DOCTORAL THESIS

Radio resource allocation algorithms for multicast OFDM systems

Alejandro de la Fuente Iglesias

Advisors:

Prof. Ana García Armada
Dr. Raquel Pérez Leal

uc3m | Universidad Carlos III de Madrid

Department of Signal Theory and Communications

November 2017

uc3m | Universidad Carlos III de Madrid

TESIS DOCTORAL

Radio resource allocation algorithms for multicast OFDM systems

Autor: Alejandro de la Fuente Iglesias

Directora: Prof.ª Ana García Armada
Directora: Dra. Raquel Pérez Leal

Firma del tribunal calificador

Presidente:	Dr. Atilio Manuel Da Silva Gameiro
Vocal:	Dra. María de Diego Antón
Secretario:	Dr. Víctor Pedro Gil Jiménez

Calificación:

Leganés, 6 de Noviembre de 2017

Preface

Imagine there's no heaven
It's easy if you try
No hell below us
Above us only sky
Imagine all the people
Living for today...

Imagine there's no countries
It isn't hard to do
Nothing to kill or die for
And no religion, too
Imagine all the people
Living life in peace...

You may say I'm a dreamer
But I'm not the only one
I hope someday you'll join us
And the world will be as one

Imagine no possessions
I wonder if you can
No need for greed or hunger
A brotherhood of man
Imagine all the people
Sharing all the world...

You may say I'm a dreamer
But I'm not the only one
I hope someday you'll join us
And the world will live as one

John Lennon

Six years ago, I imagined a change in my life.
Of course, I started a way towards my own life.
And now here we are in the half-way of a wonderful project.
In the half-way of something, as the poet said:
"*haciendo camino al andar*"

Why not?
Let's imagine that *the world will live as one*

Acknowledgement

No podría comenzar esta tesis de ninguna otra manera que agradeciéndoles todo a mis padres, su amor y apoyo incondicional que me ha acompañado y me acompaña en cada paso y cada reto de mi vida.

A Sonia, que apareciste en mi vida hace menos de dos años y te has convertido en mi apoyo, mi cómplice, mi motivación, ... en mi amor.

A los pequeños, Pablo e Isa, Hugo y Mario, porque nada resulta tan gratificante como verles disfrutar, acompañarles a crecer y quizás, solo quizás, ser una buena referencia a quien seguir.

A todos los que habéis compartido estos años de vuelta a la escuela, estos maravillosos años que me han permitido crecer, reinventarme, conoceros, disfrutaros y aprender, sin duda seguir aprendiendo. Los compañeros que ya estábais cuando empecé como Juanjo (¡qué gran maestro!), Máximo, Javier o Cecilia, que hicisteis tan fáciles, agradables y divertidos mis comienzos. También a los que han ido apareciendo y hemos compartido aficiones, momentos de diversión e incluso alguno de estrés, como Álex, Borja, Juan Carlos, Alberto ... Y por supuesto, a muchos que os conocí y disfruté fuera del laboratorio (que también hay vida más allá). A Fer con sus clases de tenis y sus maratones, a Gema con su piscina y sus clases de yoga (¡cómo os voy a echar de menos!). A Estefa, Ana, Héctor ... porque lo próximo será paracaidismo en Guatemala, Nueva Zelanda o Bagan?

Mis dos directoras de tesis Ana y Raquel me han dado la oportunidad de vivir la experiencia que han significado estos cuatro años de tesis. Gracias por vuestro apoyo y ayuda en el desarrollo de esta investigación. Y sin duda, por mantener siempre una actitud positiva y aguantar esos momentos de frustración que (supongo que no soy el único) surgen a lo largo de una tesis.

Un agradecimiento especial tengo que dedicar a Joaquín, por su enorme ayuda y sus consejos en el trabajo que hemos realizado juntos. Espero que pronto se pueda repetir y poder seguir aprendiendo a tu lado.

No solo en la UC3M hay personas a las que compartir y agradecer esta tesis. Mi colaboración y estancia en la UIB con Guillem y Felip ha sido verdaderamente especial. Trabajar con vosotros ha sido un placer y un enorme aprendizaje. Os estoy muy agredecido y espero que muy pronto continuemos compartiendo nuevas investigaciones.

Furthermore, I have been able to enjoy two great experiences in international research centers. First, I would like to thank people in Eurecom, Navid, Florian, Raymond. I spent a fantastic time there with my Greek-French colleges. After that, I spent three exciting months in London Kings College. I would like to thank Mischa, Oliver and Maria their hospitality and the great experience of researching at Kings and enjoying London.

A todos aquellos que estáis cerca, mi hermana Virgi y Charly, a mis *botones coloraos*, a mis ex Ericsson y a todos aquellos que injustamente me dejo por el olvido que imponen el tiempo y la distancia, y que sin duda sabrán que unas cañas siempre lo acercan todo.

The cover and back cover are designed by freepik.

"As we express our gratitude,
we must never forget that the highest
appreciation is not to utter words,
but to live by them."

John F. Kennedy

Abstract

Video services have become highly demanded in mobile networks leading to an unprecedented traffic growth. It is expected that traffic from wireless and mobile devices will account for nearly 70 percent of total IP traffic by the year 2020, and the video services will account for nearly 75 percent of mobile data traffic by 2022. Multicast transmission is one of the key enablers towards a more spectral and energy efficient distribution of multimedia content in current and envisaged mobile networks. It is worth noting that multicast is a mechanism that efficiently delivers the same content to many users, not only focusing on video broadcasting, but also distributing many other media, such as software updates, weather forecast or breaking news.

Although multicast services are available in Long Term Evolution (LTE) and LTE-Advanced (LTE-A) networks, new improvements are needed in some areas to handle the demands expected in the near future. Resource allocation techniques for multicast services are one of the main challenging issues, since it is required the development of novel schemes to meet the demands of their evolution towards the next generation. Most multicast techniques adopt rather conservative strategies that select a very robust modulation and coding scheme (MCS), whose characteristics are determined by the propagation conditions experienced by the worst user in the group in order to ensure that all users in a multicast group are able to correctly decode the received data. Obviously, this robustness comes at the prize of a low spectral efficiency.

This thesis presents an exhaustive study of broadcast/multicast technology for current mobile networks, especially focusing on the scheduling and resource allocation (SRA) strategies to maximize the potential benefits that multicast transmissions imply on the spectral efficiency. Based on that issue, some contributions have been made to the state of the art in the radio resource management (RRM) for current and beyond mobile multicast services.

- In the frame of LTE/LTE-A, the evolved multimedia broadcast and multicast service (eMBMS) shares the physical layer resources with the unicast transmission mode (at least up to Release 12). Consequently, the time allocation to multicast transmission is limited to a maximum of a 60 percent, and the remaining subframes (at least 40 percent) are reserved for unicast transmissions. With the aim of achieving the maximum aggregated data rate (ADR) among the multicast users, we have implemented several innovative SRA schemes that combine the allocation of multicast and unicast resources in the LTE/LTE-A frame, guaranteeing the prescribed quality of service (QoS) requirements for every user.

- In the specific context of wideband communication systems, the selection of the multicast MCS has often relied on the use of wideband channel quality indicators (CQIs), providing rather imprecise information regarding the potential capacity of the multicast channel. Only recently has the per-subband CQI been used to improve the spectral efficiency of the system without compromising the link robustness. We have proposed novel subband CQI-based multicast SRA strategies that, relying on the selection of more spectrally efficient transmission modes, lead to increased data rates while still being able to fulfill prescribed QoS metrics.

- Mobile broadcast/multicast video services require effective and low-complexity SRA strategies. We have proposed an SRA strategy based on multicast subgrouping and the scalable video coding (SVC) technique for multicast video delivery. This scheme focuses on reducing the search space of solutions and optimizes the ADR. The results in terms of ADR, spectral efficiency, and fairness among multicast users, along with the low complexity of the algorithm, show that this new scheme is adequate for real systems.

These contributions are intended to serve as a reference that motivate ongoing and future investigation in the challenging field of RRM for broadcast/multicast services in next generation mobile networks.

Resumen

La demanda de servicios de vídeo en las redes móviles ha sufrido un incremento exponencial en los últimos años, lo que a su vez ha desembocado en un aumento sin precedentes del tráfico de datos. Se espera que antes del año 2020, el tráfico debido a dispositivos móviles alcance cerca del 70 por ciento del tráfico IP total, mientras que se prevé que los servicios de vídeo sean prácticamente el 75 por ciento del tráfico de datos en las redes móviles hacia el 2022. Las transmisiones *multicast* son una de las tecnologías clave para conseguir una distribución más eficiente, tanto espectral como energéticamente, del contenido multimedia en las redes móviles actuales y futuras. Merece la pena reseñar que el *multicast* es un mecanismo de entrega del mismo contenido a muchos usuarios, que no se enfoca exclusivamente en la distribución de vídeo, sino que también permite la distribución de otros muchos contenidos, como actualizaciones *software*, información meteorológica o noticias de última hora.

A pesar de que los servicios *multicast* ya se encuentran disponibles en las redes *Long Term Evolution* (LTE) y *LTE-Advanced* (LTE-A), la mejora en algunos ámbitos resulta necesaria para manejar las demandas que se prevén a corto plazo. Las técnicas de asignación de recursos para los servicios *multicast* suponen uno de los mayores desafíos, ya que es necesario el desarrollo de nuevos esquemas que nos permitan acometer las exigencias que supone su evolución hacia la próxima generación. La mayor parte de las técnicas *multicast* adoptan estrategias conservadoras, seleccionando esquemas de modulación y codificación (MCS) impuestos por las condiciones

de propagación que experimenta el usuario del grupo con peor canal, para así asegurar que todos los usuarios pertenecientes al grupo *multicast* sean capaces de decodificar correctamente los datos recibidos. Como resulta obvio, la utilización de esquemas tan robustos conlleva el precio de sufrir una baja eficiencia espectral.

Esta tesis presenta un exhaustivo estudio de la tecnología *broadcast/multicast* para las redes móviles actuales, que se centra especialmente en las estrategias de asignación de recursos (SRA), cuyo objetivo es maximizar los beneficios que la utilización de transmisiones *multicast* potencialmente implica en términos de eficiencia espectral. A partir de dicho estudio, hemos realizado varias contribuciones al estado del arte en el ámbito de la gestión de recursos radio (RRM) para los servicios *multicast*, aplicables en las redes móviles actuales y futuras.

- En el marco de LTE/LTE-A, el eMBMS comparte los recursos de la capa física con las transmisiones *unicast* (al menos hasta la revisión 12). Por lo tanto, la disponibilidad temporal de las transmisiones *multicast* está limitada a un máximo del 60 por ciento, reservándose las subtramas restantes (al menos el 40 por ciento) para las transmisiones *unicast*. Con el objetivo de alcanzar la máxima tasa total de datos (ADR) entre los usuarios *multicast*, hemos implementado varios esquemas innovadores de SRA que combinan la asignación de los recursos *multicast* y *unicast* de la trama LTE/LTE-A, garantizando los requisitos de QoS a cada usuario.

- En los sistemas de comunicaciones de banda ancha, la selección del MCS para transmisiones *multicast* se basa habitualmente en la utilización de CQIs de banda ancha, lo que proporciona información bastante imprecisa acerca de la capacidad potencial del canal *multicast*. Recientemente se ha empezado a utilizar el CQI por subbanda para mejorar la eficiencia espectral del sistema sin comprometer la robustez de los enlaces. Hemos propuesto nuevas estrategias para SRA *multicast* basadas en el CQI por subbanda que, basándose en la selección de los modos de transmisión con mayor eficiencia espectral, conducen a mejores tasas de datos, a la vez que permiten cumplir los requisitos de QoS.

- Los servicios móviles de vídeo *broadcast/multicast* precisan estrategias eficientes de SRA con baja complejidad. Hemos propuesto una estrategia de SRA basada en subgrupos *multicast* y la técnica de codificación de vídeo escalable (SVC) para la difusión de vídeo *multicast*, la cual se centra en reducir el espacio de búsqueda de soluciones y optimizar el ADR. Los resultados obtenidos en términos de ADR, eficiencia espectral y equidad entre los usuarios *multicast*, junto con la baja complejidad del algoritmo, ponen de manifiesto que el esquema propuesto es adecuado para su implantación en sistemas reales.

Estas contribuciones pretenden servir de referencia que motive la investigación actual y futura en el interesante ámbito de RRM para los servicios *broadcast/multicast* en las redes móviles de próxima generación.

Contents

List of Tables

List of Figures

List of Algorithms

Notation

a, A	scalar.
$\boldsymbol{a}$	vector.
a_i	i-th position of the vector $\boldsymbol{a}$.
$\boldsymbol{A}$	matrix.
$\boldsymbol{a}_i$	i-th column of the matrix $\boldsymbol{A}$.
$a_{i,j}$	position of the matrix $\boldsymbol{A}$ in the i-th row and j-th column.
$\boldsymbol{I}$	identity matrix.
$(.)^*$	conjugate of a scalar, vector or matrix.
$(.)^T$	transpose of a vector or matrix.
$(.)^H$	Hermitian of a matrix.
$\mathcal{CN}\left(\mu, \sigma^2\right)$	complex Gaussian random variable with means μ and variance σ^2.
$\mathbb{E}\{p(\mathbf{u})\}$	expectation with respect to the distribution $p(\mathbf{u})$.
$\lvert a \rvert$	absolute value of the scalar a.
$\lVert \boldsymbol{a} \rVert$	Euclidean norm of the vector $\boldsymbol{a}$.
$\overline{\boldsymbol{a}}$	means of vector $\boldsymbol{a}$.

$\hat{a}$	estimation of a.
$\lceil a \rceil$	round to the closer integer higher than a.
$\wedge$	logical conjunction.
$\arg\max$	arguments of the maxima.
$\log(\cdot)$	natural (base e) logarithm.
$\sim$	approximate equality.
$\triangleq$	defined to be equal to.
$\equiv$	definition.
$\mathcal{N}$	set of N elements.
$\in$	included in a set.
$\subseteq$	subset.
$\forall$	for all.
$x : P(x)$	set of all x for which $P(x)$ is true.
$\cap$	intersection of sets.
$\emptyset$	empty set.
$\setminus$	difference of sets.
$\mathbb{C}$	field of complex numbers.
$\mathcal{O}(\cdot)$	computational cost in operations.

Acronyms

3GPP 3rd Generation Partnership Project.

4G 4th generation of mobile communications.

5G 5th generation of mobile communications.

5G-PPP 5G Infrastructure Public Private Partnership.

ADR aggregated data rate.

AL-FEC application layer - forward error correction.

AMC adaptive modulation and coding.

AR augmented reality.

AWGN additive white Gaussian noise.

BICM bit-interleaved coded modulation.

BLER block error rate.

BM-SC broadcast multicast service center.

BS base station.

C-RAN cloud radio access network.

CC component carrier.

CDF cumulative distribution function.

CFC CQI feedback cycle.

CMS conventional multicast scheme.

CoMP coordinated multi-point.

CP cyclic prefix.

CQI channel quality indicator.

CR cognitive radio.

CSI channel state information.

D2D device-to-device.

DASH dynamic adaptive streaming over HTTP.

DC direct current.

DLC data link control.

DRX discontinuous reception.

DSA dynamic spectrum access.

EESM exponential effective SINR mapping.

eMBMS evolved multimedia broadcast and multicast service.

eNodeB evolved node B.

EPA 3GPP extended pedestrian A.

ESS exhaustive search scheme.

ETU 3GPP extended typical urban.

f-OFDM filtered OFDM.

FBMC filterbank based multicarrier.

FDD frequency division duplex.

FEC forward error correction.

FFT fast Fourier transform.

FI fairness index.

FLUTE file delivery over unidirectional transport.

FSS fast search scheme.

GBR guaranteed bit rate.

GFDM generalized frequency division multiplexing.

GOC group-oriented communications.

GSCA greedy subband CQI algorithm.

ICI inter-carrier interference.

IFFT inverse fast Fourier transform.

IoT Internet of things.

ISI inter-symbol interference.

JMSUT joint multicast subgrouping and unicast transmissions.

JMUS joint multicast/unicast scheduling.

LDPC low-density parity check.

LSA licensed shared access.

LTE Long Term Evolution.

LTE-A LTE-Advanced.

LUT look-up table.

MAMVL multicast resource allocation based on multiple video layers.

MBMS-GW MBMS gateway.

MBSFN multicast/broadcast over single frequency network.

MCE multi-cell/multicast coordination entity.

MCS modulation and coding scheme.

MIESM mutual information effective SINR mapping.

MIMO multiple input multiple output.

MME mobility management entity.

MooD multicast operation on demand.

MRA multicast resource allocation.

MRC maximum ratio combiner.

MS mobile station.

MSCA minimum subband CQI algorithm.

MSML multicast subgrouping for multilayer video applications.

MT maximum throughput.

MTC machine-type communication.

OFDM orthogonal frequency division multiplexing.

OFDMA orthogonal frequency division multiple access.

OMS opportunistic multicast scheduling.

OSS optimized search scheme.

