transmission, message W_i is transmitted over $M - i + 1$ channel blocks, which allocates significantly more resources to the earlier messages. To balance the resource allocation between the messages, we consider transmitting each message over a limited window of channel blocks.

In *generalized time-sharing* transmission each message is encoded with equal time allocation over B consecutive blocks as long as the total deadline of M channel blocks is not met. Messages from W_1 to W_{M-B+1} are encoded over a window of B blocks, while messages W_i, for $i \in \{M - B + 2, M - B + 3, \ldots, M\}$ are encoded over $M - i + 1$ blocks. In particular

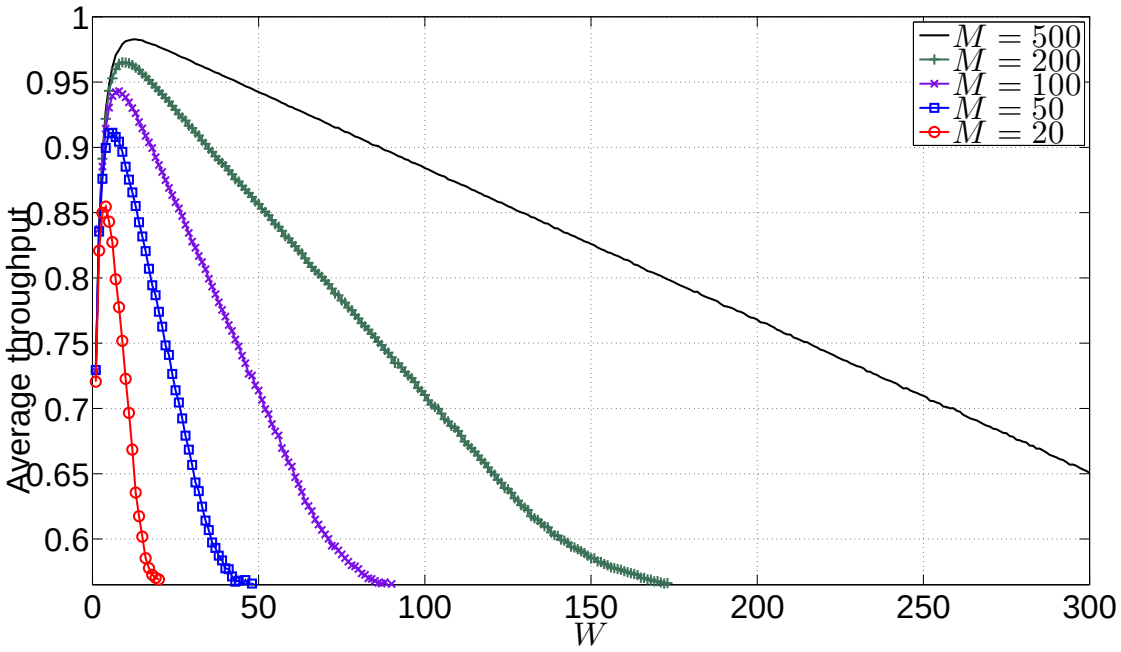


Figure 4.6: $r = 1$ bpcu, $P = 5$ dB ($\overline{C} > r$).

we focus on the effect of variable B on the average throughput $\overline{r}_{gTS}$. In case $B \ll M$, most of the messages are transmitted over B slots together with $B-1$ other messages. In this case the MI accumulated for a generic message W_i is:

$$I_i^{tot} = \frac{1}{B} \sum_{t=i}^{i+B-1} C_t. \tag{4.25}$$

By the law of large numbers, Eqn. (5.17) converges in probability to the average channel capacity $\overline{C}$ as $B \to \infty$. Thus, we expect that, when the transmission rate r is above $\overline{C}$, the gTS scheme shows poor performance for large B (and hence, large M), while almost all messages are received successfully if $\overline{C} > r$. We confirm this by analyzing the effect of B on $\overline{r}$ numerically in Fig. 4.5 for $M = 10^4$ and $r = 1$ bpcu. For $P = 0$ dB we have $\overline{C} < r$, which leads to a decreasing $\overline{r}_{gTS}$ with increasing window size B. On the other hand, for $P = 2$ dB, we have $\overline{C} > r$, and accordingly $\overline{r}_{gTS}$ approaches 1 as B increases.

The same reasoning cannot be applied if the window size is on the order of the number of messages, as the number of initial messages which share the channel with less than $B-1$ other messages and the number of final messages which share the channel with more than $B-1$ messages are no longer negligible with respect to M. In Fig. 4.6, we plot $\overline{r}_{gTS}$ as a function of B for relatively small number of messages and $\overline{C} \geq r$. As seen in the figure, for a given M an optimal value of B can be chosen to maximize $\overline{r}_{gTS}$. Optimal B increases with

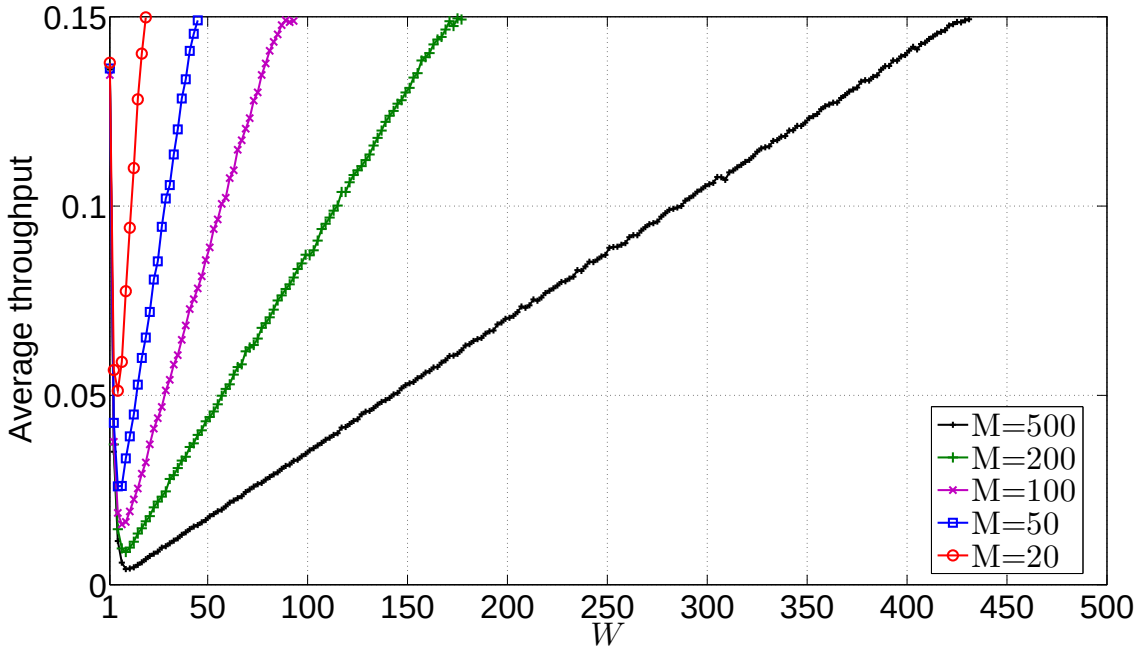


Figure 4.7: $r = 1$ bpcu, $P = -3$ dB ($\overline{C} < r$).

M when $r < \overline{C}$. We plot $\overline{r}_{gTS}$ for $\overline{C} < r$ in Fig. 4.7. From the figure we see that $\overline{r}_{gTS}$ decreases monotonically with B up to a minimum, after which it increases almost linearly. The initial decrease in the average throughput is due to the averaging effect described above, which is relevant for $B \ll M$. The following increase in $\overline{r}_{gTS}$ is because the messages which are transmitted earlier (i.e., W_i with $i \ll B$) get an increasing amount of resources as B increases, and so the probability to be decoded increases. As a matter of fact, for each finite i, the average MI accumulated for message W_i grows indefinitely with B, i.e.:

$$\lim_{B\to\infty} E\left\{\sum_{t=i}^{i+B-1} \frac{C_t}{\min\{t,B\}}\right\} = \overline{C} \lim_{B\to\infty} \sum_{t=i}^{i+B-1} \frac{1}{\min\{t,B\}} = +\infty. \tag{4.26}$$

Thus, for a fixed i, letting B go to infinity leads to an infinite average MI, which translates into a higher $\overline{r}_{gTS}$. Note that this is valid only for relatively small i and large B, i.e., only messages transmitted earlier benefit from increasing B, while the rest of the messages are penalized. If B is small compared to M, as in the plot of Fig. 4.5 for $P = 0$ dB, the fraction of messages which benefit from the increasing B (i.e., messages $W_1, \ldots, W_L$ with $L \ll B$) remains small compared to M. For this reason $\overline{r}_{gTS}$ does not increase with B for the range of B considered in Fig. 4.5, while it does for the same range in Fig. 4.7.

Although the idea of encoding a message over a fraction of the available consecutive slots (e.g., $B < M$ for message W_1 in gTS) can be applied to all the schemes considered in this

chapter, the analysis becomes quite cumbersome. Hence, we restrict our analysis to the TS scheme as explained above.

4.3.3 Superposition Transmission (ST)

In *superposition transmission (ST)* the superposition of t codewords, chosen from t independent Gaussian codebooks of size 2^{nr}, corresponding to the available messages $\{W_1, \ldots, W_t\}$ is transmitted in channel block t, $t \in \{1, \ldots, M\}$. The codewords are scaled such that the average total transmit power in each block is P. In the first block, only information about message W_1 is transmitted with average power $P_{11} = P$; in the second block we divide the total power P among the two messages, allocating P_{12} and P_{22} for W_1 and W_2, respectively. In general, over channel block t we allocate an average power P_{it} for W_i, while $\sum_{i=1}^{t} P_{it} = P$.

Let $\mathcal{S}$ be any subset of the set of messages $\mathcal{M} = \{1, \ldots, M\}$. We define $C(\mathcal{S})$ as follows:

$$C(\mathcal{S}) \triangleq \sum_{t=1}^{M} \log_2 \left(1 + \frac{\phi[t] \sum_{s \in \mathcal{S}} P_{st}}{1 + \phi[t] \sum_{s \in \mathcal{M} \setminus \mathcal{S}} P_{st}} \right). \tag{4.27}$$

This provides an upper bound on the total rate of messages in set $\mathcal{S}$ that can be decoded jointly at the user considering the codewords corresponding to the remaining messages as noise. The receiver first checks if any of the messages can be decoded alone by considering the other transmissions as noise. If a message can be decoded, the corresponding signal is subtracted and the process is repeated over the remaining signal. If no message can be decoded alone, then the receiver considers joint decoding of message pairs, followed by triplets, and so on so forth. This algorithm gives us the maximum throughput. However, it is challenging in general to find a closed form expression for the average total rate, and optimize the power allocation. Hence, we focus here on the special case of equal power allocation, where we divide the total average power P among all the available messages at each channel block. The performance of the ST scheme will be studied in Section 4.5 numerically and compared with the other transmission schemes and the upper bound which will be introduced next.

4.4 Upper Bound

We provide an upper bound on the performance by assuming that the transmitter is informed about the exact channel realizations at the beginning of the transmission. This allows the trans-

mitter to optimally allocate the resources among messages to maximize the average throughput $\overline{r}$. Assume that $C_1, \ldots, C_M$ are known by the transmitter and the maximum number of messages that can be decoded is $m \leq M$. We can always have the first m messages to be the successfully decoded ones by reordering. When the channel state is known at the transmitter, the first m messages can be decoded successfully if and only if [76],

$$ir \leq C_{m-i+1} + C_{m-i+2} + \cdots + C_M, \quad \text{for } i = 1, \ldots, m.$$

We can equivalently write these conditions as

$$r \leq \min_{i \in \{1,\ldots,m\}} \left[\frac{1}{m-i+1} \sum_{j=i}^{M} C_j \right]. \tag{4.28}$$

Then, for each channel realization $\{h[1], \ldots, h[M]\}$, the upper bound on the average throughput is given by $\frac{m^*}{M} r$, where m^* is the greatest m value that satisfies (4.28). This is an upper bound for each specific channel realization obtained by optimally allocating the resources. An upper bound on $\overline{r}$ can be obtained by averaging this over the distribution of the channel realizations.

Another upper bound on $\overline{r}$ can be found from the ergodic capacity assuming all messages are available at the encoder at the beginning and letting M go to infinity. Finally, the bound $\overline{r} \leq r$ follows naturally from the data arrival rate. Thus, $\overline{r}$ is be bounded above by $\min\left\{r, \overline{C}\right\}$.

Comparing the bound $\min\left\{r, \overline{C}\right\}$ and Eqn. (4.22) we see that the aJE scheme achieves the optimal average throughput in the limit of infinite M.

4.5 Numerical Results

In this section we provide numerical results comparing the proposed transmission schemes. For the simulations we assume that the channel is Rayleigh fading, i.e., the channel state $\phi(t)$ is exponentially distributed with parameter 1, i.e., $f_\Phi(\phi) = e^{-\phi}$ for $\phi > 0$, and zero otherwise.

In Fig. 4.8 the cumulative mass function (cmf) of the number of decoded messages is shown for the proposed transmission techniques for $r = 1$, $M = 50$ and $P = 1.44$ dB, which corresponds to an outage probability of $p = 0.5$ for the MT scheme and an average channel capacity $\overline{C} \simeq 1.07 > r$. We see that MT outperforms ST and TS schemes, as its cmf lays below the other two. The gTS scheme improves significantly compared to the ordinary TS scheme.

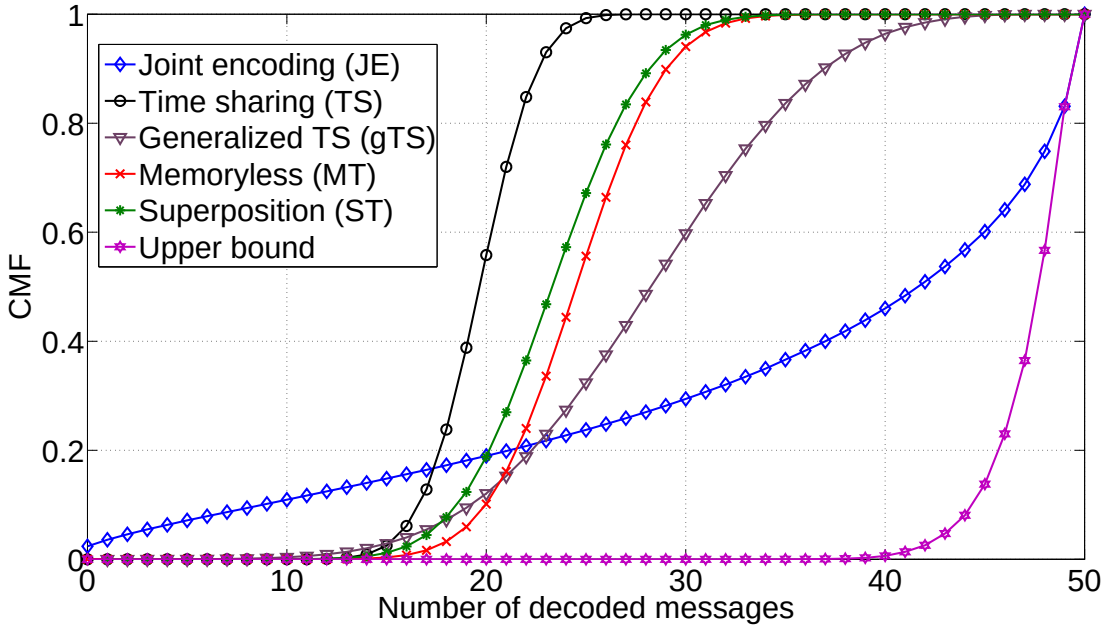


Figure 4.8: $P = 1.44$ dB ($\overline{C} > r$).

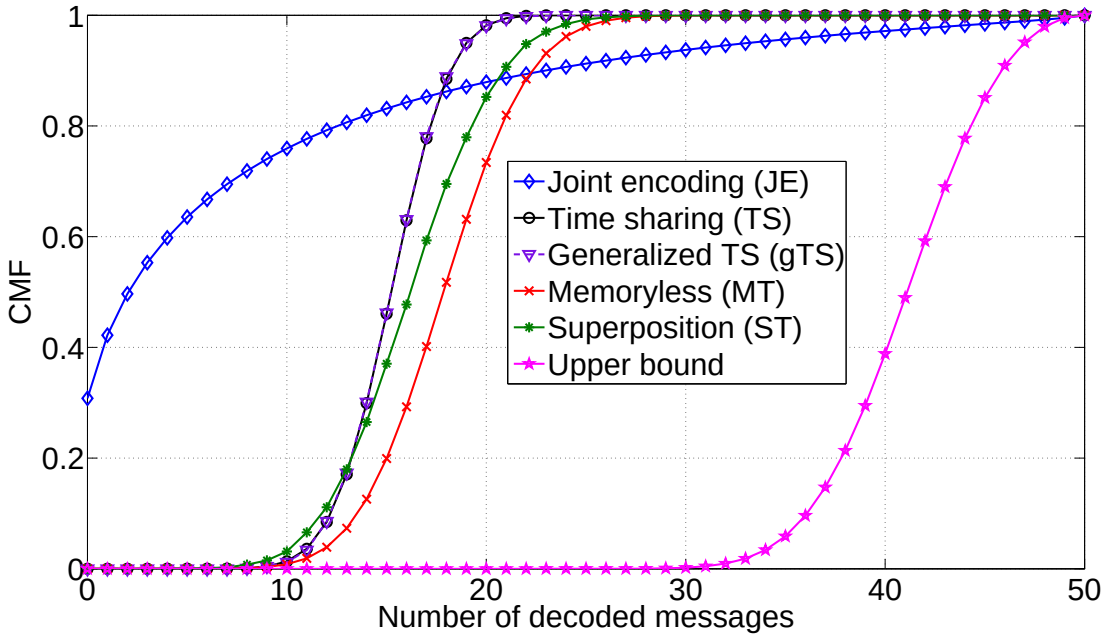


Figure 4.9: $P = 0$ dB ($\overline{C} < r$).

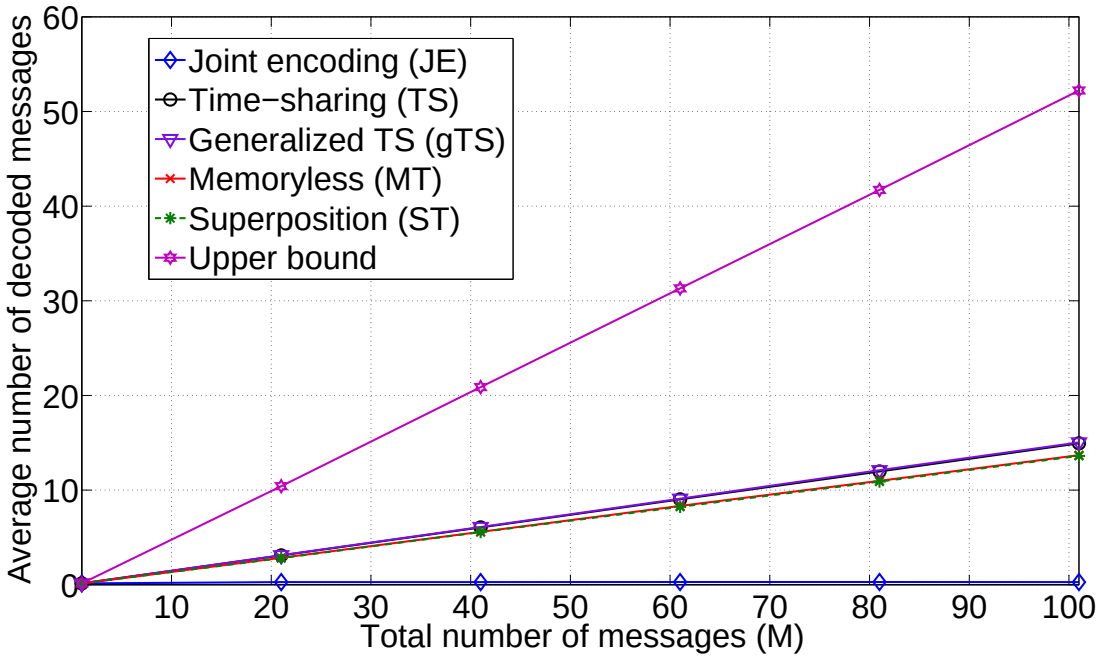


Figure 4.10: $P = -3$ dB ($\overline{C} < r$).

On the other hand, the comparison with the JE scheme depends on the performance metric we choose. For instance, JE has the lowest probability to decode more than m messages, for $m \leq 15$, while it has the highest probability for $m \geq 22$.

In Fig. 4.9 the cmf's for $P = 0$ dB are shown. In this case the average capacity is $\overline{C} \simeq 0.86$. Comparing Fig. 4.9 and Fig. 4.8, we see how the cmf of the JE scheme has different behaviors depending on whether $\overline{C}$ is above or below r. We see from Fig. 4.9 that for the JE scheme there is a probability of about 0.3 not to decode any message, while in all the other schemes such probability is zero. However, the JE scheme also has the highest probability to decode more than 23 messages. Furthermore, we note that the cmf of the gTS scheme converges to the cmf of TS scheme at low SNR. This is because, as shown in Section 4.3.2, when $\overline{C} < r$, the optimal window size B is equal to M, which is nothing but the TS scheme.

In Fig. 4.10 and Fig. 4.11 the average number of decoded messages is plotted against M for SNR values of -3 dB and 2 dB, respectively, and a message rate of $r = 1$ bpcu. While JE outperforms the other schemes at $SNR = 2$ dB, it has the poorest performance at $SNR = -3$ dB. This behavior is expected based on the threshold behavior of the JE scheme that we have outlined in Section 4.3.1. Note that the average capacity corresponding to $SNR = -3$ dB and 2 dB are $\overline{C} = 0.522 < r$ and $\overline{C} = 1.158 > r$, respectively. Note that

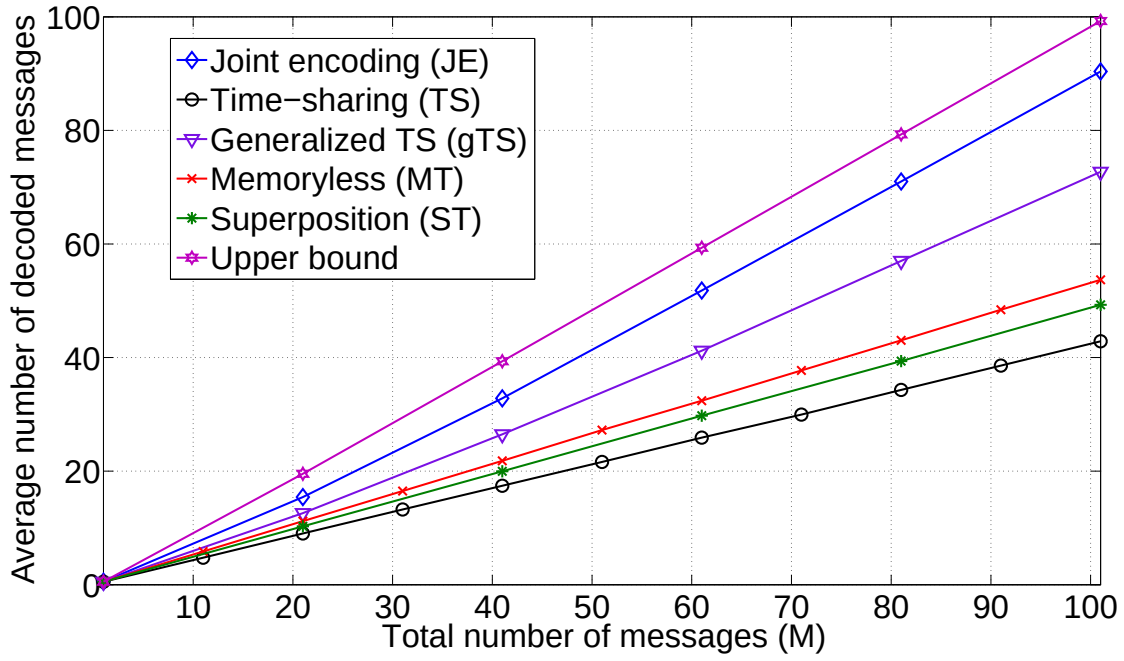


Figure 4.11: $P = 2$ dB ($\overline{C} > r$).

when $\overline{C} > r$ the average throughput of the JE scheme is close to the upper bound although the deadline constraint imposed by M is quite tight. We observe that, in the low SNR regime, i.e., when $\overline{C} < r$, all the proposed schemes other than JE perform very close to each other, and significantly below the upper bound. The large gap to the informed transmitter upper bound is mostly due to the looseness of this bound in the low SNR regime. Note from the two figures that none of the schemes dominates the others at all SNR values.

In Fig. 4.12 the average throughput $\overline{r}$ is plotted against the transmission rate r for the case of $M = 100$ and $P = 20$ dB. The aJE scheme outperforms all the other schemes, performing very close to the upper bound, illustrating the rate adaptation capability of the aJE scheme. The number M' of messages transmitted in the aJE scheme is chosen so that $\frac{M'}{M} = 0.95\frac{\overline{C}}{r}$. In the figure we also show the upper bound obtained from the ergodic capacity $\min(r, \overline{C})$. It can be seen how it closely approximates the informed transmitter upper bound for $r < 6$. The JE scheme performs better than the others up to a certain transmission rate, beyond which rapidly becomes the worst one. This is due to the phase transition behavior exposed in Section 4.3.1 in the case of asymptotically large delay, and observed here even for a relatively small M. Among the other schemes, MT achieves the highest average throughput in the region $r < 6.8$, while TS has the worst performance. The opposite is true in the region $r > 6.8$, where the curve of ST

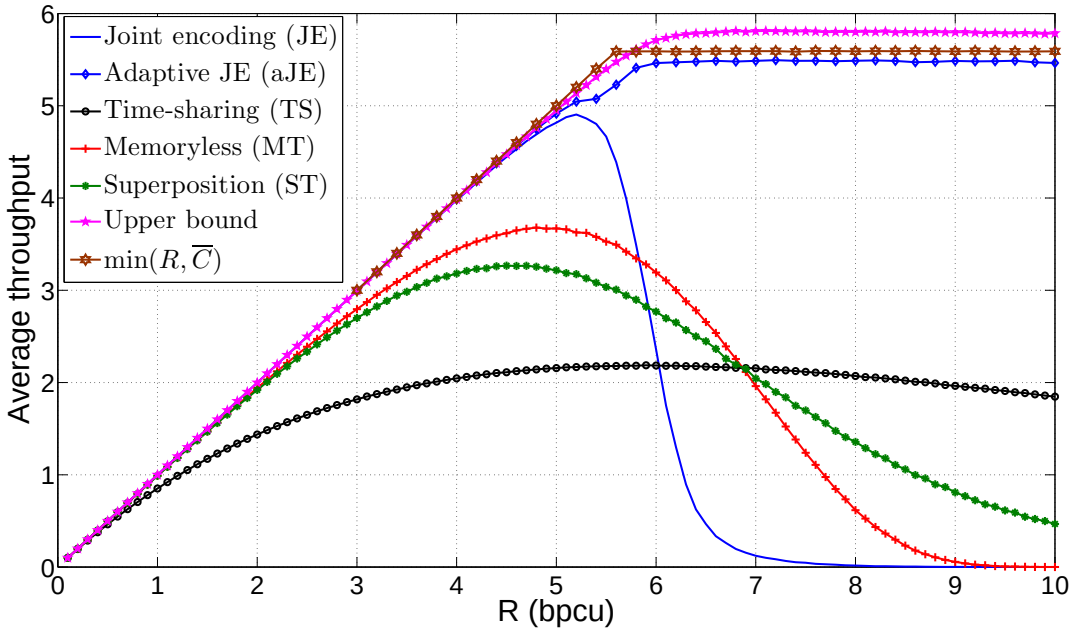


Figure 4.12: Average throughput vs r for $P = 20$ dB and $M = 100$ messages. The upper bound $\min(r, \overline{C})$ is also shown.

scheme is upper and lower bounded by the curves of the MT and TS schemes. We have repeated the simulations with different parameters (i.e., changing P and M) with similar results, that is, MT, TS, and ST schemes meet approximately at the same point, below which MT has the best performance of the three while above the intersection TS has the best performance. At the moment we have no analytical explanation for this observation, which would mean that there is always a scheme outperforming ST. We next study the performance of the considered schemes as a function of the distance from the transmitter.