PDSCH physical downlink shared channel.

PF proportional fairness.

PHY-FEC physical layer - forward error correction.

PL pluralistic licensing.

PMCH physical multicast channel.

PRB physical resource block.

PtM point-to-multipoint.

QAM quadrature amplitude modulation.

QoE quality of experience.

QoS quality of service.

RB resource block.

RLNC random linear network coding.

RRM radio resource management.

RTT round-trip-time.

SC-PTM single cell point to multipoint.

SDL supplemental downlink.

SFBC space-frequency block-coding.

SFN single frequency network.

SG-SCA second group subband CQI algorithm.

SG-WCA second group wideband CQI algorithm.

SINR signal to interference plus noise ratio.

SISO single input single output.

SRA scheduling and resource allocation.

SSCA sorted subband CQI algorithm.

SVC scalable video coding.

TD transmit diversity.

TDM time division multiplexing.

TTI transmission time interval.

TV television.

TVWS TV white spaces.

UDN ultra dense network.

UE user equipment.

UFMC universal filtered multi-carrier.

UHD ultra high definition.

UHDTV ultra high definition television.

UPA uniform power allocation.

VR virtual reality.

WCA wideband CQI algorithm.

Chapter 1

Introduction

Mobile video traffic is one of the key drivers that will lead to an unprecedented traffic growth with an increasing number of applications, such as downloading, streaming, or conferencing. According to the Cisco Visual Networking Index^{TM} [1], driven by the increased usage of smartphones and the emergence of new broadband services and applications, traffic from wireless and mobile devices will account for more than two-thirds of total IP traffic by 2020. Additionally, the Ericsson Mobility Report^{TM} [2] concludes that mobile video traffic is forecast to grow by around 50 percent annually through 2022 to account for nearly 3 quarters of all mobile data traffic. This scenario contemplates that the upcoming enhanced video services, such as ultra high definition (UHD), 4K and 3D videos, have also been launched through mobile networks taking benefits of the quality of service (QoS) breakthroughs introduced in current and beyond wireless systems [3].

Broadcast/multicast transmission represents a viable and effective solution to convey data simultaneously to a group of users through point-to-multipoint (PtM) communication, with important enhancements on the capacity and the spectrum efficiency of cellular systems. As a matter of fact, broadcast/multicast has been identified as one of the enabling technologies to cope, to some extent, with the increasing demand of mobile video data and to provide a better quality of experience (QoE) to end-users [4].

A great variety of upcoming applications need to be accommodated in the current and future mobile networks. Long Term Evolution (LTE) [5] and beyond provide the access platforms to broadband services, enabling high data rates, high spectrum efficiency, low latency, and in general high system capacity. The flat all-IP network infrastructure and the exploitation of orthogonal frequency division multiple access (OFDMA) on the radio interface are important enablers to achieve the broadband service requirements.

While standardization bodies, industry and academia are still trying to conceive what the 5th generation of mobile communications (5G) will be [6, 7, 8, 9, 10], LTE and LTE-Advanced (LTE-A) have been adopted as worldwide 4th generation of mobile communications (4G) cellular technologies [11, 12]. Moreover, it is anticipated that LTE-A will continue evolving beyond 4G in a backward compatible manner, thus benefiting from the large economies of scale established around the different releases of the LTE/LTE-A ecosystem [13]. Broadcast/multicast transmissions are fully supported on both LTE and LTE-A systems through the use of the evolved multimedia broadcast and multicast service (eMBMS) [12, 3, 14]. Furthermore, when deployed in single frequency network (SFN) mode, also known as multicast/broadcast over single frequency network (MBSFN), eMBMS can exploit a pre-existing LTE/LTE-A infrastructure to provide broadcast/multicast service coverage over an area simultaneously served by a flexible number of base stations (BSs) on the same frequency channel.

Implemented as a subsystem of LTE/LTE-A, the eMBMS shares the physical layer resources with the unicast transmission mode. In fact, LTE/ LTE-A networks (at least up to Release 12) use time division multiplexing (TDM) to aggregate broadcast/multicast and unicast services, with a time allocation to eMBMS services limited to a maximum of a 60 percent. Using a system that has been optimized for unicast services to deliver broadcast/multicast transmissions imposes some trade-offs that, in the end, compromise the system performance in terms of spectral/energy efficiency and offered QoE. Even though there seems to be a consensus in 3GPP working groups to remove this TDM-related constraint in LTE Release 14 specifications, thus opening the door to supporting efficient stand-alone eMBMS networks [13, 15], there are still some challenging issues associated to the radio resource management (RRM) processes.

Multicast applications, such as mobile TV, are typically resource hungry. This poses important challenges to the effective allocation of the scarce available spectrum and may severely limit the overall capacity of the LTE system. Consequently, the improvement of spectral efficiency is linked to designing an adequate scheduling and resource allocation (SRA) strategy for broadcast/multicast services. In an LTE/LTE-A system, the evolved node B (eNodeB) is in charge of performing RRM procedures, collecting all the channel quality indicators (CQIs) reported by the users that the eNodeB is serving. However, these procedures cannot be performed individually in multicast transmissions, so that RRM requires policies to decide how the multicast members will be served. The conventional approach to face this challenging issue has been to transmit the same data flow to all mobile stations (MSs) in the multicast group. The problem with this rather simplistic approach is that different users in the multicast group experience heterogeneous channel quality conditions and thus, as a certain QoS has to be guaranteed to the whole set of users in the group, the spectral efficiency of the multicast transmission is always determined by the users experiencing the worst channel conditions, usually located at the cell-edge. That is, users in the multicast group experiencing poor channel quality conditions force the BS to use robust modulation and coding schemes (MCSs) to guarantee a reliable transmission at the expense of wasting the potential spectral efficiency provided by the good channel conditions experienced by users near the cell-center, and consequently, the high potential of OFDMA resource allocation is only partially exploited [16]. In addition, RRM must be executed every transmission time interval (TTI), i.e. 1 ms in LTE, that leads to the development of low complexity algorithms that can be implemented in real scenarios.

The remaining of this introduction chapter details the objectives, structure and contributions of this thesis.

1.1 Objectives

This Ph.D. thesis is focused on the development of novel SRA strategies to address the important challenges that broadcast/multicast resource allocation presents. The efficient use of multicast transmission in mobile networks enables new services based on group-oriented communications (GOC).

The main goal of this Ph.D. thesis can be split into partial objectives performed during this work that can be summarized as:

- Objective 1: Study of multicast/broadcast in mobile networks. Identify new technologies/strategies to enhance the delivery of multicast services in next generation networks.
- Objective 2: Definition of a system model that allows us to present a new framework to study novel multicast SRA schemes.
- Objective 3: Implementation of new strategies to allocate resources jointly to multicast and unicast transmissions.
- Objective 4: Development of novel SRA schemes what make an efficient utilization of the subband CQI feedback reported by the multicast users.
- Objective 5: Design a low complexity SRA strategy that improves the aggregated data rate (ADR) of multicast video transmissions taking advantage of scalable video coding (SVC) technique.

1.2 Thesis structure

This thesis document is structured as follows.

Chapter 1 introduces the motivation which leads us to investigate the multicast RRM in mobile networks. The objectives, contributions and structure are also presented in this chapter.

Chapter 2 gives an overview of the architecture for broadcast/multicast in mobile networks and the new technologies for next generation mobile services, explaining their application for broadcast/multicast communications.

Chapter 3 details radio resource allocation for multicast transmissions in orthogonal frequency division multiplexing (OFDM) systems. This chapter presents the state of the art in broadcast/multicast resource allocation, the system model deployed to evaluate the SRA proposals collected in this thesis, and the group-based multicasting as a basic pillar of these proposals.

Chapter 4 is focused in joint multicast/unicast resource allocation strategies. The proposed joint multicast/unicast scheduling (JMUS) and joint multicast subgrouping and unicast transmissions (JMSUT) schemes are shown, their problem formulation are detailed, and a complete assessment of both strategies is presented.

Chapter 5 presents a novel multicast SRA framework using subband CQI feedback that aims at optimizing the ADR of multicast communications. A two-step two-subgroup scheme is proposed to overcome the complexity limitations inherent to multicast RRM. This chapter shows new algorithms implemented with an efficient use of the subband CQI which outperform those proposed in the literature.

Chapter 6 addresses the developing of a new resource allocation scheme that makes use of the SVC multiple layers to deliver multicast video services. The proposed multicast resource allocation based on multiple video layers (MAMVL) scheme maximizes the ADR and focuses on reducing the complexity of the solution space to develop an algorithm easy to implement in real systems.

Finally, some conclusions have been obtained from the analysis and development of multicast SRA strategies, and future lines of research have been presented to expand the work of this thesis.

1.3 Contributions

This work originally started from the results obtained during my M.S. thesis [17, 18], where we presented a performance analysis of eMBMS in LTE networks and the benefits of using dynamic MBSFN. Throughout the development of this thesis, I have been part of the following projects that have partially funded this research:

- LTExtreme project (IPT-2012-0525-430000) supported by the Spanish Ministry of Economy and Competitiveness, National Plan for Scientific Research, Development and Technological Innovation (INNPACTO subprogram), with the participation of Nokia Networks, Universidad Politécnica de Madrid and Universidad Carlos III de Madrid.

- Enabling technologies for Licensed and unlicensed Shared Access communications (ELISA) project (TEC2014-59255-c3-3-R)) supported by the Agencia Estatal de Investigación and Fondo Europeo de Desarrollo Regional (AEI/FEDER, UE), with the participation of Centre Tecnològic Telecomunicacions Catalunya, Universidad Carlos III de Madrid and Universitat de les Illes Balears.

According to the global objective of this thesis, the main contributions made throughout the realization of this research are summarized in the following list:

Chapter 2

- Revision and proposal of new technologies and trends for near future mobile networks, especially oriented to broadcast/multicast services.

Chapter 3

- Design of a theoretical holistic system model as base of analysis tool for multicast RRM in mobile networks.

Chapter 4

– Development and evaluation of a low complexity algorithm according to JMUS strategy.

– Development and evaluation of a SRA scheme based on JMSUT strategy.

Chapter 5

– Development of a novel multicast SRA framework based on subband CQIs.

– Implementation of a two-step two-subgroup SRA strategy.

– Assessment of a set of reduced-complexity algorithms working on two-step two-subgroup SRA scheme.

Chapter 6

– Development and evaluation of a MAMVL framework.

– Search space reduction to achieve a low complexity scheme.

– Implementation of an algorithm to solve the MAMVL scheme.

Some review and research results obtained during the realization of this thesis have been published in different international journals and magazines. A contribution to a *white paper* has been made as a result of the deep study of this relevant topic. Both international and national conference publications have been presented. Furthermore, some posters have been shown in several seminars and summer schools.

- International journals and magazines:

  1. **Alejandro de la Fuente**, Raquel Pérez Leal, and Ana García Armada, "New technologies and trends for next generation mobile broadcasting services", *IEEE Communications Magazine, Nov 2016.*

2. **Alejandro de la Fuente**, Guillem Femenias, Felip Riera-Palou, and Ana García Armada, "Subband CQI feedback-based multicast resource allocation in MIMO-OFDMA networks", *Recommended for publication in IEEE Transactions on Broadcasting.*
3. **Alejandro de la Fuente**, José Joaquín Escudero Garzás, and Ana García Armada, "Radio resource allocation for multicast services based on multiple video layers", *Submitted the revised version to IEEE Transactions on Broadcasting.*

- White paper contribution:

1. Expert Advisory Group of the European Technology platform Networld 2020, "Strategic research and innovation agenda", *Pervasive Mobile Virtual Services, July 2016.*

- International conference publications:

1. **Alejandro de la Fuente**, Carlos M. Lentisco, Luis Bellido, Raquel Pérez Leal, Ana García Armada, Encarna Pastor, and Alejandro Garcia Bolivar, "End to end measurements of multimedia streaming over LTE", *25th European Conference on Networks and Communications (EuCNC), June 2016.*
2. Oliver Holland, Adnan Aijaz, Shuyu Ping, Stan Wong, Jane Mack, Lisa Lam, and **Alejandro de la Fuente**, "Aggregation in TV white space and assessment of an aggregation-capable Wi-Fi white space device", *IEEE International Conference on Communications (ICC), May 2016.*
3. **Alejandro de la Fuente**, Carlos M. Lentisco, Luis Bellido, Raquel Pérez Leal, Encarna Pastor, and Ana García Armada, "Analysis of the impact of FEC techniques on a multicast video streaming service over LTE", *24th European Conference on Networks and Communications (EuCNC), July 2015.*
4. Carlos M. Lentisco, Luis Bellido, **Alejandro de la Fuente**, Encarna Pastor, Raquel Pérez Leal, and Ana García Armada, "A

model to evaluate MBSFN and AL-FEC techniques in a multicast video streaming service", *IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Oct 2014.*

5. **Alejandro de la Fuente**, Ana García Armada, and Raquel Pérez Leal, "Joint multicast/unicast scheduling with dynamic optimization for LTE multicast service", *20th European Wireless Conference (EW), May 2014.*
6. **Alejandro de la Fuente**, Raquel Pérez Leal, and Ana García Armada, "Performance analysis of eMBMS in LTE: dynamic MBSFN areas", *OPNETWORK, Aug 2013.*

- National conference publications:

1. **Alejandro de la Fuente**, Raquel Pérez Leal, and Ana García Armada, "Joint strategy for LTE resource allocation: multicast subgrouping & unicast transmissions", *JITEL, Oct 2015.*
2. **Alejandro de la Fuente**, Raquel Pérez Leal, and Ana García Armada, "Resource allocation management to broadcast/multicast services". *URSI XXX Symposium Nacional, Sep 2015.*
3. **Alejandro de la Fuente**, Raquel Pérez Leal, and Ana García Armada, "Análisis de prestaciones de eMBMS en LTE: redes de frecuencia única", *URSI XXVIII Simposium Nacional, Sep 2013.*

- Posters

1. **Alejandro de la Fuente**, F. Riera-Palou, G. Femenias and Ana García Armada, "Enhanced scheduling and resource allocation in multicast OFDMA systems", *Summer School on Spectrum Aggregation and Sharing for 5G Networks, Eurecom, Sophia-Antipolis, Oct 2016.*
2. **Alejandro de la Fuente**, J. Joaquín Escudero and Ana García Armada, "Multi-objective optimization for resource allocation

in broadcast/multicast systems", *IEEE Communications Society Summer School, Trento, June 2016.*

3. **Alejandro de la Fuente** and Ana García Armada, "A solution to broadcasting problem for resource allocation in LTE-A", *EURASIP-IEEE 3rd Spain Workshop on Signal Processing, Information Theory and Communications, Jan 2016.*

4. **Alejandro de la Fuente** and Ana García Armada, "Resource allocation to broadcast/multicast services", *Joint IEEE-EURASIP Spain Seminar on Signal Processing, Communication and Information Theory, Dec 2014.*

5. Carlos M. Lentisco, Luis Bellido, Encarna Pastor, and **Alejandro de la Fuente**, "Unicast and multicast streaming services over LTE networks", *European Conference on Networks and Communications (EuCNC), June 2014.*

Chapter 2

New technologies and trends for next generation broadcast/multicast services

Mobile data traffic has been growing rapidly in recent years, and this growth is expected to accelerate in the near future. For that reason, the mobile industry is preparing for 1000 times data traffic growth, where richer content will be delivered, including more video traffic used in multiple emerging applications. Furthermore, 28 billion interconnected devices are expected in 2021 [2]. It is widely accepted that this 1000 times capacity increase will be achieved by the combination of three approaches:

- Spectral efficiency improvements achieved through the introduction of new signal processing and coordination techniques.
- The optimized use of the spectrum, adequately combining licensed and unlicensed frequency bands.
- An increase in the density of base stations.

Broadcast/multicast is one of the major enablers to achieve these goals, since it provides an efficient use of the spectrum. Indeed, current trends in broadcasting will make it more dynamic, making possible the efficient use of on-demand spectrum. In addition, the deployment of small cells, thus increasing the number of base stations, enhances venue casting and allows the use of unlicensed spectrum in some areas.

This chapter presents the evolution of broadcast/multicast in the current LTE and LTE-A networks. Some novel technologies that are being envisioned to improve broadcasting are described, providing an overview of the actual challenges and potential solutions. Finally, new applications enabled by the use of 5G broadcasting and GOC are detailed.

2.1 Evolution of broadcast/multicast in mobile networks

Traditionally, mobile systems have been using unicast transmissions to every user, even to deliver certain services such as radio or television (TV) content in wide areas. However, the utilization of unicast transmissions presents clear limitations regarding the waste of resources that are individually allocated to each user which are demanding the same content. Broadcast/multicast transmissions where the same data is delivered simultaneously to a certain amount of users in a determined area have inherent advantages because of the use of common resources [19].

The 3rd Generation Partnership Project (3GPP) defines eMBMS in [5, 20] to deliver broadcast/multicast services in mobile networks, therefore combining unicast and broadcast services in the same network. The architecture required for the introduction of eMBMS in LTE/LTE-A is shown in Figure 2.1 [14], where new entities are included. The broadcast multicast service center (BM-SC) is in charge of global configuration, authorization and authentication, and billing. The MBMS gateway (MBMS-GW) manages the transmission of multicast IP packets from the BM-SC to the eNodeB. The mobility management entity (MME) handles the control signalling of the session. Finally, the multi-cell/multicast coordination entity (MCE) performs the synchronization of the different cells in the MBSFN

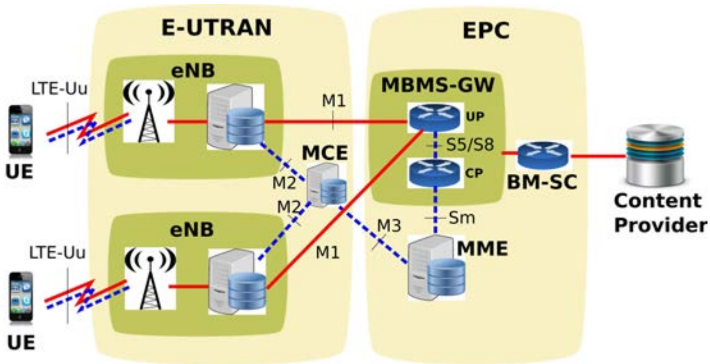


Figure 2.1: Architecture required for the introduction of eMBMS in LTE showing the different nodes that constitute the radio access and core networks.

area.

Multicast in LTE was commercially launched in South Korea in January 2014. After that, the BBC deployed an experimental system for the 2014 Commonwealth Games, and Verizon successfully trialled live eMBMS technology to a selected group of end users in the 2014 Superbowl.

2.1.1 Multicast/broadcast over single frequency network

The 3GPP proposes the introduction of MBSFN to improve the performance of eMBMS [21]. In SFN technology, all BSs that belong to the same MBSFN area transmit the same signal, at the same time, and in the same frequency to the multicast users; therefore all base stations must be tightly synchronized. Thus, the transmissions from multiple base stations are received by the user equipment (UE) as a single transmission with multi-path propagation, and destructive interference becomes constructive, combining the received signals that arrive at the UE within the OFDM cyclic prefix (CP) so as to avoid inter-symbol interference (ISI).

In order to increase the distance at which an eNodeB can be placed, MBSFN uses an extended CP where the 0.5 ms slot can accommodate six symbols, reducing the payload. Since in MBSFN-based transmissions the CP should not only cover the main part of the actual channel time dispersion, but also the main part of the timing difference between the transmissions received by the UE from the eNodeBs involved in the MBSFN transmission, the performance is increased using extended CP.

To change an eNodeB interference from destructive to constructive, we need to calculate the maximum distance at which the eNodeB can be placed. The radio cell size is an important factor in determining how many tiers of cells, around the region where the multicast users are placed, can be included in the same MBSFN area so that the signal to interference plus noise ratio (SINR) in the UE can be increased. Note that the smaller the cell is, the greater the number of eNodeBs adding their signals correctly to the UE. Thus, the multicast transmission achieves better performance.

The multicast transmission of events requires to achieve the given QoS requirements to offer the user a satisfying experience. With the aim of developing a method that optimizes the average performance, we have proposed a dynamic cluster of eNodeBs to create the MBSFN area in [17], which is based on the location of the UEs and the CQI received from each UE. Based on the CQI dinamically received from the UEs, the size of the MBSFN area can be adapted, increasing or decreasing the number of cells required to satisfy the QoS requirements to a certain percentage of users, whose radio channel conditions are also dynamically changing. The delivery of multicast transmission with high data rate requires the use of small cells, and a MBSFN area around the region where the multicast users are placed.

2.1.2 AL-FEC techniques over eMBMS

The choice of MCS in an LTE transmission implies a specific level of protection against errors, since the MCS defines a specific modulation scheme and also the forward error correction (FEC) overhead that is applied at the physical layer. Regarding multicast transmissions, it is worth considering that the radio channel conditions vary among all the users that receive the multicast service in an MBSFN area. Therefore, the block error rate (BLER) experienced by the set of multicast users can have a great variance. In order to increase the robustness and the reliability of the multicast transmissions, the 3GPP proposes an additional level of FEC redundancy at the application layer for eMBMS [22]. The solution proposed by the 3GPP to deliver video streaming over eMBMS uses the file delivery over unidirectional transport (FLUTE) protocol to send video segments with the corresponding application layer - forward error correction (AL-FEC) over multicast. Furthermore, when the AL-FEC is not able to recover a piece of information due to a high error rate, the application layer at a receiver can request a unicast delivery. In order to facilitate this combination of multicast transmission/unicast recovery, video segments are generated following the dynamic adaptive streaming over HTTP (DASH) recommendation [23].

AL-FEC is an error correction technique that sends redundant data to facilitate the recovery of the lost packets. Currently, the utilization of Raptor codes is standardized by the 3GPP for transmissions over eMBMS [22] to ensure reliable transmission over unreliable channels. Raptor codes are fountain codes, coding on-the-fly as many symbols as necessary from the symbols of the source block. The code rate, defined as the ratio between the original symbols and the symbols resulting of the encoding process is normally used to represent the amount of redundancy introduced at the transmitter. In wireless environments, with limited resources and high packet loss, it is necessary to find the value of the code rate that maximizes the useful data rate while guaranteeing a target coverage [24].

Related to this research area, we have proposed a novel perspective analyzing the trade-off between AL-FEC and physical layer - forward error correction (PHY-FEC) for achieving the maximum service data rate while limiting the percentage of unicast retransmissions, when different MBSFN area sizes are employed [25, 26]. In addition, we have implemented a testbed, along the frame of LTExtreme in Nokia Networks lab, with the aim of comparing different strategies encompassing both physical and application layer to improve the end to end delivery of video streaming content [27].

2.1.3 Multicast operation on demand

Dynamic switching between multicast and unicast transmissions makes possible the provision of multicast services on demand, taking advantage of their scalability. First, they can be geographically localized, using broadcast/multicast transmission only where it is needed. Second, multicast can be used as much as it is needed, reserving a certain amount of bandwidth resources for these transmissions. Finally, the service can be switched to multicast transmission solely in cases when it brings efficiency.

This feature is called multicast operation on demand (MooD) and has been introduced in LTE by the 3GPP [22]. MooD enables certain content that is initially delivered over the unicast network to be turned into a multicast transmission, in order to efficiently use network resources when the traffic volume exceeds a certain threshold.

2.1.4 Single cell point to multipoint transmission

The 3GPP introduces a new line of research in the evolution of eMBMS called single cell point to multipoint (SC-PTM) transmission [28]. The SC-PTM study has investigated methods of downlink multicast over physical downlink shared channel (PDSCH)[1]. More efficient resource utilization is achieved improving the flexibility of radio resource control by means of multiplexing multicast and unicast on the same physical channel. Note that if no data are available for multicast transmission is more efficient to schedule a unicast transmission on PDSCH than PMCH, since PMCH transmission always occupies the entire system bandwidth [29]. Moreover, the MCS employed on PMCH can only be changed through a reconfiguration.

2.2 New challenges for broadcast/multicast services in next generation mobile networks

Despite the fact that the evolution of broadcast/multicast in the 3GPP standards, many challenging issues are currently being considered to support next generation broadcasting services in LTE-A mobile networks and beyond. In this section, we detail some of them, summarizing their current

[1]Note that the original definition of multicast transmission employs physical multicast channel (PMCH), whereas the unicast transmissions use PDSCH.

status and ongoing research.

2.2.1 Adaptive resource allocation in multicast transmission

Resource allocation strategies have become one of the main challenges for the delivery of multicast services. New techniques are required to achieve high performance, both in terms of the ADR and fairness among all the users and services.

The different channel conditions between the users have traditionally forced system designers to adopt a conservative approach, which maximizes the fairness among the users. However, this is achieved at the cost of limiting the service throughput by the user that suffers the worst channel conditions. The conventional multicast scheme (CMS) introduces inefficiencies when some users experience poor channel conditions, and as a result the available spectrum is not fully exploited.

Recent research on multicast resource allocation in OFDM-based systems proposes solutions based on the efficient distribution of the physical resource blocks (PRBs) among different broadcast/multicast groups to improve the trade-off between service ADR and fairness. The block diagram of a packet scheduler that performs the resource allocation of an OFDM system is depicted in Figure 2.2. The frequency domain scheduler is used to split all the broadcast or multicast users into several subgroups according to the CQI reported by the terminals, and after that to design a SRA strategy based on multicast subgroups using different MCS. These strategies allow the system to improve the fairness among users with different channel conditions, delivering the service at different rates to the users that belong to each subgroup, at the time the ADR is maximized.

Practical systems demand low complexity schemes, where the time required to adapt the resource allocation to channel variations is minimized. Solutions to find optimal resource allocation in very few steps have been studied for a single cell with one multicast group in [30].

Other research is focused on resource allocation strategies for multi-flow delivery among the users in multicast environments [31]. Additionally, the

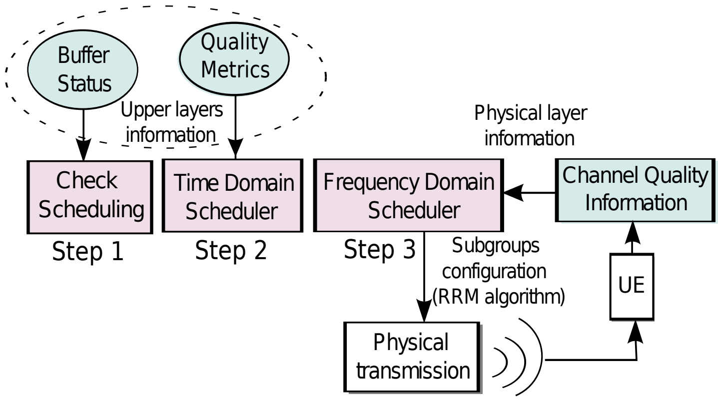


Figure 2.2: Block diagram of a packet scheduler that performs the resource allocation of an OFDM system. An adequate design of the scheduler will allow the system to improve the fairness among users while the ADR is maximized.

eNodeB requires the CQI feedback from the users, but CQI is a private information generated by the users with their own measurements and can be untruthful. Based on these statements, a mechanism to elicit the true CQI from the user is proposed.

All these works are optimizing resource allocation focusing on the trade-off between throughput and fairness. However, there are other important parameters related to the QoS, such as latency or guaranteed bit rate (GBR), that must be taken into account. Also, the 3GPP has specified QoE metrics that may be reported on a voluntary basis [22] and that may also be considered. Therefore, there is the need to develop new algorithms considering these requirements. Furthermore, the resource allocation problem must be addressed for a single-cell scenario, where a single eNodeB is delivering multiple multicast services to users. In addition, new resource allocation strategies must be developed for heterogeneous multi-cell scenarios working in a coordinated SFN that exploits the benefits of improving the channel quality of the users by using MBSFN areas.

It should be noted that the main objective of this thesis is to provide novel solutions that improve those available in the literature for multicast resource allocation. Chapter 3 presents a detailed study of the state of the art in broadcast/multicast resource allocation.

2.2.2 New waveforms for converged services

It is well known that OFDM, the waveform used in the 4G (LTE and LTE-A), presents interesting characteristics to be used in wireless networks. However, the use of a time-domain rectangular window in OFDM has the disadvantage of requiring very strict time and frequency synchronization. This means that the addition of a CP is mandatory, resulting in throughput loss. LTE is able to achieve this tight synchronization since the users are allowed to transmit, at the expense of exchanging energy-costly messages. These assumptions are no longer feasible when looking at expected use cases in future mobile communication systems such as machine-type communication (MTC) Internet of things (IoT), or user-centric deployments [32, 33, 34, 35, 36]. A system incorporating IoT devices should preferably allow them to transmit their messages without tight synchronization and using cheaper components [37].

4G deployments are based on devices connected to the network in a cell-centric way. It is worth mentioning that some important features of mobile communications such as coordinated multi-point (CoMP) [38], handover management, and offloading, may take advantage of the cell-centric concept. On the contrary, user-centric processing, when devices belong to multiple cells, leads to a disparity in the distances between the device and all access points whose respective carrier frequencies are also different. Consequently, tight synchronization may not be possible or cost-effective in a user-centric system.

These upcoming trends are leading the international community to the conclusion that the air interface of 5G [39, 40] needs to lower the degree of synchronization that OFDM is currently demanding. As a consequence, in the last few years some alternatives to OFDM have been proposed, such as filterbank based multicarrier (FBMC), universal filtered multi-carrier (UFMC), or generalized frequency division multiplexing (GFDM) [37].

Nevertheless, the way these new waveforms can be used to improve broadcasting in range, capacity, robustness, and utility remains today an open issue. For instance, longer ranges will be needed to provide converged

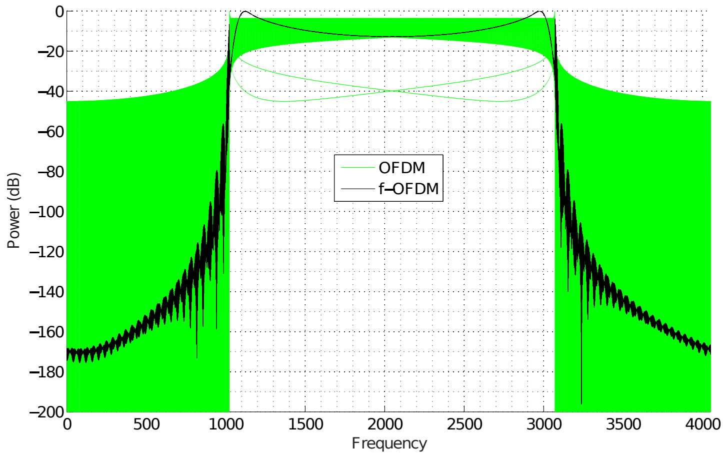


Figure 2.3: Comparison of the spectrum of classical OFDM and f-OFDM showing a better frequency confinement for f-OFDM.

TV services in 5G networks. To achieve the delivery of the TV service up to a range of 60 km using the traditional OFDM waveform, an extended cyclic prefix of 200 μs would be required, not only when traditional macro cells are used, but also when massive small cells are deployed, in order to benefit from SFN gain. With the current LTE parameters, this would imply only one OFDM symbol per subframe, consequently decreasing the LTE data rate more than ten times. How this can be managed with the new waveforms is yet to be defined. An alternative waveform that may be able to cope with the requirements of IoT, and at the same time maintain the benefits of the CP for broadcasting scenarios, is the recently proposed filtered OFDM (f-OFDM) [41]. In Figure 2.3, the spectrum of classical OFDM and f-OFDM are compared, showing that the latter has better frequency confinement, which is more suitable for asynchronous IoT communications.

2.2.3 Spectrum sharing and aggregation

The traditional regulation of spectrum use is based on two ends: exclusive use and license-exempt access. The idea of flexible licensing provides new opportunities in spectrum use for 5G systems by reusing parts of unused spectrum. New strategies are needed to support the variety and density of

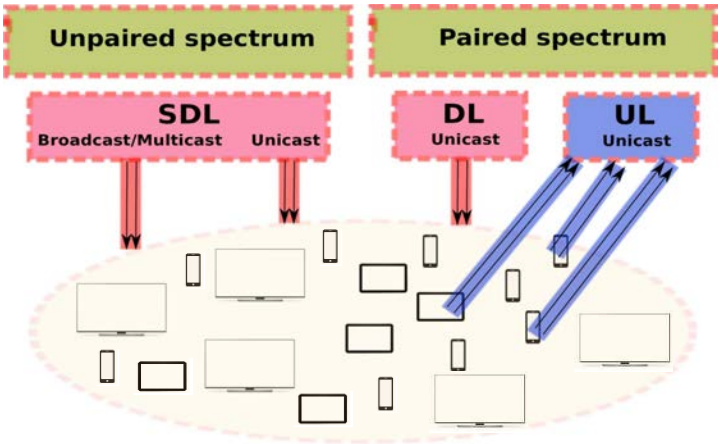


Figure 2.4: SDL deployment with paired and unpaired spectrum to deal with asymmetric downlink traffic in broadcasting services and increase the downlink capacity.

the upcoming wireless services and users, taking into account a good trade-off between cost and interference resilience, while ensuring service priority and spectrum availability. In particular, broadcasting presents the ability to use unpaired spectrum for the delivery of mass media or content. Therefore, spectrum sharing using flexible licensing allows the system to share some spectral resources to support the upcoming demand with the required QoS. At the same time, it opens up access to the unused spectrum bands in any location to be used in an opportunistic way by other 5G actors. Different regulatory solutions to achieve such flexibility have been proposed, such us light licensing, licensed shared access (LSA), and pluralistic licensing (PL) [42].

The light licensing approach consists of coordinated sharing between primary users (those with higher priority or legacy rights on spectrum usage) and secondary users (those allowed to use the spectrum without interfering with primary users). The scalability needed in mobile services makes this approach inappropriate due to its limitations in transmission power [42]. As a consequence, this strategy can only be used in systems where interference is controlled.