We scale the average received power at the receiver with $d^{-\alpha}$, where d is the distance from the transmitter to the receiver and α is the path loss exponent. The results are shown in Fig. 4.13 for $P = 20$ dB, $M = 100$, $r = 1$ bpcu and a path loss exponent $\alpha = 3$. The dependence of $\overline{r}$ on the distance is important, for instance, in the context of broadcast transmission in cellular networks, in which case the receiving terminals may have different distances from the transmitter. In such a scenario the range of the average channel SNR values at the receivers becomes important, and the transmitter should use a transmission scheme that performs well over this range. For instance, in a system in which all users have the same average SNR, which is the case for a narrow-beam satellite system where the SNR within the beam footprint has variations of at most a few dB's on average [77], the transmission scheme

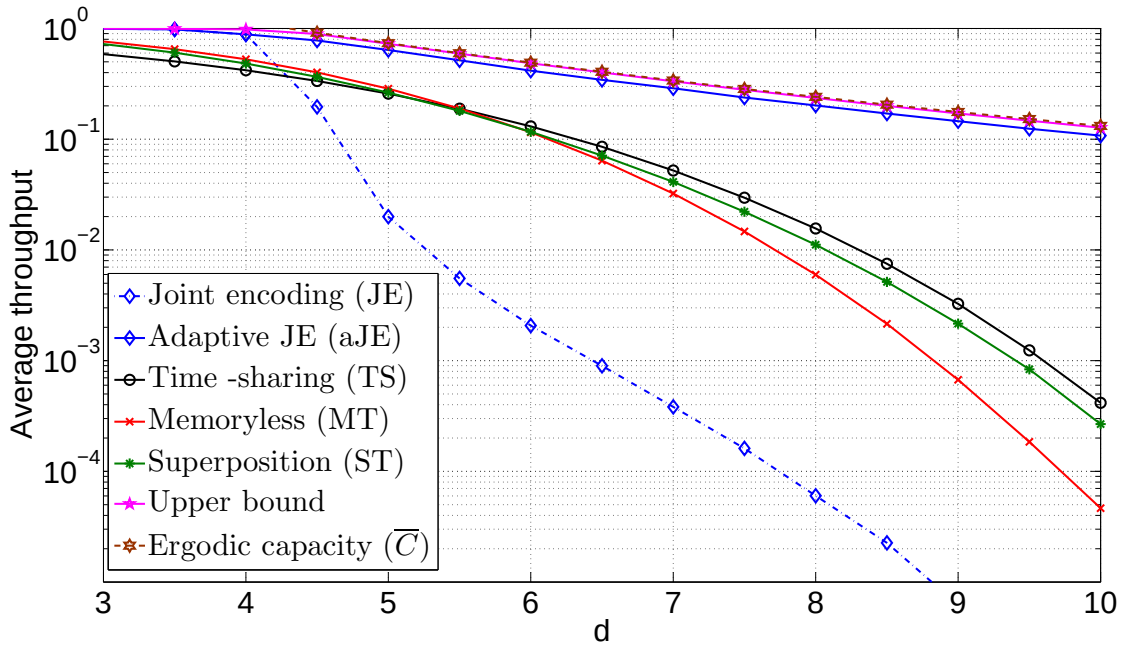


Figure 4.13: Average throughput $\overline{r}$ vs distance from the transmitter for $r = 1$ bpcu, $M = 100$, $P = 20$ dB and $\alpha = 3$.

should perform well around the average SNR of the beam. A similar situation may occur in a microcell, where the relatively small radius of the cell implies a limited variation in the average SNR. Instead, in the case of a macrocell, in which the received SNR may vary significantly from the proximity of the transmitter to the edge of the cell, the transmitter should adopt a scheme which performs well over a larger range of SNR values. In the range up to $d = 4$ the JE scheme achieves the highest average throughput while for $d \geq 6$ the TS scheme outperforms the others. The drop in the average throughput in the JE scheme when passing from $d = 4$ to $d = 5$ is similar to what we observe in Fig. 4.12 when the rate increases beyond $r = 6$ bpcu. In both cases the transition takes place as the transmission rate surpasses the average channel capacity. The aJE scheme, which selects the fraction of messages to transmit based on $\overline{C}$, outperforms all other schemes and gets relatively close to the informed transmitter upper bound and the ergodic capacity. It is interesting to observe that the behavior of the JE and the aJE schemes in case of finite delay constraint ($M = 100$) closely follows the results shown for the asymptotic case in Section 4.3.1. The aJE scheme adapts the average transmission rate at message level to the average channel capacity. We recall that, in the aJE scheme, the transmitter only has a statistical knowledge of the channel, and yet gets pretty close to the performance of a genie-aided transmitter even for a reasonably low number of channel blocks. We further

notice how the adaptive JE scheme closely approaches the ergodic capacity, even though data arrives gradually at the transmitter during the transmission, instead of being available at the beginning, which is generally assumed for the achievability of the ergodic capacity [61]. We should note that in Fig. 4.12 the average transmission rate is optimized for each given distance for the aJE scheme, while such optimization is not done for the other schemes. Thus, in case two (or more) terminals have different distances from the transmitter, the optimization can no longer be performed and a tradeoff between the throughputs of the two nodes would be needed. The performance can be improved by considering a combination of the aJE scheme with the TS or ST schemes. The plots in Fig. 4.12 show how TS, MT and ST schemes are more robust compared to the JE scheme, as their average throughputs decreases smoothly with the distance, unlike the JE scheme, which has a sudden drop. This provides robustness in the case of multiple receivers with different average SNRs or when the channel statistics information at the transmitter is not accurate.

4.6 Conclusions

We have considered a transmitter streaming data to a receiver over a block fading channel, such that the transmitter is provided with an independent message at a fixed rate at the beginning of each channel block. We have used the average throughput as our performance metric. We have proposed several new transmission schemes based on joint encoding, time-division and superposition encoding. A general upper bound on the average throughput has also been introduced assuming the availability of CSI at the transmitter.

We have shown analytically that the joint encoding (JE) scheme has a threshold behavior. It performs well when the target rate is below the average channel capacity $\overline{C}$, while its performance drops sharply when the target rate surpasses $\overline{C}$. To adapt to an average channel capacity that is below the fixed message rate r, the adaptive joint encoding (aJE) scheme transmits only some of the messages. We have proved analytically that the aJE scheme is asymptotically optimal, achieving the ergodic capacity as the number of channel blocks goes to infinity, even though data arrives gradually over time at a fixed rate, rather than being available initially. We have also shown numerically that, even for a finite number of messages, the aJE scheme outperforms other schemes in all the considered settings and performs close to the upper bound.

The schemes based on the joint encoding of the messages (JE and aJE) create an M-block long concatenated code, such that either all or none of the messages can be decoded. This is useful when the underlying application has a minimum rate requirement that needs to be satisfied over M channel blocks, or when the average SNRs of the users vary over a limited range of SNR values. Independent encoding made in time sharing-based schemes (TS, gTS, MT), instead, makes each message less strictly dependent on the decoding of the others, allowing some of them to be decoded also at low average SNR but, on the other hand, implying the possibility not to decode some of them in when the average SNR is high. The ST scheme, based on message superposition, collocates itself between JE- and TS-based schemes, as messages are encoded independently, but the probability of correctly decoding each one of them is affected by the decoding of the others.

We conclude that the aJE scheme is advantageous in systems with a single receiver or with multiple receivers having similar average SNR values, as the performance of the user with the highest average SNR is limited by the user with the lowest average SNR. On contrast, the gTS and ST schemes can be attractive when broadcasting to multiple users with a wide range of SNR values, or in a point-to-point system with inaccurate CSI, as their average throughputs decrease gradually with decreasing SNR.

Appendix 4

Proof of Theorem 4.1

Let B_k denote the event "the first k messages can be decoded at the end of channel block k", while B_k^c denotes the complementary event. The event B_k holds if and only if

$$C_{k-i+1} + C_{k-i+2} + \cdots + C_k \geq ir \tag{4.29}$$

is satisfied for all $i = 1, \ldots, k$. Let $E_{k,j}$ denote the event "the j-th inequality needed to decode the first k messages in k channel blocks is satisfied", that is:

$$E_{k,j} \triangleq \{C_{k-j+1} + \cdots + C_k \geq jr\}, \tag{4.30}$$

for $j = 1, \ldots, k$, while $E_{k,j}^c$ denotes the complementary event.

Note that in the JE scheme if m messages are decoded these are the first m messages. Let n_d denote the number of decoded messages at the end of channel block M. Then the average

throughput is

$$\overline{r}_{JE} = r[Pr\{n_d \geq 1\} + Pr\{n_d \geq 2\} + \cdots + \Pr\{n_d \geq M-1\} + \Pr\{n_d \geq M\}]. \quad (4.31)$$

The k-th term in the sum of Eqn. (4.31) is the probability of decoding *at least* k (i.e. k or more) messages. Each term in Eqn. (4.31) can be expressed as the sum of two terms as:

$$Pr\{n_d \geq k\} = Pr\{B_k, n_d \geq k\} + Pr\{B_k^c, n_d \geq k\} \quad (4.32)$$

The first term in Eqn. (4.32) is the probability of "decoding k messages at the end of channel block k *and* decoding at least k messages at the end of M channel blocks". If B_k holds, the event "decode at least k messages at the end of channel block M" is satisfied; hence, we have:

$$Pr\{B_k, n_d \geq k\} = Pr\{B_k\} = Pr\{E_{k,1}, \ldots, E_{k,k}\}. \quad (4.33)$$

As for the second term in Eqn. (4.32), it is the probability of decoding at least k messages but *not* k at the end of channel block k. It can be further decomposed into the sum of two terms:

$$Pr\{B_k^c, n_d \geq k\} = Pr\{B_k^c, B_{k+1}, n_d \geq k\} + Pr\{B_k^c, B_{k+1}^c, n_d \geq k\}. \quad (4.34)$$

The event $n_d \geq k$ holds if the condition B_{k+1} is satisfied (i.e., if $k+1$ messages are decoded at the end of block $k+1$, then more than k messages are decoded at the end of channel block M); hence, we have:

$$Pr\{B_k^c, B_{k+1}, n_d \geq k\} = Pr\{B_k^c, B_{k+1}\}.$$

Plugging these into Eqn. (4.32), we obtain

$$Pr\{n_d \geq k\} = Pr\{B_k\} + Pr\{B_k^c, B_{k+1}\} + Pr\{B_k^c, B_{k+1}^c, n_d \geq k\}. \quad (4.35)$$

We can continue in a similar fashion, so that, the event "at least k messages are decoded" can be written as the union of the disjoint events ("k messages are decoded in k slots") $\bigcup$ ("k messages are not decoded in k slots but $k+1$ messages are decoded in $k+1$ slots") $\bigcup \cdots \bigcup$ ("no message can be decoded before slot M but M messages are decoded in slot M"). Hence, by the law of total probability, we have:

$$Pr\{n_d \geq k\} = \sum_{j=k}^{M} Pr\{B_k^c, B_{k+1}^c, \ldots, B_{j-1}^c, B_j\}. \quad (4.36)$$

Note that each term of the sum in Eqn. (4.36) says nothing about what happens to messages beyond the j-th, which can either be decoded or not. Plugging Eqn. (4.36) in Eqn. (4.31) we find:

$$\begin{aligned} E[m] &= \sum_{k=1}^{M}\sum_{j=k}^{M} Pr\{B_k^c, B_{k+1}^c, \ldots, B_{j-1}^c, B_j\} \\ &= \sum_{j=1}^{M}\sum_{k=1}^{j} Pr\{B_k^c, B_{k+1}^c, \ldots, B_{j-1}^c, B_j\}. \end{aligned} \tag{4.37}$$

We can rewrite each of these events as the intersection of events of the kind $E_{k,i}$ and $E_{k,i}^c$. Each term of the sum in Eqn. (4.37) can be written as:

$$\begin{aligned} Pr\{B_k^c, B_{k+1}^c, \ldots, B_{j-1}^c, B_j\} = Pr\{&E_{k,1}, B_k^c, B_{k+1}, \ldots, B_{j-1}^c, B_j\} \\ &+Pr\{E_{k,1}^c, B_k^c, B_{k+1}^c, \ldots, B_{j-1}^c, B_j\}. \end{aligned} \tag{4.38}$$

As the event $E_{k,1}^c$ implies the event B_k^c, this can be removed from the second term on the right hand side of Eqn. (4.38). Note that, in general, the event $E_{k,i}^c$, $i \in \{1, \cdots, k\}$ implies the event B_k^c. In order to remove the event B_k^c from the first term as well, we write it as the sum of probabilities of two disjoint events: one intersecting with $E_{k,2}$ and the other with $E_{k,2}^c$. Then we get:

$$\begin{aligned} Pr\{B_k^c, B_{k+1}^c, \ldots, B_{j-1}^c, B_j\} &= Pr\{E_{k,1}, E_{k,2}, B_k^c, \ldots, B_{j-1}^c, B_j\} \\ &+ Pr\{E_{k,1}, E_{k,2}^c, B_k^c, \ldots, B_{j-1}^c, B_j\} \\ &+ Pr\{E_{k,1}^c, B_{k+1}^c, \ldots, B_{j-1}^c, B_j\}. \end{aligned} \tag{4.39}$$

Now B_k^c can be removed from the second term of the sum thanks to the presence of $E_{k,2}^c$. Each of the terms in the right hand side of Eqn. (4.39) can be further written as the sum of the probabilities of two disjoint events, and so on so forth. The process is iterated until all the B_d^c, $d < j$ events are eliminated and we are left with the intersections of events only of the type $E_{p,q}$ and $E_{p,q}^c$, for some $p, q \in \{k, k+1, \ldots, M\}$ and B_j. The iteration is done as follows:

For each term of the summation, we take the B_l^c event with the lowest index. If any $E_{l,j}^c$ event is present, then B_l^c can be eliminated. If not, we write the term as the sum of the two probabilities corresponding to the events which are the intersections of the B_l^c event with $E_{l,d+1}$ and $E_{l,d+1}^c$, respectively, where d is the highest index j among the events in which $E_{l,j}$ is already present. The iteration process stops when $l = j$.

At the end of the process all the probabilities involving events $B_k^c, \ldots, B_{j-1}^c$ will be removed and replaced by sequences of the kind:

$$\{E_{k,1}, E_{k,2}, \ldots, E_{k,i_k}^c, E_{k+1,i_k+1}, \ldots, E_{k+1,i_{k+1}}^c, \ldots, E_{j-1,i_{j-2}+1}, E_{j-1,i_{j-1}}^c, B_j\}, \quad (4.40)$$

where $i_{j-1} \in \{j-1-k, \ldots, j-1\}$ is the index corresponding to the last inequality required to decode $j-1$ messages, which is not satisfied. Note that exactly one $E_{l,r}^c$ event for each B_l^c is present after the iteration.

For B_j to hold, all the events $E_{j,1}, \ldots, E_{j,j}$ must hold. It is easy to show that, after the iterative process used to remove the B_l^c's, the event $E_{j,i_{j-1}+1}$ ensures that all the events required for B_j with indices lower than or equal to i_{j-1} automatically hold. Thus, we can add the events $\{E_{j,i_{j-1}+1}, \ldots, E_{j,j}\}$ to guarantee that B_j holds, and remove it from the list. It is important to notice that the term $E_{j,j}$ is always present. At this point we are left with the sum of probabilities of events, which we call *E-events*, each of which is the intersection of events of the form $E_{i,j}$ and $E_{i,j}^c$. Thus, an E-event S_k^j has the following form:

$$\begin{aligned} S_k^j \triangleq \{&E_{k,1}, E_{k,2}, \ldots, E_{k,i_k}^c, E_{k+1,i_k+1}, \ldots, E_{k+1,i_{k+1}}^c, \\ &\ldots, E_{j-1,i_{j-2}+1}, E_{j-1,i_{j-1}}^c, E_{j,i_{j-1}+1}, \ldots, E_{j,j}\}. \end{aligned} \quad (4.41)$$

By construction, the number of E-events for the generic term j of the sum in Eqn. (4.37) is equal to the number of possible dispositions of $j-k$ E^c's over $j-1$ positions. As the number of events of type E^c is different for the E-events of different terms in Eqn. (4.37), the E-events relative to two different terms of Eqn. (4.37) are different. We define $\mathcal{S}_j$ as the set of all E-events which contain the event $E_{j,j}$. The elements of $\mathcal{S}_j$ correspond to all the possible ways in which j messages can be decoded at the end of block number j. The cardinality of $\mathcal{S}_j$ is equal to:

$$|\mathcal{S}_j| = \sum_{k=1}^{j} \frac{(j-1)!}{(k-1)!(j-k)!} = 2^{j-1}. \quad (4.42)$$

Now we want to prove that

$$\sum_{S_k^j \in \mathcal{S}_j} Pr\{S_k^j\} = Pr\{E_{j,j}\}. \quad (4.43)$$

Note that each $E_{k,l}$ corresponds to a different event if the index k is different, even for the same index l; thus, the law of total probability can not be directly applied to prove Eqn. (4.43). However, we will prove in the following preposition that $Pr\{E_{k_1,l}\} = Pr\{E_{k_2,l}\}$, $\forall k_1, k_2$.

Proposition 1: Let us consider a set of random variables $C_1, \cdots, C_j$ that are conditionally i.i.d. given U. Given any two ordering vectors $\mathbf{i} = i_1, i_2, \cdots, i_j$ and $\mathbf{l} = l_1, l_2, \cdots, l_j$, we have

$$\begin{aligned} &Pr\{C_{i_1} \gtrless r, \ldots, C_{i_1} + \cdots + C_{i_j} \gtrless jr\} \\ &= Pr\{C_{l_1} \gtrless r, \ldots, C_{l_1} + \cdots + C_{l_j} \gtrless jr\}. \end{aligned} \tag{4.44}$$

Proof: The left hand side of Eqn. (4.44) can be rewritten as:

$$\begin{aligned} &Pr\{C_{i_1} \gtrless r, \ldots, C_{i_1} + \cdots + C_{i_j} \gtrless jr\} \\ &= \int_{-\infty}^{+\infty} du \int_{\theta_1^{low}}^{\theta_1^{up}} dc_{i_1} \ldots \int_{\theta_j^{low}}^{\theta_j^{up}} dc_{i_j} f_{\mathbf{C_i}|U}(\mathbf{c_i}|u) f_U(u), \end{aligned} \tag{4.45}$$

where $\mathbf{C_i} = C_{i_1}, \ldots, C_{i_j}$ and $\mathbf{c_i} = c_{i_1}, \ldots, c_{i_j}$, while θ_h^{low} and θ_h^{up} are the lower and upper extremes of the integration interval. θ_h^{low} is either equal to $-\infty$ or to $hr - c_{i_1} - \cdots - c_{i_{h-1}}$, $\forall h \in \{1, \ldots, j\}$, depending on whether there is a $<$ or a $\geq$ in the h-th inequality within brackets in Eqn. (4.45), respectively, while θ_h^{up} is either equal to $hr - c_{i_1} - \cdots - c_{i_{h-1}}$ or to $+\infty$ depending on whether there is a $<$ or a $\geq$ in the h-th inequality of Eqn. (4.45), respectively. By using Eqn. (4.8) and Eqn. (4.9) we can write:

$$\begin{aligned} Pr\{C_{i_1} \gtrless r, \ldots, C_{i_1} + \cdots + C_{i_j} \gtrless jr\} &= \int_{-\infty}^{+\infty} du f_U(u) \int_{\theta_1^{low}}^{\theta_1^{up}} dc_{i_1} \ldots \int_{\theta_j^{low}}^{\theta_j^{up}} dc_{i_j} f_{\mathbf{C}_i|U} \\ &= \int_{-\infty}^{+\infty} du f_U(u) \int_{\theta_1^{low}}^{\theta_1^{up}} dc_{l_1} \ldots \int_{\theta_j^{low}}^{\theta_j^{up}} dc_{l_j} f_{\mathbf{C}_l|U} \\ &= Pr\{C_{l_1} \gtrless r, \ldots, C_{l_1} + \cdots + C_{l_j} \gtrless jr\}, \end{aligned} \tag{4.46}$$

where we defined $f_{\mathbf{C}_i|U} \triangleq f_{C_{i_1}|U}(c_{i_1}|u) \cdots f_{C_{i_j}|U}(c_{i_j}|u)$ and $f_{\mathbf{C}_l|U} \triangleq f_{C_{l_1}|U}(c_{l_1}|u) \times \cdots \times f_{C_{l_j}|U}(c_{l_j}|u)$ The proposition above guarantees that, although these events do not partition the whole probability space of $E_{j,j}$, their probabilities add up to that of $E_{j,j}$, i.e.:

$$\sum_{k=1}^{2^{j-1}} Pr\{S_k^j\} = Pr\{E_{j,j}\} = Pr\{C_1 + \cdots + C_j \geq jr\}. \tag{4.47}$$

Finally, plugging Eqn. (4.47) into Eqn. (4.37) we can write:

$$E[m] = \sum_{j=1}^{M} \sum_{k=1}^{j} Pr\{B_k^c, B_{k+1}^c, \ldots, B_{j-1}^c, B_j\} \tag{4.48}$$

$$= \sum_{j=1}^{M} \sum_{S_k^j \in \mathcal{S}_j} Pr\{S_k^j\}$$

$$= \sum_{j=1}^{M} Pr\{C_1 + \cdots + C_j \geq jr\}. \tag{4.49}$$

Achievability of ergodic capacity with aJE

In the following we prove that the average throughput of the aJE scheme $\overline{r}_{aJE}$ approaches αr for large M if $\overline{C} > \alpha r$. Similarly to the JE scheme, it is sufficient to prove that, if $\overline{C} > \alpha r$,

$$\lim_{M\to\infty} \sum_{m=1}^{M\alpha} a_m^* = c, \tag{4.50}$$

for some $0 < c < \infty$, where $a_m^* \triangleq Pr\left\{\frac{C_1^*+\cdots+C_m^*}{m} < r\right\}$. We can rewrite a_m^* as:

$$a_m^* = Pr\left\{ l_m > \frac{\overline{C}/\alpha - r}{\sigma_c\sqrt{\left(\frac{1}{m} + \frac{1-\alpha}{M\alpha^2}\right)}} \right\}, \tag{4.51}$$

where

$$l_m \triangleq \frac{\overline{C}/\alpha - \frac{C_1+\cdots+C_m}{m} - \frac{(1-\alpha)}{\alpha}\frac{1}{M(1-\alpha)}\sum_{j=M\alpha+1}^{M} C_j}{\sigma_c\sqrt{\frac{1}{m} + \frac{1-\alpha}{M\alpha^2}}} \tag{4.52}$$

is a random variable with zero mean and unit variance. From the law of large numbers applied to Eqn. (4.51), we have $\lim_{m\to+\infty} a_m^* = 0$. First we show that

$$\lim_{m\to+\infty} \left(\frac{a_m^*}{d_m}\right) = c', \tag{4.53}$$

for some $0 < c' < +\infty$ where we have defined:

$$d_m \triangleq Pr\left\{ l'_m > \frac{\overline{C}/\alpha - r}{\sigma_c\sqrt{\left(\frac{1}{m} + \frac{1-\alpha}{m\alpha^2}\right)}} \right\}, \tag{4.54}$$

and

$$l'_m \triangleq \frac{\overline{C}/\alpha - \frac{C_1+\cdots+C_m}{m} - \frac{(1-\alpha)}{\alpha}\frac{1}{m(1-\alpha)}\sum_{j=M\alpha+1}^{M\alpha+m} C_j}{\sigma_c\sqrt{\frac{1}{m} + \frac{1-\alpha}{m\alpha^2}}} \tag{4.55}$$

such that l'_m is a random variable with zero mean and unit variance. We have

$$\begin{aligned}
\lim_{m\to+\infty} \left(\frac{a_m^*}{d_m}\right) &= \lim_{m\to+\infty} \frac{Pr\left\{ l_m > \frac{\overline{C}/\alpha - r}{\sigma_c\sqrt{\left(\frac{1}{m} + \frac{1-\alpha}{M\alpha^2}\right)}} \right\}}{Pr\left\{ l'_m > \frac{\overline{C}/\alpha - r}{\sigma_c\sqrt{\left(\frac{1}{m} + \frac{1-\alpha}{m\alpha^2}\right)}} \right\}} \\
&= \lim_{m\to+\infty} \frac{Q\left(\frac{\overline{C}/\alpha - r}{\sigma_c\sqrt{\left(\frac{1}{m} + \frac{1-\alpha}{M\alpha^2}\right)}}\right)}{Q\left(\frac{\overline{C}/\alpha - r}{\sigma_c\sqrt{\left(\frac{1}{m} + \frac{1-\alpha}{m\alpha^2}\right)}}\right)} \qquad (4.56) \\
&\leq 1, \qquad (4.57)
\end{aligned}$$

where inequality (4.57) follows from the fact that $m < M$ and $Q(x)$ is monotonically decreasing in x. Then we show that

$$\lim_{M\to\infty} \sum_{m=1}^{M\alpha} d_m = c'', \tag{4.58}$$

for some $0 < c'' < +\infty$. To prove the convergence of the series sum we show that $\lim_{m\to+\infty} \frac{d_{m+1}}{d_m} = \lambda'$, for some $0 < \lambda' < 1$. From the central limit theorem we can write:

$$\begin{aligned}
\lim_{m\to+\infty} \frac{d_{m+1}}{d_m} &= \lim_{m\to+\infty} \frac{Pr\left\{ l_{m+1} > \frac{\overline{C}/\alpha - r}{\sigma_c\sqrt{\left(\frac{1}{m+1}+\frac{1-\alpha}{(m+1)\alpha^2}\right)}} \right\}}{Pr\left\{ l_m > \frac{\overline{C}/\alpha - r}{\sigma_c\sqrt{\left(\frac{1}{m}+\frac{1-\alpha}{m\alpha^2}\right)}} \right\}} \\
&= \lim_{m\to+\infty} \frac{Q\left(\frac{\overline{C}/\alpha - r}{\sigma_c\sqrt{\left(\frac{1}{m+1}+\frac{1-\alpha}{(m+1)\alpha^2}\right)}} \right)}{Q\left(\frac{\overline{C}/\alpha - r}{\sigma_c\sqrt{\left(\frac{1}{m}+\frac{1-\alpha}{m\alpha^2}\right)}} \right)} \\
&\leq \lim_{m\to+\infty} \frac{\frac{\sigma_c\sqrt{\left(\frac{1}{m+1}+\frac{1-\alpha}{(m+1)\alpha^2}\right)}}{(\overline{C}/\alpha - r)\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\overline{C}/\alpha - r}{\sigma_c\sqrt{\left(\frac{1}{m+1}+\frac{1-\alpha}{(m+1)\alpha^2}\right)}} \right)^2}}{\frac{\frac{\overline{C}/\alpha - r}{\sigma_c\sqrt{\left(\frac{1}{m}+\frac{1-\alpha}{m\alpha^2}\right)}}}{1+\left(\frac{\overline{C}/\alpha - r}{\sigma_c\sqrt{\left(\frac{1}{m}+\frac{1-\alpha}{m\alpha^2}\right)}} \right)^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\overline{C}/\alpha - r}{\sigma_c\sqrt{\left(\frac{1}{m}+\frac{1-\alpha}{m\alpha^2}\right)}} \right)^2}} \qquad (4.59) \\
&= \lim_{m\to+\infty} \frac{\sqrt{\left(\frac{1}{m+1}+\frac{1-\alpha}{(m+1)\alpha^2}\right)}}{(\overline{C}/\alpha - r)^2} \frac{\left[\sigma_c^2\left(\frac{1}{m}+\frac{1-\alpha}{m\alpha^2}\right) + (\overline{C}/\alpha - r)^2\right]}{\sqrt{\frac{1}{m}+\frac{1-\alpha}{m\alpha^2}}} e^{-\frac{(\overline{C}/\alpha - r)^2}{2\sigma_c^2}\left(\frac{\alpha^2}{\alpha^2-\alpha+1}\right)} \\
&= e^{-\frac{(\overline{C}/\alpha - r)^2}{2\sigma_c^2}\left(\frac{\alpha^2}{\alpha^2-\alpha+1}\right)} \\
&< 1, \qquad (4.60)
\end{aligned}$$

where inequality (4.59) follows from Eqn. (4.20).

From Eqn. (4.60) it follows that $\lim_{M\to\infty} \overline{r}_{aJE} = r$ if $\overline{C} > \alpha r$. Similarly, it can be easily shown that $\lim_{M\to\infty} \overline{r}_{aJE} = 0$ if $\overline{C} < \alpha r$.

Chapter 5

Throughput and Delay Analysis in Video Streaming over Block-Fading Channels

"Bottomless wonders spring from simple rules, which are repeated without end."

Benoît Mandelbrot

Contents

5.1 Introduction

In Chapter 4 we studied the problem of real time video streaming over block fading channel with a global delay constraint. In the present chapter we focus on non-real time video

streaming with a per-message delay constraint. Consider a wireless terminal (e.g., a mobile terminal with a satellite connection or a smartphone) *streaming* a video file from a wireless server. In a streaming application, the user starts watching the video before the entire video file is downloaded; hence, the video packets need to be received in the order of display, adding individual delay constraints for different video packets as opposed to traditional video downloading. The goal is to transmit as many video frames as possible within the corresponding deadlines. However, in a video application, in addition to the average throughput, the quality of user experience (QoE) depends also on the delay between decoded video frames at the receiver, i.e., the *inter-decoding delay*. Therefore, in this chapter we consider both the average throughput and the inter-decoding delay in a video streaming application over a wireless fading channel.

Delay constraints are common in multimedia communications as end-to-end delay is an important aspect of QoE. As described in the previous chapter, if the delay constraint is short compared to the channel coherence time and the CSI is available only at the receiver, outages become unavoidable [68]. In such a scenario, throughput or outage capacity can be the appropriate performance metrics. There is an extensive literature on delay limited transmission over wireless fading channels (see [59] and references therein).

When we focus exclusively on channel coding, the transmitter can adapt to the average channel statistics through rate splitting and superposition transmission as in the broadcast approach of [65]. It is shown in [78] and [79] that in a video transmission application, such a superposition approach improves the end-to-end video quality significantly. However, this kind of fine adaptation is not viable in practical multimedia communication systems in which the encoding rate is fixed by a higher layer application[1]. Moreover, in most current systems the design is strictly layered and the channel encoder is oblivious to the video coding scheme used by the application layer; and hence, rate adaptation is not possible at the code level. The encoder receives video packets already encoded at a fixed rate; and cannot split the packets or change the encoding rate. On the other hand the encoder can choose to drop some of the video packets, and achieve rate adaptation at the packet level at the expense of inter-decoding delay.

In the Moving Picture Experts Group (MPEG) standard, video signals are encoded into group of pictures (GOP), each consisting of an I- frame and a number of P- and B-frames [80].

[1]Some streaming protocols, such as HTTP Live Streaming, allow rate adaption among a limited number of available rates.

Hence, each GOP can be decoded and displayed independent of the previous and following GOP's. Focusing on a block fading channel model, we assume that a whole GOP (or an integer number of GOP's) forms one video packet, and the video coding rate is such that one video packet needs to be transmitted within each channel block, i.e., channel coherence time is equal to the time between two consecutive GOP's. In the streaming scenario, this imposes a different decoding deadline for each video packet, i.e., first packet needs to be received after the first channel block, second packet after the second block, and so on so forth. Modeling the decoder at each block as a distinct virtual receiver, this channel can be seen as a physically degraded fading broadcast channel with as many virtual users as the number of channel blocks. The loss of a video packet implies a blockage in the display process which lasts until the next packet is received. Hence, both the average throughput and the maximum inter-decoding delay are considered to quantify the QoE. Both metrics have previously been considered as measures of QoE [81].

We propose four different transmission schemes based on time-sharing. We exclusively focus on time-sharing transmission mostly because of its applicability in practical systems, as it leads to lower complexity decoding schemes with respect to, for example, successive interference cancelation, which is required in the case of superposition transmission. Moreover, the throughput and delay analysis is not completely understood even for this relatively simpler transmission scheme. In particular, we will consider *memoryless transmission (MT)*, *equal time-sharing (eTS)*, *pre-buffering (PB)* and *windowed time-sharing (wTS)* schemes. We also consider an informed transmitter bound on the achievable throughput and delay performances assuming perfect CSI at the transmitter. We compare the achievable schemes and the informed transmitter bound in terms of both throughput and maximum inter-decoding delay. Our results provide fundamental performance bounds as well as an insight for the design of practical video streaming systems over wireless fading channels.

While there is an extensive literature on the higher layer analysis of video streaming applications [82], research on the physical layer aspects of streaming focus mostly on code construction [70], [71], [72]. The diversity-multiplexing trade-off for a data streaming system is studied in [73]. The channel model we study in this chapter can be seen as the dual of the streaming transmitter model we have studied in the previous chapter and in [74], in which the data packets, rather than being available at the transmitter in advance, arrive at the transmitter gradually over time.