The LSA approach authorizes additional licensed users to access primary users' spectrum under tight controls to prevent interference. The main goal of LSA is to provide additional shared spectrum usage in specific bands, while QoS for all rights holders is guaranteed.

PL is an innovative spectrum licensing approach that takes into account the requirements of primary and secondary users, providing fair use to both of them. With this technique, primary users can choose from a set of licenses, according to different rules that depend on the amount of interference they can tolerate. On the other hand, secondary users access the band in a cognitive way, observing the primary users' requirements.

These licensing approaches, in particular PL, make use of cognitive radio (CR) technology to improve spectrum use by means of dynamic spectrum access (DSA), where the unlicensed users access unoccupied licensed bands in an opportunistic way. It is worth noting that wide continuous spectrum bands are not often available, due to current regulations and policies. In this scenario, spectrum aggregation is an interesting solution. Different algorithms have been proposed to optimize spectrum assignment, maximizing the number of users, fulfilling the bandwidth requirements for secondary users, or combining with adaptive modulation and coding (AMC) to achieve higher network throughput [43]. However, these algorithms have not been designed with multicast applications in mind. Indeed, TV broadcasting has traditionally been considered a primary use of licensed bands. Broadcast/multicast, as part of future mobile services, opens up a new paradigm of secondary use of the spectrum, providing a more efficient use.

This cognitive approach can facilitate the use of small parts of unpaired spectrum. In the context of mobile broadcasting, as Figure 2.4 illustrates, the use of supplemental downlink (SDL) has been proposed to increase downlink capacity by aggregating paired and unpaired spectrum [44]. Consequently, the use of CR technology and spectrum aggregation brings new opportunities to enhance the capacity of mobile broadcasting.

Along this thesis, I have been part in the investigation of aggregation in TV white spaces (TVWS) and the testing of IEEE 802.11 white space devices that are capable of aggregating, contiguously or non-contiguously, up to 4 TV channels [45].

2.2.4 Small cells with broadcasting for venue casting

The deployment of small cells in a SFN, under the umbrella of the existing macro cells, enhances venue casting performance. This deployment, which is presented in Figure 2.5, yields a robust coverage across the venue, maximizing the SFN gain by using more overlapping cells and improving the coverage offered by macro cells. This coverage improvement enables the transmission of higher-order MCS in the SFN area, hence increasing capacity.

The introduction of ultra dense networks (UDNs) in multicast scenarios offers more advantages, such as the opportunity to localize a specific multicast service, requiring only small cells for multicasting and freeing up macro cells for unicast transmissions [46]. Small cells enhance users' experience, providing better streaming performance consisting of more channels and content availability. The relevance of small cells when trying to provide ubiquitous coverage is shown in [2], where a new approach to coverage and subscribers QoE analysis, the so called application coverage, is presented. As they show, different coverage is achieved depending on the application taken into account, with video being the most critical application. It is shown that with a macro cell deployment, it is not possible to offer the required quality for the video service. Including a micro cell improves the situation with 21 percent coverage of video, while the use of indoor small cells achieves 100 percent coverage of video services.

These enhancements in user experience affect not only venue users but also macro cell users of different services. Indeed, small cells can offload traffic from nearby macro cells, improving the availability of radio resources.

To fully leverage the advantages of small cells, adaptive SRA techniques must be designed for these specific scenarios. In the context of using SFN in heterogeneous deployments, new constraints and conditions arise as compared to the macro cell scenario. When users are close to any of the small

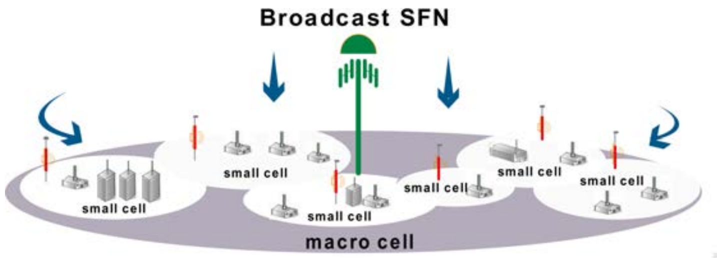


Figure 2.5: The deployment of small cells surrounded by existing macro cells enhances venue casting performance.

cell base stations, they will all share a high quality link. Consequently, the location information of the users, if available, can be an interesting input to enhance resource allocation strategies. In addition, this multi-tier deployment offers the flexibility of allocating some tiers to multicast transmissions while some others are restricted to unicast. This flexibility must be considered when designing SRA algorithms to fully exploit these potential benefits.

RRM may be facilitated by the emerging cloud radio access network (C-RAN) architecture [9, 47, 40], where baseband data are processed in a centralized way and distributed through a fiber or wireless front-haul to the small base stations. The possibility to centrally coordinate transmission has clear advantages for the deployment of SFNs. On the other hand, the limited capacity of the front-haul/back-haul may increase the latency and constitute a bottleneck for multicasting services. The joint optimization of caching and multicasting is a potential solution to improve the efficiency of massive content dissemination as shown in [48], where spatial content diversity is achieved in a large-scale heterogeneous network with back-haul constraints.

2.2.5 New communication paradigms

There are a number of emerging communication paradigms that will have a major impact on next generation mobile networks. Device-to-device (D2D) and MTC are two of the most representatives.

D2D communication has an increasing interest among the research community as a means to extend coverage and overcome the limitations of conventional cellular systems [49]. The introduction of D2D-enhanced multicast increases the service coverage, enhances the performance for cell-edge users, allows the system to increase the MCS in the downlink, and enhances the spectral efficiency by using single-frequency D2D communications. The 3GPP has recently introduced the support of D2D communications over LTE-A to enable direct connections among mobile devices in mutual proximity. Two approaches are currently investigated for D2D communications. On the one hand, inband D2D communications exploiting cellular spectrum and, on the other hand, outband D2D communications exploiting unlicensed spectrum. Lots of research are still on-ongoing to use multicast with each of the two D2D approaches.

MTC is one of the most promising solutions to enable a wide range of novel applications and services that 5G systems are targeted to provide. MTC has a set of unique and challenging characteristics [50], such as low or no mobility, time controlled, time-tolerant, secure connection, which require technically advanced solutions currently under investigation by academia, industry and standard bodies, such as the 3GPP [51, 52]. Specifically, MTC aims to provide short connection time, low-energy data transmission, discontinuous reception (DRX), reduction of signalling, control traffic, and high reliability to the thousands of devices which is foreseen to be camped in each cell [53]. In the context of cooperation between MTC and eMBMS, machine-oriented group services pose additional challenges with respect to human traffic. In particular, high latency during group formation and data delivery jointly with high-energy consumption for MTC devices have to be taken into account.

2.3 New applications enabled by 5G broadcasting

According to the Ericsson Mobility Report^{TM} [2], mobile video traffic is forecast to grow by around 50 percent annually through to 2022, when it will account for nearly 75 percent of all mobile data traffic. Moreover, the total mobile video traffic over the next six years will be more than 22 times that of the last six. The first 5G trials will be related to broadcasting of global events: in 2018, when the XXIII Winter Olympic Games begin in Pyeong Chang, South Korea plans the introduction of a prototype 5G network and Japan intends to launch a 5G trial network for the Summer Olympic Games in 2020 in Tokyo. Given these numbers it is clear that video services are driving the evolution towards the 5G.

On this context, broadcast based video service has many benefits offering anytime and anywhere connection, across all kind of devices. First of all, a single network can be used for mobile and fixed devices accessing TV content, which means both network and assets are shared for all devices and TVs. Consequently, this deployment with a single network, content and assets is easier to manage. Secondly, broadcasting is a most effective way to deliver TV content to mobile terminals, due to the fact that an inherent broadcast support in devices is leveraged without the need for any new modem. Finally, broadcasting allows the operator to provide a uniform user experience across multiple devices. The streaming experience can be improved by eliminating unicast congestion constraints. As a consequence, a great opportunity for more cost-effective data plans for mobile TV can be developed.

Broadcast/multicast based video service opens up new business opportunities. A new class of services is introduced with interactive and personalized TV services with cost-efficiency. This yields new business models based on revenue share, leased/hosted network model and others. These new business models can create new partnerships between operators, content owners, spectrum holders and advertisers.

In addition, 5G broadcasting can be extended for public safety applications, due to the fact that mobile networks can bring these services efficiently in critical situations. This service can be accessed by using push-to-voice/data/text. The recipients can be dynamically moved between broadcast and unicast depending on which transmissions achieve higher efficiency.

User groups can be preconfigured or formed on the fly according to the service provided. Public safety requirements must be fulfilled, being the end-to-end set-up time less than 300 ms and the end-to-end transport delay less than 150 ms.

The new 5G architecture and radio interface will enable enhanced quality video broadcasting with ultra high definition television (UHDTV). Online content providers such as Netflix, Amazon and YouTube have announced their plans for releasing series and films in initial UHDTV formats such as 4K and the evolution path towards 8K is already defined. Some insights from a number of technology validation tests and demonstrations that have been held to understand the requirements for UHDTV broadcasting to mobile phones and tablets are presented in [54]. While it is possible to enable regular UHDTV experiences on tablets and smartphones, this experience can be enriched by allowing end-users to freely extract a region-of-interest and navigate around the ultra-high resolution video, or add scalable augmented reality (AR)-style overlays to the video.

Indeed, emerging applications related to video, either complementing UHDTV or independently, such as telepresence, AR and virtual reality (VR) will be enabled in the near future and are considered as one of the key use cases driving the requirements for 5G [55].

AR is a concept of supplementing the real world with the virtual world [56]. Although it uses a virtual environment created by computer graphics, its main playground is the real environment. Overlaying digital information onto the real world, viewed through a camera-phone, is already possible today with multiple applications, but the business models and usage patterns are still evolving.

It is worth mentioning that telepresence is a form of innovation that has gained a lot of attention in recent years as providing new options for helping people become more engaged and collaborative. The mobile telepresence is a part of a growing consumer robotics industry that is slated to become a $6.5 Billion market by 2017, according to a 2013 report of Oyster Bay. Double Robotics, Beam+, PadBot are examples of available products that make use of today's limited cellular networks and Verizon showcases VGo as a robotic telepresence solution that provides educational solutions for children that cannot attend a traditional classroom due to illness, socio-economics, or geographical separation. Telepresence applications require very fast response time, below 10 ms in order to avoid cyber-sickness, which requires a round-trip-time (RTT) in the radio interface below 1-2 ms. The 5G Infrastructure Public Private Partnership (5G-PPP) has set the requirement of peak data rates in the order of 10 Gb/s for the evolution towards 5G, that will be required to support services such as 3D telepresence on mobile devices [9].

VR services can help people experience their presence in an imaginative world which looks real, while giving them also a chance to communicate with that world. VR is analyzed as one of the use cases in METIS project [57] when considering the evolution towards 5G. Since VR emulates the real world, the real-time video streaming should be very high in quality, so a need of 4-28 Gb/s data rate to transmit such video over the air is estimated, given that no or slight compression would be feasible. In that sense, it is pointed out in [57] that any sort of major compression introduces delays and therefore should be avoided for VR. Since the 5G air interface should be capable of handling such data rate with low latency, it is considered that future 5G mm-wave small cells can be a good enabler for such scenarios.

As pioneering work towards the implementation of these technologies, it is worth pointing out that SK Telecom has organized a technology forum on July 22, 2015 in collaboration with Google and Microsoft and has started to work with international firms in the field of AR and VR. SK Telecom has announced to be soon going to commercialize a product named T-AR for Tango, combining AR and VR.

It is worth noticing that all these new applications require very stringent requirements in terms of data rate and latencies that eMBMS can help to achieve. VR and, especially AR, require the utilization of the IoT paradigm to be implemented, where the eMBMS service may be employed jointly with disseminated machine and sensors (i.e., through the MTC paradigm) which can benefit of the multicast machine-oriented applications.

2.4 Summary

Broadcast/multicast is an important enabler to achieve an efficient use of the spectrum in the future mobile networks. The evolution of the multicast service in the 3GPP standards have become it more dynamic and useful. These enhancements bring the opportunity to offer new applications enabled by 5G broadcasting. Additionally, this chapter has detailed some challenging issues that need to be continuously enhanced in order to achieve the goals of next generation broadcasting.

- Dynamic resource allocation strategies for multicast are being developed to maximize the benefits of using multicast and broadcast transmissions in heterogeneous networks, where the service is delivered to users with different SINR.

- New waveforms that improve the inefficiencies of traditional OFDM for some foreseen use cases of the upcoming 5G networks. The development of a 5G standard to deal with both mobile and broadcast industry demands requires the analysis of the implications of the waveforms in the broadcasting performance.

- Novel schemes to enable the spectrum sharing for an optimized spectrum usage. The use of a coordinated spectrum access and sharing infrastructure are required, instead of using a competitive and interfering access. In addition, an efficient use of unpaired spectrum can be carried out with broadcasting services.

- Small cells which are deployed together with the existing macro cells enhances venue casting, since they improve the coverage and thereby the capacity everywhere, including the opportunity of better utilization of unlicensed spectrum.

- New communication paradigms based on D2D and MTC, which are receiving an increasing interest as potential enablers of many 5G applications.

From this point onwards, this thesis is focused on multicast RRM. Nevertheless, it should be noted that the knowledge of the other new challenges for broadcast/multicast services places in context the development of multicast SRA strategies towards next generation mobile networks.

This chapter is mainly based on the following published work:

- **Alejandro de la Fuente**, Raquel Pérez Leal, and Ana García Armada, "New technologies and trends for next generation mobile broadcasting services", *IEEE Communications Magazine, Nov 2016.*

The following works have also been referenced in this chapter:

- **Alejandro de la Fuente**, Raquel Pérez Leal, and Ana García Armada, "Performance Analysis of eMBMS in LTE: Dynamic MBSFN Areas", *OPNETWORK, Aug 2013.*

- **Alejandro de la Fuente**, Carlos M. Lentisco, Luis Bellido, Raquel Pérez Leal, Encarna Pastor, and Ana García Armada, "Analysis of the impact of FEC techniques on a multicast video streaming service over LTE", *24th European Conference on Networks and Communications (EuCNC), July 2015.*

- Carlos M. Lentisco, Luis Bellido, **Alejandro de la Fuente**, Encarna Pastor, Raquel Pérez Leal, and Ana García Armada, "A model to evaluate MBSFN and AL-FEC techniques in a multicast video streaming service", *IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Oct 2014.*

- **Alejandro de la Fuente**, Carlos M. Lentisco, Luis Bellido, Raquel Pérez Leal, Ana García Armada, Encarna Pastor, and Alejandro Garcia Bolivar, “End to End Measurements of Multimedia Streaming Over LTE”, *25th European Conference on Networks and Communications (EuCNC), June 2016.*

- Oliver Holland, Adnan Aijaz, Shuyu Ping, Stan Wong, Jane Mack, Lisa Lam, and **Alejandro de la Fuente**, “Aggregation in TV White Space and Assessment of an Aggregation-Capable Wi-Fi White Space Device”, *IEEE International Conference on Communications (ICC), May 2016.*

Chapter 3

Radio resource allocation for mobile broadcast/multicast

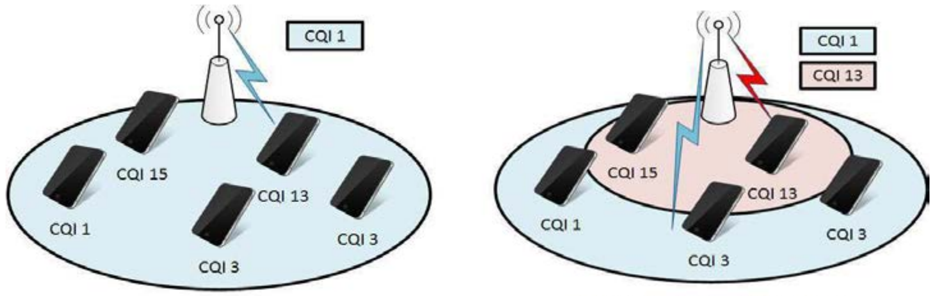
In an LTE/LTE-A system, the RRM procedure at the BSs, based on the CQIs received from the MSs, is in charge of determining which is the best allocation of resources (i.e., PRBs and MCSs) to broadcast/multicast traffic flows according to certain optimization criteria. The conventional approach to face this challenging issue has been to transmit the same data flow to all MSs in the multicast group. The problem with this rather simplistic approach is that different users in the multicast group experience heterogeneous channel quality conditions and thus, as a certain QoS has to be guaranteed to the whole set of users in the group, the spectral efficiency of the multicast transmission is always determined by the users experiencing the worst channel conditions, usually located at the cell-edge. That is, users in the multicast group experiencing poor channel quality conditions force the BSs to use robust MCSs to guarantee a reliable transmission at the expense of wasting the potential spectral efficiency provided by the good channel conditions experienced by users near the cell-center. Consequently, RRM processes for mobile broadcast/multicast present today some important challenges.

First, this chapter presents the state of the art in broadcast/multicast resource allocation. Next, the system model deployed to evaluate the different multicast SRA proposals collected in this thesis is detailed. Finally, the group-based multicasting strategy, that is a basic pillar of the proposals explained in the following chapters, is described.

3.1 State of the art in broadcast/multicast resource allocation

Although eMBMS over LTE networks is able to provide enhanced capabilities for multimedia data delivery, one of the main challenges in integrating this technology within the current cellular infrastructure is represented by the management of the licensed radio spectrum. Indeed, RRM is in charge of performing an efficient link adaptation procedure regarding the multicast transmissions according to the channel conditions experienced by the multicast users. Further, in a multicast scenario such adaptation has to be accomplished on a per-group basis thus taking into account channel information of all the users registered to a given multicast service. However, the presence of cell-edge users, which experience poor channel conditions and consequently cannot support high data rates, influences strongly the QoS that the cellular infrastructure could provide. In addition, the recent avalanche of new services over the Internet video applications (e.g., Spotify, Netflix, Facebook), requiring large amount of bandwidth, poses serious problems regarding the spectrum utilization and the coexistence with other cellular transmissions (e.g., normal phone calls).

Generally speaking, in a standard LTE system the BS, namely eNodeB, is in charge of performing RRM procedures. Indeed, once that all the CQIs of all the multicast members have been collected by the BS, a dynamic scheduling and link adaptation of the available radio resources is performed. Then, QoS management, admission control, and a persistent scheduling are accomplished. RRM works for multicast services were initially split based on the transmission type: single-rate and multi-rate transmissions [58]. In single-rate the BS transmits to all the users in each multicast group at the



(a) Single-rate multicast transmission **(b)** Multi-rate multicast transmission

Figure 3.1: Types of multicast transmissions.

same data rate. In multi-rate, instead, the BS transmits to each user at different data rates exploiting users' frequency diversity, according to the heterogeneity of the wireless channels. For the sake of clarity, the two types of multicast transmissions are illustrated in Figure 3.1.

3.1.1 Single-rate multicast transmission

Single-rate schemes have been quite popular due to their implementation simplicity. While some approaches aim to maximize the percentage of users served in a given instant of time (i.e. conservative approaches), other solutions maximize the performance without considering the satisfaction of the users (i.e. opportunistic approaches).

Multicast conservative approach

Multicast conservative approach aims to deliver the same content (or service) to each multicast member at the same time by exploiting a shared channel. This is made to avoid network capacity reduction and quality degradation that may happen with the use of one unicast link between each single user and the BS. In addition, it is worth noticing that with the shared channel the amount of required radio resources has no direct relation to the number of multicast members. Therefore, the spectral efficiency (measured in b/s/Hz) increases and this is more evident in scenarios where the number of multicast terminals is high.

Based on this, a primary basic SRA solution considered over LTE systems is identified by the conservative approach. In particular, this is based on the idea that a given QoS is guaranteed to all the multicast members. The way how the QoS is chosen by the mobile network strictly depends on the channel qualities experienced during the multicast session. In particular, going into details the conservative scheme adaptively sets the multicast transmission parameters (i.e., MCS) to suit the user with the worst channel quality [59], usually located at the cell-edge or in a coverage hole. Consequently, such a solution exploits poorly the spectral efficiency benefits of multicast, since users experiencing good channel conditions (usually located close to the BS) are severely hindered from utilizing an MCS that fully exploits their good channel state.

Multicast opportunistic approach

In order to overcome the limitations belonging to the conservative approach and to exploit efficiently the multi-user diversity, thus providing a more effective selection of the MCS based on the users' channel information, the opportunistic multicasting [60] has been proposed in the literature as a possible solution. The idea behind the opportunistic strategy is to select the most suitable MCS in each time slot of the multicast session and then select, accordingly, the portion of multicast users that may be served. This is because each user experiences independent fading over different time slots. However, different strategies may be exploited for the selection of the most suitable portion of users to serve, and, generally, can be summarized in two main categories:

- *Pre-defined fixed MCS*: with this strategy the MCS for serving the multicast session is fixed over the time and it cannot change during the entire multicast session delivery. In such a case, a multicast member will be able to receive the content of interest only if the channel quality is good enough to support the MCS adopted by the BS. The selection of the MCS may be performed with the aim *(i)* to achieve a given spectral efficiency or *(ii)* to maximize either the spectral efficiency or the achievable system throughput. Therefore, the former approach presents severe limitations regarding the achievable session coverage (limited number of users) that remains fixed for the entire multicast session.

- *Average group throughput*: following this strategy the multi-user diversity enables the BS to transmit towards a given multicast group based on the average throughput history of the group [61]. In evaluating the average throughput history, the BS may consider *(i)* the average throughput achieved among all the multicast members (i.e. multicast content will be transmitted by the BS with the MCS supported by half (50 percent) of all group members) or *(ii)* to select the most appropriate data transmission rate based on the exponential moving average of throughput values of the involved terminals.

However, even if the aforementioned approaches are able to extend the capability of the BS to dynamically change the portion of users that are served in every time slot, some issues and challenges related to the opportunistic multicasting have to be considered. In particular, a very challenging issue related to opportunistic multicast scheduling (OMS) is that it cannot guarantee adequate fairness among multicast members on a short term. Above mentioned issues are exacerbated when considering real-time video streams characterized by strict QoS constraints in terms of delivery delay and jitter.

3.1.2 Multi-rate multicast transmission

Multi-rate schemes take into account the intrinsic heterogeneous channel characteristics of wireless networks. Multi-rate multicast allows each user to receive multimedia traffic based on their own capabilities. In particular, this solution is in the middle between the conservative and the opportunistic multicast approaches. This scheme, also known as multicast subgrouping, is able to reduce the limitations of the conservative strategy (where the presence of users with poor channel conditions strongly affects the overall system spectral efficiency), and the issues related to the opportunistic multicasting that prefers the maximization of the system performance without considering the percentage of users served during the multicast session. In particular two techniques may be considered for providing multi-rate multicast transmission: *(i)* the stream splitting and *(ii)* the group splitting.

The stream splitting is based on dividing high-rate multimedia contents in multiple substreams of lower data rate [62]. In such technique, the information quality improves as users receive more substreams. A base substream, receivable by all the multicast users, is transmitted in order to accomplish full coverage of the multicast group. Afterwards, users with good channel conditions receive additional enhancement streams to improve information quality.

The group splitting technique, instead, divides multicast members into different subgroups (i.e., smaller portions of the overall multicast group), each one formed by users experiencing similar channel conditions, in such a case the MCS is selected differently for each subgroup [63]. Thus, the goal of subgrouping is to serve all the multicast members every time slot by guaranteeing improved session quality, coverage, and spectrum efficiency. Concerning the subgroup formation, this may be performed by following different strategies. In particular, a basic approach splits all the multicast members into two subgroups, thus defining a separate multicast transmission for each subgroup [64]. Each multicast stream is delivered using different MCS, power level, and, consequently, with a QoS that strictly depends on these system parameters and the channel conditions of the users within the two subgroups. Another strategy that may be exploited is to design the subgroup formation according to an optimization problem. In doing this, the most suitable subgroup configuration (i.e., number of subgroups with the related MCS, portion of users, assigned resources, and data rate for each enabled subgroup) is dynamically selected by the BS based on the optimization of a given utility function which takes into account the users' CQI values and the QoS constraints of the multicast session [65]. In addition, the optimization problem can be formulated in order to achieve different goals, for instance the maximization of the system throughput, spectral efficiency, or energy efficiency. A further strategy is to recast the subgroup creation phase in order to minimize the amount of radio resources (i.e., in the LTE system represented by the PRBs) required for the delivery of the multicast session. Of course, what may be maximized or minimized is not limited to the radio resources or other system-level parameters, but can also involve fairness, power consumption, and session delivery time.

A novel interesting approach consists in the combination of group splitting with the joint use of SVC using stream splitting to deliver the multicast video services split into several layers to different subgroup of users according to their MCS. In [16] Condoluci et al. introduce a link adaptation procedure based on the subband CQI feedback reported by the multicast users. The proposed SRA scheme is implemented for multilayer video applications, splitting each multicast group into different subgroups, where two different cost functions are analyzed both achieving high spectral efficiency and utilization. In [66] Orsino et al. propose a SRA technique based on group splitting with the joint use of SVC, where the scheduling decision is taken through the use of a multi-criteria decision making method exploited in a real-time way providing a good trade-off between throughput, fairness and user's satisfaction index.

Considering the exploitation of short-range links (i.e., D2D communications), the subgrouping approach can be further enhanced to improve the multicast service performance according to the users channel qualities, while simultaneously improving the system capacity and reducing the consumption of system resources [67]. In such a situation, proximity-based transmissions may be supported over licensed (e.g., LTE-Direct) or unlicensed (e.g., WiFi-Direct) bands. In detail, by exploiting the D2D paradigm, a subset of multicast members (within a subgroup or not) is expected to act as relay nodes thus receiving the data content from the BS and then forwarding the same to multicast members that are in proximity. In doing this, the logic followed by the BS to select the relays is to activate the short-range links for nodes with worse channel conditions thus maximizing the overall system performance.

Recent works in RRM focus on a fair resource distribution between multicast and unicast traffic [68, 69, 70]. In [68], Pizzi et al. propose a novel heuristic RRM policy achieving a fair resource distribution between unicast and multicast traffic. A subgrouping strategy is applied to both multicast and (virtual) unicast groups that, based on reported wideband CQI metrics, assembles users in subgroups showing similar channel characteristics. In [69], Christodoulou et al. focus on dynamically balancing the resource allocation between unicast and multicast services based on the status of the users' bearer queues. In [70], Chen et al. develop an optimal solution using convex optimization and dynamic programming to fairly allocate the rate across multiple cells to multicast and unicast users based on the channel conditions reported in the wideband CQI. In contrast, other recent works focus on the benefits of implementing SRA strategies that use random linear network coding (RLNC) to deliver a multicast layered service [71, 72]. In [71], Tassi et al. propose a resource allocation framework based on a multi-rate approach that aims at minimizing the total amount of required radio resources to deliver a multicast layered service using RLNC. The resource allocation goal is fulfilled by jointly optimizing both the transmission parameters and the employed coding scheme using a subgrouping strategy that is based on the behaviour of the data queues of the users served in each scheduling interval. Different to [71], Chau et al. in [72] use a channel-aware subgrouping strategy that is based on the use of conservative approaches relying on the wideband CQI reported by the MSs.

Although, a large amount of ongoing research are available in multicast RRM literature, we have investigated on three aspects that, to the best of our knowledge, are not completely resolved:

- The existing SRA strategies to deliver a multicast service (combined or not with unicast services) only consider the resources dedicated for multicast transmission. We have investigated the combination of multicast and unicast resource allocation according to the LTE/LTE-A frame to deliver a multicast service.

- Most of the existing SRA algorithms for broadcast/multicast services are based on the use of either oversimplistic wideband CQI, or very conservative low complexity approaches using subband CQI. We have developed a new SRA framework based on an elaborate use of subband CQI.
- The existing SRA schemes to deliver multicast video services based on stream splitting techniques result in an extremely large solution space, with high complexity for real implementations. We have complemented the available solutions in the literature with a novel SRA strategy that maximizes ADR with the focus on minimizing the solution space.

3.2 System model

Let us consider the downlink of an LTE-like multiple input multiple output (MIMO)-OFDMA multicellular system consisting of K BSs in which the reference BS (also denoted as the BS of interest or the tagged BS) is providing a multicast service to M MSs distributed over the corresponding coverage area[1]. The sets $\mathcal{K} = \{0, \ldots, K-1\}$ and $\mathcal{M} = \{1, \ldots, M\}$ will be used throughout this work to index BSs and MSs, respectively. Furthermore, from this point onwards it will be assumed that BS 0 is the reference BS multicasting to the M users under study. A general setup will be assumed in which BS 0, transmitting with a total power P_T, is equipped with N_T transmit antennas, and MS $m \in \mathcal{M}$ is equipped with N_{Rm} receive antennas.

The traffic flow that has to be multicast to the M users arrives from higher layers to the data link control (DLC) layer of the BS where it is

[1] The analytical framework presented in this work could readily be extended to a MB-SFN with several BSs simultaneously transmitting the same signal over the same frequency channel. The multi-cell multicast scenario employed in Chapter 4 uses this MBSFN extension.

buffered in a queue. In order to simplify the analytical framework, a full-buffer traffic model will be considered in which, at any given time, the serving BS is assumed to have data to multicast. This simple model allows the assessment of performance metrics related to the spectral efficiency, fairness or error probability irrespective of the actual traffic distribution model. Furthermore, it will also be assumed that the objective of the reference BS is to provide maximal spectral efficiency with constraints on the minimum data rate $\tau_{\min}$ (measured in b/s) that has to be guaranteed to the whole set of multicast MSs in the coverage area and on a BLER less or equal than a target value BLER_0. Hence, depending on the channel state information (CSI) received from all the served MSs, the task of the RRM procedure will be to allocate resources (e.g., power, data rates, frequency subbands and/or time-slots) to the multicast data flow in such a way that the data symbols, jointly with added signaling information, can be mapped onto the time-frequency plane by using appropriate MIMO-OFDM-based processing strategies.

In the time-domain, the downlink of an LTE-like system using frequency division duplex (FDD) is structured into radio frames with a duration of 10 ms. Every radio frame is divided into ten equally-sized subframes, also known as TTIs, defined as the duration of the transmission of the physical layer encoded packets over the radio air-interface, consisting of two consecutive time slots of duration $T_{\text{slot}} = 0.5$ ms each. Following LTE specifications, every time slot consists of a fixed number N_s of OFDM symbols which can be either equal to seven, when using normal CP, or equal to six, when using extended CP. For a normal CP, the first symbol has a CP length of $T_{\text{CPlong}} = 5.2$ μs and the remaining six OFDM symbols have a CP of length $T_{\text{CPshort}} = 4.7$ μs. For the extended mode, instead, the CP duration is $T_{\text{CPextended}} = 16.67$ μs. Irrespective of the CP in use, the useful OFDM symbol time is $T_u = 66.7$ μs. The normal CP is typically used in urban cells, while the extended CP is used in special cases like multi-cell multicast and in very large cells.

In the frequency-domain, slotted transmissions employ a bandwidth

B_{FFT} ranging from 1.4 MHz to 20 MHz with the number of subcarriers in the OFDM signal N_{FFT} ranging from 128 to 2048. The most commonly used bandwidth deployments are those corresponding to 5, 10 and 20 MHz with 512, 1024 and 2048 orthogonal subcarriers, respectively. LTE Release-10 introduced the capability to aggregate up to five component carriers (CCs) with a maximum aggregated bandwidth of 100 MHz. The subcarrier spacing is $\Delta f = 1/T_{OFDM} = 15$ kHz. Out of the N_{FFT} subcarriers, the direct current (DC) subcarrier is not used, a non-negligible amount of them are left empty in order to set guard frequency bands on either side of the system bandwidth and only N_c subcarriers are used to transmit data, pilots, synchronization and control channels. In particular, for the bandwidth deployments using $B_{\text{FFT}} = 5$, 10 or 20 MHz, the number of useful subcarriers is $N_c = 300$, 600 or 1200, respectively.

The basic time-frequency resource unit is a resource block (RB), which is composed of one time slot (0.5 ms) and twelve contiguous subcarriers (180 kHz). However, following the RRM procedure specified in LTE/LTE-A, it will be assumed in this work that the RRM process takes place over a TTI of two consecutive time slots ($T_{\text{TTI}} = 1$ ms) and thus, the minimum non-null amount of resources that can be allocated to a given traffic flow is composed of two consecutive (in time) RBs, also known as a PRB. Hence, in deployments using $B_{\text{FFT}} = 5$, 10 or 20 MHz, the number of available PRBs per TTI is $N_{\text{PRB}} = 25$, 50 or 100 PRBs, respectively.

3.2.1 Transmitter/Receiver

Our aim in this subsection is to present a formulation for a MIMO-OFDM scheme that provides a description of the physical layer in terms of the SINR experienced at the output of the MS detector. MIMO technology comprises a great variety of techniques that can be used to exploit the multiple propagation paths between the N_T transmit antennas at the BS and the N_{Rm} receive antennas at a served MS m (see [73] for a review). In a multicast service scenario, however, MIMO transmit diversity (TD) schemes are the most suitable ones[2]. In this work, only transmit diversity schemes for two transmit antennas and one codeword will be considered, and will be compared with the single transmit antenna scheme.