5.2 System Model

We consider a video streaming system over a block fading channel. The channel is constant for a block of n channel uses and changes in an i.i.d. manner from one block to the next. We assume that the file to be streamed to the receiver consists of M independent packets denoted by $W_1, \ldots, W_M$, all available at the transmitter. The receiver wants to decode these packets gradually as the transmitter continues the transmission. We assume that the packet W_t needs to be decoded by the end of channel block t, $t = 1, \ldots, M$, otherwise it becomes useless. The data packets all have the same size; and it is assumed that each packet is generated at rate r bits per channel use (bpcu) which is fixed by the application layer, i.e., W_t is chosen randomly with uniform distribution from the set $\mathcal{W}_t = \{1, \ldots, 2^{nr}\}$. The channel in block t is given by

$$\mathbf{y}[t] = h[t]\mathbf{x}[t] + \mathbf{z}[t],$$

where $h[t]$ is the channel state, $\mathbf{x}[t]$ is the length-n channel input vector, $\mathbf{z}[t]$ is a vector of i.i.d. zero mean unit-variance Gaussian noise, and $\mathbf{y}[t]$ is the length-n channel output vector at the receiver. Instantaneous channel gains are known only at the receiver. We have a short-term average power constraint of P, i.e., $\mathrm{E}[\mathbf{x}[t]\mathbf{x}[t]^\dagger] \leq nP$ for $t = 1, \ldots, M$, where $\mathbf{x}[t]^\dagger$ represents the Hermitian transpose of $\mathbf{x}[t]$.

The channel from the source to the receiver can be seen as a physically degraded broadcast channel, in which the decoder at each channel block acts as a virtual receiver trying to decode the packet corresponding to its channel block. See Fig. 5.1 for an illustration of this channel model. We denote the instantaneous channel capacity over channel block t by C_t:

$$C_t \triangleq \log_2(1 + \phi[t]P), \tag{5.1}$$

where $\phi[t] = |h[t]|^2$ is a random variable distributed according to a probability density function (pdf) $f_\Phi(\phi)$.

We define the average throughput, $\overline{r}$, as the average rate at the end of M channel blocks:

$$\overline{r} \triangleq \frac{r}{M} \sum_{m=1}^{M} m \cdot \eta(m), \tag{5.2}$$

where $\eta(m)$ is the probability of decoding exactly m messages out of M. In addition to the average throughput, we also study the maximum number of consecutive channel blocks in which no message is decoded, denoted by D^{max}. Since D^{max} is also a random variable whose

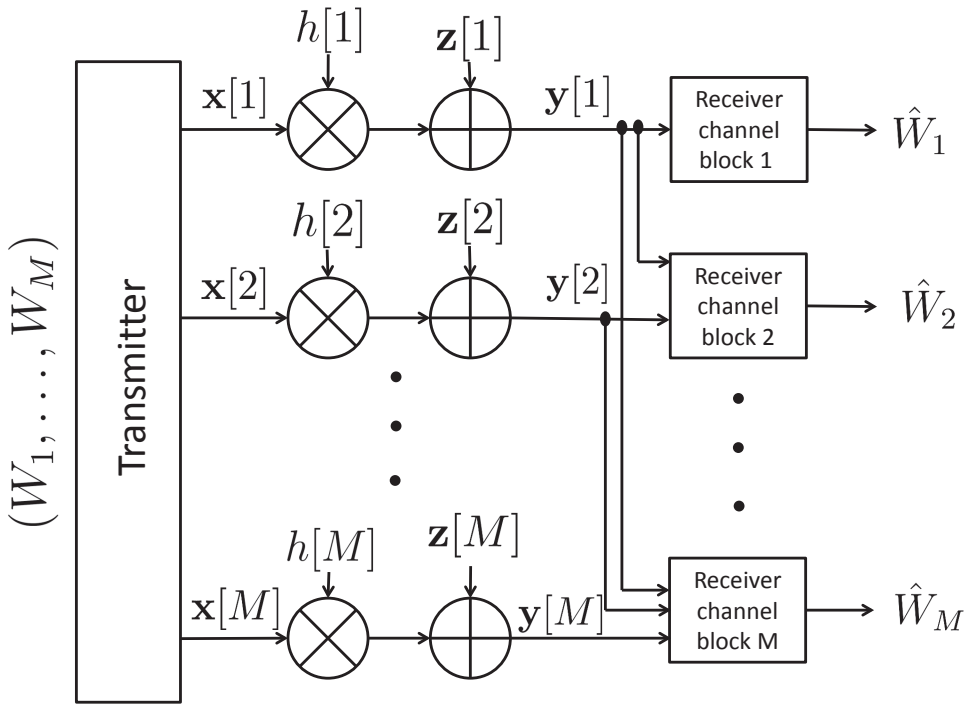


Figure 5.1: Equivalent channel model for streaming a video file composed of M packets over M blocks of the fading channel to a single receiver.

realization depends on the channel state, we consider the *average maximum delay* $\overline{D}^{max}$ as our performance measure. We have:

$$\overline{D}^{max} \triangleq \sum_{d=1}^{M} d \cdot Pr\{D^{max} = d\} = \sum_{d=1}^{M} Pr\{D^{max} \geq d\}. \tag{5.3}$$

The second equality in Eqn. (5.3) derives from the definition of mean value. We show this for the case with $M = 3$, the extension to any value of M is straightforward. By definition of mean value we have:

$$\begin{aligned}
\overline{D}^{max} &\triangleq \sum_{d=1}^{3} d \cdot Pr\{D^{max} = d\} \\
&= 1 \cdot Pr\{D^{max} = 1\} + 2 \cdot Pr\{D^{max} = 2\} + 3 \cdot Pr\{D^{max} = 3\} \\
&= (Pr\{D^{max} = 1\} + Pr\{D^{max} = 2\} + Pr\{D^{max} = 3\}) \\
&\quad + (Pr\{D^{max} = 2\} + Pr\{D^{max} = 3\}) \\
&\quad + Pr\{D^{max} = 3\} \\
&= Pr\{D^{max} \geq 1\} + Pr\{D^{max} \geq 2\} + Pr\{D^{max} \geq 3\} \\
&= \sum_{d=1}^{3} Pr\{D^{max} \geq d\}.
\end{aligned} \tag{5.4}$$

5.3 Informed Transmitter Bound

We first provide an upper bound on the achievable average throughput and a lower bound on the average maximum delay by assuming that the transmitter is informed about the exact channel realization over all M channel blocks from the very beginning. This allows the transmitter to optimally allocate the available resources among the messages. In particular, knowing the channels *a priori* the transmitter can choose the optimal subset S_{opt} of messages to be transmitted. Note that power allocation is not possible due to short-term power constraint. In order to find the set of messages S_{opt} that minimizes the average maximum delay, we first need to find the maximum number of decodable messages given the channel realization. It follows from the physically degraded broadcast channel model depicted in Fig. 5.1 that the total number of messages that can be decoded up to channel block t, denoted as $\Psi_d(t)$, $t = 1, \ldots, M$, is bounded as:

$$\Psi_d(t) \leq \min \left\{ t, \left\lfloor \frac{I^{tot}(t)}{r} \right\rfloor \right\}, \tag{5.5}$$

where $I^{tot}(t) = \sum_{i=1}^{t} C_i$, is the total mutual information (MI) accumulated until channel block t, while $\lfloor x \rfloor$ is the largest integer smaller than or equal to x. At each channel block t, we check whether we can decode packet W_t on top of the packets that have already been decoded. Note that there is no gain in decoding a packet in advance of its decoding deadline. Let $v_t^d \in \{0, 1\}$ denote whether W_t is decoded or not, i.e., $v_t^d = 1$ if W_t is decoded and $v_t^d = 0$ if not. We have $\Psi_d(t) = v_1^d + \cdots + v_t^d$, and

$$v_{t+1}^d = \begin{cases} 1 & \text{if } I^{tot}(t+1) \geq (\Psi_d(t) + 1)\, r, \\ 0 & \text{otherwise.} \end{cases} \tag{5.6}$$

This recursion returns $\mathbf{V}_d = [v_1^d \cdots v_M^d]$, an M-length binary vector, or equivalently, the transmission scheme that maximizes the throughput, but may lead to a suboptimal result in terms of maximum delay. From a delay perspective it may be a better choice not to transmit some of the packets even if enough mutual information is accumulated by their deadlines, and instead transmitting packets that are further in time. This is equivalent to shift rightwards some of the ones in $\mathbf{V}_d$ so as to minimize the number of consecutive zeroes in the vector. Note that this process leaves the throughput unchanged. Let us consider the following example. Assume that, with reference to Fig. 5.2, the iterative process described by Eqn. (5.6) returns the sequence

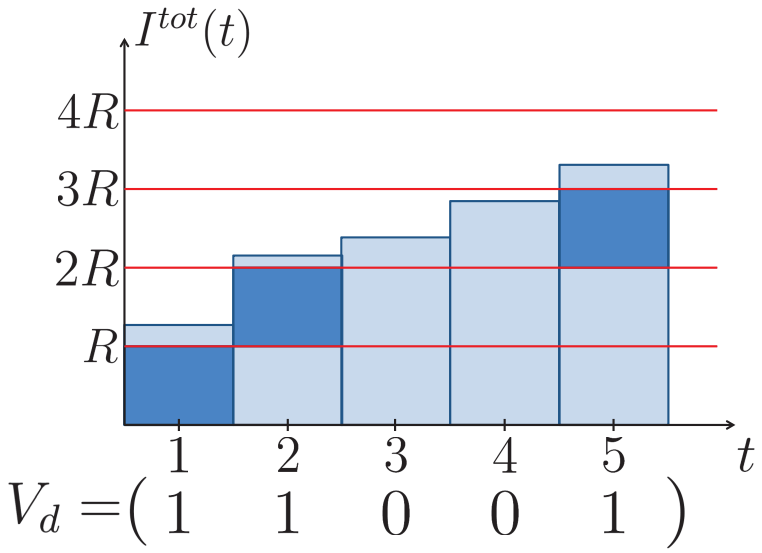


Figure 5.2: I^{tot} plotted against t and corresponding vector $\mathbf{V}_d$.

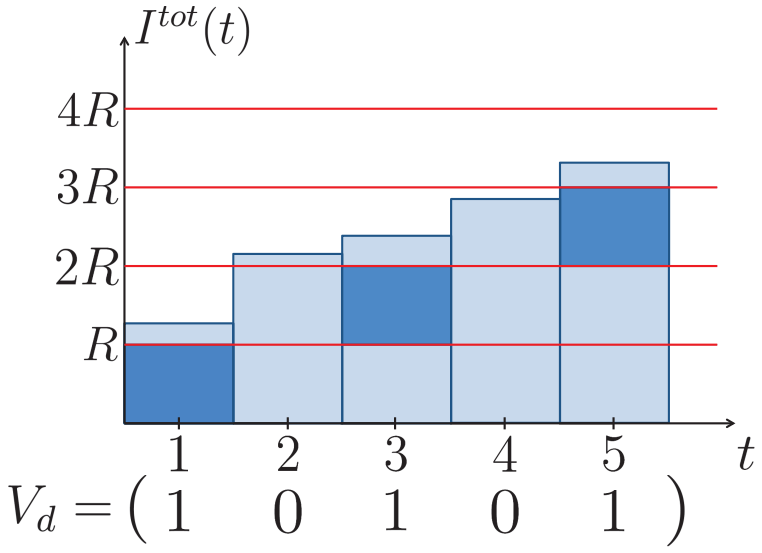


Figure 5.3: I^{tot} plotted against t and corresponding vector $\mathbf{V}_d$ in case of delay-optimal message choice.

$\mathbf{V}_d = [11001]$. This allocation strategy achieves a throughput of $3/5$ and a maximum delay of 2. However, a better choice for the transmitter is to transmit message W_3 instead of message W_2, as shown in Fig. 5.3. This gives the new allocation strategy $\mathbf{V}'_d = [10101]$, which has the same throughput as $\mathbf{V}_d$ but a maximum delay of 1.

In order to minimize the maximum delay, the transmitter can choose to drop a message even if it could be decoded with high probability. Instead, resources could be allocated to a

message with a higher index, which, if decoded, would lead to a lower maximum delay. Note that the maximum delay is optimized without decreasing the average throughput. The algorithm `Min_Del_Max_Rate` which optimizes both $\overline{r}$ and $\overline{D}^{max}$ can be obtained as follows.

Definition 5.1 $\mathbf{V}_{lb}(D)$, $D \leq M$, is the binary string of length M with the lowest maximum delay, i.e., maximum distance between consecutive zeros, attainable with $\left\lfloor M \cdot \frac{D}{D+1} \right\rfloor$ zeros that has the smallest decimal representation. $\mathbf{V}_{lb}(D)$ can be constructed by taking a sequence of M bits all equal to 0, and, starting from the $D+1$-th most significant bit, substituting a 0 with a 1, every D bits.

Definition 5.2 We define $\Psi^d(n) = \sum_{i=1}^{n} v_i^d$ and $\Psi_D^{lb}(n) = \sum_{i=1}^{n} v_i^{lb}(D)$, where v_i^d and $v_i^{lb}(D)$ are the i-th bits starting from the most significant bit of $\mathbf{V}_d$ and $\mathbf{V}_{lb}(D)$, respectively. In other words, $\Psi^v(n)$ and $\Psi_D^{lb}(n)$ are the cumulative sum of the elements of v_i^d and $v_i^{lb}(D)$, respectively, up to the n-th most significant bit.

Algorithm 1 `Min_Del_Max_Rate(`r, M, $N_d(M)$, $\mathbf{V}_d$`)`

$D_{IT}^{max} = \min\left\{D \mid \Psi^d(n) \geq \Psi_D^{lb}(n), \forall n \in \{1, \ldots, M\}\right\}$
$1\_index = 0$
$0\_index = 0$
$\xi_0 = [0, \ldots, 0]$
$\xi_1 = [0, \ldots, 0]$
for $k = 1$ to M **do**
  if $v_i^{lb}(D_{IT}^{max}) - v_i^d == 1$ **then**
    $0\_index = 0\_index + 1$
    $\xi_0[0\_index] = \xi_0[0\_index] + 1$
  end if
  if $v_i^{lb}(D_{IT}^{max}) - v_i^d == -1$ **then**
    $1\_index = 1\_index + 1$
    $\xi_1[1\_index] = \xi_1[1\_index] + 1$
  end if
end for
for $j = 1$ to length(ξ_0) **do**
  $V_d[\xi_0[j]] = 1$
  $V_d[\xi_1[length(\xi_0) - j]] = 0$
end for
$\mathbf{S}_{opt} = \mathbf{V}_d$
return $\mathbf{S}_{opt}$

Theorem 5.1 The minimum achievable maximum delay in a given realization is D_{IT}^{max}, such that $\Psi^d(n) \geq \Psi^{lb}_{D_{IT}^{max}}(n)$, $\forall n \in \{1,\ldots,M\}$, and $\exists m : \Psi^d(m) < \Psi^{lb}_{D_{IT}^{max}-1}(m)$.

Proof We start by showing that

$$\Psi^d(n) \geq \Psi^{lb}_{D^*}(n), \forall n \in \{1,\ldots,M\} \Rightarrow D_{IT}^{max} \geq D^*.$$

$\Psi^{lb}_{D^*}(n)$ is the total number of 1's present among the leftmost n bits of the tentative sequence $\mathbf{V}_{lb}(D^*)$. Thus $\Psi^d(n) \geq \Psi^{lb}_{D^*}(n)$, $\forall n \in \{1,\ldots,M\}$ implies that $\mathbf{V}_d$ has at least as many 1's as $\mathbf{V}_{lb}(D^*)$ on the left of $n, \forall n \in \{1,\ldots,M\}$. Thus $\mathbf{V}_d$ can be always turned into $\mathbf{V}_{lb}(D^*)$ by shifting some of the 1s to the right. If $\Psi^d(M) > \Psi^{lb}_{D^*}(M)$, instead, $\mathbf{V}_d$ can be turned into $\mathbf{V}_{lb}(D^*)$ up to a certain n^* and from n^*+1 to M it can be padded with ones, thus achieving a maximum delay that is no greater than D^*. Now we show that if $\exists m : \Psi^d(m) < \Psi^{lb}_{D^*-1}(m)$, then $D_{IT}^{max} > D^* - 1$. $\Psi^d(m) < \Psi^{lb}_{D^*-1}(m)$ implies that there are less 1's in positions $1 \rightarrow m$ of $\mathbf{V}_d$ than there are in the same positions of $\mathbf{V}_{ls}(D^*-1)$. Thus it is not possible to have the delay of $\mathbf{V}_d$ as low as $\mathbf{V}_{ls}(D^*)$ in the first m positions as the 1's can only be shifted rightwards; thus the maximum delay in the first part of $\mathbf{V}_d$ is larger than $D^* - 1$.

Using Theorem 5.1 the `Min_Del_Max_Rate` algorithm (Algorithm 1) has been obtained. The algorithm takes as inputs r, M, $N_d(M)$ and $\mathbf{V}_d$. $N_d(M)$ and $\mathbf{V}_d$ can be obtained using the recursion in Eqn. 5.6. The output of the algorithm is the set of messages $\mathbf{S}_{opt}$ that constitutes the optimal choice in terms of both throughput and maximum delay.

5.4 Transmission Schemes

In this section we introduce four different transmission schemes based on time-sharing. Each channel block is divided among the messages for which the deadline has not yet expired. Thus, while the first channel block is divided among all the messages $W_1,\ldots,W_M$, the second channel block is divided among messages $W_2,\ldots,W_M$, as the deadline of message W_1 expires at the end of the first block. In general the encoder divides the channel block t into $M-t+1$ portions $\alpha_{tt},\ldots,\alpha_{Mt}$, such that $\alpha_{mt} \geq 0$ and $\sum_{m=t}^{M} \alpha_{mt} = 1$. In channel block t, $\alpha_{mt}n$ channel uses are allocated for the transmission of message W_m. We assume that Gaussian codebooks are used in each portion for each message, and the corresponding codelengths are sufficient

to achieve the instantaneous capacity. Then the total amount of received mutual information relative to message W_m is:

$$I_m^{tot} \triangleq \sum_{t=1}^{m} \alpha_{mt} C_t. \tag{5.7}$$

The proposed schemes differ in the way the channel uses are allocated among the messages for which the deadline has not yet expired. Different time allocations lead to different average throughput and average maximum delay performances.

5.4.1 Memoryless Transmission (MT)

In *memoryless transmission (MT)* each message is transmitted only within the channel block just before its expiration, that is, message W_t is transmitted over channel block t. Equivalently we have $\alpha_{mt} = 1$, if $t = m$, and $\alpha_{mt} = 0$, otherwise. In MT, message W_t can be decoded if and only if $C_t \geq r$. Due to the i.i.d. nature of the channel state over blocks, the successful decoding probability $p \triangleq Pr\{C_t \geq r\}$ is constant over messages. The probability that exactly m messages are decoded is given by:

$$\eta(m) \triangleq \binom{M}{m} p^m (1-p)^{M-m}. \tag{5.8}$$

The average number of decoded messages for the MT scheme is $\overline{r}_{MT} = Mp$.

Next we derive the exact expression for the average maximum delay for MT, denoted by $\overline{D}_{MT}^{max}$. Consider Eqn. (5.3). The term $Pr\{D^{max} \geq d\}$ in the sum is the probability that a sequence of M Bernoulli random variables with parameter p contains at least d consecutive zeros. This probability can be evaluated by modeling the number of consecutive zeros as a Markov chain, and finding the probability of reaching the final absorbing state of d consecutive zeros. This probability is given in the following theorem:

Theorem 5.2: Let $x_1, \cdots, x_M$ be a sequence of i.i.d. Bernoulli random variables with parameter $p = \mathrm{E}\{x_i = 1\}$. The probability of having at least d consecutive zeros in the sequence is given by:

$$Pr\{D \geq d\} = a_{k0} + \sum_{i=1}^{d} a_{di} \binom{M - r_i - 1}{r_i - 1} \left(\frac{1}{\varphi_{di}}\right)^M, \tag{5.9}$$

where $d \in \{0, \ldots, M\}$, $k \leq d$ is the number of distinct zeros of the polynomial in z:

$$q_d(z) = 1 - p \sum_{j=1}^{k} z^j (1-p)^{j-1}, \tag{5.10}$$

φ_{di}, $i \in \{1, \ldots, k\}$ are the zeros of the polynomial in Eqn. (5.10) with multiplicity r_i, while a_{di}, $i \in \{1, \ldots, d\}$, are constants with values:

$$a_{di} = \lim_{z \to \varphi_{di}} \frac{(z - \varphi_{di})^{r_i} [z(1-p)]^d}{(1-z) q_d(z)}, \tag{5.11}$$

and

$$a_{d0} = \frac{p^d}{q_d(1)}.$$

Proof: See Appendix 5.

Finally, by plugging (5.9) into (5.3) we find:

$$\overline{D}_{MT}^{max} = \sum_{m=1}^{M} \left[a_{m0} + \sum_{i=1}^{m} a_{mi} \left(\frac{1}{\alpha_{mi}} \right)^M \right]. \tag{5.12}$$

5.4.2 Equal Time-Sharing Transmission (eTS)

In the equal time-sharing (eTS) transmission scheme each channel block is equally divided among all the messages whose deadline has not expired yet, that is, for, $m = 1, \ldots, M$, we have $\alpha_{mt} = \frac{1}{M-t+1}$ for $t = 1, \ldots, m$, and $\alpha_{mt} = 0$, for $t = m+1, \ldots, M$.

In eTS, messages whose deadlines are later in time are allocated more resources; and hence, are more likely to be decoded. We have $I_i^{tot} < I_j^{tot}$ for $1 \leq i < j \leq M$. Hence, the probability of decoding exactly m messages is:

$$\eta(m) \triangleq Pr\{I_m^{tot} \geq r \geq I_{m-1}^{tot}\}, \tag{5.13}$$

for $m = 0, 1, \ldots, M$, where we define $I_0^{tot} = 0$ and $I_{M+1}^{tot} = \infty$. As the decoded messages in eTS are always the last ones, we can express the average maximum delay of eTS, $\overline{D}_{eTS}^{max}$ as a function of its average throughput $\overline{r}_{eTS}$ as follows:

$$\begin{aligned} \overline{D}_{eTS}^{max} &\triangleq \sum_{m=0}^{M} (M-m) \cdot \eta(m) \\ &= \sum_{m=0}^{M} M \cdot \eta(m) - \sum_{m=0}^{M} m \cdot \eta(m) \\ &= M \left(1 - \frac{\overline{r}_{eTS}}{r} \right). \end{aligned} \tag{5.14}$$

5.4.3 Pre-Buffering Transmission (PB)

In most practical streaming systems the receiver first accumulates video frames in the play-out buffer and then starts displaying them at a constant frame rate after a sufficient amount of frames has been received, in order to compensate for the delay jitter of arriving packets [83]. We call this type of streaming transmission with buffering in advance *pre-buffering (PB)* transmission, in which only the last B messages are transmitted. The receiver accumulates information relative to the last B messages during the first $M-B+1$ channel blocks.

The initial buffering phase introduces a start-up delay of $M-B$ channel blocks. On the other hand, if a sufficiently large buffering period is chosen, all the transmitted messages can be received correctly, achieving an average throughput of $r\frac{B}{M}$.

Transmitted messages are encoded with equal time allocation over the first $M-B+1$ blocks. Due to the delay constraint, message W_{M-B+1} is transmitted up to channel block $M-B+1$. Hence, in block $M-B+2$ the last $B-1$ messages are transmitted with equal time allocation. The process continues up until channel block M, in which only message W_M is transmitted. In the following we call $\overline{r}(B)$ and $\overline{D}^{max}(B)$ the average throughput and the average maximum delay achieved by the scheme using a buffering period of B channel blocks, respectively. The number B_{opt} of messages to be transmitted is chosen so that

$$B_{opt} = \underset{B\in\{1,\ldots,M\}}{\arg\min} \left\{\overline{D}^{max}(B)\right\}. \tag{5.15}$$

In the following we show that the B_{opt} of (5.15) also maximizes the average throughput. The average throughput when transmitting only the last B messages is given by:

$$\begin{aligned}\overline{r}_{PB}(B) &= \frac{r}{M}\sum_{m=1}^{B} Pr\left\{\text{decode at least } m \text{ messages}\right\} \\ &= \frac{r}{M}\sum_{m=1}^{B} Pr\left\{I_{M-m+1}^{tot} \geq r\right\},\end{aligned} \tag{5.16}$$

where the mutual information accumulated by the receiver for message W_m, for $m = M-B+1,\ldots,M$, is given by:

$$I_m^{tot} = \frac{1}{B}\sum_{t=1}^{M-B+1} C_t + \sum_{t=M-B+2}^{m}\frac{C_t}{M-t+1}. \tag{5.17}$$

From Eqn. (5.16) we have:

$$\begin{aligned}\overline{r}_{PB}(B) &= \frac{r}{M}\left[B - \sum_{m=1}^{B} Pr\left\{I_{M-m+1}^{tot} < r\right\}\right] \\ &= \frac{r}{M}\left[B - \sum_{m=1}^{B} Pr\left\{D^{max} \geq M-m+1\right\}\right].\end{aligned} \tag{5.18}$$

The average maximum delay when only the last B messages are transmitted is:

$$\overline{D}_{PB}^{max}(B) = M - B + \sum_{d=1}^{B} Pr\left\{D^{max} \geq M - B + d\right\}. \tag{5.19}$$

From (5.18) and (5.19) we find

$$\overline{r}(B) = r\left(1 - \frac{\overline{D}^{max}(B)}{M}\right),$$

and finally

$$\underset{B\in\{1,\cdots,M\}}{\arg\min}\left\{\overline{D}_{PB}^{max}(B)\right\} = \underset{B\in\{1,\cdots,M\}}{\arg\max}\left\{\overline{r}_{PB}(B)\right\}. \tag{5.20}$$

This proves that the average throughput and the maximum delay can be optimized simultaneously. Although it is not straightforward to come up with an analytical expression for the optimal B value of the PB scheme for the general case, in the following theorem we derive the optimal fraction of messages $\alpha_{opt} = \frac{B_{opt}}{M}$ such that almost all of the transmitted messages can be decoded with probability that approaches 1 asymptotically as M goes to infinity if a fraction $\alpha' < \alpha_{opt}$ of the messages is transmitted, while at most a fraction smaller than α_{opt} of the messages can be decoded if $\alpha' > \alpha_{opt}$.

Theorem 5.3 An average throughput of $r\alpha$ can be achieved in the limit of infinite M by transmitting $M\alpha + o(M)$ messages as long as

$$\alpha < \alpha_{opt} = \frac{1}{\frac{r}{\overline{C}} + 1},$$

while if $\alpha > \alpha_{opt}$, an average throughput smaller than $r\alpha_{opt}$ is achieved.

Proof Assume the last B messages i.e., $W_M - B + 1, \ldots, W_M$, are transmitted, with $B = M\alpha + o(M)$, $\alpha \leq 1$. Message W_{M-B+1}, for which the deadline expires first, is the one that gets the least amount of mutual information, that is:

$$I_{M-B+1} = \frac{1}{B}\sum_{t=1}^{M-B+1} C_t. \tag{5.21}$$

The probability of decoding a fraction $\alpha = \frac{B}{M}$ of the messages is then:

$$\begin{aligned} Pr\left\{I_{M-B+1} \geq r\right\} &= Pr\left\{\frac{1}{B}\sum_{t=1}^{M-B+1} C_t \geq r\right\} \\ &= Pr\left\{\sum_{t=1}^{M-B+1}\frac{C_t}{M-B} - \overline{C} \geq \frac{B}{M-B}r - \overline{C}\right\} \\ &= Pr\left\{S_{M-B} - \overline{C} \geq \frac{B}{M-B}r - \overline{C}\right\}, \end{aligned} \tag{5.22}$$

where $S_{M-B} \triangleq \sum_{t=1}^{M-B+1} \frac{C_t}{M-B}$ is the sample mean of the instantaneous channel capacity over the first $M-B$ channel blocks. By the low of large numbers follows that:

$$\lim_{M\to\infty} Pr\left\{\left|S_{M\left(1-\alpha-\frac{o(M)}{M}\right)} - \overline{C}\right| > \delta\right\} = 0,\ \forall\delta > 0. \tag{5.23}$$

Using equations (5.22) and (5.23) we find:

$$Pr\left\{I_{M-B+1} \geq r\right\} = \begin{cases} 1, \text{ if } \lim_{M\to\infty} \frac{B}{M-B} r < \overline{C} \\ 0, \text{ if } \lim_{M\to\infty} \frac{B}{M-B} r > \overline{C}. \end{cases} \tag{5.24}$$

We can write:

$$\begin{aligned} \lim_{M\to\infty} \frac{B}{M-B} r &= \lim_{M\to\infty} \frac{M\alpha + o(M)}{M - M\alpha + o(M)} r \\ &= \frac{\alpha}{1-\alpha} r. \end{aligned} \tag{5.25}$$

Finally, using Eqn. (5.25) in Eqn. (5.24) we find:

$$Pr\left\{I_{M-B+1} \geq r\right\} = \begin{cases} 1, \text{ if } \alpha < \frac{1}{\frac{r}{C}+1} \\ 0, \text{ if } \alpha > \frac{1}{\frac{r}{C}+1}. \end{cases} \tag{5.26}$$

Eqn. (5.26) implies that if a fraction of messages α' larger than α_{opt} is transmitted, then the average throughput is less than $r\alpha'$. We now prove that, if $\alpha' > \alpha_{opt}$, then the average throughput is lower than $r\alpha_{opt}$, i.e., there is no advantage in transmitting a fraction of messages larger than α_{opt}. Assume $M\alpha' + o(M)$ messages are transmitted, with $\alpha' = \alpha_{opt} + \chi_1 + \chi_2 \leq 1$, $\chi_1 > 0$ and $\chi_2 > 0$. We want to prove that not even a fraction $\alpha'' = \alpha_{opt} + \chi_1 < \alpha'$ of the messages can be decoded. The mutual information accumulated for message number $M\left[1 - \left(\alpha'' + \frac{o(M)}{M}\right)\right]$ is:

$$\begin{aligned} I_{M\left[1-\left(\alpha''+\frac{o(M)}{M}\right)\right]} &= \sum_{t=1}^{M\left[1-\left(\alpha'+\frac{o(M)}{M}\right)\right]+1} \frac{C_t}{M\left(\alpha' + \frac{o(M)}{M}\right)} + \sum_{t=M\left[1-\left(\alpha'+\frac{o(M)}{M}\right)\right]+2}^{M\left[1-\left(\alpha''+\frac{o(M)}{M}\right)\right]} \frac{C_t}{M-t} \\ &\leq \sum_{t=1}^{M\left[1-\left(\alpha'+\frac{o(M)}{M}\right)\right]+1} \frac{C_t}{M\left(\alpha'' + \frac{o(M)}{M}\right)} + \sum_{t=M\left[1-\left(\alpha'+\frac{o(M)}{M}\right)\right]+2}^{M\left(\alpha''+\frac{o(M)}{M}\right)} \frac{C_t}{M\left(\alpha'' + \frac{o(M)}{M}\right)} \\ &= \sum_{t=1}^{M\left[1-\left(\alpha''+\frac{o(M)}{M}\right)\right]+1} \frac{C_t}{M\left(\alpha'' + \frac{o(M)}{M}\right)}, \end{aligned} \tag{5.27}$$

where the inequality follows from the fact that $\alpha'' < \alpha'$. At this point it is sufficient to note that the last term of Eqn. (5.27) is the mutual information accumulated for the first message in

case a fraction α'' of the messages is transmitted. Thus, using the result of the first part of the theorem, we conclude that a fraction of the messages α'' can not be decoded even if a larger fraction is transmitted.

In Section 5.5, we provide a numerical optimization of the PB scheme, and compare it with the other proposed transmission strategies and the upper bound. As we will see from the numerical results, this buffering approach can improve the average throughput significantly as it provides rate adaptation at the packet level by eliminating some of the packets, and thus increasing the correct decoding probability of the remaining packets.

5.4.4 Windowed Time Sharing (wTS)

We have seen in the PB scheme that transmitting only a subset of the messages can improve the system throughput by allowing rate adaptation at the packet level. However, in the PB scheme only the last B packets are transmitted leading to a minimum delay of $M - B$ channel blocks. In the next scheme, called the windowed time-sharing (wTS) scheme, only a fraction $\lceil M/B \rceil$ of the messages is transmitted, where $\lceil x \rceil$ is the smallest integer larger than or equal to x; however, unlike in PB, the transmitted messages are distributed among the whole set of available messages, that is, one packet from each consecutive B packets is transmitted over B consecutive channel blocks. So, for instance, if $B = 3$, the first message to be transmitted is W_3, which is repeated in channel blocks 1, 2 and 3, followed by message W_6, which is transmitted in the next three channel blocks, and so on.

B is optimized according to two different criteria, namely average throughput maximization and delay minimization, which lead to the two variants *throughput-wTS (T-wTS)* and *delay-wTS (D-wTS)*, respectively. In wTS a message is decoded with the probability p_B given below:

$$p_B = Pr\left\{I_{kB} \geq r\right\} = Pr\left\{\sum_{t=kB-W+1}^{\min\{kB,M\}} C_k \geq r\right\}, \tag{5.28}$$

for $k \in \{1, \ldots, \left\lceil \frac{M}{B} \right\rceil\}$. A lower bound on $\overline{D}_{wTS}^{max}$ can be found by substituting $\left\lfloor \frac{M}{B} \right\rfloor$ for M in Eqn. (5.12), p_B for p in equations (5.10) and (5.11) and multiplying Eqn. (5.12) with B. An upper bound can be found in a similar way by using $\left\lceil \frac{M}{B} \right\rceil$ instead of $\left\lfloor \frac{M}{B} \right\rfloor$. Similarly, an upper and a lower bound on $\overline{r}_{wTS}$ are given by $\left\lceil \frac{M}{B} \right\rceil \cdot p_B$ and $\left\lfloor \frac{M}{B} \right\rfloor \cdot p_B$, respectively.

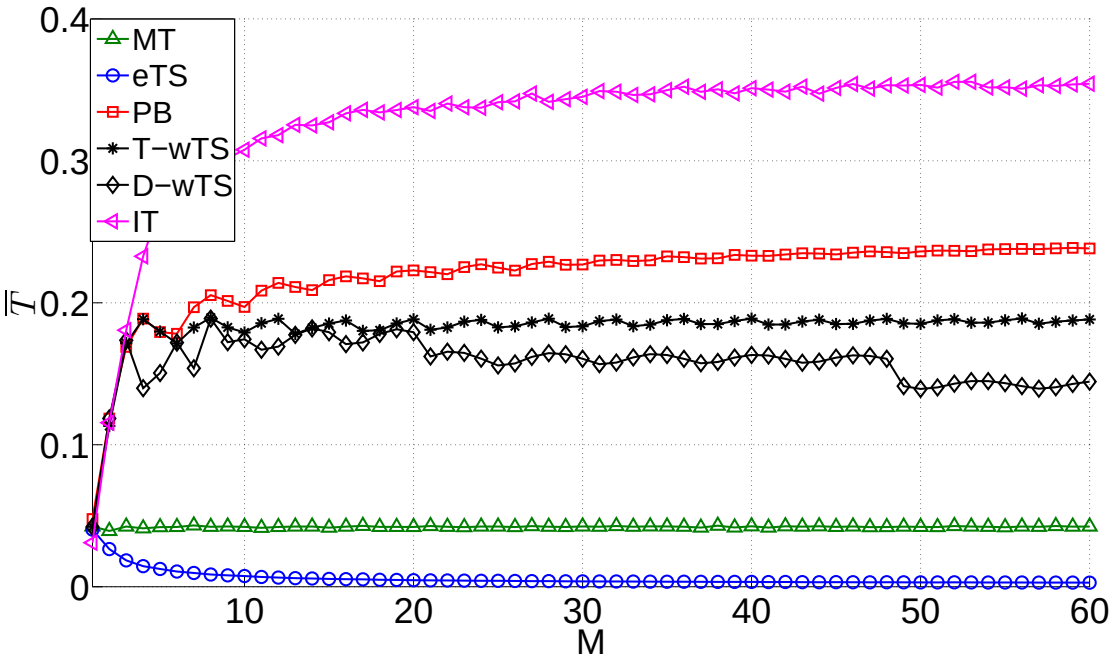


Figure 5.4: Throughput $\overline{r}$ plotted against the number of messages transmitted for $SNR = -5$ dB and $r = 1$ bpcu.

5.5 Numerical Results

In this section we compare the average throughput and the average maximum delay of the proposed schemes numerically. Fig. 5.4 and Fig. 5.5 show the average throughput and the average maximum delay for the proposed schemes, respectively, for rate $r = 1$ and $\mathrm{SNR} = -5$ dB. Both variants of the wTS scheme perform close to the informed transmitter lower bound in terms of maximum delay, while the PB scheme is the one with the highest average throughput, followed by T-wTS and D-wTS. The eTS scheme shows quite poor performance in both delay and number of decoded messages. From the plots it emerges that wTS in its two variants T-wTS and D-wTS, can help reduce the delay while achieving a relatively good average throughput in the low SNR regime. Fig. 5.6 and Fig. 5.7 show the average throughput and the average maximum delay, respectively, for the proposed schemes for rate $r = 1$ and $\mathrm{SNR} = 5$ dB. Also for this SNR the two variants of the wTS scheme perform close to the informed transmitter lower bound in terms of maximum delay. The highest average throughput is achieved by the T-wTS scheme together with the MT scheme, followed by the PB, D-wTS and eTS schemes. From Fig. 5.6 and Fig. 5.7 we see that, when the SNR is high, the MT scheme, together with the T-wTS scheme, achieves the best performances in terms of both delay and average

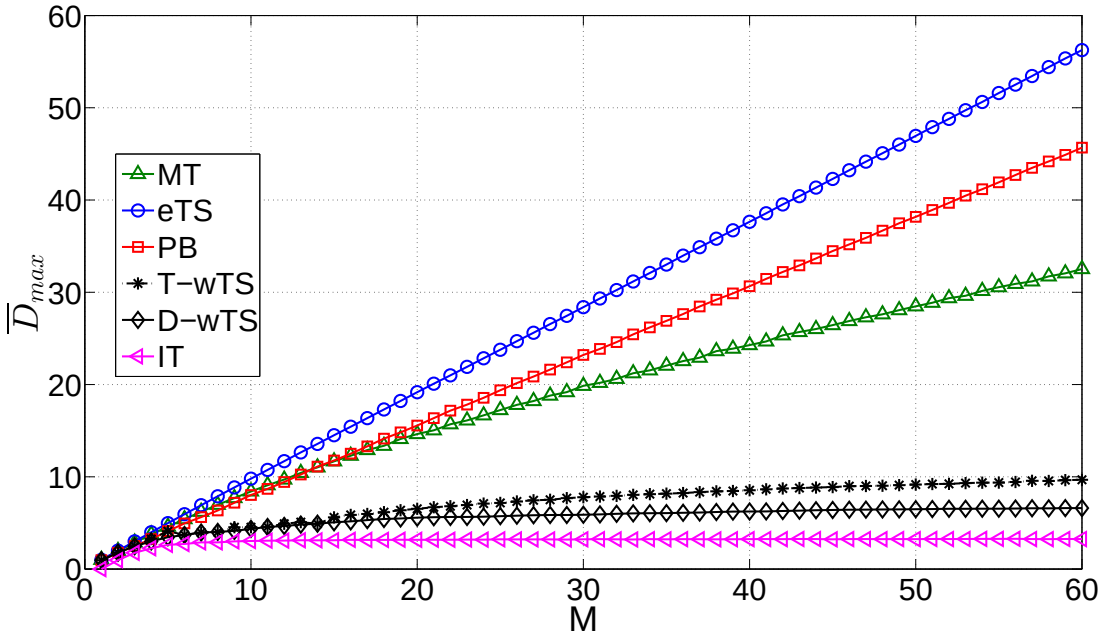


Figure 5.5: Average maximum delay $\overline{D}^{max}$ plotted against the number of transmitted messages messages for $SNR = -5$ dB and $r = 1$ bpcu.

throughput. This suggests that a simple memoryless approach is sufficient when the channel SNR is sufficiently high, while at low SNR more complex encoding techniques can help to significantly improve the performance.

The D-wTS scheme shows a sudden decrease in the average throughput, which, with reference to Fig. 5.4, also corresponds to a decrease in the slope of the curve at points corresponding to $M = 7$, $M = 20$ and $M = 48$. This is due to the optimization of the window size B. We recall that in D-wTS the window size represents the number of channel blocks dedicated to a message and is optimized so as to achieve the minimum average maximum delay. While a large B leads to a high decoding probability, it implies a small number of transmitted messages, which bounds from below the minimum delay to B. As a matter of fact, only $\lceil \frac{M}{B} \rceil$ messages are transmitted in the wTS scheme, which implies that the maximum delay, in a given realization, is a multiple of B. If, for instance, $B = 2$ and $m = 3$ consecutive messages are lost, the corresponding delay is $m \cdot B = 6$. Formally, given a window size B^* there is a certain probability p_B^l of not decoding a message. For any fixed $m \in \{0, \dots, M\}$, using Eqn. (5.9) it can be easily shown that the probability of losing at least m consecutive messages increases with M. Thus a value B^* which is optimal for a certain M, may not be the optimal for a larger number of messages, as the probability that more than one consecutive messages get lost

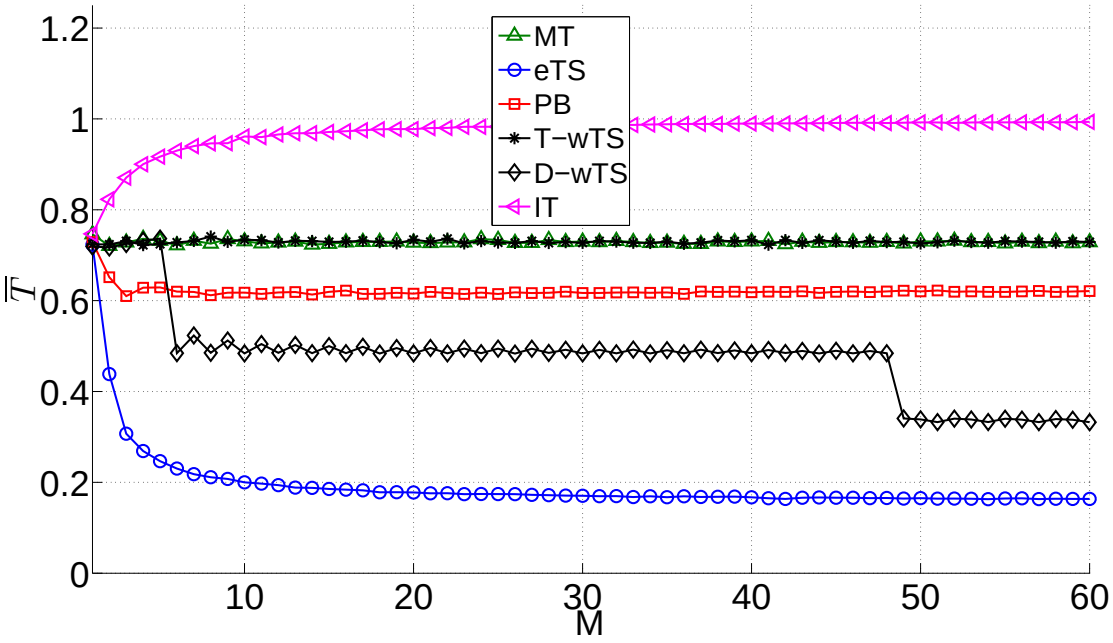


Figure 5.6: Throughput $\overline{r}$ plotted against the number of messages transmitted for $SNR = 5$ dB and $r = 1$ bpcu.

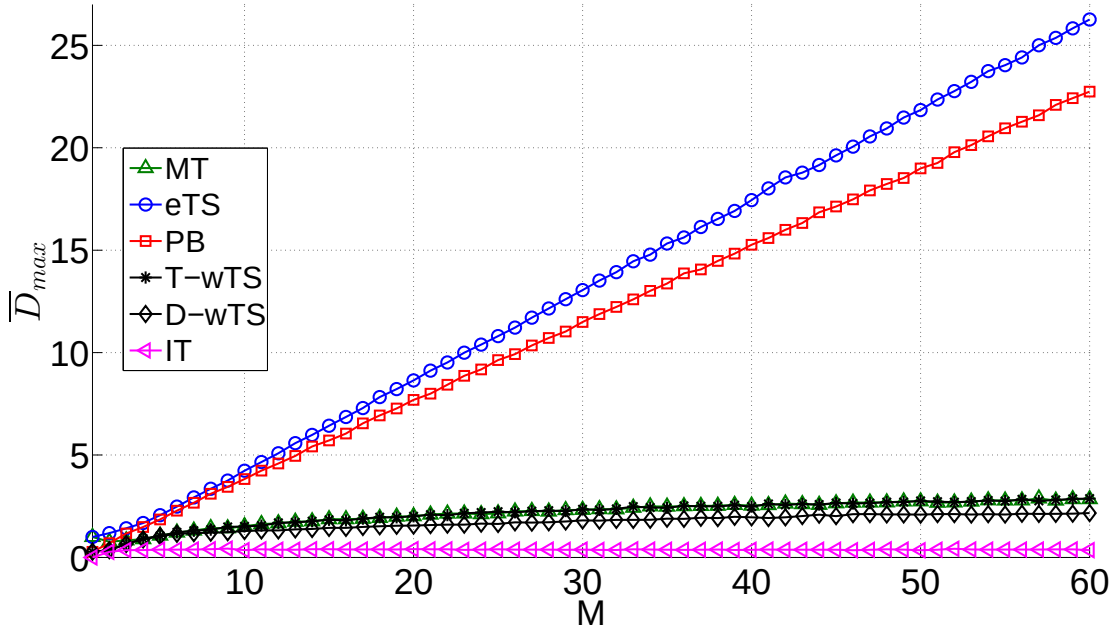


Figure 5.7: Average maximum delay $\overline{D}^{max}$ plotted against the number of transmitted messages messages for $SNR = 5$ dB and $r = 1$ bpcu.

increases with M. The optimal choice may be to increase B, so that the probability of losing consecutive messages is decreased. This is confirmed by Fig. 5.8, where the optimal window

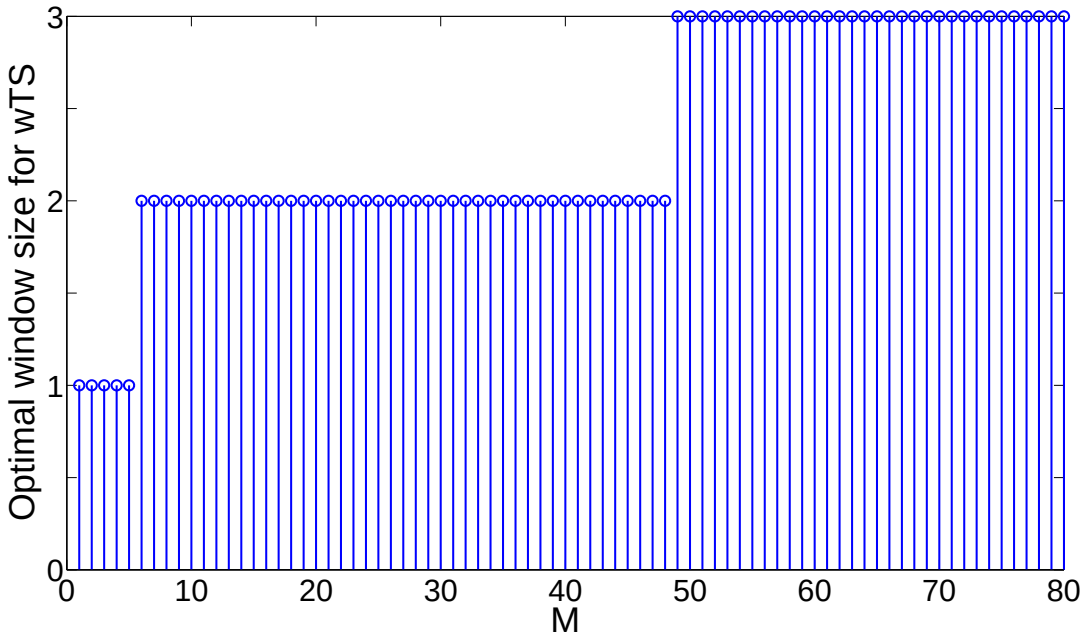


Figure 5.8: Optimal window size for the T-wTS scheme plot against the total number of messages M for $SNR = 5\ dB$.

size, obtained numerically, is plotted against the total number of messages. An increase in B implies a decrease in the slope of the average number of decoded messages, as a smaller fraction of messages is transmitted, as shown in the plots. The T-wTS scheme, in which B is optimized so as to achieve the maximum average throughput, shows a good tradeoff between average throughput, which, unlike D-wTS, is almost independent from number of messages, and average maximum delay, performing close to the D-wTS scheme.

5.6 Conclusions

We have studied the problem of video streaming over slow fading channels with per-packet delay constraints. We have proposed four different transmission schemes based on time-sharing. We have carried out theoretical as well as numerical analysis for the average throughput and maximum delay performances. We have also derived bounds on both the average throughput and maximum delay by introducing an informed transmitter scheme, in which the transmitter

is assumed to know the channel states in advance. We have seen that the wTS scheme can provide a good trade-off between the average throughput and the maximum delay by deciding on the proportion of transmitted video packets. In practice this corresponds to reducing the coding rate of the video at the packet level. We also proved that in the PB scheme almost all transmitted messages can be decoded with probability that goes to 1 as M goes to infinity if a fraction of the messages smaller than a threshold value depending on the transmission rate and the average channel capacity are transmitted.

Appendix 5

Proof of Theorem 5.2

The probability of having a run of at least d, $d \in \{0, \ldots, M\}$, consecutive zeros in the sequence is equivalent to finding the probability of state d after M steps in the Markov chain depicted in Fig. 5.9. The state d is an absorbing state, i.e., once the process reaches that state, it remains

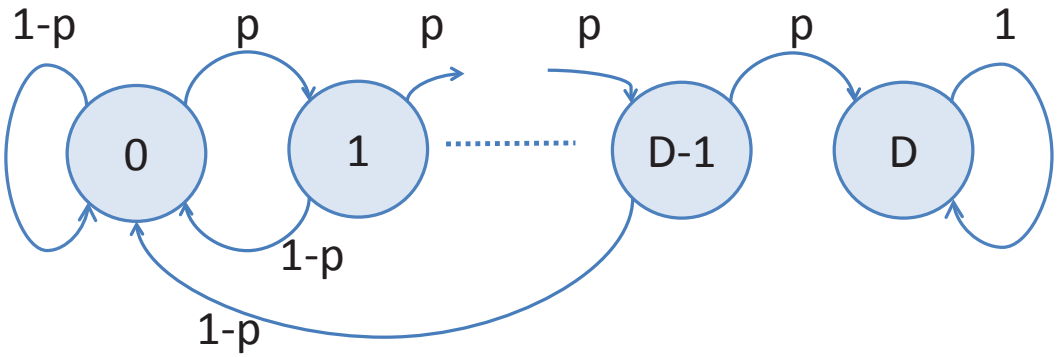


Figure 5.9: Markov chain for the calculation of the average maximum delay in memoryless transmission.

there with probability 1. Let $\mathbf{p}_t$ be a d-length probability mass function, where $\mathbf{p}_t(i)$, $i = 0, \ldots, d$, denotes the probability of being in state i at step t. The vector $\mathbf{p}_t$ of state occupancy at step t for the Markov chain in Fig. 5.9 can be obtained as:

$$\mathbf{p}_t = \mathbf{p}_{t-1}\mathbf{H} = \mathbf{p}_0\mathbf{H}^t \tag{5.29}$$

where $\mathbf{p}_0 = [1\ 0\ \cdots\ 0]$ and $\mathbf{H}$ is the $(D+1) \times (D+1)$ transition matrix of the chain which can be obtained from Fig. 5.9:

$$\mathbf{H} = \begin{pmatrix} 1-p & p & 0 & 0 & \cdots & 0 & 0 \\ 1-p & 0 & p & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \\ 1-p & 0 & 0 & 0 & \cdots & 0 & p \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}. \tag{5.30}$$

The probability of being in state d after M steps $\mathbf{p}_M(d)$ can be found from Eqn. (5.29). Since $\mathbf{p}_0 = [1\ 0\ \cdots\ 0]$ we have:

$$\mathbf{p}_M(d) = \mathbf{H}^M(1, d+1). \tag{5.31}$$

In order to evaluate $\mathbf{H}^M(1, D+1)$, we apply the *Z-transform* to Eqn. (5.29), taking into account that the recursive formula is defined only for $t \geq 1$. The Z-transform $\mathcal{P}(z)$ of a discrete vectorial function $\mathbf{p}_t$ is defined as:

$$\mathcal{P}_z = \sum_{t=0}^{+\infty} \mathbf{p}_t z^t. \tag{5.32}$$

To account for the fact that $m \geq 1$ in Eqn. (5.29) we can write:

$$\sum_{t=1}^{+\infty} \mathbf{p}_t z^t = \sum_{t=0}^{+\infty} \mathbf{p}_t z^t - \mathbf{p}_0 = \mathcal{P}_z - \mathbf{p}(0), \tag{5.33}$$

and

$$\begin{aligned} \sum_{t=1}^{+\infty} \mathbf{p}_{t-1}\mathbf{H} z^t &= z \sum_{t=1}^{+\infty} \mathbf{p}_{t-1}\mathbf{H} z^{t-1} \\ &= z \sum_{t=0}^{+\infty} \mathbf{p}_t \mathbf{H} z^t \\ &= z\mathcal{P}_z\mathbf{H}. \end{aligned} \tag{5.34}$$

Plugging Eqn. (5.33) and Eqn. (5.34) into Eqn. (5.29) we find:

$$\mathcal{P}_z = \mathbf{p}(0)\left(\mathbf{I} - z\mathbf{H}\right)^{-1}, \tag{5.35}$$

where $\mathbf{I}$ is the $(D+1) \times (D+1)$ identity matrix. By comparing Eqn. (5.35) with Eqn. (5.29), we see that $(\mathbf{I} - z\mathbf{H})^{-1}$ is the Z-transform of the matrix of functions in the discrete variable t $\mathbf{H}^t$. The Z-transform $\mathcal{C}_z$ of a matrix of functions in the discrete variable t $\mathbf{C}_t$ is defined as:

$$\mathcal{C}_z = \sum_{t=0}^{+\infty} \mathbf{C}_t z^t. \tag{5.36}$$

Note that in the sum of Eqn. (5.36) the term z^t is a scalar function of z and t which is multiplied to each of the elements of matrix $\mathbf{C}_t$. We now look for $(\mathbf{I} - z\mathbf{H})^{-1}$, which is the inverse of:

$$(\mathbf{I} - z\mathbf{H}) = \begin{pmatrix} 1-z(1-p) & -zp & 0 & 0 & \cdots & 0 & 0 \\ -z(1-p) & 1 & -zp & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \\ -z(1-p) & 0 & 0 & 0 & \cdots & 1 & -zp \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1-z \end{pmatrix}. \tag{5.37}$$

Once $(\mathbf{I} - z\mathbf{H})^{-1}$ is known, it is sufficient to inversely transform it and get $\mathbf{H}^t$. We find the inverse of matrix (5.37) for a generic d using Gauss-Jordan elimination. As we only need the element $\mathbf{H}^M(1, d+1)$, we only report here the first row of $(\mathbf{I} - z\mathbf{H})^{-1}$:

$$(\mathbf{I} - z\mathbf{H})^{-1}_{[1,:]} = \frac{\begin{pmatrix} 1-z & (1-z)(zp) & (1-z)(zp)^2 & \cdots & (1-z)(zp)^{d-1} & (zp)^d \end{pmatrix}}{(1-z)q_d(z)}, \tag{5.38}$$

where

$$q_d(z) = 1 - p\sum_{j=1}^{d} z^j(1-p)^{j-1}. \tag{5.39}$$

The probability of being in state d at step M is the inverse Z-transform of element $[1, d+1]$ of matrix $(\mathbf{I} - z\mathbf{H})^{-1}$, i.e.:

$$\mathbf{p}_M(d+1) = Z^{-1}\left\{\frac{(zp)^d}{(1-z)q_d(z)}\right\}, \tag{5.40}$$

where $Z^{-1}\{\mathcal{P}_z\}$ is the inverse Z-transform of $\mathcal{P}_z$ defined as:

$$Z^{-1}\{\mathcal{P}_z\} = \frac{1}{2\pi j}\oint_{\gamma} \mathcal{P}_z z^{t-1} dz = \mathbf{p}_t, \tag{5.41}$$

γ being a counterclockwise-oriented circle around the origin of the complex plane. An easier way to solve Eqn. (5.40) is to decompose the Z-transform using partial fraction decomposition, i.e., rewriting it as:

$$\mathcal{P}(1, D+1) = a_{d,0} + \sum_{i=2}^{k} a_{d,i}\left(\frac{1}{1-\frac{z}{\alpha_{d,i}}}\right)^d, \tag{5.42}$$

where $\alpha_{d,i}$, $i \in \{1, \ldots, k\}$, are the $k \leq d$ distinct zeros of the polynomial in z:

$$q_d(z) = 1 - p\sum_{j=1}^{k} z^j(1-p)^{j-1}, \tag{5.43}$$

with multiplicity r_i, while $a_{d,i}$ $i \in \{2, \ldots, m\}$ are constants assuming values:

$$a_{d,i} = \lim_{z \to +\alpha_{d,i}} \frac{(z - \alpha_{d,i})^{r_i} [z(1-p)]^d}{(1-z) q_d(z)}, \tag{5.44}$$

and

$$a_{d,0} = \frac{p^d}{q_d(1)}. \tag{5.45}$$

Once in the form of Eqn. (5.42), $\mathbf{p}_{1,D+1}(z)$ can be inversely transformed using the linearity of the inverse Z-transform and the fact that:

$$Z^{-1}\left\{ \left(\frac{1}{1 - \frac{z}{\alpha_{d,i}}} \right)^{r_i} \right\} = \binom{d - r_i - 1}{r_i - 1} \left(\frac{1}{\alpha_{d,i}} \right)^d. \tag{5.46}$$

Eqn. (5.46) follows from the fact that:

$$\begin{aligned} Z\left[\left(\frac{1}{\alpha} \right)^d \right] &\triangleq \sum_{d=0}^{\infty} \left(\frac{1}{\alpha} \right)^d z^d \\ &= \sum_{d=0}^{\infty} \left(\frac{z}{\alpha} \right)^d \\ &= \frac{1}{1 - z/\alpha}, \end{aligned} \tag{5.47}$$

for $|z| < \alpha$. Finally, using Eqn. (5.47) and Eqn. (5.42), we find:

$$Pr\{D \geq d\} = a_{d,0} + \sum_{i=2}^{k} a_{d,i} \binom{M - r_i - 1}{r_i - 1} \left(\frac{1}{\alpha_{d,i}} \right)^M. \tag{5.48}$$

Chapter 6

Network-Coded Diversity Protocol for Collision Recovery in Slotted ALOHA Networks

"A parte la follia di ucciderci l'un l'altro per motivi irrilevanti, eravamo felici."

Sergio Atzeni

Contents

6.1 Introduction

The throughput of Slotted ALOHA (SA) systems is limited by the collisions that take place when more than one node accesses the channel in the same time slot. This limitation is particularly problematic in satellite networks with random access, where the long round-trip time (RTT) greatly limits feedback from the receiver, for example to perform load control or to request a retransmission. Techniques like Diversity Slotted ALOHA (DSA) [84], in which each packet is transmitted more than once, have been proposed in order to increase the probability of successful detection. The spectral efficiency of SA systems can be increased by exploiting the collided signals. In [85] a novel scheme called network-assisted diversity multiple access (NDMA), inspired by signal separation principles borrowed from signal processing, has been presented. In NDMA the collisions are recovered through successive retransmissions, assuming feedback from the receiver. The receiving terminal interprets the signals observed over consecutive transmission slots as a matrix, which is processed in the analog domain so that the single bursts can be recovered if the matrix is full rank. In Contention Resolution Diversity Slotted ALOHA (CRDSA) [86] the transmissions are organized in frames, each constituted by a fixed number of transmission slots and no feedback is assumed from the receiver. The collided signals are exploited using an iterative interference cancelation (IIC) process. In CRDSA each packet is transmitted more than once and uncollided packets are subtracted from the slots in which their replicas are present. In [87] a packet-level forward error correction (FEC) code has been applied to CRDSA, while in [88] a convergence analysis and optimization of CRDSA has been proposed.

Another technique that allows to extract information from colliding signals is physical layer network coding (PHY NC). PHY NC was originally proposed to increase spectral efficiency in two-way relay communication [16] by having the relay decoding the collision of two signals under the hypothesis of symbol, frequency and phase synchronism. Several studies have been reported in the literature about synchronization issues, gain analysis and ad-hoc

modulation techniques for PHY NC in the case of two colliding signals [17], [18], [19]. In [89] a cooperative relaying protocol that leverages on PHY NC and IIC has been proposed, while in [20] PHY NC has been applied in the satellite context for pairwise node communication. In [21] and [22] it has been proposed to apply PHY NC to determine the identity of transmitting nodes in case of acknowledgement (ACK) collision in multicast networks by using energy detection and ad-hoc coding schemes, under the hypothesis of phase synchronous signal superposition at the receiver. In [23] an overview of the state of the art on PHY NC has been presented from an information theoretical point of view. In [90] PHY NC has been applied for collision resolution in ALOHA systems with feedback from the receiver, under the assumption of frequency synchronous transmitters.