Let us denote by $\boldsymbol{s}_{t,q} = \left[s_{t,q,0}, \ldots, s_{t,q,N_{t,q}-1}\right]^T$ the complex-valued vector of data symbols transmitted during OFDM symbol q of TTI t (we assume zero-mean uncorrelated unit-energy symbols, i.e. $\mathbb{E}\left\{\boldsymbol{s}_{t,q}\boldsymbol{s}_{t,q}\right\} = \boldsymbol{I}_{N_{t,q}}$). Note that the number of data symbols transmitted during different OFDM symbols of different TTIs can vary. In fact, in an LTE/LTE-A frame there are OFDM symbols and/or OFDM subcarriers that are dedicated to the transmission of pilots, synchronization, and control channels. In order to simplify the mathematical notation used in this work, from this point onwards we will assume that on each TTI there are a fixed amount of N_i OFDM symbols completely dedicated to data transmission and that each of this OFDM symbols contains N_d data subcarriers. This simplifying assumption can produce very slight quantitative differences with respect to results obtained using the real frame structure specified by LTE/LTE-A but, certainly, the qualitative conclusions drawn from both approaches would be exactly the same. Furthermore, using this simplifying assumption, the analytical framework can focus on a generic OFDM symbol of a generic TTI and thus, from this point onwards, except otherwise stated, the subindexes t and q will be omitted without sacrificing analytical accuracy.

Depending on the number of transmit antennas, the vector of data symbols $\boldsymbol{s} \in \mathbb{C}^{N_d \times 1}$, which has to be transmitted on a generic OFDM symbol of a generic TTI, is split into N_d/N_T blocks that are encoded using an orthogonal space-frequency block-coding (SFBC) Alamouti scheme [74] to obtain a matrix $\boldsymbol{C} = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{N_d}] \in \mathbb{C}^{N_T \times N_d}$ of SFBC encoded symbols, with $\boldsymbol{c}_s = [c_{1,s}, \ldots, c_{N_T,s}]$. For instance, assuming a per-subcarrier uniform power allocation (UPA), the power allocated to each of the N_c useful subcarriers is $P_s = P_T/N_c$ and the nth block of SFBC symbols, with

[2]Using spatial multiplexing in a multicast scenario requires that the MIMO matrices characterizing the propagation between the BS and all the MSs in the multicast group have a rank equal or greater than the number of spatial streams being multicast.

$n \in \{0, \ldots, N_d/N_T - 1\}$, is obtained as

$$c_{1,n} = \sqrt{P_s} s_s, \tag{3.1}$$

for the trivial case in which $N_T = 1$, and

$$\begin{bmatrix} c_{1,2n} & c_{1,2n+1} \\ c_{2,2n} & c_{2,2n+1} \end{bmatrix} = \sqrt{\frac{P_s}{2}} \begin{bmatrix} s_{2n} & s_{2n+1} \\ -s^*_{2n+1} & s^*_{2n} \end{bmatrix}, \tag{3.2}$$

for $N_T = 2$.

The symbols $\boldsymbol{C}$ in the frequency domain are converted to the time domain by using a per-transmit antenna inverse fast Fourier transform (IFFT) and furthermore, a CP is added to each OFDM symbol. After propagation through the wireless MIMO channel, the CP part is removed from the time-domain received signal and, again, a per-receive antenna fast Fourier transform (FFT) is used to obtain the frequency domain received samples. Very common simplifying assumptions used in the analysis of OFDM-based systems are the consideration of ideal synchronization and sampling processes at the receiver side and the use of CPs of duration greater than the maximum delay spread of the channel impulse response. Under these assumptions, the system is not affected neither by ISI nor by inter-carrier interference (ICI). Hence, the received samples at the output of the N_{Rm} processing stages of MS m can be expressed by the $N_{Rm} \times N_d$ complex-valued matrix $\boldsymbol{Y}_m$, whose sth column contains the vector of N_{Rm} signal samples received by user m on subcarrier s, which can be expressed as

$$\boldsymbol{y}_{m,s} = \boldsymbol{H}_{m,s}\boldsymbol{c}_s + \boldsymbol{\zeta}_{m,s} + \boldsymbol{\eta}_{m,s}, \tag{3.3}$$

where

$$\boldsymbol{\zeta}_{m,s} = \sum_{k=1}^{K-1} \boldsymbol{H}^{(k)}_{m,s}\boldsymbol{c}^{(k)}_s \tag{3.4}$$

represents the multi-cell interference term, $\boldsymbol{\eta}_{m,s} \sim \mathcal{CN}_{N_{Rm},1}\left(\boldsymbol{0}, \sigma^2_\eta \boldsymbol{I}_{N_{Rm}}\right)$ denotes the additive white Gaussian noise (AWGN) vector and $\boldsymbol{H}^{(k)}_{m,s} =$

$\left[\boldsymbol{h}_{m,s,1}^{(k)}, \ldots, \boldsymbol{h}_{m,s,N_T}^{(k)}\right]$ is the complex-valued $N_{R_m} \times N_T^{(k)}$ MIMO channel matrix characterizing the propagation loss (including transmit and receive antenna gains), large-scale shadow fading and small-scale time/frequency/space selective fading between the kth BS and MS m on subcarrier s. For later convenience, $h_{m,s,n_t}^{(k)}$ is defined as the channel vector characterizing the propagation between the transmit antenna n_t of the BS k and the MS m on subcarrier s. Furthermore, note that, in order to simplify notation, the superindex has been removed from the channel matrix $\boldsymbol{H}_{m,s}^{(0)}$ and the matrix of SFBC symbols $\boldsymbol{C}^{(0)}$ corresponding to the reference BS. The particular characteristics of the MIMO channel models used in this work will be fully specified in the numerical results of the developed SRA proposals in the following chapters.

Finally, each of the N_d/N_T blocks of received samples on the N_{Rm} receive antennas of user m are processed using a properly designed space-frequency block decoder and a maximum ratio combiner (MRC) in order to obtain the vector of estimated data symbols $\boldsymbol{z} \in \mathbb{C}^{N_d \times 1}$. In fact, we have that

$$z_s^{(m)} = \sqrt{P_s} s_s \sum_{n_r=1}^{N_{Rm}} \left|h_{m,s,n_r,1}\right|^2 + \sum_{n_r=1}^{N_{Rm}} h_{m,s,n_r,1}^{*} \left(\zeta_{m,s,n_r} + \eta_{m,s,n_r}\right) \tag{3.5}$$

for the single transmit antenna case,

$$\begin{aligned} z_{2n}^{(m)} = {} & \sqrt{P_s} s_{2n} \sum_{n_r=1}^{N_{Rm}} \left(\left|h_{m,2n,n_r,1}\right|^2 + \left|h_{m,2n+1,n_r,2}\right|^2 \right) \\ & + \sqrt{P_s} s_{2n+1}^{*} \sum_{n_r=1}^{N_{Rm}} \left(h_{m,2n+1,n_r,1}^{*} h_{m,2n+1,n_r,2} - h_{m,2n,n_r,1}^{*} h_{m,2n,n_r,2} \right) \\ & \quad + \sqrt{2} \sum_{n_r=1}^{N_{Rm}} \Big(h_{m,2n,n_r,1}^{*} \left(\zeta_{m,2n,n_r} + \eta_{m,2n,n_r}\right) \\ & \qquad + h_{m,2n+1,n_r,2} \left(\zeta_{m,2n+1,n_r}^{*} + \eta_{m,2n+1,n_r}^{*}\right) \Big) \end{aligned} \tag{3.6}$$

for the symbol transmitted on subcarrier $s = 2n$ when $N_T = 2$, and

$$
\begin{aligned}
z_{2n+1}^{(m)} = & \sqrt{P_s} s_{2n+1} \sum_{n_r=1}^{N_{Rm}} \left(\left| h_{m,2n+1,n_r,1} \right|^2 + \left| h_{m,2n,n_r,2} \right|^2 \right) \\
& + \sqrt{P_s} s_{2n}^* \sum_{n_r=1}^{N_{Rm}} \left(h_{m,2n,n_r,1}^* h_{m,2n,n_r,2} - h_{m,2n+1,n_r,1}^* h_{m,2n+1,n_r,2} \right) \\
& + \sqrt{2} \sum_{n_r=1}^{N_{Rm}} \Big(h_{m,2n+1,n_r,1}^* \left(\zeta_{m,2n+1,n_r} + \eta_{m,2n+1,n_r} \right) \\
& + h_{m,2n,n_r,2} \left(\zeta_{m,2n,n_r}^* + \eta_{m,2n,n_r}^* \right) \Big)
\end{aligned} \tag{3.7}
$$

for the symbol transmitted on subcarrier $s = 2n + 1$ when $N_T = 2$. Notice that, except for channels with strong frequency selectivity, $h_{m,2n,n_r,n_t} \approx h_{m,2n+1,n_r,n_t}$ and therefore, the second term in (3.6) and (3.7), corresponding to the self-interference term produced by the SFBC scheme, is practically negligible.

Assuming a block-fading channel [75], where the duration of a block corresponds to one TTI, all symbols received by user m on subcarrier s during a generic TTI are characterized by the same SINR that can be expressed as

$$
\gamma_s^{(m)} = \frac{P_s \left\| h_{m,s,1} \right\|^4}{h_{m,s,1}^H \left(\sum_{k=1}^{K-1} \frac{P_s}{N_T^{(k)}} H_{m,s}^{(k)} H_{m,s}^{(k)H} \right) h_{m,s,1} + \sigma_\eta^2 \left\| h_{m,s,1} \right\|^2} \tag{3.8}
$$

for the single transmit antenna case,

$$
\gamma_{2n}^{(m)} = \frac{\frac{P_s}{2} \left(\left\| h_{m,2n,1} \right\|^2 + \left\| h_{m,2n+1,2} \right\|^2 \right)}{\sigma_{C m,2n}^2 + \sigma_{M m,2n}^2 + \sigma_{N m,2n}^2} \tag{3.9}
$$

for the symbol transmitted on subcarrier $s = 2n$ when $N_T = 2$, where

$$
\sigma_{C m,2n}^2 = \frac{P_s}{2} \left| h_{m,2n+1,1}^H h_{m,2n+1,2} - h_{m,2n,1}^H h_{m,2n,2} \right|^2 , \tag{3.9a}
$$

$$\sigma^2_{Mm,2n} = h^H_{m,2n,1}\left(\sum_{k=1}^{K-1}\frac{P_s}{N_T^{(k)}}H^{(k)}_{m,2n}H^{(k)H}_{m,2n}\right)h_{m,2n,1} + h^H_{m,2n+1,2}\left(\sum_{k=1}^{K-1}\frac{P_s}{N_T^{(k)}}H^{(k)}_{m,2n+1}H^{(k)H}_{m,2n+1}\right)h_{m,2n+1,2}, \tag{3.9b}$$

and

$$\sigma^2_{Nm,2n} = \sigma^2_\eta\left(\|h_{m,2n,1}\|^2 + \|h_{m,2n+1,2}\|^2\right) \tag{3.9c}$$

are used to denote the power of the self-interference term produced by the SFBC process, the power of the intercell interference term and the power of the AWGN term, respectively. Similar SINR expression corresponding to the symbol transmitted on subcarrier $s = 2n+1$ when $N_T = 2$ has also been obtained as

$$\gamma^{(m)}_{2n+1} = \frac{\frac{P_s}{2}\left(\|h_{m,2n+1,1}\|^2 + \|h_{m,2n,2}\|^2\right)}{\sigma^2_{Cm,2n+1} + \sigma^2_{Mm,2n+1} + \sigma^2_{Nm,2n+1}} \tag{3.9}$$

where

$$\sigma^2_{Cm,2n+1} = \frac{P_s}{2}\left|h^H_{m,2n,1}h_{m,2n,2} - h^H_{m,2n+1,1}h_{m,2n+1,2}\right|^2, \tag{3.10a}$$

$$\sigma^2_{Mm,2n+1} = h^H_{m,2n+1,1}\left(\sum_{k=1}^{K-1}\frac{P_s}{N_T^{(k)}}H^{(k)}_{m,2n+1}H^{(k)H}_{m,2n+1}\right)h_{m,2n+1,1} + h^H_{m,2n,2}\left(\sum_{k=1}^{K-1}\frac{P_s}{N_T^{(k)}}H^{(k)}_{m,2n}H^{(k)H}_{m,2n}\right)h_{m,2n,2}, \tag{3.10b}$$

and

$$\sigma^2_{Nm,2n+1} = \sigma^2_\eta\left(\|h_{m,2n+1,1}\|^2 + \|h_{m,2n,2}\|^2\right). \tag{3.10c}$$

3.2.2 Modeling CSI feedback

Decisions taken by the RRM procedure at the multicasting BS are grounded on the CSI fed back by the MSs on standardized signaling/control channels.

Due to feedback channel limitations, the amount of CSI that standards allow to be sent from the MSs to the BS is quite low. In particular, the MSs in LTE/LTE-A networks can be configured to report CQIs to assist the BS in scheduling users and allocating resources. The CQI reports are derived from the received SINRs typically measured on the downlink reference signals and their granularity is determined by defining a number of subbands spanning the entire system bandwidth, each consisting of a fixed number of contiguous PRBs. In cases were only one subband (spanning the whole system bandwidth) is used, the single CQI reported by the MS is denoted as a wideband CQI. Otherwise, the different CQI reports are denoted as subband CQIs. Full feedback CQI reporting is obtained using subbands containing a single PRB. As will be shown in the numerical results obtained with the SRA proposal presented in Chapter 5, the higher the resolution of CQI reporting in the frequency domain the better the exploitation of the frequency selective RRM gain. It will be assumed in this work that the BS can choose from a set of $J = 15$ MCSs, with their corresponding CQIs, whose main characteristics, very similar to those of some of the MCSs used in LTE/LTE-A systems, are listed in Table 3.1 (results presented in this table have been obtained from [76, Chapter 5]).

Let us model the CSI feedback of a generic user m that has split the system bandwidth into $N_S^{(m)}$ subbands. To this end, denote by N_{PRB} the number of PRBs spanning the system bandwidth, then the number of PRBs per subband of user m can be obtained as ${N_{\text{PRB}}}_S^{(m)} = N_{\text{PRB}}/N_S^{(m)}$, where we have assumed that the involved parameters have been selected in such a way that this is an integer division. Now, let us define the set $\mathcal{S}_b^{(m)}$ as the one containing all the subcarriers conforming subband b of user m, where $b \in \{1, \ldots, N_S^{(m)}\}$. Each of the subcarriers $s \in \mathcal{S}_b^{(m)}$ may experience a different SINR $\gamma_s^{(m)}$ and hence, this particular subband is affected by a multistate channel characterized by the vector of SINRs

$$\boldsymbol{\gamma}^{(m)} \left(\mathcal{S}_b^{(m)} \right) \triangleq \left[\gamma_s^{(m)} \right]_{\forall \ s \in \mathcal{S}_b^{(m)}} . \tag{3.11}$$

Table 3.1: Modulation scheme, effective code rate, spectral efficiency and SINR boundaries for a $\mathrm{BLER}_0 = 0.1$ when using $N_{\mathrm{PRB}}{}_{S}^{(m)} = 1$, 4 or 25 PRBs.

MCS	**Modulation**	τ_{c_j}	**b/s/Hz**	$\Gamma_j(1)$	$\Gamma_j(4)$	$\Gamma_j(25)$
0	–	–	–	-∞	-∞	-∞
1	QPSK	0.076	0.152	-6.76	-6.93	-7.58
2	QPSK	0.120	0.234	-4.85	-5.21	-5.78
3	QPSK	0.190	0.396	-2.66	-3.31	-3.76
4	QPSK	0.300	0.602	-0.92	-1.45	-1.77
5	QPSK	0.440	0.878	1.00	0.48	0.17
6	QPSK	0.590	1.176	2.85	2.35	2.07
7	16QAM	0.370	1.476	4.69	4.23	3.99
8	16QAM	0.480	1.912	6.48	6.06	5.83
9	16QAM	0.600	2.408	8.43	8.03	7.91
10	64QAM	0.450	2.730	10.31	9.86	9.76
11	64QAM	0.550	3.324	12.13	11.75	11.64
12	64QAM	0.650	3.900	14.03	13.64	13.65
13	64QAM	0.750	4.524	16.01	15.54	15.58
14	64QAM	0.850	5.118	17.93	17.47	17.48
15	64QAM	0.930	5.556	19.87	19.39	19.46

Based on $\gamma^{(m)}\left(\mathcal{S}_b^{(m)}\right)$, a CQI metric is selected that can be used by the BS to determine the best MCS on this subband with the objective of maximizing the spectral efficiency subject to a maximum target BLER. Mathematically, denoting by $\mathcal{J}$ the set of available MCSs in the system, the instantaneous BLER experienced by user m when receiving a transport block on the set of $N_{\mathrm{PRB}}{}_{S}^{(m)}$ PRBs conforming subband $\mathcal{S}_b^{(m)}$ and using MCS $j \in \mathcal{J}$, can be expressed as

$$\mathrm{BLER}_j^{(m)}\left(\mathcal{S}_b^{(m)}\right) = \mathcal{F}_j\left(\gamma^{(m)}\left(\mathcal{S}_b^{(m)}\right)\right), \tag{3.12}$$

where $\mathcal{F}_j(\cdot)$ represents an MCS-dependent mapping function relating the

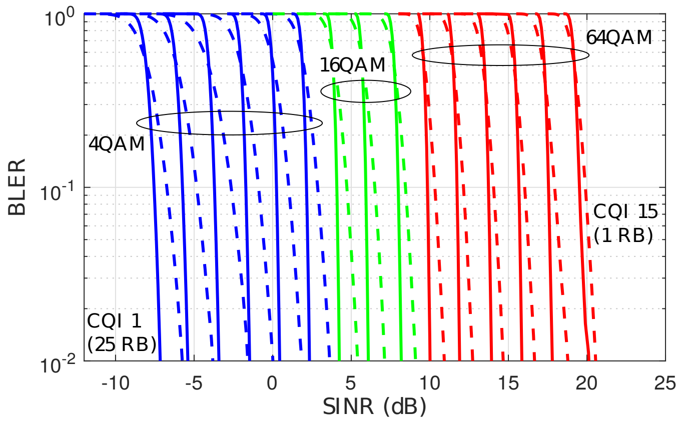


Figure 3.2: AWGN BLER curves versus SINR for the 15 LTE CQIs without HARQ (dashed and solid lines correspond, respectively, to $N_{\mathrm{PRB}}{}_S^{(m)} = 1$ PRB and $N_{\mathrm{PRB}}{}_S^{(m)} = 25$ PRBs) [76].

multistate channel vector to a BLER level. Thus, the CQI for subband b of user m can be determined by solving the optimization problem

$$\begin{aligned} \alpha_{S_b^{(m)}}^{(m)} &= \arg\max_{j \in \mathcal{J}} \tau_j \\ &\text{subject to } \mathcal{F}_j\left(\gamma^{(m)}\left(\mathcal{S}_b^{(m)}\right)\right) \le \mathrm{BLER}_0, \end{aligned} \tag{3.13}$$

where τ_j denotes the data rate (measured, for example, in b/s/PRB) of MCS j. If none of the MCSs in $\mathcal{J}$ can satisfy the BLER constraint, then $\alpha_{S_b^{(m)}}^{(m)} = 0$. Obviously, any PRB in subband b of user m will be characterized by the CQI of this subband. For later use, let us define the set $\mathcal{P}_p^{(m)}$ as the one containing all the subcarriers in PRB p of user m. Using this definition, the CQI of the PRBs in subband b of user m can be obtained as $\alpha_{\mathcal{P}_p^{(m)}} = \alpha_{\mathcal{S}_b^{(m)}}^{(m)}\ \forall \mathcal{P}_p^{(m)} \subseteq \mathcal{S}_b^{(m)}$.