In this chapter we present a new scheme named Network-Coded Diversity Protocol (NCDP), that leverages on PHY NC over an extended Galois field (EGF) for recovering collisions in symbol-synchronous SA systems. Once the PHY NC is applied to decode the collided bursts, the receiver uses common matrix manipulation techniques over finite fields to recover the original messages, which results in a high-throughput scheme. The proposed scheme and analysis differ from previous works on collision resolutions at both system level and physical level:

System level:

1. Unlike in [85] and [90], we assume that transmissions are organized in frames. We consider two different setups. In the first setup the nodes do not receive any feedback from the receiver. The absence of feedback leads to a *best-effort* scheme, in which there is no guarantee for a message to be received. On the other end, the absence of feedback from the receiver notably simplifies the system architecture and decreases the total amount of energy spent per received packet. In the second setup that we consider, instead, feedback is allowed from the receiver. In particular, we consider an ARQ scheme, in which a node receives an ACK or a negative acknowledgement (NACK) from the receiver in case a message is or is not correctly received, respectively. A message for which a NACK has been received is retransmitted in a different frame. The retransmission process goes on until the message is acknowledged.

2. We evaluate jointly the spectral efficiency (average number of messages successfully received per slot) and the energy consumption (average amount of energy needed for

a message to be correctly received) of the proposed scheme and compare it with other collision resolution schemes previously proposed in the literature.

Physical level:

1. We use an EGF, i.e., $GF(2^n)$ with $n > 2$, instead of $GF(2)$, which is generally used in PHY NC. This allows to efficiently exploit the diversity of the system, leading to an increased spectral efficiency and, depending on the system load, to an increased energy efficiency.

2. We take into account frequency and phase offsets at the transmitters when applying PHY NC for an arbitrary number of colliding signals. Up to our knowledge, the issue of frequency offsets in PHY NC has been previously addressed only for the case of two colliding signals. See, e.g., [91], [92] and references therein.

3. We show the feasibility of channel estimation for PHY NC in the presence of more than two colliding signals, unlike previous works where only two colliding signals were considered (see, e.g., [93]).

4. We study the effect of non perfect symbol synchronism on the decoder FER for an arbitrary number of colliding signals and propose four different methods to compensate for such effect.

The rest of the chapter is organized as follows. In Section 6.2 we present the system model. Section 6.3 describes how the channel decoding takes place in case of a generic number of colliding signals with independent frequency and phase offsets. In Section 6.4 the proposed scheme is described, while a theoretical analysis of its performance in terms of both spectral and energy efficiency is performed in Section 6.5. Section 6.6 deals with issues such as channel estimation and error detection, which are fundamental for a practical implementation of the proposed scheme. Section 6.7 is dedicated to the effect of and possible countermeasures to imperfect symbol synchronization on the decoder performance in case of multiple colliding signals. In Section 6.8 we present the numerical results, while Section 6.9 contains the conclusions.

6.2 System Model

Let us consider the return link (i.e., the link from a user terminal to the satellite/base station) of a multiple access system with M transmitting terminals, $T_1, ..., T_M$, and one receiver R. Packet arrivals at each transmitter are modeled as a Poisson process with rate $\frac{G}{M}$, which is independent from one transmitter to the other. Each packet $\mathbf{u}_i = [u_i(1), ..., u_i(K)]$ consists of K binary symbols of information $u_i(\xi) \in \{0, 1\}$, for $\xi = 1, \ldots, K$. We assume that, upon receiving a message, each terminal T_i uses the same linear channel code of fixed rate $r = \frac{K}{N}$ to protect its message $\mathbf{u}_i$, obtaining the codeword $\mathbf{x}_i = [x_i(1), ..., x_i(N)]$, where $x_i(l) \in \{0, 1\}$ for $l = 1, \ldots, N$. For ease of exposition a BPSK modulation is considered. Each codeword $\mathbf{x}_i$ is BPSK modulated (using the mapping $0 \rightarrow -1$, $1 \rightarrow +1$), thus obtaining the transmitted signal

$$s_i(t) = \sum_{l=1}^{N} b_i(l) g(t - lT_s), \tag{6.1}$$

where T_s is the symbol period, $b_i(l)$ is the BPSK mapping of $x_i(l)$ and $g(t)$ is the square root raised cosine (SRRC) pulse. The signal $s_i(t)$ is called *burst*.

In the following we will refer to a time division multiple access (TDMA) scheme. However, the techniques proposed in the following can be also applied to other access schemes, such as multi-frequency-TDMA (MF-TDMA), in which a frame may include several carriers [26], or code division multiple access (CDMA), where NCDP can be used to recover collisions in each of the code sub-channels. It should be noted that the proposed technique still relies on single carrier transmission of each user terminal. From the user terminal perspective no significant change is required. Transmissions are organized in frames. Each frame is divided into S time slots. The number S of time slots in a frame is fixed, i.e., it does not change from one frame to the other. The duration of each slot is equal to about N burst symbols. When more than one terminal transmits its burst in the same slot a collision occurs at the receiver. A collision involving k transmitters is said to have size k. We assume symbol-synchronous transmissions, i.e., in case of a collision, the signals from the transmitters add up with symbol synchronism at the receiver R. The received signal before matched filtering and sampling at R in case of a collision of size k (assuming, without loss of generality, the first k terminals

collide), is:

$$y(t) = h_1(t)s_1(t) + ... + h_k(t)s_k(t) + w(t), \tag{6.2}$$

where $s_i(t)$ is the burst transmitted by user i, $w(t)$ is a complex AWGN process while $h_i(t)$ takes into account the channel from terminal i to the receiver. $h_i(t)$ can be expressed as:

$$h_i(t) = A_i e^{j(2\pi\Delta\nu_i t + \varphi_i)}, \tag{6.3}$$

where $A_i = |h_i(t)|$ is a log-normally distributed random variable modeling the channel amplitude of transmitter i, while $\Delta\nu_i$ and φ_i are the frequency and phase offsets with respect to the local oscillator in R, respectively. We assume that the amplitude A_i and the frequency offset $\Delta\nu_i$ remain constant within one frame while φ_i is a random variable uniformly distributed in $[-\pi, +\pi]$ that changes independently from one slot to the other due to the phase noise at the transmitting terminals [86]. Assuming that the frequency offset is small compared to the symbol rate $1/T_s$ (i.e., $\Delta\nu T_s \ll 1$), the sample taken at time t_l after matched filtering signal $y(t)$ is:

$$r(t_l) = h_1(t_l)q_1(t_l) + ... + h_k(t_l)q_k(t_l) + n(t_l), \tag{6.4}$$

where $q(t) = s(t) \otimes g(-t)$, $\otimes$ being the convolution operator, while $n(t_l)$'s are i.i.d. zero mean complex Gaussian random variables with variance N_0 in each component. Note that, even in case a BPSK modulation is used, as we are assuming, both the in-phase (I) and quadrature (Q) components of the received signal are considered by the receiver. This is because the phases of the users have random relative offsets and thus both components carry information relative to the useful signal. The frequency and phase relative offsets must be taken into account by the decoder, as they cannot be eliminated by the demodulator. We consider this more in detail in Section 6.3.

We assume that the receiver has knowledge of the nodes that are transmitting, as well as the full channel state information at each time slot. As we are considering a random access scheme, the knowledge about node identities cannot be available *a priori* at the receiver. Instead, it must be determined by R, directly from the received signal, even in case a collision occurs. This can be achieved by having the transmitting nodes add an orthogonal preamble in each transmitted burst, designed such that the probability that two nodes use the same preamble is negligible [86]. We discuss the issue of node identification and channel estimation more in detail in Section 6.6.

6.3 Multi-User Physical Layer Network Coding

In this section we describe the way the received signal is processed by the receiver R in case of a collision.

When a collision of size k occurs, i.e., k bursts collide in the same slot, the receiver tries to decode the bit-wise XOR of the k transmitted messages. This can be done by feeding the channel decoder (e.g., a turbo decoder or an LDPC decoder) with the log-likelihood ratio log-likelihood ratio (LLR) for the received signal. The calculation of the LLR's for a collision of generic size k in case of BPSK modulation was presented in [90]. In the following we include the effect of frequency offset in the calculation of the LLR's, which was not taken into account in [90] (see [17] and [19] for an extension to higher order modulations).

Starting from the samples $r(t_l)$ the receiver R wants to decode the codeword $\mathbf{x}_s \triangleq \mathbf{x}_1 \oplus \mathbf{x}_2 \oplus \ldots \oplus \mathbf{x}_k$, where $\oplus$ denotes the bit-wise XOR. In order to do this we must feed the decoder of R with the vector $\mathbf{L}^{\oplus} = \{L^{\oplus}(1), ..., L^{\oplus}(N)\}$ of LLR's for $\mathbf{x}_s$. We have:

$$\begin{aligned} L^{\oplus}(l) &\triangleq \ln\left\{\frac{Pr\left[x_s(l)=1|r(t_l)\right]}{Pr\left[x_s(l)=0|r(t_l)\right]}\right\} \\ &= \ln\left\{\frac{Pr\left[r(t_l)|x_s(l)=1\right]}{Pr\left[r(t_l)|x_s(l)=0\right]}\right\}. \end{aligned} \tag{6.5}$$

The last equality follows from the symmetry of the XOR operator provided that $x_i(l)$'s are i.i.d. with $Pr[x_i(l)=1] = Pr[x_i(l)=0] = \frac{1}{2}$. Eqn. (6.5) reduces to the calculation of the ratio of the likelihood functions of $r(t_l)$ for the cases $x_s(l)=1$ and $x_s(l)=0$. We indicate these functions as $p_1(r(t_l))$ and $p_0(r(t_l))$ respectively. Functions $p_0(r(t_l))$ and $p_1(r(t_l))$ are Gaussian mixtures:

$$p_1(r(t_l)) = \frac{2^{-k}}{\sqrt{2\pi N_0}} \sum_{i=1}^{\lfloor \frac{k+1}{2} \rfloor} \sum_{m=1}^{\binom{k}{2i-1}} e^{-\frac{\left|r(t_l)-\mathbf{d}^o(2i-1,m)^T\mathbf{h}(t_l)\right|^2}{2N_0}}, \tag{6.6}$$

$\mathbf{h}(t_l)$ being a column vector containing the channel coefficients of the k transmitters at time t_l (that change at each sample due to frequency offsets), while $\mathbf{d}^o(2i-1,m)$ is a column vector containing one (the m-th) of the $\binom{k}{2i-1}$ possible permutations over k symbols (without repetitions) of an odd number $(2i-1)$ of symbols with value "$+1$". As for the case with $x_s = 0$ we have:

$$p_0(r(t_l)) = \frac{2^{-k}}{\sqrt{2\pi N_0}} \sum_{i=1}^{\lfloor \frac{k+1}{2} \rfloor} \sum_{m=1}^{\binom{k}{2i}} e^{-\frac{\left|r(t_l)-\mathbf{d}^e(2i,m)^T\mathbf{h}(t_l)\right|^2}{2N_0}}, \tag{6.7}$$

where $\mathbf{d}^e(2i,m)$ is a column vector containing one (the m-th) of the $\binom{k}{2i}$ possible permutations over k symbols (without repetitions) of an even number ($2i$) of symbols with value "$+1$". Finally using equations(6.6) and (6.7) in Eqn. (6.5) we find the following expression for the LLR:

$$L^{\oplus}(l) = \ln\left\{\frac{\sum_{i=1}^{\lfloor\frac{k+1}{2}\rfloor}\sum_{m=1}^{\binom{k}{2i-1}} e^{-\frac{\left|r(t_l)-\mathbf{d}^o(2i-1,m)^T\mathbf{h}(t_l)\right|^2}{2N_0}}}{\sum_{i=1}^{\lfloor\frac{k+1}{2}\rfloor}\sum_{m=1}^{\binom{k}{2i}} e^{-\frac{\left|r(t_l)-\mathbf{d}^e(2i,m)^T\mathbf{h}(t_l)\right|^2}{2N_0}}}\right\}. \tag{6.8}$$

If the decoding process is successful, R obtains the message $\mathbf{u}_s \triangleq \mathbf{u}_1 \oplus \ldots \oplus \mathbf{u}_k$. In Section 6.6 the FER curves for different collision sizes obtained using these LLR values are shown.

6.4 Network Coded Diversity Protocol

In this section we present our network-coded diversity protocol (NCDP) which aims at increasing the throughput and reducing packet losses in Slotted ALOHA multiple access systems. In the first part of the section we describe the NCDP at the transmitter side, while in the second part the receiver side is considered.

6.4.1 NCDP: Transmitter Side

We call *active terminals* the nodes that have packets to transmit in a given frame. Each message is transmitted more than once within a frame, i.e., several replicas of the same message are transmitted. We will give details about the number of replicas transmitted within a frame in the next section. Assume that node i has a message $\mathbf{u}_i$ to deliver to R during frame f, i.e., node T_i is an active terminal. Before each transmission, node i pre-encodes $\mathbf{u}_i$ as depicted in Fig. 6.1. The pre-encoding process works as follows. $\mathbf{u}_i$ is divided into $L = \frac{K}{n}$ blocks of n bits each. At each slot a different coefficient α_{ij},$j \in \{1,\ldots,S\}$, is drawn randomly according to a uniform distribution in $GF(2^n)$. If $\alpha_{ij} = 0$, terminal T_i does not transmit in slot j. Each of the L blocks $\mathbf{u}_i^r$, $r \in \{1,\ldots,L\}$, is interpreted as an element in $GF(2^n)$ and multiplied by α_{ij}. We call $\mathbf{u}'_{ij}$ the message $\mathbf{u}_i$ after the multiplication by α_{ij}. $\mathbf{u}'_{ij}$ is then channel encoded, generating the codeword $\mathbf{x}_{ij} = C(\mathbf{u}'_{ij})$. After channel coding, a header p_i is added to $\mathbf{x}_{ij}$. Such header is chosen within a set of orthogonal codewords (e.g., Walsh-Hadamard). The same header p_i is used for all transmissions of node T_i within frame f. Once the header is attached, $\mathbf{x}_{ij}$ is BPSK

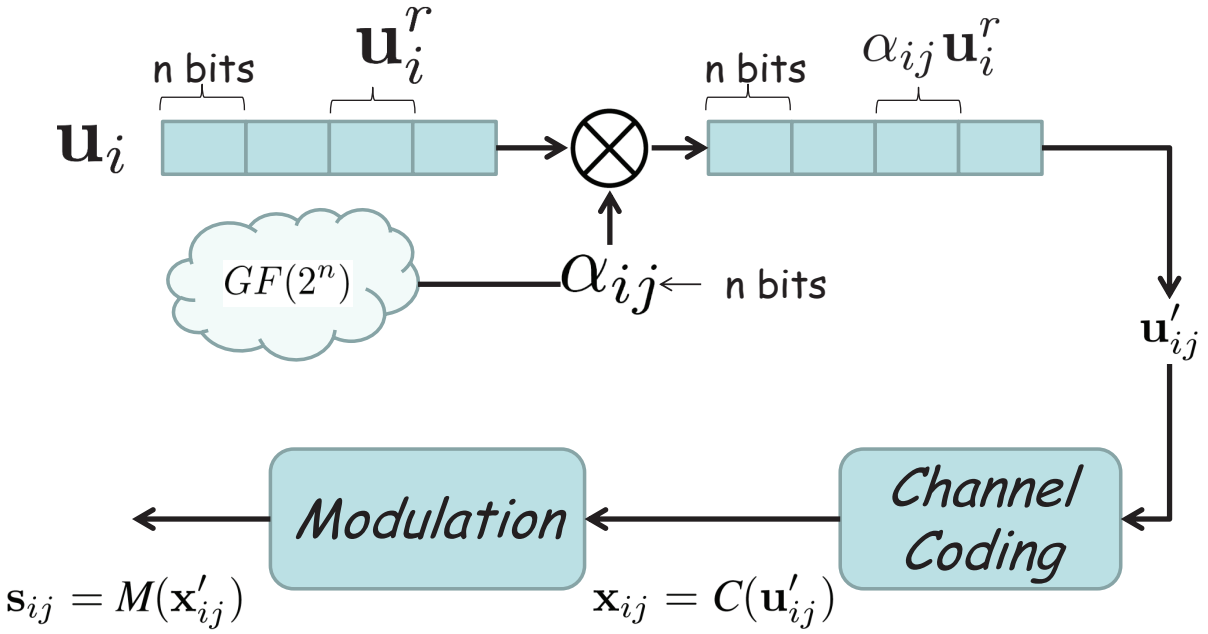


Figure 6.1: NCDP pre-encoding, channel coding and modulation scheme at the transmitter side. The message to be transmitted is divided into sub-blocks of n bits each. Each sub-block is multiplied by a coefficient $\alpha_{ij} \in GF(2^n)$, $j \in \{1, \ldots, S\}$. Coefficients α_{ij} are chosen at random in each time slot. After the multiplication, the message is channel-encoded, a header is attached and the modulation takes place.

modulated and transmitted.

The choice of the coefficients and of the header is done as follows. Node T_i draws a random number μ[1]. μ is used to feed a pseudo-random number generator in $GF(2^n)$, which is the same for all terminals and is known at R. The first S outputs of the generator are used as coefficients. The header is uniquely determined by μ, i.e., there is a one-to-one correspondence between the set of values that can be assumed by μ and the set of available orthogonal headers. The orthogonality of the preambles allows the receiver to know which of the active terminals in frame f is transmitting in each time slot. Moreover, as the header univocally determines μ and thus the set of coefficients used by each node, R is able to know which coefficient is used by each transmitter in each slot. As we we will see in Section 6.4.2, this is of fundamental importance for the decoding process. As said before, the set of headers is a set of orthogonal words, such as those usually adopted in CDMA. The fundamental difference with respect to a CDMA system is that in such a system the orthogonality of the codes is used to orthogonalize the channels and expand the spectrum, while in NCDP the orthogonality of the preamble is used only for determining the identity of the transmitting node, which is obtained without any

[1] μ can be, for instance, a function of the node's id.

spectral expansion, as the symbol rate $1/T_s$ is equal to the chip rate (i.e., the rate at which the modulated symbols are transmitted over the channel) [86].

6.4.2 NCDP: Receiver Side

The decoding scheme at the receiver side is illustrated with an example in Fig. 6.2 and Fig. 6.3. In the example, a frame with $S = 4$ slots and $N^{tx} = 3$ active terminals is considered. In

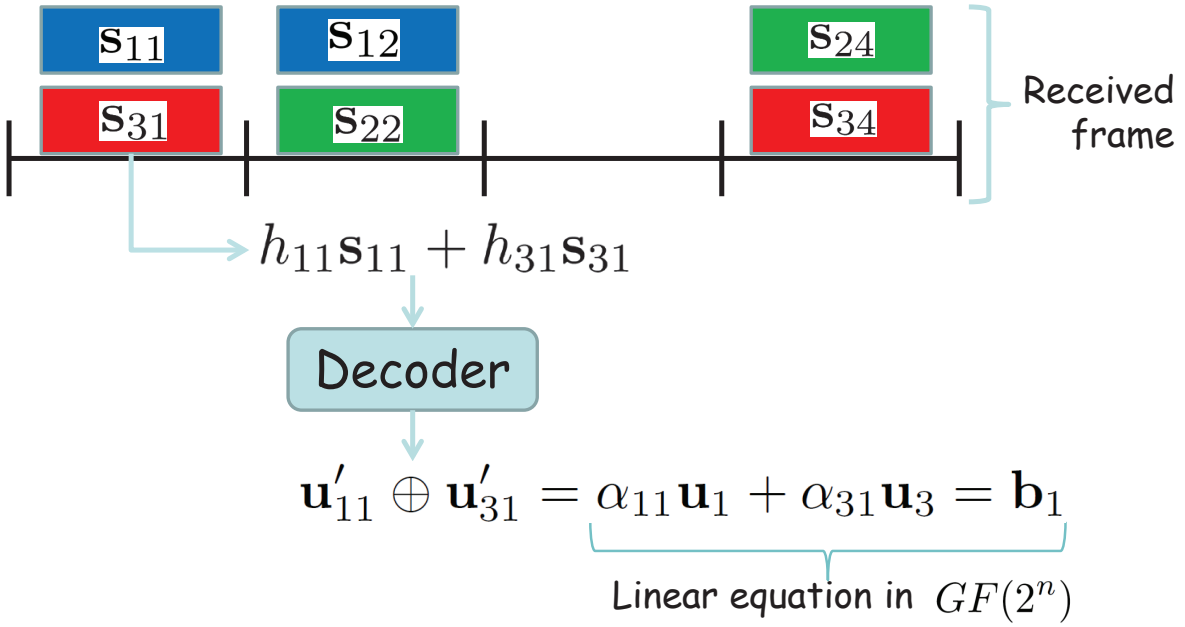


Figure 6.2: In each slot the receiver uses the orthogonal preambles to determine which nodes are transmitting. With the same preamble the channel from each of the transmitters in the slot to R is estimated. The channel $h_{ij}, j \in \{1, \ldots, S\}$ changes at each slot due to phase noise, according to the channel model described in Section 6.2. Once the channels have been estimated, the decoder applies MU PHY NC to calculate the bitwise XOR of the transmitted messages. The bitwise XOR corresponds to a linear equation in $GF(2^n)$ with coefficients α_{ij} which are known to the receiver through the header. In the figure only bursts with non-zero coefficients are shown. In order to simplify the notation, in the figure we indicated the vector $\mathbf{u}'_{ij} = \left[\alpha_{ij}\mathbf{u}_i^1, \ldots, \alpha_{ij}\mathbf{u}_i^L\right]$, representing the network coded packet, as $\alpha_{ij}\mathbf{u}_i$.

each slot the receiver uses the orthogonal preamble of each burst to determine which node is transmitting and which coefficient has been used for that burst. As described in Section 6.4.1, the coefficients used by a node in each burst are univocally determined by the preamble. The preamble can be determined at R using a bank of correlators which calculates in parallel the correlation of the received signal with each element in the set of available preambles. The preamble is also used by R to estimate the channel for each of the transmitters. The details

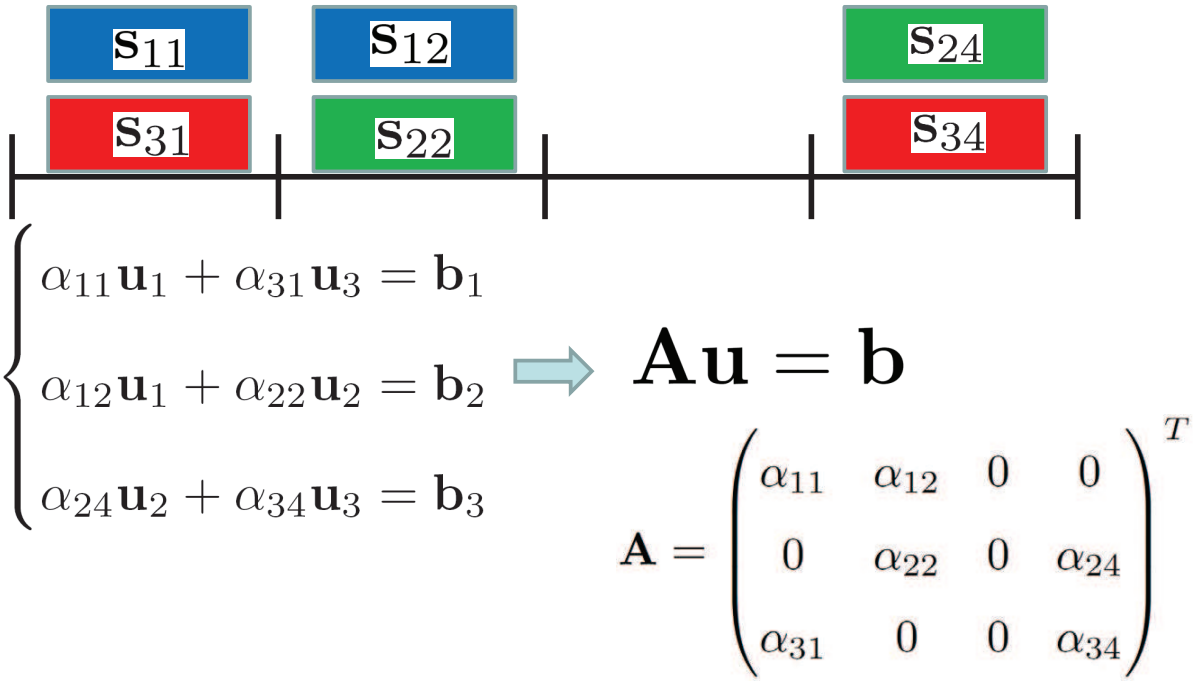


Figure 6.3: The receiver tries to channel-decode all the occupied slots, thus obtaining a system of equations in $GF(2^n)$. At this point, if the matrix $\mathbf{A}$ is full rank, R can obtain all the original messages. If $\mathbf{A}$ is rank deficient, R can decode the "clean" bursts (i.e., the bursts that did not experience collision), then subtract them from the slots where their replicas are. The procedure goes on until there are no more clean bursts. In the figure, ^{T} represents the transpose operator. In the figure only bursts with non-zero coefficients are shown. In order to simplify the notation, in the figure we indicated the vector $\mathbf{u}'_{ij} = [\alpha_{ij}\mathbf{u}_i^1, \ldots, \alpha_{ij}\mathbf{u}_i^L]$, representing the network coded packet, as $\alpha_{ij}\mathbf{u}_i$.

about the channel estimation are given in Section 6.6.1. Once the channel has been estimated, the receiver applies PHY NC decoding to calculate the bitwise XOR of the transmitted messages, as detailed in Section 6.3. According to what is stated in Section 2.1.1 and Section 6.4.1, the bitwise XOR is interpreted as a sum in $GF(2^n)$. Thus the slots that have been correctly decoded are interpreted as a system of linear equations in $GF(2^n)$ with coefficients α_{ij}, which are known to the receiver through the headers (see Fig. 6.2). At this point, if the coefficient matrix $\mathbf{A}$ has full rank, R can recover all the original messages using common matrix manipulation techniques in $GF(2^n)$ (see Fig. 6.3). If $\mathbf{A}$ is not full rank, not all the transmitted packets can be recovered. However, a part of them can still be retrieved using Gaussian elimination. The decoding process in case of rank deficient coefficient matrix is analyzed in Section 6.5. Note that, while in [85] the coefficient matrix $\mathbf{A}$ (called *mixing matrix*) is a complex matrix whose elements are the terminals channel gains, in NCDP $\mathbf{A}$ is a matrix in an EGF. In NCDP each slot is processed only once in the complex domain (PHY NC decoding), while all matrix manipu-

lations are done in $GF(2^n)$. In [85], instead, the matrix $\mathbf{A}$ is processed entirely in the complex domain. Operating in $GF(2^n)$ has an important advantage in terms of complexity, as all the processing can be done in the digital domain, and avoids numerical stability problems that may derive from using a complex matrix. If, on the one hand, using a complex coefficient matrix leads to a high probability of having full rank (which, however, also depends on the precision of the quantization in the sampling process), on the other hand in NCDP a relatively small field size (e.g., $GF(2^8)$) already achieves almost the same performance in terms of throughput as in the case of a complex matrix, as we show analytically in Section 6.5.

6.5 Throughput Analysis

During each frame users buffer packets to be transmitted in the following frame. Each node transmits its packet more than once within a frame, randomly choosing a new coefficient in $GF(2^n)$ in an independent fashion at each transmission. As described in the previous section, the coefficients can be generated using a pseudo-random number generator fed with a seed which is univocally determined by the chosen orthogonal preamble. Using the preamble the receiver can build up a coefficient matrix $\mathbf{A}$ for each frame, with $\mathbf{A}_{i,j} = \alpha_{ij}$, $\alpha_{ij} \in \{1, \ldots, 2^n - 1\}$, such as the one represented in Table 6.1. Columns represent time slots while rows represent

Table 6.1: Example of access pattern for three nodes transmitting in a frame with $S = 4$ slots. $\alpha_{ij} \in GF(2^n)$ is the coefficient used by node i in slot j. Each coefficient can assume one of $q = 2^n$ possible values, including value 0, which corresponds to the case in which the terminal does not transmit.

	Slot 1	Slot 2	Slot 3	Slot 4
T_1	α_{11}	α_{12}	α_{13}	α_{14}
T_2	α_{21}	α_{22}	α_{23}	α_{24}
T_3	α_{31}	α_{32}	α_{33}	α_{34}

the active terminals, i.e., the terminals that transmit in present frame. If $\alpha_{ij} = 0$, terminal i does not transmit in slot j. During time slot j, R receives the sum of the bursts with $\alpha_{ij} \neq 0$. From the received signal, R tries to obtain the bit-wise XOR of the encoded messages as described in Section 6.2. The XOR is interpreted by R as a linear equation in $GF(2^n)$, the coefficients

of which are derived through the orthogonal preamble as described in Section 6.4. If N^{tx} is the number of active terminals in a frame and assuming that all the received signals are decoded correctly, a linear system of equations in $GF(2^n)$ is obtained with S equations and N^{tx} variables. Each variable corresponds to a different source message. If **A** has rank equal to N^{tx}, then all the messages can be obtained by R. A necessary condition for **A** to be full rank is $N^{tx} \leq S$, i.e., the number of active terminals in a frame must be lower than the number of slots in a frame. Assuming Poisson arrivals with aggregate intensity G, the probability of such event is:

$$Pr\{N^{tx} \leq S\} = \sum_{n=0}^{S} \frac{(GS)^n e^{-GS}}{n!}, \tag{6.9}$$

which includes also the case in which there are no active terminals during a frame. For instance, in case of $S = 100$ slots and $G = 0.8$ the probability expressed by Eqn. (6.9) is on the order of 0.99. Even if $N^{tx} < S$, however, it can still happen that **A** is not full rank, i.e., not all the messages can be recovered. The probability that **A** is full rank for a given $N^{tx} < S$ depends on the MAC policy, and particularly on the probability distribution used to choose the coefficients.

One possibility is to use a uniform distribution for the coefficients (i.e., each coefficient can assume any value in $\{0, \ldots, 2^n - 1\}$ with probability 2^{-n}). In this case the number d of transmitted replicas is a random variable, and the probability that **A** is full rank is [94]:

$$P(S, N_{tx}) = \prod_{k=0}^{N_{tx}-1} \left(1 - \frac{1}{2^{n(S-k)}}\right). \tag{6.10}$$

Using equations (6.9) and (6.10) we find the expression for the normalized throughput:

$$\begin{aligned}
\Phi &= \frac{1}{S} \sum_{m=1}^{S} m \frac{(GS)^m e^{-GS}}{m!} P(S, m) \\
&= \frac{1}{S} \sum_{m=1}^{S} \frac{(GS)^m e^{-GS}}{(m-1)!} \prod_{k=0}^{m-1} \left(1 - \frac{1}{2^{n(S-k)}}\right) \\
&= \frac{1}{S} \sum_{m=0}^{S-1} \frac{(GS)^{m+1} e^{-GS}}{(m)!} \prod_{k=0}^{m} \left(1 - \frac{1}{2^{n(S-k)}}\right) \\
&= G \sum_{m=0}^{S-1} \frac{(GS)^m e^{-GS}}{m!} \prod_{k=0}^{m} \left(1 - \frac{1}{2^{n(S-k)}}\right).
\end{aligned} \tag{6.11}$$

From Eqn. (6.11) we can see that Φ grows with n, which means that the system throughput increases with the size of the considered finite field. The throughput achievable in case of an

asymptotic large field size n is:

$$\lim_{n\to\infty}\Phi = \lim_{n\to\infty}\left[G\sum_{m=0}^{S-1}\frac{(GS)^m e^{-GS}}{m!}\prod_{k=0}^{m}\left(1-\frac{1}{2^{n(S-k)}}\right)\right]$$
$$= G\sum_{m=0}^{S-1}\frac{(GS)^m e^{-GS}}{m!}. \tag{6.12}$$

Thus, the normalized throughput Φ tends to the probability of having less than S transmitters in a frame as $n \to \infty$. Note that this is the same performance that would be achieved by a scheme that uses coefficient matrix in the complex domain, as in [85]. Further in this section we show that almost the same performance can be achieved by NCDP using a finite and relatively small field size.

The MAC scheme we just analyzed presents one main drawback in terms of energy efficiency. As a matter of fact, given the frame length S, a node transmits each message on average $E[d] = S \times p$ times, $p = (1 - 2^{-n})$ being the probability to choose a non-zero coefficient, i.e., the average number of transmissions grows linearly with S. In order to decrease the energy consumption, the probability of choosing the zero coefficient may be increased. However, a reduction in the transmission probability p may affect the system throughput. In order to understand the relationship between the probability p and the throughput Φ, we refer to some results in random matrix theory. The problem can be formulated as follows: consider an $N^{tx} \times S$ random matrix $\mathbf{A}$ over $GF(2^n)$ with i.i.d. entries, each of which assumes value 0 with probability $1 - p$ while with probability p it assumes values in $\{1, \ldots, 2^n - 1\}$. We are interested in the relationship between p and the probability that $\mathbf{A}$ is full rank. In [95] the authors show that, in order to achieve a rank $N^{tx} - O(1)$ with high probability, then, for N^{tx} large, p cannot be lower than the threshold probability $\frac{\ln(N^{tx})}{N^{tx}}$. At high loads (i.e., $G \simeq 1$), on average $N^{tx} \simeq S$, which means that, setting $p = \frac{\ln(S)}{S}$, the average number of transmissions (and so the energy consumption) for each node is $E[d] = \ln(S)$, i.e., it grows logarithmically with the number of slots in a frame. On the other side, S must be kept large enough, as this increases the decoding probability (see Eqn. 6.12). With reference to the example considered earlier in this section the average number of transmissions corresponding to the minimum required p for $S = 100$ is equal to about 4.6. We evaluated numerically the effect a reduction of p has on Φ for the case of $S = 100$ and $q = 2^8$. We considered three cases. In the first one the transmission probability in each slot has been set to $p = 1 - 2^{-n} = 0.9961$, which corresponds to the case studied in the first part of this section and for which the throughput is given by Eqn. (6.11). In

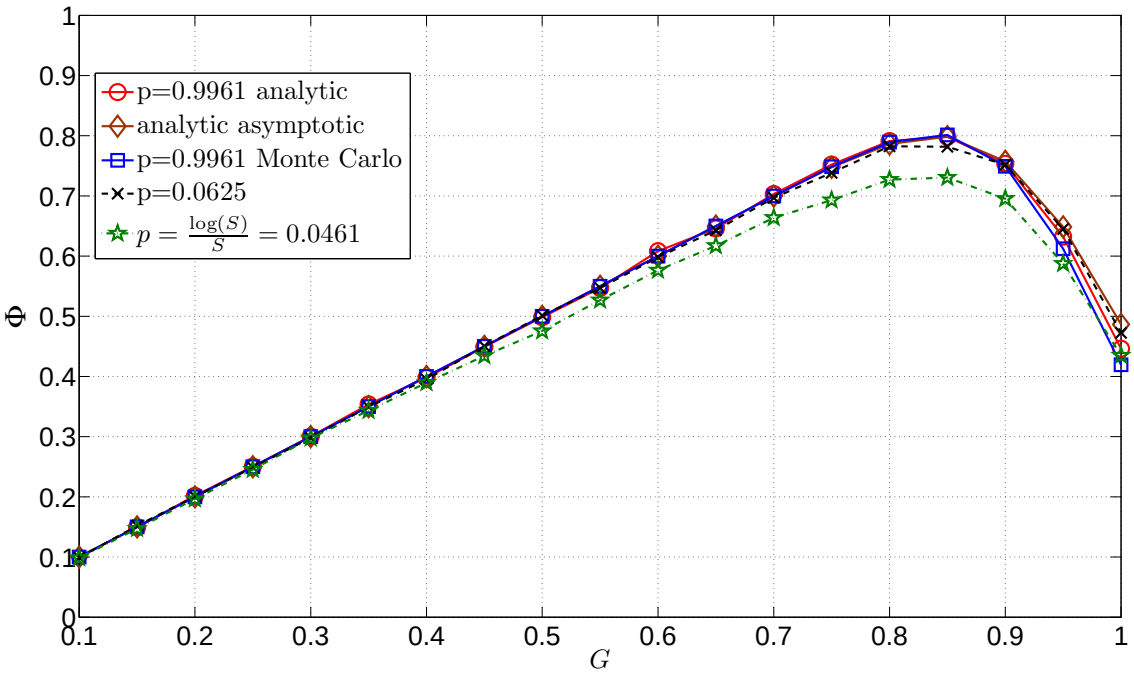


Figure 6.4: Normalized throughput plotted against the normalized offered load for different values of the transmission probability p. We set $S = 100$ slots while the coefficients were chosen in $GF(2^8)$. The asymptotic analytical curve (Eqn. 6.12) is also plotted.

the second case we set p just above the threshold, i.e., $p = 0.0625 > \frac{\ln(S)}{S} = 0.0461$, while in the last case p has been set exactly equal to the threshold probability. Fig. 6.4 shows the results together with the numerical validation of Eqn. (6.11). It is interesting to note how passing from $p = 0.9961$ to $p = 0.0628$, with a reduction in transmission probability (or, equivalently, in average energy per received message) of about 93.7%, leaves the throughput unchanged, while a further decrease of p of just another 1.5% leads to a 10% reduction in the maximum throughput with respect to the case $p = 0.9961$. The asymptotic analytical curve described by Eqn. 6.12 is also plotted in Fig. 6.4. Such curve represents the throughput of a system where coefficients are chosen in a finite field with asymptotically large field size. It also represents the throughput of a system derived from the NDMA scheme proposed in [85], i.e., the coefficient matrix is complex and is processed in the complex domain. It can be seen that the performance in terms of throughput is almost the same for NDMA and for NCDP with $p = 0.0625$, i.e., NCDP does not lose significantly with respect to NDMA, while saves in complexity by doing all the processing in a finite field instead of the complex domain.

To further lower the energy consumption and control the number of repetitions d (which,

being a Bernoulli random variable, can theoretically assume values as large as S), an alternative is to fix the number of transmitted replicas *a priori*. Although this solution may decrease the probability of decoding all the transmitted messages (because the resulting $\mathbf{A}$ matrix would be a subset of all possible matrix of the same size), it may still be possible to recover part of them by using Gaussian elimination. In order to increase the number of available sequences of coefficients and to avoid the problems deriving from an eventual unsuccessful decoding of some of the slots, the system can be designed so that, for a given preamble, a different coefficient is used for the burst transmitted in a given slot. In this way the sequence of coefficients associated to the preamble changes depending on the transmission slots. The total number of different sequences associated to a given preamble is, thus, equal to the number of possible dispositions of the d repetitions over the S slots of the frame, that is, $\binom{S}{d}$. Note that the use of a preamble is not a peculiarity of NCDP, as usually practical systems make use of a preamble to perform channel estimation.

6.6 Implementation Aspects

In this section we address several aspects related to a practical implementation of the NCDP scheme. In particular we consider the issue of node identification and channel estimation, the complexity and the detection of errors in case of unsuccessful decoding.

6.6.1 Node Identification and Channel Estimation

For each frame the receiver R needs to know which of the active terminals is transmitting in each slot and must have CSI for each of the users. Both needs are addressed including an orthogonal preamble, such as the spreading codes used in CDMA, at the beginning of each transmitter's burst. As described in Section 6.4, during a given frame each of the active terminals randomly chooses an orthogonal preamble and appends it to the message to be transmitted after the channel encoding. This leads to some additional complexity due to the need to store the whole preambles set. However, the low price and size of memories makes such additional complexity not an issue. More observations on this as well as on computational complexity analysis can be found in [86] and [96], respectively.

The use of an orthogonal preamble was proposed in [86] for the estimation of the phase

in collided bursts. In [86] frequency offset and channel amplitude are derived from the clean bursts (i.e., bursts that did not experience collision) and assumed to remain constant over the whole frame. Unlike in [86], the method we propose does not rely only on clean bursts. Thus the frequency offset and the amplitude of each transmitter must be estimated using the collided bursts. Although the performances of the estimator are likely to degrade with respect to the clean burst case, especially in case of high order collisions, the estimation can leverage in the information of all the collided bursts, which improves the estimation. For instance, if a packet is transmitted twice during a given frame and experiences collisions of size 2 in the first transmission and 4 in the second, the two estimations can be combined to obtain a better estimation of amplitude and frequency offset, which are constant during the whole frame.

In order to prove the feasibility of channel estimation in such conditions we show the results we obtained using the Estimate Maximize (EM) algorithm. We adopt the approach described in [97], where the EM algorithm is used to estimate parameters from superimposed signals. In [97] two examples are presented related to multipath delay and direction of arrival estimation. We apply the same approach to estimate amplitudes, phases and frequency offsets from the baseband samples of the received signal in case of a collision of size k. The algorithm is divided into an E step, in which each signal is estimated, and an M step, in which the mean square error between the estimation made at the E step of current iteration and the signal reconstructed using parameters calculated in previous iteration is minimized with respect to the parameters to estimate. Formally, once initialized the parameters with randomly chosen values, at each iteration we have the following two steps:

Estimation step - for $i = 1, \ldots, k$ calculate

$$\hat{p}_i^{(n)}(t) = b_i(t)\hat{A}_i^{(n)} e^{j(2\pi\widehat{\Delta\nu}_i^{(n)} T_s t + \hat{\varphi}_i^{(n)})} + \beta_i \left[r(t) - \sum_{l=1}^{k} b_l(t)\hat{A}_l^{(n)} e^{j(2\pi\widehat{\Delta\nu}_l^{(n)} T_s t + \hat{\varphi}_l^{(n)})} \right], \quad (6.13)$$

$$(6.14)$$

Maximization step - for $i = 1, \ldots, k$ calculate

$$\min_{A', \Delta\nu', \varphi'} \sum_{t=1}^{N^{pre}} \left| b_i(t)\hat{p}_i^{(n)}(t) - A' e^{j(2\pi\Delta\nu' T_s t + \varphi')} \right|^2, \quad (6.15)$$

$$(6.16)$$

where $p_i(t)$ is the preamble of burst i after the matched filter, A', $\Delta\nu'$ and φ' are tentative values for the parameters to be estimated, N^{pre} is the preamble length, $b_i(t) \in \{\pm 1\}$ is the t-th symbol in the preamble of the i-th node and T_s is the sampling period, taken equal to the symbol rate. β_i, $i = 1, \ldots, k$, are free parameters that we arbitrarily set to $\beta_i = 0.8$, for $i = 1, \ldots, k$.

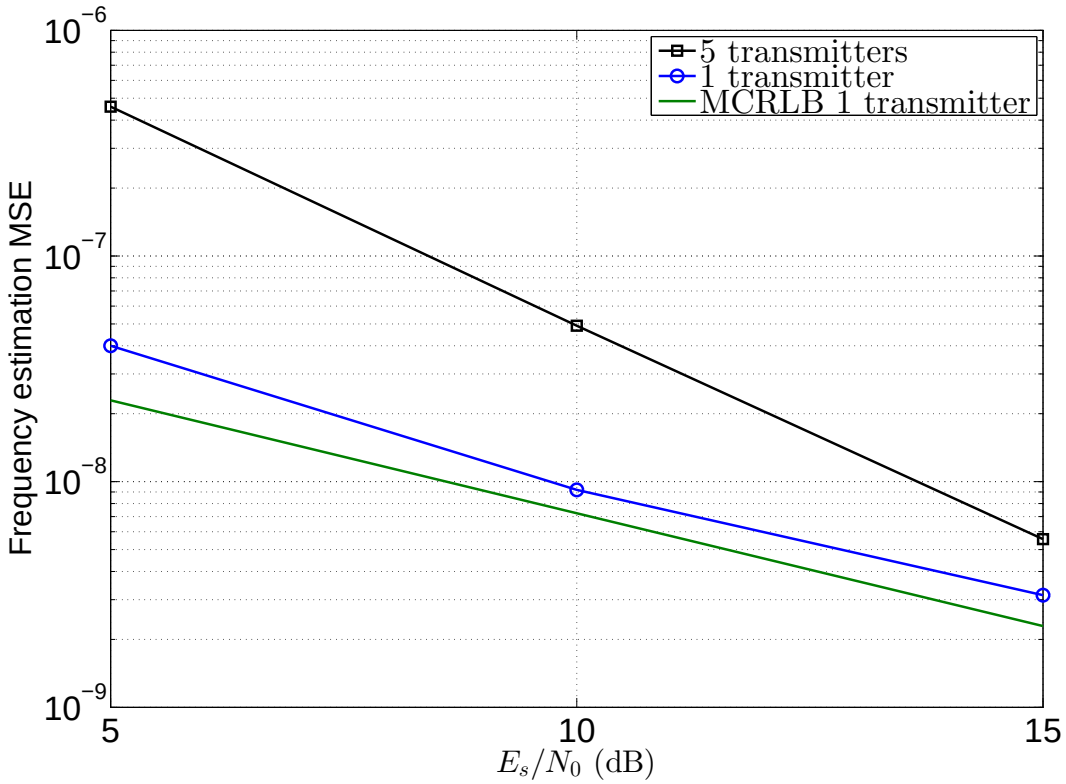


Figure 6.5: MSE of the frequency offset estimation ($E[|\widehat{\Delta\nu} - \Delta\nu|^2]$). E_s is the average energy per symbol for each node. The modified Cramer-Rao lower bound (MCRLB) for the case of one transmitter is also shown for comparison.

We evaluated numerically the performance of the EM estimator assuming that phase offsets are uniformly distributed in $[-\pi, +\pi]$, frequency offsets are uniformly distributed in $[0, \Delta\nu^{max}]$ with $\Delta\nu^{max}$ equal to 1% of the symbol rate on the channel ($1/T_s$), and amplitudes are log-normally distributed, having a mean value and a standard deviation of the associated normal variable equal to 0 and 0.27, respectively. Figures 6.5, 6.6 and 6.7 show the mean squared error (MSE) of the estimation error for frequency, phase and amplitude, respectively. Amplitude error is normalized to the actual amplitude value while phase error is normalized to π. In the simulations we used as preambles Walsh-Hadamard words of length 128 symbols.

The EM algorithm was run twice starting from randomly chosen initial values of the pa-

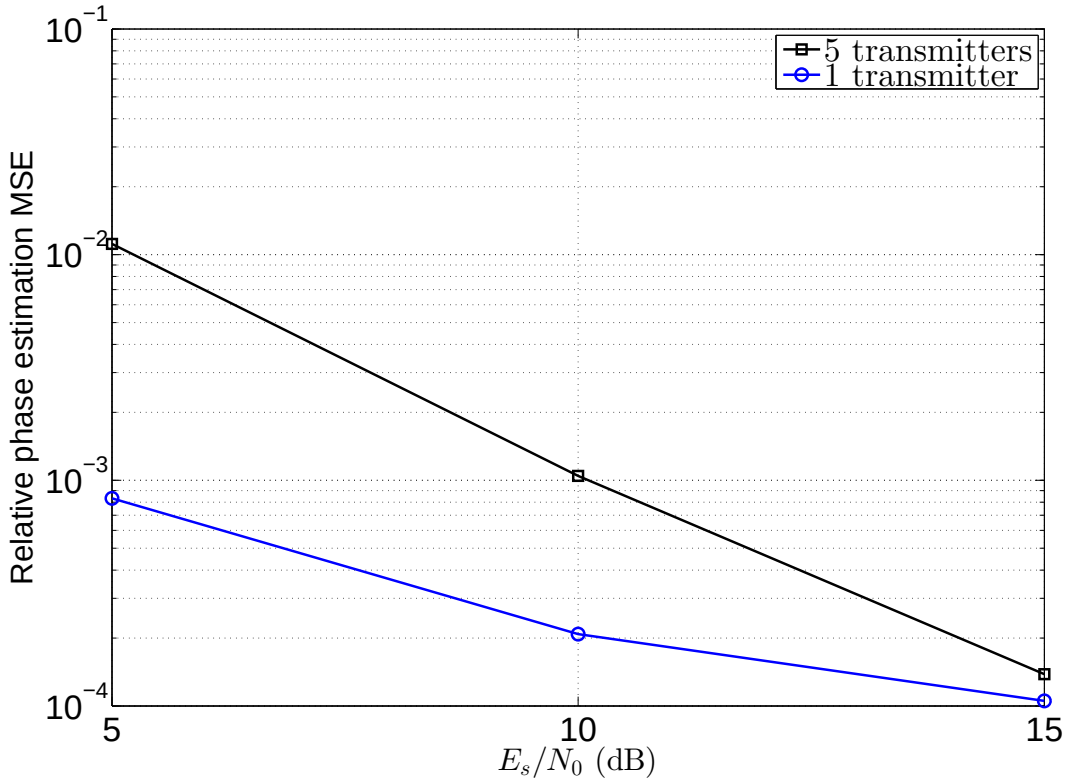


Figure 6.6: MSE of the phase offset estimation normalized to π ($E[|\hat{\varphi} - \varphi|^2]/\pi^2$). E_s is the average energy per symbol for each node.

rameters and taking as result the values of the parameters that lead to the minimum of the sum across the signals of the error calculated in the last E step. This was done in order to reduce the probability to choose a "bad" local maximum, which is a problem that affects all the "hill-climbing" algorithms. For each run of the EM algorithm 6 iterations were made.

In Fig. 6.8 the FER curves for different collision sizes obtained using the LLR values calculated in Section 6.3 are shown. The plots are obtained using a tail-biting duo-binary turbo code with rate $1/2$ and codeword length equal to 192 symbols. The same offset values used for the channel estimation were chosen. The FER curves for the case of estimated channels using the EM algorithm are also shown.

6.6.2 Complexity Analysis

From Eqn. (6.8) it is clear that the computational complexity of a MU PHY NC decoder is higher than that for a typical channel decoder and that grows as the collision size grows.

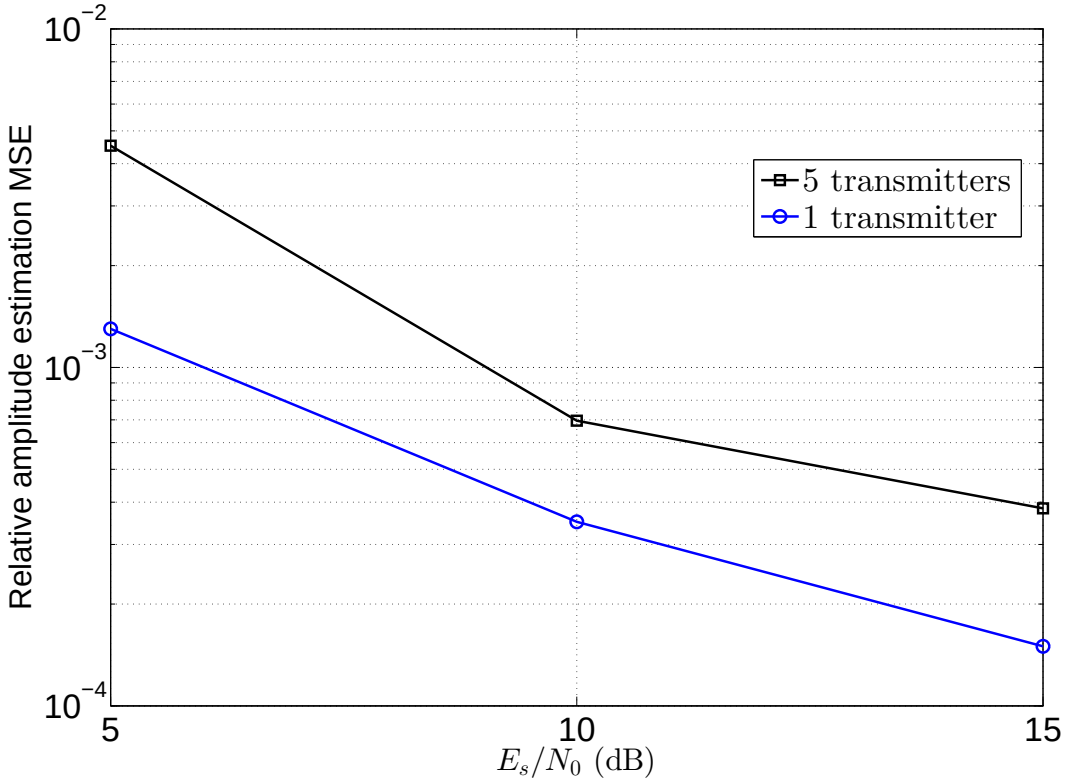


Figure 6.7: MSE of the amplitude estimation normalized to the actual amplitude of the channel ($E[|\hat{A} - A|^2/A^2]$). E_s is the average energy per symbol for each node.

Although complexity is not a main issue if the receiver is located at a satellite ground station, it may be interesting to evaluate the complexity of the decoder in order to properly dimension the receiver and evaluate the applicability of MU PHY NC in systems with low-complexity receivers. More specifically, we are interested in the additional complexity of a MU PHY NC decoder with respect to a standard one. By complexity we mean the number of elementary operations (real-valued addition and multiplication) needed to decode a codeword. As the only difference between a normal channel decoder and a MU PHY NC decoder is the calculation of the LLR's, we focus on this. In particular, we look for the scaling law of the complexity per received symbol as a function of the collision size k. With reference to the numerator of Eqn. (6.8), we start by considering the argument of the exponential function. Fixing a sample (i.e., fixing l) the term $\mathbf{d}^o(2i-1,m)^T\mathbf{h}(t_l)$ is the scalar product of a real-valued k-dimensional vector by a complex-valued k-dimensional vector. This requires a total of $2k$ elementary operations for element-wise multiplications plus 2(k-1) operations for the complex additions, for a total of $4k-2$ operations. Adding $r(t_l)$ up to the complex number $\mathbf{d}^o(2i-1,m)^T\mathbf{h}(t_l)$ requires 2 more

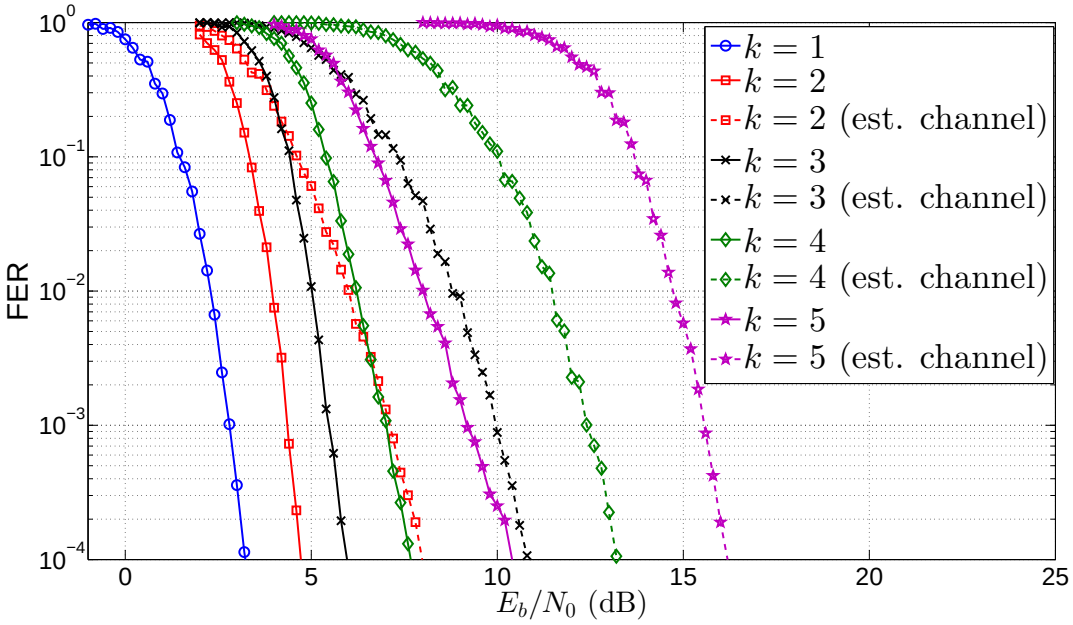


Figure 6.8: FER curves for the XOR of transmitted messages for different numbers of transmitters. E_b is the energy per information bit for each node. A tail-biting duo-binary turbo code with rate $1/2$ and codeword length 192 symbols is used by each node. Phase offsets are uniformly distributed in $[-\pi, +\pi]$, frequency offsets are uniformly distributed in $[0, \Delta\nu^{max}]$ with $\Delta\nu^{max}$ equal to 1% of the symbol rate on the channel. Amplitudes are constant and equal to 1. The FER curves for the case of estimated channels using the EM algorithm are also shown.

operations, while the modulus squared needs 3 more operations. One more operation is needed to divide the real number calculated up to now by $2N_0$. We are now left with the evaluation of an exponential function in a real point. Let us indicate with N_{exp} the number of operations needed to do this. Note that N_{exp} does not depend on k. Thus, a total of $2k+2(k-1)+2+3+1+N_{exp} = 4k + 4 + N_{exp}$ operations are needed to calculate each of the terms in the sum at the numerator of Eqn. (6.8). The analysis of the denominator is exactly the same and leads to the same result. At this point we need to know the number of elements in the sum at the numerator and at the denominator. As the sum at the numerator is done over all possible permutations of any odd number of $+1$'s in a string of k symbols, it can be easily seen that the number of terms in the sum is always equal to 2^{k-1}. The same is true for the denominator. Thus, considering both the numerator and the denominator, the per-symbol complexity for the calculation of the LLR in

Eqn. (6.8) can be written as:

$$\begin{aligned} &2(4k+4+N_{exp})2^{k-1}+1 \\ &=(4k+4+N_{exp})2^{k}+1, \end{aligned} \tag{6.17}$$

where the "$+1$" comes from the real-valued division of the numerator by the denominator. From Eqn. (6.17) we see that the complexity scales as $4k2^k$ for large k. The complexity scaling law is plotted in Fig 6.9. From Fig. 6.9 we see how the complexity rapidly grows (faster

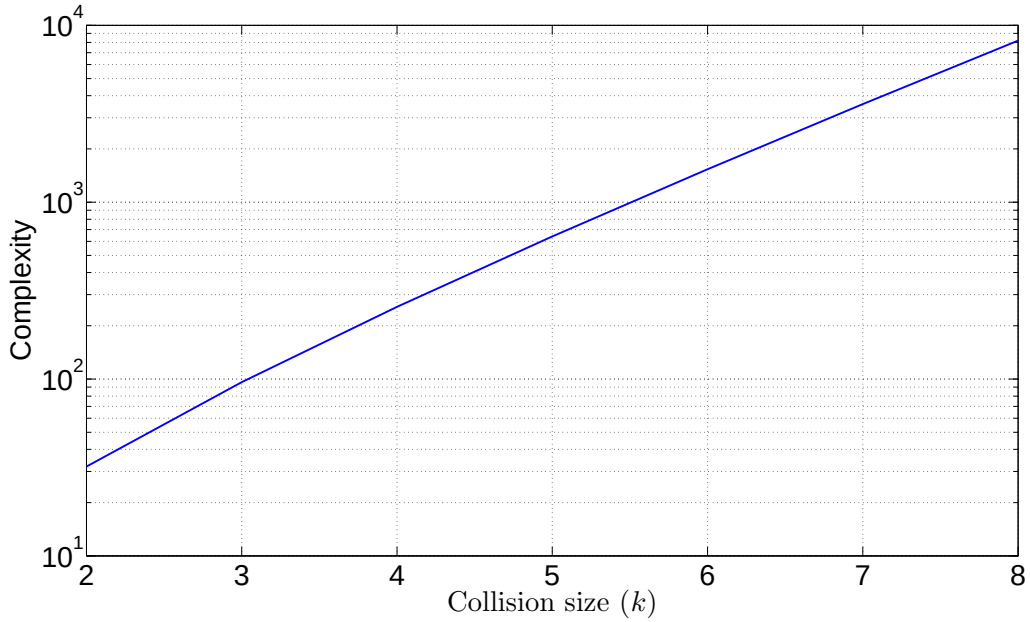


Figure 6.9: Scaling law ($4k2^k$) of MU PHY NC LLR's calculation complexity plotted against the collision size.

than exponential) with the collision size. From this, and from constraints that may come from the physical level (e.g.: maximum input power at receiver's front end, see [90] for further examples), it is likely that, in a real system, the maximum decodable collision size would be bounded. Such constraint must be accounted for at higher levels when designing the multiple access system in which MU PHY NC is applied. The maximum decodable collision size would depend on parameters such as the receiver front-end dynamic range, the computational capabilities of the receiver and eventual processing delay constraints, which may change from one system to the other. Further analysis on this aspect is out of the scope of this chapter.