Unfortunately, it is very difficult to find closed-form mapping functions $\mathcal{F}_j(\cdot)$ that accurately relate the vector of SINRs $\gamma^{(m)}\left(\mathcal{S}_b^{(m)}\right)$ to a single BLER value for an arbitrary multistate channel realization and a given MCS $j \in \mathcal{J}$. In order to overcome this problem, several link abstraction techniques have been developed that, taking as input the vector of SINRs, obtain a single scalar quality value $\gamma_{\mathrm{eff}}{}_j^{(m)}\left(\mathcal{S}_b^{(m)}\right)$, denoted as effective SINR,

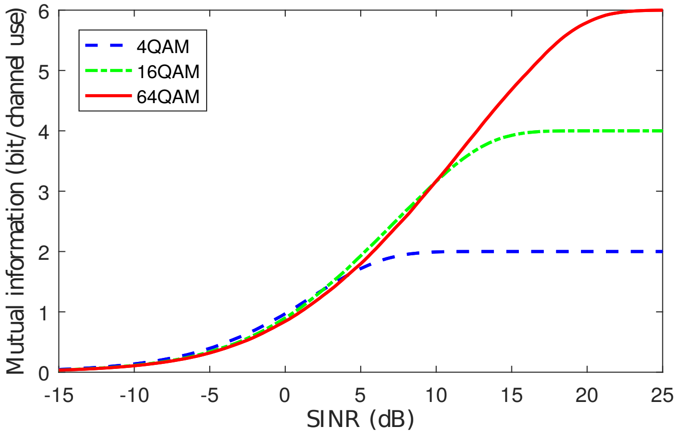


Figure 3.3: BICM capacity as a function of the SINR for the 4QAM, 16QAM and 64QAM modulation formats.

that can be easily associated to $\text{BLER}_j^{(m)}\left(\mathcal{S}_b^{(m)}\right)$ using a set of reference look-up tables (LUTs) containing the BLER values associated to the different CQIs and subband sizes in AWGN channels [77]. As an example, Figure 3.2 shows the AWGN BLER curves of the transmission modes representing the 15 LTE/LTE-A CQIs for $N_{\text{PRB}}{}_S^{(m)} = 1$ PRB (dashed lines) and $N_{\text{PRB}}{}_S^{(m)} = 25$ PRBs (solid lines). Note that, the larger the subband the higher is the turbo code block size that can be applied and the steeper is the slope of the waterfall region of the corresponding BLER curves.

One of the most popular choices for link abstraction is the mutual information effective SINR mapping (MIESM), which has the advantage over alternative methods, such as exponential effective SINR mapping (EESM), that it does not require an empirical off-line calibration step as long as codes that perform close to capacity are employed (e.g., turbo codes or low-density parity check (LDPC) codes) [78]. Using the MIESM-based approach, the effective SINR experienced by user m when uses MCS mode j on a multistate channel characterized by $\boldsymbol{\gamma}^{(m)}\left(\mathcal{S}_b^{(m)}\right)$ can be obtained as

$$\gamma_{\text{eff}}{}_j^{(m)}\left(\mathcal{S}_b^{(m)}\right) \quad = J_{\mathfrak{m}(j)}^{-1}\left(\frac{1}{N_d N_{\text{PRB}}{}_S^{(m)}} \sum_{\forall\, s \in \mathcal{S}_b^{(m)}} J_{\mathfrak{m}(j)}\left(\gamma_s^{(m)}\right)\right), \qquad (3.14)$$

where $\mathfrak{m}(j)$ denotes the modulation format used in MCS j and $J_{\mathfrak{m}(j)}(\cdot)$ denotes the bit-interleaved coded modulation (BICM) capacity for the modulation format $\mathfrak{m}(j)$, which can be obtained by numerically evaluating [79, eq. (14)] via the Monte Carlo method. Figure 3.3 shows the BICM capacity versus SINR for quadrature amplitude modulation (QAM), 16QAM and 64QAM over the AWGN channel. It is worth mentioning that all MCSs sharing the same modulation format experience the same effective SINR.

Using the previously described strategy, the optimization problem posed in (3.13) simplifies to

$$\begin{aligned} \alpha^{(m)}_{S^{(m)}_b} = &\arg\max_{j \in \mathcal{J}} \quad \tau_j \\ &\text{subject to } \psi^{(j)}_{N_{\mathrm{PRB}}{}^{(m)}_S}\left({\gamma_{\mathrm{eff}}}^{(m)}_j\left(\mathcal{S}^{(m)}_b\right)\right) \leq \mathrm{BLER}_0, \end{aligned} \tag{3.15}$$

where $\psi^{(j)}_{N_{\mathrm{PRB}}{}^{(m)}_S}\left({\gamma_{\mathrm{eff}}}^{(m)}_j\left(\mathcal{S}^{(m)}_b\right)\right)$ represents the curve used to map the effective SINR ${\gamma_{\mathrm{eff}}}^{(m)}_j\left(\mathcal{S}^{(m)}_b\right)$ onto a predicted BLER value for a given MCS j used on a subband with ${N_{\mathrm{PRB}}}^{(m)}_S$ PRBs (see the curves plotted in Figure 3.2). Notice that this problem can be simplified as

$$\begin{aligned} \alpha^{(m)}_{S^{(m)}_b} = &\arg\max_{j \in \mathcal{J}} \quad \tau_j \\ &\text{subject to } {\gamma_{\mathrm{eff}}}^{(m)}_j\left(\mathcal{S}^{(m)}_b\right) \geq \Gamma_j\left({N_{\mathrm{PRB}}}^{(m)}_S\right), \end{aligned} \tag{3.16}$$

where $\Gamma_j\left({N_{\mathrm{PRB}}}^{(m)}_S\right)$ denotes the effective SINR value (also known as effective SINR boundary) for which MCS j provides a BLER equal to BLER_0 when transmitting on ${N_{\mathrm{PRB}}}^{(m)}_S$ PRBs. As an example, the SINR boundaries for the 15 MCSs that will be used in this work are listed in Table 3.1 for ${N_{\mathrm{PRB}}}^{(m)}_S = 1$, 4 and 25 PRBs and $\mathrm{BLER}_0 = 0.1$ (see also the results presented in Figure 3.2).

3.2.3 Scheduling and resource allocation for multicast service using subgroup optimization

As previously stated in this section, when dealing with a multicast service, the SRA strategies can be described as decision making algorithms that, at the beginning of a given TTI and based on the CQI received from the served MSs, aim at the optimization of utility functions related to spectral and/or energy efficiency with prescribed QoS constraints. In particular, it will be assumed in this work that the reference BS (or the reference MBSFN with several BSs) aims at providing maximal ADR with constraints, first, on the minimum sustained data rate that has to be allocated to the whole set of users in the multicast group and, second, on the maximum BLER that the multicast data flow can tolerate at any of the MSs. The ADR of a multicast group can be obtained as the product of the number of group members that can correctly decode the multicast information by the effective data rate (measured in b/s) allocated to that particular group. The effective data rate basically depends on the number of PRBs allocated to the multicast group and the data rate (measured in b/s/PRB) characterizing the MCS also allocated to that group. Hence, for a given number of allocated PRBs, if the BS could use a given MCS irrespective of the number of users in the group, the higher the number of group members the higher the ADR would be. The problem, however, is that the distribution of users on the coverage area of the BS can be quite heterogeneous and thus, different members of the multicast group can be confronted by very dissimilar radio link qualities. Although users experiencing good channel propagation conditions could use MCSs providing high transmission rates, the multicast group members experiencing poor channel propagation conditions can only support the use of very robust MCSs providing low transmission rates. Consequently, as the group member experiencing the worst channel conditions is the one that determines the MCS selected by the multicasting BS and, furthermore, the probability of having a multicast group member experiencing very poor channel propagation conditions increases with the number of users in the group, increasing the number of multicast group members does not always translate onto a proportional ADR increase. A solution to this problem consists of splitting the multicast group into disjoint subgroups, allocating disjoint sets of PRBs to each subgroup and using an appropriate MCS to multicast information to the members in each subgroup. Obviously, this should be done with the aim of maximizing the ADR while, at the same time, providing the minimum required data rate to the worst multicast subgroup and guarantying a BLER less or equal than the target one to all the members in the multicast group.

Let us assume a particular SRA period in which the whole set of users $\mathcal{M}$ has been split into M_G disjoint multicast subgroups indexed by the sets $\mathcal{M}^{(g)}$, with $g \in \{1, \ldots, M_G\}$. Let us also define the PRB and MCS allocation sets as $\mathcal{N}_B = \left\{\mathcal{N}_B^{(1)}, \ldots, \mathcal{N}_B^{(M_G)}\right\}$ and $\boldsymbol{j} = \left\{j^{(1)}, \ldots, j^{(M_G)}\right\}$, with $\mathcal{N}_B^{(g)}$ and $j^{(g)}$ denoting, respectively, the set of PRBs (or equivalently, the set of subcarriers in these PRBs) and the MCS allocated to the gth multicast subgroup. Using these definitions, the ideal SRA optimization problem at hand could be formally posed as

$$\begin{aligned} \max_{M_G, \mathcal{N}_B, \boldsymbol{j}} \; & \sum_{g=1}^{M_G} d^{(g)} \sum_{\forall m \in \mathcal{M}_g} \left[1 - \mathrm{BLER}_{j^{(g)}}^{(m)}\left(\mathcal{N}_B^{(g)}\right)\right] \\ \text{subject to } & d^{(g)} \geq \tau_{\min} \quad \forall\, g \in \{1, \ldots, M_G\} \\ & \mathcal{M}^{(g)} \cap \mathcal{M}^{(h)} = \emptyset \quad \forall\, g \neq h \\ & \mathcal{N}_B^{(k)} \cap \mathcal{N}_B^{(i)} = \emptyset \quad \forall\, k \neq i \\ & \mathrm{BLER}_{j^{(g)}}^{(m)}\left(\mathcal{N}_B^{(g)}\right) \leq \mathrm{BLER}_0 \quad \forall\, m,\; g \end{aligned} \tag{3.17}$$

where, denoting by $N_{\mathrm{PRB}}^{(g)}$ the number of PRBs allocated to multicast subgroup g, and the data rate per PRB corresponding to the MCS j assigned to multicast subgroup g as $\tau_{j^{(g)}}$, the corresponding data rate (measured in b/s) of subgroup g can be obtained as

$$d^{(g)} = N_{\mathrm{PRB}}^{(g)}\, \tau_{j^{(g)}}. \tag{3.18}$$

The first constraint in (3.17) ensures that all the users in the system are allocated a data rate greater or equal than τ_{min}, the second constraint guarantees that a given MS belongs to a single multicast subgroup, the third constraint enforces a given PRB to be allocated to a single multicast subgroup and the fourth one is used to ensure that the average BLER experienced by a given user is always upper bounded by the target value BLER_0. Using the concept of effective SINR and applying a similar approach to that described in Section 3.2.2, optimization problem (3.17) can be rewritten as

$$\begin{aligned} \max_{M_G, \mathcal{N}_B, \boldsymbol{j}} \; & \sum_{g=1}^{N_G} d^{(g)} \sum_{\forall m \in \mathcal{M}^{(g)}} \left[1 - \psi_{N_{\text{PRB}}^{(g)}}^{(j^{(g)})} \left({\gamma_{\text{eff}}}_{j^{(g)}}^{(m)} \left(\mathcal{N}_B^{(g)} \right) \right) \right] \\ \text{subject to } & d^{(g)} \geq \tau_{\text{min}} \quad \forall\, g \in \{1, \ldots, M_G\} \\ & \mathcal{M}^{(g)} \cap \mathcal{M}^{(h)} = \emptyset \quad \forall\, g \neq h \\ & \mathcal{N}_B^{(k)} \cap \mathcal{N}_B^{(i)} = \emptyset \quad \forall\, k \neq i \\ & {\gamma_{\text{eff}}}_{j^{(g)}}^{(m)} \left(\mathcal{N}_B^{(g)} \right) \geq \Gamma_{j^{(g)}} \left(N_{\text{PRB}}^{(g)} \right) \quad \forall\, m,\; g. \end{aligned} \tag{3.19}$$

Unfortunately, there are two major difficulties that have to be overcome when trying to solve this problem:

- On the one hand, RRM decisions are taken at the multicasting BS and the per-subcarrier SINR values needed to calculate the effective SINRs are only available at the receiver side. The only CSIs available at the BS are the wideband and/or subband CQIs and thus, if needed, a mechanism should be devised to obtain an estimation of the effective SINRs from the CQIs.

- On the other hand, the only way to optimally solve problem (3.19) is to carry out an exhaustive search over all possible combinations of multicast subgroups and allocations of PRBs and MCSs in order to find the constrained allocation maximizing the global ADR. Such an exhaustive search happens to be computationally intractable even for unrealistically basic multicast systems with very low numbers of PRBs and/or users. Hence, assumptions, tools and algorithms that eventually provide us with the necessary means to tackle this problem in a simpler yet effective way should be devised.

3.3 Group-based multicasting

As it has been previously mentioned, multicast subgrouping is a solution studied in the literature to RRM in broadcast/multicast transmissions. The main contributions included in this thesis have been developed to enhance the state of the art in RRM using group-based multicasting (see Chapters 4 to 6) and stream splitting techniques (see Chapter 6). The group-based multicasting strategy is based on the creation of different multicast subgroups to deliver one multicast service, and it has been adopted to overcome the limitations of both CMS and OMS [63]. This strategy allocates the available resources into different subgroups, minimizing the negative effects of users with poor channel conditions and serving all multicast members in the same time slot. The group-based multicasting strategy can be split into three different phases:

- **CQI collection:** first, the BS collects the CQI feedback from the MSs placed in the MBSFN area and which are demanding the multicast service. For each CQI feedback cycle (CFC), the resource allocation procedure carried out in one of the the BSs of the SFN, creates a vector $\boldsymbol{\alpha} = \{\alpha_m;\ m = 1, \ldots, M\}$ with the CQIs reported by all the MSs.

- **Subgroup creation:** the multicast members are split into several multicast subgroups. Each multicast subgroup delivers the service using different MCS and, consequently, serving the users that support the decoding of this scheme fulfilling the target BLER requirements, e.g. LTE-LTE-A standard establishes a target $\text{BLER}_0 = 0.1$ [5].

- **Resource allocation:** the resource allocation procedure works in such a way that UE that reports a CQI will be served by the multicast subgroup closer to the reported CQI and whose MCS can be decoded by the user. Resources must be allocated in such a way that every user is served by a multicast subgroup whose MCS the user can correctly decode.