6.6.3 Error Detection

An important issue in slotted ALOHA is the capability of the receiver to determine whether the received bursts are correctly decoded or not. This is particularly important in NCDP, where the error made in the decoding of a collision can lead to the loss of a whole frame. A common practice in packet networks is the use of a cyclic redundancy check (CRC), that allows to detect a wrong decoding with a certain probability. Some CRC's are based on a field which is appended to the message before channel coding, called *CRC field*. As the CRC operations are done in $GF(2)$ and by the linearity of the channel encoder, the CRC field in the message obtained by decoding a collision of size k is a good CRC for $\mathbf{u}_s$, which is the bitwise XOR of the messages encoded in the k collided signals. This allows to detect decoding errors, within the limits of the CRC capabilities, also in collided bursts. The implementation aspect of what type of CRC should be used is out of scope of this paragraph.

6.7 Performance of Multi User Physical Layer Network Coding with Imperfect Symbol Synchronization

In Section 6.2 we assumed that signals from different receivers add up with symbol synchronism at the receiver in case of a collision. In Fig. 6.10 an example is shown of received signal and sampling instants in the case of three nodes transmitting with no timing offsets. The transmitted signals, which are also shown, modulate the sets of symbols [-1 1 -1], [-1 1 1] and [-1 -1 -1]. The situation depicted in the figure is an illustrative one, as in a real system both I and Q signal components are present, signals may have different amplitudes, phase and frequency offsets for each of the bursts and the signal is immersed in thermal noise. In a real system there is likely to be a certain symbol misalignment, which grows larger as the resources dedicated to the synchronization phase diminish (see, e.g., [98] and references therein for examples of synchronization algorithms). Being able to cope with non perfect symbol synchronism can bring important advantages, such as less stringent constraints on signal alignment, with consequent savings in terms of network resources needed for the synchronization. In this section we study the effect of non perfect symbol synchronization and propose possible countermeasures. Let us consider a slotted multiple access with k nodes transmitting in the same slot. We assume that each transmitter has its own phase and frequency offsets. We further assume that each

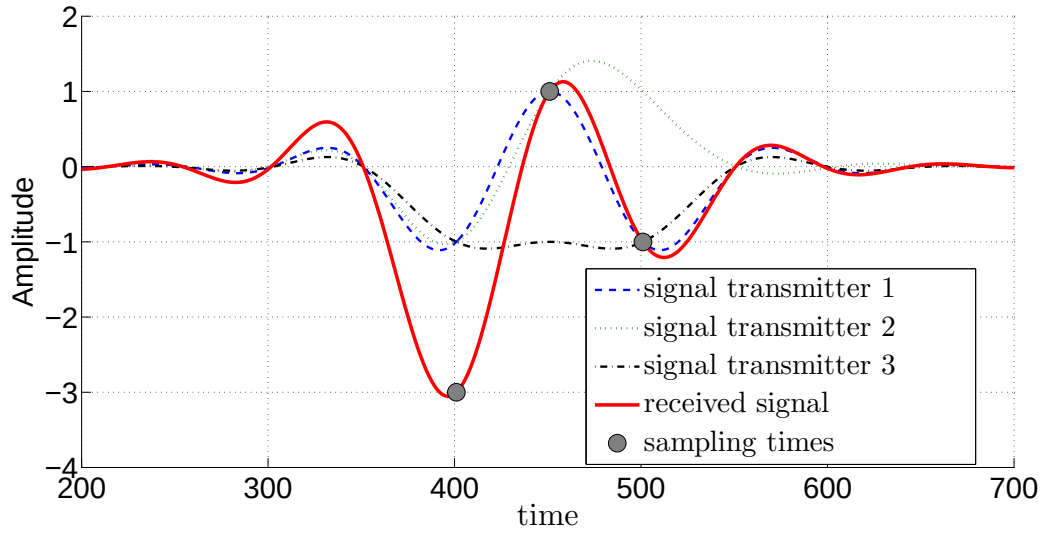


Figure 6.10: Received signal after the matched filter in case of three colliding bursts with no timing offsets, i.e., $\Delta T_1 = \Delta T_2 = \Delta T_3 = 0$. The transmitted signals after the matched filter in case of collision-free reception are also shown. The transmitted symbols are: [-1 1 -1], [-1 1 1] and [-1 -1 -1] for transmitter 1, 2 and 3, respectively. For sake of clarity, frequency and phase offsets as well as channel amplitudes were not included in the plot and the signals were considered as real. The samples, shown with grey circles in the figure, are taken at instants corresponding to the optimal sampling instants for each of the signals as if they were received without experiencing collision.

burst falls completely within the boundaries of a time slot, i.e., no burst can fall between two consecutive time slots. With reference to the burst that first arrives at R, we call T' the time at which the peak of the first symbol is seen by the receiver. In other words, T' is the time instant at which R sees the peak of the first symbol among all bursts and symbols transmitted within a given slot. We define the *relative delay* (RD) ΔT_i of node i as the temporal distance between the peak value of the first pulse of burst i and T'. In other words, the burst which arrives first at the receiver is used as reference, i.e., has RD equal to 0. We assume SRRC pulses with roll off factor α are used. We further assume that all RD's belong to the interval $[0, \Delta T^{max}]$, with $0 \leq \Delta T^{max} \leq T_s/2$.

In case of a collision of k bursts, the received signal before the matched filter is:

$$y(t) = \sum_{i=1}^{k} s_i(t) + w(t), \tag{6.18}$$

where

$$s_i(t) = A_i \sum_{l=1}^{N} b_i(l) g(t - lT_s - \Delta T_i) e^{j(2\pi \Delta \nu_i t + \varphi_i)}, \tag{6.19}$$

N being the number of symbols in the burst, $g(t)$ is the square root raised cosine pulse and $w(t)$ represents an AWGN process. The samples taken after the matched filter at times t_l are:

$$r(t_l) = y(t) \otimes g(-t) \mid_{t=t_l} = \sum_{i=1}^{k} q_i(t_l) + n(t_l), \tag{6.20}$$

where,

$$q_i(t_l) = A_i \sum_{l=1}^{N} b_i(l) p(t_l - lT_s - \Delta T_i) e^{j(2\pi \Delta\nu_i t_l + \varphi_i)}, \tag{6.21}$$

$p(t)$ being the raised cosine pulse and $n(t_l)$ is the noise process after filtering and sampling. Note that in (6.21) the exponential term is treated as a constant. This approximation is done under the assumption that $\Delta\nu T_s \ll 1$, i.e., the exponential term is almost constant over many symbol cycles.

The sampled signal is then sent to the channel decoder. It is not clear at this point which is the optimal sampling time for the received signal, as the optimal sampling time for each of the bursts taken singularly may be different. Moreover, sampling the signal just once may not be the optimal choice. Actually, as we will show in next section, the performance of the decoder is quite poor in case a single sample per symbol is taken.

In the following we propose four different methods to mitigate the impairment due to imperfect symbol synchronization. We assume that R has knowledge of the relative delays of all the transmitters, which can be derived through the orthogonal preambles. We further assume that R has perfect CSI for each of the transmitters. Without loss of generality and for ease of exposition, from now on we will refer to the sampling time for the first symbol.

6.7.1 Single Sample

Mean Delay The first method we consider is the *Mean Delay* (MD) method, which we use as our benchmark. In MD the received signal is sampled just once per symbol. The sampling time is chosen to be the mean of the relative delay, i.e.:

$$T^{MD} = \frac{1}{k} \sum_{m=1}^{k} \Delta T_m. \tag{6.22}$$

The sample $r(T^{MD})$ is then used to calculate the LLR's as in Eqn. (6.8). The inter-symbol interference (ISI) is not taken into account.

6.7.2 Multiple Samples

In the following we describe four different methods that use k samples per symbol, k being the collision size.

We start by describing two methods in which the symbol is sampled k times in correspondence of the RD's. Due to the non perfect synchronization, when the signal is sampled in ΔT_i the sample obtained is the sum of the first symbol of each of the users, weighted by the relative channel coefficient, plus a term of ISI due to signals $s_j, j \in \{1, \ldots, k\}, j \neq i$, which are sampled at non ISI-free instants. As the LLR calculation needs the channels of each of the users, the ISI should be taken into account. However, the ISI is a function of many (theoretically all) symbols, and can not be taken into account exactly. In Fig. 6.11 the received

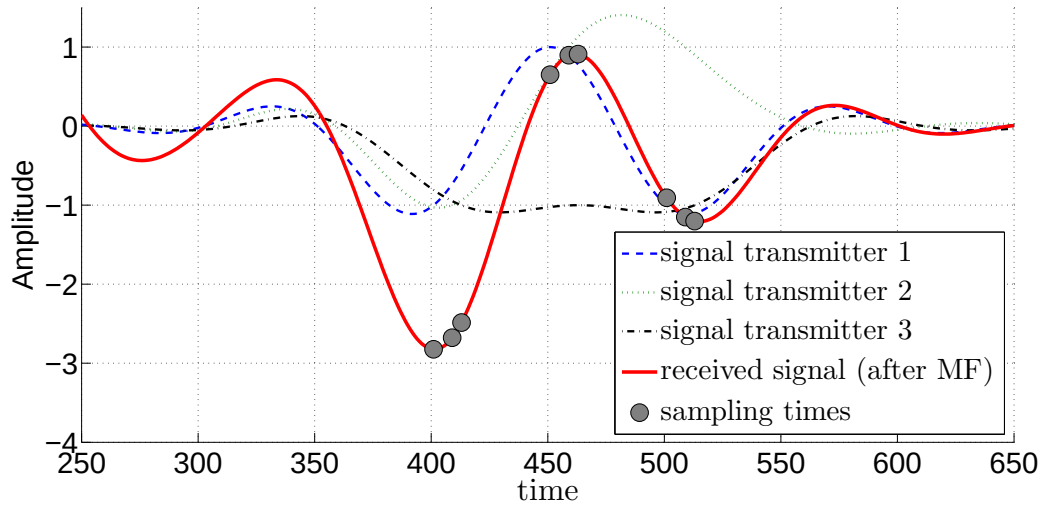


Figure 6.11: Received signal after the matched filter in case of three colliding bursts with timing offsets $\Delta T_1 = 0, \Delta T_2 = T_s/6$ and $\Delta T_3 = T_s/4$. The transmitted signals after the matched filter in the case of collision-free reception are also shown. The transmitted symbols are: [-1 1 -1], [-1 1 1] and [-1 -1 -1] for transmitter 1, 2 and 3, respectively. The samples, shown with grey circles in the figure, are taken at instants corresponding to the optimal sampling instants for each of the signals as if they were received without experiencing collision. Unlike in the case of perfect symbol alignment, here more than one sample per symbol is taken.

signal after the matched filter is shown in the case of three colliding bursts with timing offsets $\Delta T_1 = 0, \Delta T_2 = T_s/6$ and $\Delta T_3 = T_s/4$. The transmitted signals after the matched filter in the case of collision-free reception are also shown. The symbols transmitted by each terminal are the same as in Fig. 6.10. The samples, shown with grey circles in the figure, are taken in correspondence of the RD's, which coincide with the optimal sampling instants for each of the

signals as if they were received without experiencing collision.

Mean LLR In *Mean LLR* (MLLR) the received signal is sampled k times in the instants corresponding to ΔT_i, $i = 1, \ldots, k$. For each of the samples the LLR's are calculated as in (6.8). Then the average of the k LLR's is passed to the decoder.

Mean Sample As in MLLR, also in *Mean Sample (MS)* $r(t)$ is sampled k times in correspondence of the relative delays. The difference between the two methods is that in MS the samples are averaged out to obtain the mean sample:

$$\overline{r}(t) = \frac{1}{k} \sum_{m=1}^{k} r(\Delta T_m). \tag{6.23}$$

Finally, $\overline{r}(t)$ is used in Eqn. (6.8) instead of $r(t)$.

Uniform Sampling In the *Uniform Sampling (US)* method the signal is sampled k times as in previous methods, but the sampling times do not correspond to the RD's. The sampling times are chosen uniformly in $[0, \Delta T^{max}]$, i.e., in case of k transmitters the samples are taken at intervals of $\Delta T^{max}/(k-1)$. Then, as in MS, the samples are averaged out and used in the calculation of the LLRs. This method has the advantage that receiver does not need the knowledge of the RD's in order to perform the decoding and the sampling itself is simplified as it is done uniformly in each symbol.

Equivalent Channel The received signal is sampled k times in the instants corresponding to ΔT_i, $i = 1, \ldots, k$. In the method *Equivalent Channel (EC)* the amplitude variation of the channel of each user due to imperfect timing is taken into account for the current symbol. Note that the ISI is not taken into account, only the variation in amplitude of present symbol due to imperfect timing is accounted for. Assuming that the received signal is sampled at time $t = \Delta T_i$, then the channel coefficient of burst q that is used in the LLR is:

$$h_q^{eq}(t) = A_q e^{j(2\pi \Delta\nu_q T_s \Delta T_i + \varphi_q)} p(\Delta T_i - \Delta T_q), \tag{6.24}$$

$p(t)$ being the raised cosine pulse. After the sampling, the k samples per symbol are averaged together and used in the LLR instead of $r(t)$. This sampling procedure is equivalent (apart from the ISI) to filtering the received signal using a filter which is matched not to the single pulse,

but to the pulse resulting from the delayed sum of M pulses. In Fig. 6.12 the frame error rate is shown for the case of 5 transmitters with delays uniformly distributed in $[0, T_s/4]$. Constant channel amplitudes were considered, while phases and frequency offsets are i.i.d. random variables in $[0, 2\pi]$ and $[0, \Delta\nu_{max}]$ respectively, where $\Delta\nu_{max}$ is equal to $1/(100T_s)$. The results for the 5 different methods are shown together with the FER for the case of ideal symbol synchronism. The methods that use more than one sample per symbol perform significantly better than MD, which uses only one sample per symbol. Among the methods based on oversampling, MS and EC perform slightly better than the other two. The FER of all methods present a lower slope with respect to the ideal case. The loss is about 1 dB at $FER = 10^{-2}$ for the methods that use oversampling. Among the considered methods, the MS and the EC achieve slightly better performances than the others in terms of FER. Both of them use the information about the relative delays of the colliding signals. The MLLR method also uses information about the relative delays, but is not able to use it properly, performing roughly as the US method, which only uses information about the minimum and the maximum of the delays.

6.8 Numerical Results

Our performance metrics are the normalized throughput Φ defined as:

$$\Phi = G(1 - \Upsilon), \tag{6.25}$$

where $\Upsilon \in [0, 1]$ is the average packet loss rate (i.e., the ratio of the number of lost packets to the total number of packets that arrive at the transmitters), and the average energy consumption per received message η, defined as the average number of transmissions needed for a message to be correctly received by R. We consider two benchmarks. The first one is the CRDSA scheme, which has been proposed in [86]. In CRDSA a node transmits two or more copies of a burst (twin bursts) in different slots randomly chosen within a frame. Each of the twin bursts contains information about the position of the other twin bursts in the frame. If one of the twin bursts does not experience a collision (i.e., it is *clean*) and can be correctly decoded, the position of the other twin bursts is known. These bursts may or may not experience a collision with other bursts. If a collision occurs, these are removed through interference cancelation using the decoded bursts. In order to do this R memorizes the whole frame, decodes the clean bursts, reconstructs the modulated signals and, once the effect of each user's channel has been

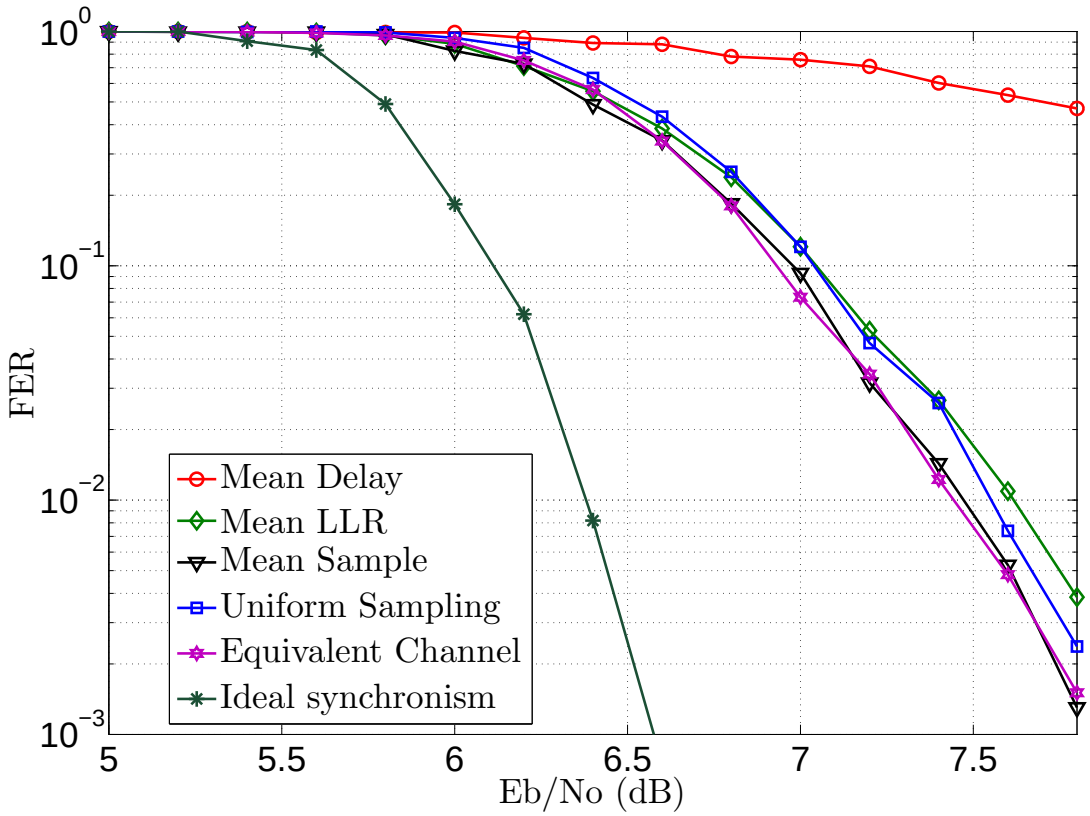


Figure 6.12: FER for decoding a collision of size 5 with independent frequency and phase offsets across the transmitters and delays uniformly distributed in $[0, T_s/4]$. A roll-off factor of $\alpha = 0.35$ was used. Perfect CSI is assumed. A tail-biting duo-binary turbo code with rate $1/2$ and codeword length 1504 symbols is used by each node. The results for the 5 different methods are shown together with the FER for the case of ideal symbol synchronism. Oversampling significantly improves the FER with respect to the case of a single sample. The two methods that exploit knowledge of relative delays, i.e., MS and EC, perform slightly better than the others. The FER of all methods present a lower slope w.r.t. the ideal case, losing about 1 dB at $FER = 10^{-2}$ for the methods that use more than one sample.

included in the reconstruction, they are subtracted from the slots in which their replicas are located. The IIC process is iterated for a number N^{iter} of times, at each time decoding the bursts that appear to be "clean" after the previous IIC iteration.

The second benchmark we consider is a SA system. In a SA system each burst is transmitted only once. Burst are received correctly if only one is transmitted in each slot. If more than one node transmits, a collision is declared. The capture effect is not considered in any of the schemes.

We consider two different simulation setups. In the first one the nodes do not receive any

feedback from the receiver, while in the second setup R gives some feedback to the active terminals. For this last case we consider an ARQ scheme, in which a node receives an ACK or a NACK from the receiver in case a message is or is not correctly received, respectively. The amount of feedback is limited to one ACK/NACK message per node and per frame. An alternative to the NACK is to having the transmitters using a counter for each transmitted packet, indicating the time elapsed since it has been transmitted. If the timer exceeds a threshold value (which depends on the system's RTT), the message is declared to be lost. A node that receives a NACK (or whose timer exceeds the threshold vale) enters a *backlog state*. Backlogged nodes retransmit the message for which they received the NACK in another frame, uniformly chosen at random among the next B frames. We call B the *maximum backlog time*. The process goes on until the message is acknowledged [99]. In both setups we assume a very large population of users. Furthermore, we assume that the average SNR is high enough so that the FER at the receiver is negligible.

In the first setup, in which no feedback is provided by the receiver, the average amount of energy spent by a node for each message which is correctly received does not change with the system load G, and is equal to the average number of times a message is repeated within a frame. In Fig. 6.13 the normalized throughput Φ is plotted against the normalized traffic load G. The normalized traffic load is the average rate at which the *new messages* (i.e., messages which are being transmitted for the first time) are injected in the network, and is independent from the number of times a message is repeated within a slot. In the figure the throughput curves of NCDP and CRDSA schemes for $d = 2$ and $d = 3$ replicas are shown. The throughput curve for NCDP in case of a constant retransmission probability $p = 0.0453$ is also shown. Note that this probability is above the threshold value we mentioned in Section 6.5, as for $S = 150$ we have $\log(S)/S = 0.0334$. The scheme with $p = 0.0453$ outperforms all the others in terms of throughput, achieving a peak value of more than 0.8. The precoding coefficients of NCDP are drown uniformly in $GF(2^8)$ in all but one case. The normalized throughput for NCDP with $d = 3$ and $GF(2)$ is also plotted in order to show the gain deriving from using an extended field. It can be seen from the figure that there is no gain with respect to CRDSA if a field of size 2 is used, while the gain is significant (about 15%) if a field of size 2^3 is used. Furthermore, is interesting to note how increasing the number of transmissions per message (and so the energy consumption) leads to an increase in the peak throughput of the system. However, Φ increases about 0.2 when passing from $d = 2$ to $d = 3$ repetitions, while the

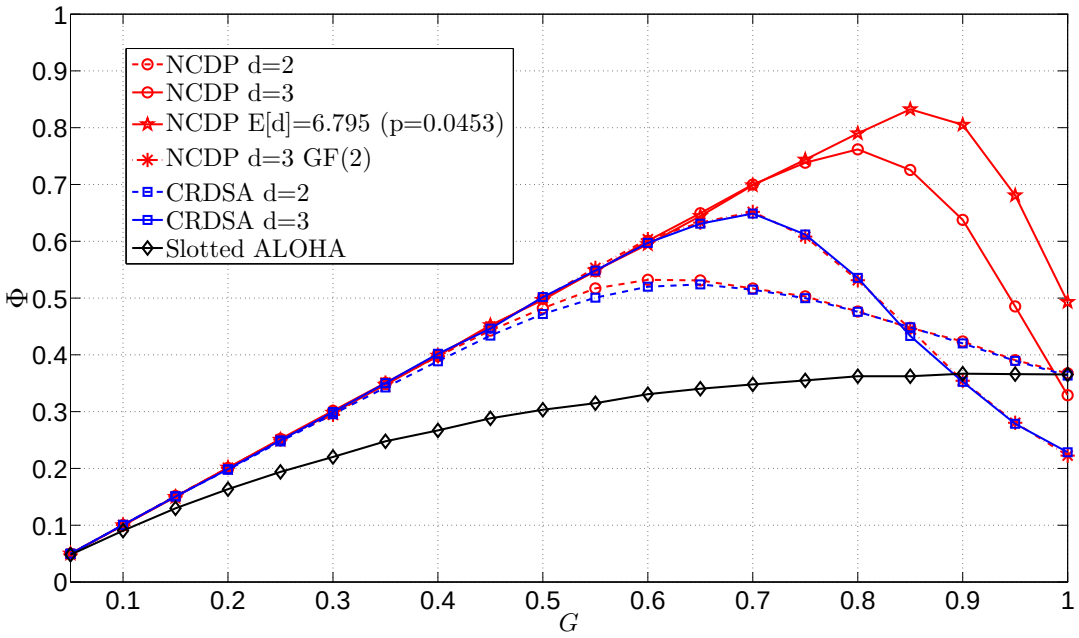


Figure 6.13: Normalized throughput Φ vs normalized traffic load G. The normalized traffic load is the average rate at which new messages are injected in the network, and is independent from the number of times a message is repeated within a slot. In the simulation the frame size has been set to $S = 150$ slots. The field size for the coefficients of NCDP has been set to 2^8 in all but one case (indicated in the legend) for which $GF(2)$ has been used. No feedback has been assumed from the receiver.

increase in the peak throughput is only about 0.05 when passing from $d = 2$ repetitions per message to an average of $E[d] = 6.795$ in case of a fixed transmission probability.

In the second setup, in which retransmissions are allowed, we evaluate jointly the spectral efficiency (average number of messages successfully received per slot) and the energy consumption (average number of transmissions needed for a message to be correctly received) of the schemes under study. In Fig. 6.14, Φ is plotted against G for a frame size $S = 150$ slots and a maximum backlog time $B = 50$ frames. The precoding coefficients of NCDP are drawn uniformly in $GF(2^8)$. The figure shows how Φ increases linearly with G up to a threshold load value. Such threshold increases with the (average) number of repetitions and corresponds to the maximum network load for which the throughput in the setup without feedback (Fig. 6.13) has a linear behavior, i.e., there are no losses in the system. This indicates that, if the load is such that a non negligible fraction of the messages are not decoded at the first attempt, the retransmissions saturate the channel, blocking both the iterative cancelation process of CRDSA and the Gaussian elimination decoding in NCDP. Note that this does not happen in the SA system,

coherently to what shown in [99] for the case of large backlog time. The Φ curve of NCDP bounds from above that of CRDSA. The reason for this lies in the way the decoding process is carried out by the receiver in NCDP. R first tries to decode the whole frame, which is feasible if the coefficient matrix $\mathbf{A}$ has rank N^{tx}. If the whole frame can not be decoded, then R applies Gaussian elimination on $\mathbf{A}$, in order to recover as many messages as possible. It can be easily verified that Gaussian elimination in NCDP is the equivalent, in a finite field, of the IIC process of CRDSA, which is applied in the analog domain. In order to compare jointly the spectral and

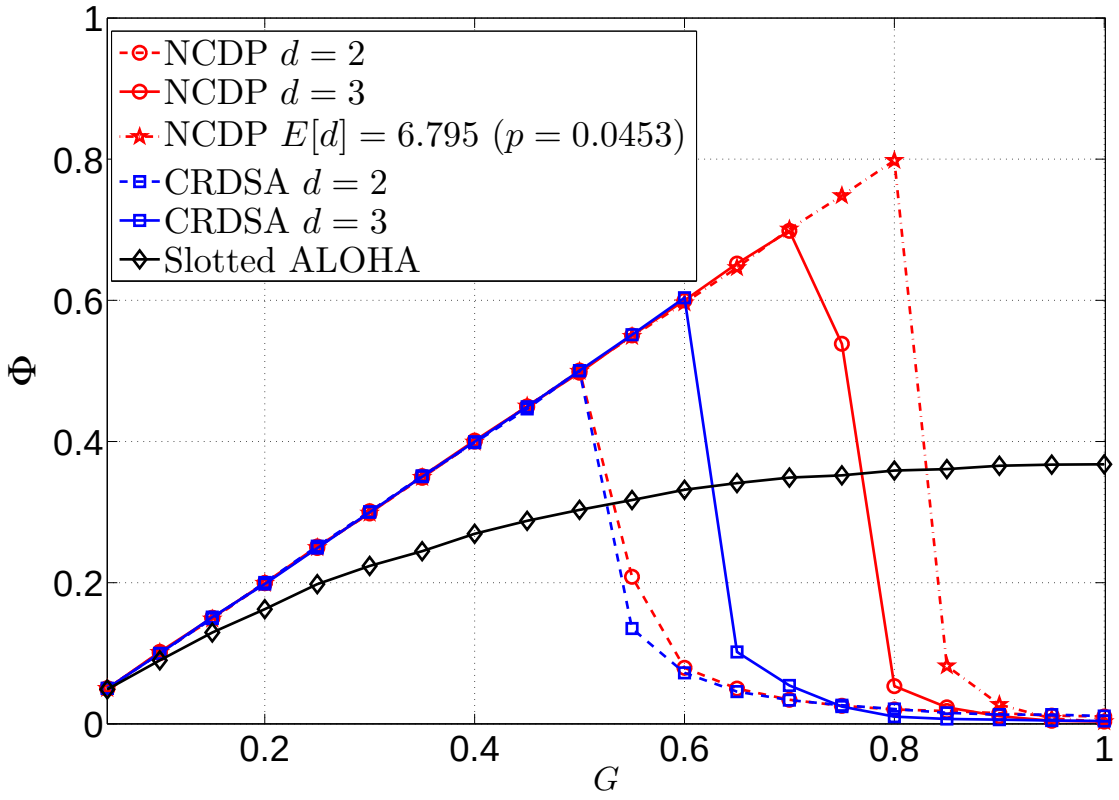


Figure 6.14: Normalized throughput Φ vs normalized traffic load G in a system with retransmission. In the simulation the frame size was set to $S = 150$ slots while the maximum backlog time was set to $B = 50$ frames. The field size for the coefficients of NCDP has been set to 2^8.

the energy efficiency of the different schemes, we plot the curves for the normalized throughput vs the average energy consumption per received message η, which is shown in Fig. 6.15. The increase in the number of repetitions corresponds to an increase in throughput but also to a higher energy consumption for a given transmitter in a given frame. However, as shown in Fig. 6.15, this does not necessarily imply a loss in energy efficiency. From the figure it can be seen that that there is no scheme that consistently outperforms the others in terms of both energy and

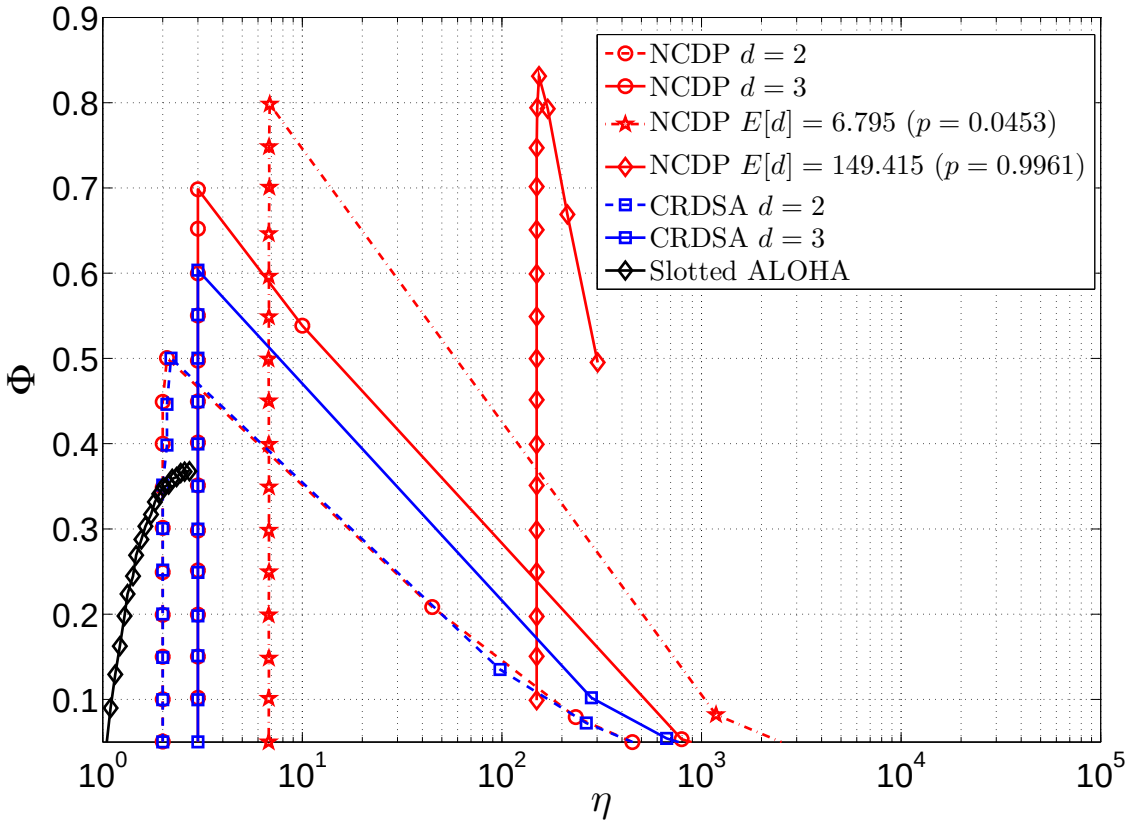


Figure 6.15: Normalized throughput vs average energy consumption per decoded message for $S = 150$ and $B = 50$ frames. The field size for the coefficients of NCDP has been set to 2^8.

spectral efficiency for any desired throughput. Determining which scheme is best depends on the maximum throughput we want to achieve. SA achieves a higher throughput with a lower energy consumption with respect to the other schemes in the region $\Phi < 0.35$, while in the region $\Phi > 0.35$ both NCDP and CRDSA achieve a higher throughput with lower energy consumption with respect to SA. NCDP and CRDSA behave almost in the same way in the case of 2 repetitions, achieving a maximum throughput of 0.5 for an average energy consumption of 2. In the case of 3 repetitions NCDP achieves a maximum Φ of 0.7, higher than CRDSA, for which the peak value is 0.6, for $\eta = 3$. In the NCDP scheme with a retransmission probability of $p = 0.0453$ a peak throughput of 0.8 is achieved in correspondence of an average energy consumption of $\eta = 6.795$. For comparison, we also show the throughput-energy curve for NCDP in case of $p = 0.9961$, i.e., the precoding coefficients are chosen uniformly in $GF(2^8)$. The high p leads to a high throughput, but also to a high energy consumption, with a minimum of $\eta = 149.415$. Moreover, we note that the gain with respect to the scheme with $p = 0.0453$

is negligible (about 5%), especially when compared to the energy saving of about 95% of this last one.

6.9 Conclusions

We have proposed a new collision recovery scheme for symbol-synchronous slotted ALOHA systems based on PHY layer NC over extended Galois fields. The extended Galois fields allows to efficiently exploit the diversity of the system, leading to an increased spectral efficiency and, depending on the system load, to an increased energy efficiency. We have compared the proposed scheme with two benchmarks in two different setups. One is a best-effort setup, in which the nodes do not receive any feedback from the receiver, while the other is a setup in which feedback is allowed from the receiver and an ARQ mechanism is assumed. In the second setup we have evaluated jointly the spectral efficiency and the energy consumption of the proposed scheme. Once the PHY layer NC is applied to decode the collided bursts, the receiver applies common matrix manipulation techniques over finite fields, which results in a high-throughput scheme. We showed that NCDP achieves a higher spectral efficiency with respect to the considered benchmarks, while there is not a single scheme that outperforms the others in terms of both energy and spectral efficiency, but the best scheme depends on the maximum achievable throughput. Rather than being an alternative to existing collision resolution methods based on interference cancelation, NCDP can complement such schemes in specific setups such as symbol synchronous systems with terminals having similar channel gains.

Furthermore, we carried out an analysis of several physical layer issues related to multi-user PHY NC. We extended the analysis on and proposed countermeasures against the effects of physical layer impairments on the FER when applying PHY NC for a generic number of colliding signals. In particular, we took into account frequency and phase offsets at the transmitters which, up to our knowledge, have been previously addressed only for the case of two colliding signals. Finally, we showed the feasibility of channel estimation for PHY NC in the presence of more than two colliding signals and studied the effect of non perfect symbol synchronism on the decoder FER, proposing four different methods to compensate for such effect. Up to our knowledge, this kind of analysis has been carried out only for the case of two

colliding signals and mainly in the context of two-way relay communication.

Chapter 7

Summary of Contributions and Future Work

"[...] e Atene hat sos sette sapientes
Cleobulo e Solone sos pius mannos
Chilone, Periandru cun Talete
Bìante cun Pittaccu e faghen sette."

Antonio Pazzola

In the present dissertation we considered several communication issues related to the forward and the reverse link of satellite networks, we proposed countermeasures to channel impairments and analyzed their performance either analytically, semi-analytically or numerically. The proposed solutions were mainly inspired by the principle of network coding, according to which encoding at packet level is allowed in the network. The techniques we proposed stem from a theoretical analysis of the problems and a successive adaptation to practical scenarios.

In the following we provide a summary of the main contributions for each of the chapters.

In **Chapter 3** we addressed the problem of missing coverage in urban environment for heterogeneous LMS networks, adopting a cooperative approach. We carried out an analytical study considering a mathematically tractable and yet practically interesting network model, in which fading and shadowing effects in the communication channels as well as the medium access mechanism of the cooperating nodes have been taken into account. By applying the Max-flow Min-cut theorem we derived an analytical lower bound on the achievable coverage as a function of both the information rate at physical layer and the rate of innovative packets

injected in the network per unit-time. Our results give a tradeoff between the coverage and the rate at which the information can be injected in the network, and at the same time quantify the gain derived from node cooperation through the short range interface. We showed that the diversity gain grows with the number of terminals, which indicates that important improvements in terms of transmission rate at the satellite can be achieved through cooperation.

Starting from the considered theoretical model, a practical cooperative scheme based on network coding has been proposed for enhancing coverage in heterogeneous satellite vehicular LMS systems over DVB-SH. Our numerical results, based on physical layer abstraction, showed that the proposed cooperative system can bring important benefits in terms of coverage with respect to a relaying scheme in which network coding is not used.

In **Chapter 4** and **Chapter 5** we considered the problem of video streaming over satellite. Two different streaming typologies were considered, namely real-time and non-real-time streaming, having different delay constraint. In the first setup we consider a transmitter streaming data to a receiver over a block fading channel, such that the transmitter is provided with an independent message at a fixed rate at the beginning of each channel block. We have used the average throughput as our performance metric. We have proposed several new transmission schemes based on joint encoding, time-division and superposition encoding. A general upper bound on the average throughput has also been introduced assuming the availability of CSI at the transmitter.

We have proved analytically that the aJE scheme is asymptotically optimal as the number of channel blocks goes to infinity, even though data arrives gradually over time at a fixed rate, rather than being available initially as usually assumed for the achievability of the ergodic capacity. We have also shown numerically that, even for a finite number of messages, the aJE scheme outperforms other schemes in all the considered settings and performs close to the upper bound.

The schemes based on the joint encoding of the messages (JE and aJE) create an M-block long concatenated code, such that either all or none of the messages can be decoded. This is useful when the underlying application has a minimum rate requirement that needs to be satisfied over M channel blocks, or when the average SNRs of the users vary over a limited range of SNR values. Independent encoding made in time sharing-based schemes (TS, gTS, MT),

instead, makes each message less strictly dependent on the decoding of the others, allowing some of them to be decoded also at low average SNR but, on the other hand, implying the possibility not to decode some of them in when the average SNR is high. The ST scheme, based on message superposition, collocates itself between JE- and TS-based schemes, as messages are encoded independently, but the probability of correctly decoding each one of them is affected by the decoding of the others.

The aJE scheme has proved to be advantageous in systems with a single receiver or with multiple receivers having similar average SNR values, as the performance of the user with the highest average SNR is limited by the user with the lowest average SNR. On contrast, the gTS and ST schemes can be attractive when broadcasting to multiple users with a wide range of SNR values, or in a point-to-point system with inaccurate CSI, as their average throughputs decrease gradually with decreasing SNR.

The setup studied in Chapter 5 models the problem of non-real-time video streaming over slow fading channels with per-packet delay constraints. We have proposed four different transmission schemes based on time-sharing. We have carried out a theoretical as well as a numerical analysis of the average throughput and the maximum delay of the proposed schemes. We have also derived bounds on both the throughput and maximum delay by introducing an informed transmitter scheme, in which the transmitter is assumed to know the channel states in advance. We have seen that the wTS scheme can provide a good trade-off between the throughput and the maximum delay by deciding on the proportion of transmitted video packets. In practice this corresponds to reducing the coding rate of the video at the packet level. In our future work we will consider more advanced transmission techniques such us superposition encoding together with buffering.

Finally, in **Chapter 6** we considered the problem of collisions in the return link. We proposed a collision recovery scheme for symbol-synchronous slotted ALOHA systems based on PHY layer NC over extended Galois fields. The extended Galois fields allows to efficiently exploit the diversity of the system, leading to an increased spectral efficiency and, depending on the system load, to an increased energy efficiency with respect to other systems. We have compared the proposed scheme with two benchmarks in two different setups. One is a best-effort setup, in which the nodes do not receive any feedback from the receiver, while the other is

a setup in which feedback is allowed from the receiver and an ARQ mechanism is assumed. In the second setup we have evaluated jointly the spectral efficiency and the energy consumption of the proposed scheme. Once the PHY layer NC is applied to decode the collided bursts, the receiver applies common matrix manipulation techniques over finite fields, which results in a high-throughput scheme. We showed that NCDP achieves a higher spectral efficiency with respect to the considered benchmarks, while there is not a single scheme that outperforms the others in terms of both energy and spectral efficiency, but the best scheme depends on the maximum achievable throughput. Rather than being an alternative to existing collision resolution methods based on interference cancelation, NCDP can complement such schemes in specific setups such as symbol synchronous systems with terminals having similar channel gains.

Furthermore, we carried out an analysis of several physical layer issues related to multi-user PHY NC. We extended the analysis on and proposed countermeasures against the effects of physical layer impairments on the FER when applying PHY NC for a generic number of colliding signals. In particular, we took into account frequency and phase offsets at the transmitters which, up to our knowledge, have been previously addressed only for the case of two colliding signals. Finally, we showed the feasibility of channel estimation for PHY NC in the presence of more than two colliding signals and studied the effect of non perfect symbol synchronism on the decoder FER, proposing four different methods to compensate for such effect. Up to our knowledge, this kind of analysis has been carried out only for the case of two colliding signals and mainly in the context of two-way relay communication.

As future work we plan to enhance the channel estimation, so that channel codes with longer codewords can be adopted, with consequent gain in terms of spectral efficiency. We also plan to consider Nested Lattice Codes [100], for which promising theoretical result were obtained in both the point to point and the multiple access scenario.

Bibliography

[1] R. Ahlswede, C. Ning, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. on Info. Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[2] A. R. Lehman and E. Lehman, "Complexity classification of network information flow problems," in *ACM-SIAM Symp. on Discrete Algorithms*. Philadelphia, PA, U.S.A.: Society for Industrial and Applied Mathematics, 2004, pp. 142–150.

[3] A. Sprintson, *Network Coding and its Applications in Communication Networks*, Texas A&M University, College Station, TX, U.S.A., http://cegroup.ece.tamu.edu/spalex/files/tutorial.pdf.

[4] R. Koetter and M. Medard, "An algebraic approach to network coding," in *IEEE Int'l Symp. on Info. Theory*, Washington, DC, U.S.A., June 2001.

[5] C. Fragouli and E. Soljanin, "Network coding fundamentals," *Foundations and Trends in Networking*, vol. 2, no. 1, pp. 1–133, 2007.

[6] T. Ho, "Networking from a network coding perspective," Ph.D. dissertation, Massachussetts Institute of Technology, http://www.its.caltech.edu/ tho/ho-thesis.pdf, 2004.

[7] T. Ho, R. Koetter, M. Médard, D. R. Karger, and M. Effros, "The benefits of coding over routing in a randomized setting," in *Proc. 2003 IEEE Int'l Symp. on Info. Theory*, Yokohama, Japan, June-July 2003.

[8] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *IEEE Allerton Conf. on Communication, Control, and Computing*, Urbana-Champaign, IL, U.S.A., Oct. 2003.

[9] F. Vieira and J. Barros, "Network coding multicast in satellite networks," in *Next Generation Internet Networks, 2009*, Aveiro, Portugal, July 2009.

[10] F. Vieira, M. A. V. Castro, and J. Lei, "Datacast transmission architecture for DVB-S2 systems in railway scenarios," in *IEEE Int'l Workshop on Signal Processing for Space Comm.*, Rhodes Island, Greece, Oct. 2008.

[11] S. Katti, H. Rahul, W. Hu, D. Katabi, M. Médard, and J. Crowcroft, "Xors in the air: Practical wireless network coding," *IEEE/ACM Trans. on Networking*, vol. 16, no. 3, pp. 497–510, June 2008.

[12] S. E. Rouayheb, A. Sprintson, and C. Georghiades, "On the relation between the index coding and the network coding problems," in *IEEE Int'l Symp. on Info. Theory,*, Toronto, Canada, July 2008.

[13] M. L. A. Sprintson, "On the hardness of approximating the network coding capacity," in *IEEE Int'l Symp. on Info. Theory*, Toronto, Canada, July 2008.

[14] A. Agarwal and M. Charikar, "On the advantage of network coding for improving network throughput," in *IEEE Info. Theory Workshop*, Oct. 2004.

[15] T. Ho and D. S. Lun, *Network Coding: an introduction*. Cambridge University Press, 2008.

[16] S. Zhang, S. Liew, and P. Lam, "Physical layer network coding," in *ACM/IEEE International Conference on Mobile Computing and Networking*, Los Angeles, CA, U.S.A., Sep. 2006.

[17] F. Rossetto and M. Zorzi, "On the design of practical asynchronous physical layer network coding," in *IEEE Workshop on Signal Proc. Advances in Wireless Comm.*, Perugia, Italy, June 2009.

[18] R. H. Y. Louie, Y. Li, and B. Vucetic, "Practical physical layer network coding for two-way relay channels: Performance analysis and comparison," *IEEE Trans. on Wireless Comm.*, vol. 9, no. 2, pp. 764–777, Feb. 2010.

[19] J. H. Sorensen, R. Krigslund, P. Popovski, T. Akino, and T. Larsen, "Physical layer network coding for FSK systems," *IEEE Comm. Letters*, vol. 13, no. 8, Aug. 2009.

[20] F. Rossetto, "A comparison of different physical layer network coding techniques for the satellite environment," in *IEEE Advanced Satellite Multimedia Systems Conf.*, Cagliari, Italy, Sep. 2010.

[21] M. Durvy, C. Fragouli, and P. Thiran, "Towards reliable broadcasting using ACKs," in *IEEE Int'l Symp. on Info. Theo.*, Nice, France, June 2007.

[22] C. H. Foh, J. Cai, and J. Qureshi, "Collision codes: Decoding superimposed BPSK modulated wireless transmissions," in *IEEE Consumer Comm. and Networking Conf.*, Las Vegas, NV, U.S.A., Jan. 2010.

[23] B. Nazer and M. Gastpar, "Reliable physical layer network coding," *Proceedings of the IEEE*, vol. 99, no. 3, pp. 438–460, Mar. 2011.

[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., J. W. . Sons, Ed., 2006.

[25] European Telecommunications Standards Institute, "ETSI TS 102 584 V1.2.1, Digital Video Broadcasting (DVB); DVB-SH Implementation Guidelines Issue 2." Jan. 2011.

[26] ——, "ETSI EN 301 790 V1.5.1, Digital Video Broadcasting (DVB); Interaction channel for satellite distribution systems." May 2009.

[27] Exalted Project, "First report on LTE-M algorithms and procedures," http://www.ict-exalted.eu, Aug. 2011.

[28] D. S. Lun, M. Médard, R. Koetter, and M. Effros, "On coding for reliable communication over packet networks," *Physical Comm.*, vol. 1, no. 1, pp. 3–20, 2008.

[29] Y. Tseng, S. Ni, Y. Chen, and J. Sheu, "The broadcast storm problem in a mobile ad hoc network," *Wireless Networks*, vol. 8, pp. 153–167, 2002.

[30] J. Wu and F. Dai, "A generic distributed broadcast scheme in ad hoc wireless networks," *IEEE Trans. on Computers*, vol. 53, no. 10, pp. 1343–1354, Oct. 2004.

[31] A. Vanelli-Coralli, G. E. Corazza, G. K. Karagiannidis, P. T. Mathiopoulos, D. S. Michalopoulos, C. Mosquera, S. Papaharalabos, and S. Scalise, "Satellite communications: Research trends and open issues," in *Int'l Workshop on Satellite and Space Comm.*, Toulouse, France, Sept. 2007.

[32] S. Morosi, E. D. Re, S. Jayousi, and R. Suffritti, "Hybrid satellite/terrestrial cooperative relaying strategies for DVB-SH based communication systems," in *European Wireless Conf.*, May 2009.

[33] G. Cocco, C. Ibars, and O. D. R. Herrero, "Cooperative satellite to land mobile gap-filler-less interactive system architecture," in *IEEE Advanced Satellite Mobile Systems Conf.*, Cagliari, Italy, Sep. 2010.

[34] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. on Info. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.

[35] M. Sardari, F. Hendessi, and F. Fekri, "Infocast: A new paradigm for collaborative content distribution from roadside units to vehicular networks," in *Annual IEEE Comm. Society Conf. on Sensor, Mesh and Ad Hoc Comm. and Networks*, Rome, Italy, June 2009.

[36] P. Cataldi, A. Tomatis, G. Grilli, and M. Gerla, "A novel data dissemination method for vehicular networks with rateless codes," in *IEEE Wireless Comm. and Networking Conf. (WCNC)*, Budapest, Hungary, Apr. 2009.

[37] L. Wan, S. Tsai, and M. Almgren, "A fading-insensitive performance metric for a unified link quality model," in *IEEE Wireless Comm. and Networking Conf.*, vol. 4, Las Vegas, NV, U.S.A., Apr. 2006.

[38] Sony, Intel, "TGn sync proposal MAC simulation methodology," IEEE 21 802.11-04/895r2, Nov. 2004.

[39] "Draft IEEE 802.16m evaluation methodology." http://www.ieee802.org/16/tgm/docs/80216m-07_037r2.pdf: Institute of Electrical and Electronics Engineers (IEEE), Dec. 2007.

[40] *ANSI/IEEE Std 802.11, 1999 Edition (R2003)*, Institute of Electrical and Electronics Engineers (IEEE), http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=9543, 1999.

[41] S. Deb, M. Effros, T. Ho, D. R. Karger, R. Koetter, D. S. Lun, M. Médard, and N. Ratnakar, "Network coding for wireless applications: A brief tutorial," in *IEEE Int'l Workshop on Wireless Ad-hoc Networks*, London, UK, May 2005.

[42] H. Suzuki, "A statistical model for urban radio propogation," *IEEE Trans. on Comm.*, vol. 25, no. 7, pp. 673–680, July 1977.

[43] E. Lutz, D. Cygan, M. Dippold, F. Dolainsky, and W. Papke, "The land mobile satellite communication channel-recording, statistics, and channel model," *IEEE Trans. on Vehicular Technology*, vol. 40, no. 2, pp. 375–386, May 1991.

[44] F. Perez-Fontan, M. A. Vazquez-Castro, S. Buonomo, J. P. Poiares-Baptista, and B. Arbesser-Rastburg, "S-band LMS propagation channel behaviour for different environments, degrees of shadowing and elevation angles," *IEEE Trans. on Broadcasting*, vol. 44, no. 1, pp. 40–76, Mar. 1998.

[45] "Digital Video Broadcasting (DVB); Upper Layer Forward Error Correction in DVB. DVB Document A148," http://www.dvb.org/, Mar. 2010.

[46] European Telecommunications Standards Institute, "ETSI TS 102 472 V1.1.1, Digital Video Broadcasting (DVB); IP Datacast over DVB-H: Content Delivery Protocols," June 2006.

[47] ——, "DVB-SH implementation guidelines, DVB BlueBook A120," http://www.dvb-h.org/, May 2008.

[48] I. Sen and D. W. Matolak, "Vehicle-vehicle channel models for the 5-GHz band," *IEEE Trans. on Intelligent Transportation Systems*, vol. 9, no. 2, pp. 235–245, June 2008.

[49] G. Acosta-Marum and M. A. Ingram, "Six Time- and Frequency-Selective Empirical Channel Models for Vehicular Wireless LANs," in *IEEE Vehicular Technology Conf.*, Baltimore, MD , U.S.A., Sept. 2007.

[50] G. Acosta, K. Tokuda, and M. A. Ingram, "Measured joint Doppler-delay power profiles for vehicle-to-vehicle communications at 2.4 GHz," in *IEEE Global Telecomm. Conf.*, Dallas, TX, U.S.A., Nov. 2004.

[51] L. Cheng, B. E. Henty, , D. D. Stancil, F. Bai, and P. Mudalige, "Mobile vehicle-to-vehicle narrow-band channel measurement and characterization of the 5.9 GHz dedicated short range communication (DSRC) frequency band," *IEEE Journal on Selected Areas in Comm.*, vol. 25, no. 8, pp. 1501–1516, Oct. 2007.

[52] Y. C. Cheng, J. Bellardo, P. Benkö, A. C. Snoeren, G. M. Voelker, and S. Savage, "Jigsaw: Solving the puzzle of enterprise 802.11 analysis," in *Special Interest Group on Data Comm. Conf.*, Pisa, Italy, Sep. 2006.

[53] J. Lei, M. A. Vazquez-Castro, and T. Stockhammer, "Link-layer FEC and cross-layer architecture for DVB-S2 transmission with QoS in railway scenarios," *IEEE Trans. on Vehicular Technology*, vol. 58, no. 8, pp. 4265–4276, Oct. 2009.

[54] European Telecommunications Standards Institute, "ETSI EN 302 583 V 1.2.1, Digital Video Broadcasting (DVB); Framing Structure, channel coding and modulation for Satellite Services to Handheld devices (SH) below 3 GHz." Ago. 2008.

[55]

[56] D. Tse, R. Yates, and L. Zang, "Fading broadcast channels with state information at the receivers," in *IEEE Allerton Conf. on Comm., Control, and Computing*, Monticello, IL, U.S.A., Sep. 2008.

[57] European Telecommunications Standards Institute, "ETSI TR 102 376 V1.1.1 (2005-02) Digital Video Broadcasting (DVB) user guidelines for the second generation system for broadcasting, interactive services, news gathering and other broadband satellite applications (DVB-S2)," Feb. 2005.

[58] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. on Info. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[59] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[60] A. Goldsmith, *Wireless Communication*. Cambridge University Press, 2005.

[61] A. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. on Info. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[62] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Comm.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[63] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. on Info. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[64] A. S. Avestimehr and D. N. C. Tse, "Outage capacity of the fading relay channel in the low snr regime," *IEEE Trans. on Info. Theory*, vol. 53, no. 4, pp. 1401–1415, Apr. 2007.

[65] S. Shamai, "A broadcast strategy for the Gaussian slowly fading channel," in *IEEE Int'l Symp. on Info. Theory*, Ulm, Germany, June-July 1997.

[66] S. Shamai and A. Steiner, "A broadcast approach for a single user slowly fading mimo channel," *IEEE Trans. on Info. Theory*, vol. 49, no. 10, pp. 2617–2635, Oct. 2003.

[67] D. Gündüz and E. Erkip, "Opportunistic cooperation by dynamic resource allocation," *IEEE Trans. on Wireless Comm.*, vol. 6, no. 4, pp. 1446–1454, Apr. 2007.

[68] S. V. Hanly and D. N. C. Tse, "Multiaccess fading channels. II. Delay-limited capacities," *IEEE Trans. on Info. Theory*, vol. 44, no. 7, pp. 2816–2831, Nov. 1998.

[69] O. Oyman and S. Singh, "Quality of experience for HTTP adaptive streaming services," *IEEE Comm. Magazine*, vol. 50, no. 4, pp. 20–27, Apr. 2012.

[70] M. C. O. Bogino, P. Cataldi, M. Grangetto, E. Magli, and G. Olmo, "Sliding-window digital fountain codes for streaming multimedia contents," in *Int'l Symp. on Circuits Systems*, New Orleans, LA, U.S.A., May 2007.

[71] A. Badr, A. Khisti, and E. Martinian, "Diversity embedded streaming erasure codes (DE-SCo): Constructions and optimalty," in *IEEE Global Telecomm. Conf.*, Miami, FL, U.S.A., Dec. 2010.

[72] D. Leong and T. Ho, "Erasure coding for real-time streaming," in *IEEE Int'l Symp. Info. Theory (ISIT)*, Boston, MA, U.S.A., July 2012.

[73] A. Khisti and S. C. Draper, "Streaming data over fading wireless channels: The diversity-multiplexing tradeoff," in *IEEE Int'l Symp. on Info. Theory*, St. Petersburg, Russia, Aug. 2011.

[74] G. Cocco, D. Gunduz, and C. Ibars, "Real-time broadcasting over block-fading channels," in *Int'l Symp. on Wireless Comm. Systems*, Aachen, Germany, Nov. 2011.

[75] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. on Vehicular Technology*, vol. 43, no. 2, pp. 359–378, May 1994.

[76] V. V. Prelov, "Transmission over a multiple access channel with a special source hierarchy," *Problemy Peredachi Informatsii*, vol. 20, no. 4, pp. 3–10, Oct.-Dec. 1984.

[77] G. Maral and M. Bousquet, *Satellite Communications Systems - systems, techniques and technologies*. John Wiley & Sons Ltd, 2009.

[78] C. T. K. Ng, D. Gunduz, A. J. Goldsmith, and E. Erkip, "Distortion minimization in gaussian layered broadcast coding with successive refinement," *IEEE Trans. on Info. Theory*, vol. 55, no. 11, pp. 5074–5086, Nov. 2009.

[79] D. Gunduz and E. Erkip, "Joint source-channel codes for MIMO block-fading channels," *IEEE Trans. on Info. Theory*, vol. 54, no. 1, pp. 116–134, Jan. 2008.

[80] Y. Wang, J. Ostermann, and Y.-Q. Zhang, *Video processing and communications*, A. V. Oppenheim, Ed. Prentice Hall, 2002.

[81] W. Zeng, C. T. K. Ng, and M. Medard, "Joint coding and scheduling optimization in wireless systems with varying delay sensitivities," in *IEEE Comm. Society Conf. on Sensor, Mesh and Ad Hoc Comm. and Networks (SECON)*, Seoul, Korea, June 2012.

[82] M. van der Schaar, S. Krishnamachari, C. Sunghyun, and X. Xiaofeng, "Adaptive cross-layer protection strategies for robust scalable video transmission over 802.11 WLANs," *IEEE Journal on Selected Areas in Comm.*, vol. 21, no. 10, pp. 1752–1763, Dec. 2003.

[83] T. H. Luan, L. X. Cai, and S. Xuemin, "Impact of network dynamics on user's video quality: Analytical framework and QoS provision," *IEEE Transactions on Multimedia*, vol. 12, no. 1, pp. 64–78, Jan. 2010.

[84] G. Choudhury and S. Rappaport, "Diversity ALOHA - a random access scheme for satellite communications," *IEEE Trans. on Comm.*, vol. 31, no. 3, pp. 450–457, Mar. 1983.

[85] M. K. Tsatsanis, R. Zhang, and S. Banerjee, "Network-assisted diversity for random access wireless networks," *IEEE Trans. on Signal Processing*, vol. 48, no. 3, pp. 702–711, Mar 2000.

[86] E. Casini, R. De Gaudenzi, and O. d. R. Herrero, "Contention resolution diversity slotted ALOHA (CRDSA): An enhanced random access scheme for satellite access packet networks," *IEEE Trans. on Wireless Comm.*, vol. 6, no. 4, pp. 1408–1419, Apr. 2007.

[87] H.-C. Bui, J. Lacan, and M.-L. Boucheret, "An enhanced multiple random access scheme for satellite communications," in *IEEE Wireless Telecomm. Symp.*, London, U.K., Apr. 2012.

[88] G. Liva, "Graph-based analysis and optimization of contention resolution diversity slotted ALOHA," *IEEE Trans. on Comm.*, vol. 59, no. 2, pp. 477–487, Feb. 2011.

[89] A. Argyriou and A. Pandharipande, "Cooperative protocol for analog network coding in distributed wireless networks," *IEEE Trans. on Wireless Comm.*, vol. 9, no. 10, pp. 3112–3119, Oct. 2010.

[90] G. Cocco, C. Ibars, D. Gündüz, and O. d. R. Herrero, "Collision resolution in slotted ALOHA with multi-user physical-layer network coding," in *IEEE Vehicular Technology Conf. (VTC Spring)*, Budapest, Hungary, May 2011.

[91] D. Maduike, H. D. Pfister, and A. Sprintson, "Design and implementation of physical-layer network-coding protocols," in *IEEE Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, U.S.A., Nov. 2009.

[92] L. Lu and S. C. Liew, "Asynchronous physical-layer network coding," *IEEE Trans. on Wireless Comm.*, vol. 11, no. 2, pp. 819–831, Feb. 2012.

[93] M. Jain, S. L. Miller, and A. Sprintson, "Parameter estimation and tracking in physical layer network coding," in *IEEE Global Telecomm. Conf.*, Houston, TX, U.S.A., Dec. 2011.

[94] O. Trullols-Cruces, J. M. Barcelo-Ordinas, and M. Fiore, "Exact decoding probability under random linear network-coding," *IEEE Comm. Letters*, vol. 15, no. 1, pp. 67–69, Jan. 2011.

[95] J. Blömer, R. Karp, and E. Welzl, "The rank of sparce random matrices over finite fiels," *Random Structures and Algorithms*, vol. 10, no. 4, pp. 407–419, July 1997.

[96] G. Cocco, N. Alagha, C. Ibars, and S. Cioni, "Practical issues in multi-user physical layer network coding," in *IEEE Advanced Satellite Mobile Systems Conf.*, Baiona (Spain), Sep. 2012.

[97] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 36, no. 4, pp. 477–489, Apr. 1988.

[98] R. D. J. V. Nee, "Timing aspects of synchronous CDMA," in *IEEE Int'l Symp. on Personal, Indoor and Mobile Radio Comm.*, The Hague, The Netherlands, Sep. 1994.

[99] L. Kleinrock and S. Lam, "Packet switching in a multiaccess broadcast channel: Performance evaluation," *IEEE Trans. on Comm.*, vol. 23, no. 4, pp. 410–423, Apr. 1975.

[100] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. on Info. Theory*, vol. 48, no. 6, pp. 1250–1276, June 2002.