The utilization of a cost function allows the second and the third steps of group-based multicasting to make the subgroup creation and the resource allocation based on the optimization of such cost function. Multiple parameters may be taken into account to develop the most adequate cost function. The most widely used cost functions in the literature [63] are the following:

- Maximum throughput (MT) that is employed to maximize the service ADR and can be formulated in the addressed scenario as

$$\Pi^{MT} = \arg\max_{r_j,\, x_{m,j}} \left\{ \sum_{j=1}^{J} \tau_j r_j \sum_{m=1}^{M} x_{m,j} \right\}. \tag{3.20}$$

- Proportional fairness (PF) that improves the fairness among the users at the time a high service ADR is achieved, and can be formulated in the addressed scenario as

$$\Pi^{PF} = \arg\max_{r_j,\, x_{m,j}} \left\{ \sum_{j=1}^{J} \sum_{m=1}^{M} \log\left(\tau_j r_j x_{m,j}\right) \right\}, \tag{3.21}$$

  where the vector $\boldsymbol{\tau} = \{\tau_j;\ j = 1, \ldots, J\}$ denotes the data rate per PRB (measured in b/s/PRB) corresponding to each available MCS j, the vector $\boldsymbol{r} = \{r_j;\ j = 1, \ldots, J\}$ the number of PRBs allocated to the multicast subgroup assigned to MCS j, and the binary matrix $\boldsymbol{X} = \{x_{m,j};\ m = 1, \ldots, M;\ j = 1, \ldots, J\}$ is introduced to describe which MCS j user m is assigned to, from those MCSs which the user is able to decode, such that

$$x_{m,j} = \begin{cases} 1, & \text{if user } m \text{ is assigned to MCS } j \\ 0, & \text{otherwise.} \end{cases} \tag{3.22}$$

Next, Chapters 4 to 6 of this thesis present novel SRA strategies that employ the system model described in the Section 3.2. These strategies have been developed to maximize the ADR (see MT utility function in Equation (3.20)) taking advantage of the following environments:

- Joint resource allocation for multicast and unicast transmissions (Chapter 4).
- Multicast resource allocation based on subband CQI feedback (Chapter 5).
- Multicast resource allocation based on multiple video layers (Chapter 6).

3.4 Summary

This chapter has presented an exhaustive analysis of the state of the art in broadcast/multicast resource allocation solutions. Single-rate and multi-rate multicast transmission schemes have been referenced, citing the most relevant works available in the literature.

The system model employed in the development and analysis of the SRA contributions made in this thesis has been detailed. An abstraction of the physical layer has been derived by means of the effective SINR, which allows as to model the CSI feedback. This model is employed to formulate the optimization problem that maximizes the ADR of a multicast service using subgroups.

The multicast SRA strategy based on subgroups, which is employed in the contributions made in this thesis, is described. Additionally, the most widely employed utility functions with this approach are presented.

The system model presented in this chapter is detailed on the following submitted work:

- **Alejandro de la Fuente**, Guillem Femenias, Felip Riera-Palou, and Ana García Armada, "Subband CQI feedback-based Multicast Resource Allocation in MIMO-OFDMA Networks", *Submitted to IEEE Transactions on Broadcasting.*

Chapter 4

Joint multicast/unicast resource allocation strategies

The growing demand for video services in mobile networks poses new challenges in the design of techniques to improve the data rate and the delay required to provide those services. These techniques must guarantee the scalability for large amount of users and the reliability of the transmission to everyone, every time and everywhere. Broadcast and multicast transmissions are fully supported on both LTE and LTE-A systems through the use of the eMBMS [3, 14]. Furthermore, when deployed in SFN mode, also known as MBSFN, eMBMS can exploit a pre-existing infrastructure to provide broadcast/multicast service coverage over an area simultaneously served by a flexible number of BSs transmitting on the same frequency channel.

Implemented as a subsystem of LTE/LTE-A, the eMBMS shares the physical layer resources with the unicast transmission mode. In fact, LTE/LTE-A networks (at least up to Release 12) use TDM to aggregate broadcast/multicast and unicast services using the PMCH with a time allocation to eMBMS services limited to a maximum of a 60 percent [21]. The remaining frames (at least 40 percent) are reserved for unicast transmissions using the PDSCH. The LTE/LTE-A frame/subframe structure for FDD is

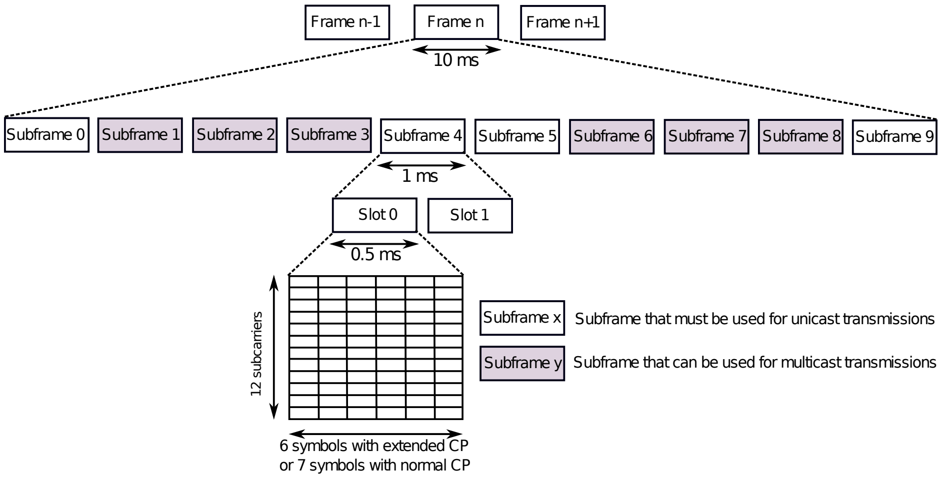


Figure 4.1: LTE-FDD frame/subframe structure

illustrated in Figure 4.1.

It is worth mentioning that using a system that has been optimized for unicast services to deliver broadcast/multicast transmissions imposes some trade-offs that, in the end, compromise the system performance in terms of spectral/energy efficiency and offered QoE. Even though there seems to be a consensus in 3GPP working groups to remove this TDM-related constraint in LTE Release 14 specifications, thus opening the door to supporting efficient stand-alone eMBMS networks [13, 15], there are still some challenging issues associated to the RRM processes.

One of the most important challenge that the research community are currently working on consists in the development of resource allocation strategies, as it has been detailed in the prior chapter, with the aim of optimizing the utilization of multicast transmissions. This chapter presents SRA strategies that combine the allocation of multicast and unicast resources in the LTE/LTE-A frame in order to achieve the maximum ADR among all the multicast users of the service.

4.1 Joint multicast/unicast scheduling with dynamic optimization

We have developed a JMUS strategy to maximize the ADR delivered to all the multicast users in the MBSFN area [80][1]. The proposed technique combines unicast and multicast transmissions to guarantee a minimum data rate, τ_{min}, for all the users demanding a multicast service[2].

Let us consider a multicast service that is delivered in an MBSFN area using a dedicated bandwidth. An LTE-like system can use multicast or unicast transmissions to provide the service to all the users. In such a way, this JMUS proposal aims to achieve the optimal compromise between unicast and multicast transmissions to maximize the ADR of the multicast group, guaranteeing the given QoS requirements. To this end, the optimal MCS and the optimal number of subframes reserved for multicast transmission are obtained each LTE frame; furthermore, the unicast QoS-aware metric proposed in [81] to guarantee the data rate is employed to allocate the remaining resources.

4.1.1 JMUS problem formulation

The proposed JMUS strategy takes into account the data rate achieved both in the multicast and the unicast transmissions in order to maximize the ADR.

On the one hand, it is required to calculate the data rate of the users which are able to receive the multicast service based on the selected MCS $j^{(\mu)}$ satisfying the target $\text{BLER}_0 = 0.1$ requirements. The multicast data

[1]This work was started prior to the beginning of this PhD thesis.

[2]Note that by multicast service we refer to a streaming or downloading service simultaneously delivered to all the users in the system, while we denote by multicast transmission when the BS uses the PMCH (at least up to Release 12) to send the same data to all the users and by unicast transmissions when BS uses PDSCH to send the data to each MS [5].

rate for MCS $j^{(\mu)}$ can be calculated as

$$d^{(j^{(\mu)})} = \frac{u^{(\mu)}}{N_f} \tau_{j^{(\mu)}} N_{\text{PRB}} \tag{4.1}$$

where $d^{(j^{(\mu)})}$ denotes the multicast data rate (measured in b/s) corresponding to MCS $j^{(\mu)}$, $u^{(\mu)} \in \{0, \dots, N_m\}$ the number of subframes used for the multicast transmission, $\tau_{j^{(\mu)}}$ the data rate (measured in b/s/PRB) of MCS $j^{(\mu)}$ and N_{PRB} the number of available PRBs. Note that $N_f = 10$ and $N_m = 6$ denote the total number of subframes and the maximum number of them that can be used for multicast transmissions respectively, as the LTE/LTE-A standards define [21].

On the other hand, the remaining subframes in the LTE frame, $N_f - u^{(\mu)}$, are allocated to unicast transmissions. The data rate for the unicast transmission to user m that can correctly decode the MCS $j^{(m)}$ is given as

$$d_m^{(v)} = \frac{u_m}{N_f}^{(v)} r_m^{(v)} \tau_{j^{(m)}} \tag{4.2}$$

where the vector $\boldsymbol{d}^{(v)} = \{d_m^{(v)};\ m = 1, \dots, M^{(v)}\}$ denotes the data rate (measured in b/s) of the $M^{(v)}$ users served with unicast transmissions, the vectors $\boldsymbol{r}^{(v)} = \{r_m^{(v)};\ m = 1, \dots, M^{(v)}\}$ and $\boldsymbol{u}^{(v)} = \{u_m^{(v)};\ m = 1, \dots, M^{(v)}\}$ denote the number of PRBs and the amount of subframes assigned to each user during the unicast transmissions, respectively.

The optimization problem results in maximizing the ADR for the multicast service, taking into account several constraints. The maximization problem can be expressed as

$$\underset{j^{(\mu)}, u^{(\mu)}, r_m^{(v)}, u_m^{(v)}}{\text{maximize}} \quad M^{(\mu)} d^{(j^{(\mu)})} + \sum_{m=1}^{M^{(v)}} d_m^{(v)} \tag{4.3}$$

$$\text{subject to} \quad M^{(\mu)} + M^{(v)} = M \tag{4.3a}$$

$$u^{(\mu)} \in \{0, \dots, N_m\} \tag{4.3b}$$

$$u_m^{(v)} \in \{0, \ldots, N_f\} \tag{4.3c}$$

$$d^{(j^{(\mu)})} \geq \tau_{min} \tag{4.3d}$$

$$d_m^{(v)} \geq \tau_{min} \qquad \forall m \in \mathcal{M}^{(v)} \tag{4.3e}$$

$$\sum_{m=1}^{M^{(v)}} r_m^{(v)} \frac{u_m^{(v)}}{N_f} \leq N_{\text{PRB}}(1 - \frac{u^{(\mu)}}{N_f}) \tag{4.3f}$$

where M is the total number of users that are split between those served using multicast, $M^{(\mu)}$, and those with unicast transmissions and $M^{(v)}$ (4.3a). Both $M^{(\mu)}$ and $M^{(v)}$ depend on the MCS, denoted as $j^{(\mu)}$, that is selected for multicast transmission. The number of multicast subframes are limited by the LTE standard and is given as (4.3b), whereas the number of subframes that can be dedicated for unicast transmissions to each user m is given as (4.3c). The minimum data rate requirements for multicast and unicast transmissions, respectively, are established in (4.3d) and (4.3e). Finally, (4.3f) limits the resources allocated to the vectors $\boldsymbol{r}^{(v)}$ and $\boldsymbol{u}^{(v)}$.

The goal of the JMUS strategy is to find the optimal values of $j^{(\mu)}$ and $u^{(\mu)}$ that maximize the ADR at each frame transmission. The unicast QoS-aware scheduling must also guarantee the optimal allocation of vectors $\boldsymbol{r}^{(v)}$ and $\boldsymbol{u}^{(v)}$ to serve with unicast transmissions the users that experience worst channel conditions.

4.2 Joint strategy for LTE resource allocation: multicast subgrouping and unicast transmissions

The JMUS strategy detailed above can be enhanced taking advantage of the multicast subgrouping technique. We have proposed a novel JMSUT strategy [82], so that it is in charge of the RRM to provide a multicast service in an LTE-like system. This approach combines the unicast transmissions using the LTE subframes reserved for that purpose, and multicast subgrouping for the transmission in the remaining subframes. The goal of the JMSUT strategy, as well as JMUS, is to maximize the ADR whereas it guarantees the fulfillment of the QoS requirements of every user.

Considering the delivery of a multicast service in the conditions detailed in prior section, that is an MBSFN area of an LTE-like system using a dedicated bandwidth, the proposed JMSUT strategy employs multicast and unicast transmissions to provide the service to all the users maximizing the ADR. On the one hand, the JMSUT scheme searches the optimal resource allocation in the multicast subframes, splitting the users into multicast subgroups where the service is delivered using different MCSs. On the other hand, the JMSUT employs the unicast QoS-aware scheduling proposed in [81] to deliver the service using unicast transmissions to the MSs experiencing worst channel conditions. Consequently, the JMSUT aims to maximize the service ADR and, at the same time, guarantees the QoS requirements for all the users demanding the service.

The proposed JMSUT strategy can be split into the same three phases described in Section 3.3 for group-based multicasting:

- **CQI collection:**

  For each CFC, the BS creates a vector $\boldsymbol{\alpha} = \{\alpha_m;\ m = 1, \ldots, M\}$ with the CQIs reported by all the MSs.

- **Multicast subgroup creation:**

  The proposed JMSUT splits the multicast members into different multicast subgroups, and the members with worst channel conditions can be attended using unicast transmissions. Each multicast subgroup delivers the service using different MCS.

- **Joint multicast and unicast resource allocation:**

  Radio resources available to deliver the service depend on the bandwidth reserved for that purpose. These PRBs are allocated in multicast subframes (up to 6 subframes of the LTE frame) and unicast subframes.

> The proposed SRA scheme works so that if a MS reports a CQI that is equal or greater than the lowest multicast subgroup with allocated PRBs, this user will be served by the multicast subgroup closer to the reported CQI and whose MCS can be decoded by the user. On the other hand, only the users which report a CQI lower than the MCS employed in the most robust multicast subgroup will be served by means of the reserved subframes for unicast transmissions in the LTE frame.

Consequently, this strategy is based on the MT optimization problem (see (3.20)). This problem presents several constraints, such as the minimum data rate requirements per user, or the number of available PRBs to deliver the service. The following subsection briefly describes the formulation of the proposed JMSUT.

4.2.1 JMSUT problem formulation

The proposed JMSUT strategy aims to maximize the achieved ADR taking into account the data rate both in the multicast and the unicast transmissions. This strategy is based on the maximization of a utility function that consists in the sum of the data rate of all the members demanding the multicast service. Furthermore, some QoS requirements must be guaranteed regarding the τ_{min} per every user.

On the one hand, it is required to calculate the data rate of the users which are able to receive the multicast service from any of the available multicast subgroups satisfying the target $\mathrm{BLER}_0 = 0.1$ requirements. Each subgroup delivers the service using a different MCS, and the vector $\boldsymbol{r}^{(\mu)} = \{r_j^{(\mu)};\ j = 1, \ldots, J\}$ denotes the number of PRBs assigned to each MCS j. The ADR for multicast transmissions can be calculated as

$$d^{(\mu)} = u^{(\mu)} \sum_{j=1}^{J} \tau_j r_j^{(\mu)} \sum_{m=1}^{M^{(\mu)}} x_{m,j}. \tag{4.4}$$

Recall that $u^{(\mu)} \in \{0, \ldots, N_m\}$ denotes the number of subframes used for the multicast transmission, the vector $\boldsymbol{\tau} = \{\tau_j;\ j = 1, \ldots, J\}$ the data rate per PRB (measured in b/s/PRB) corresponding to each available MCS j, and the binary matrix $\boldsymbol{X} = \{x_{m,j};\ m = 1, \ldots, M;\ j = 1, \ldots, J\}$ describes which MCS j is user m assigned to (see (3.22)).

On the other hand, the remaining time resources, $N_f - u^{(\mu)}$, in the LTE frame are allocated to unicast transmissions. The data rate for the unicast transmission to user m that can correctly decode the MCS $j^{(m)}$ is given as JMUS problem (see (4.2)).

Consequently, the optimization problem results in maximizing the ADR for the multicast service, taking into account not only the constraints explained before in JMUS, but also the introduction of multicast subgrouping results in new constraints Hence, the JMSUT maximization problem can be expressed as

$$\underset{r_j^{(\mu)}, u^{(\mu)}, r_m^{(\upsilon)}, u_m^{(\upsilon)}}{\text{maximize}} \quad d^{(\mu)} + \sum_{m=1}^{M^{(\upsilon)}} d_m \tag{4.5}$$

$$\text{subject to} \quad M^{(\mu)} + M^{(\upsilon)} = M \tag{4.5a}$$

$$u^{(\mu)} \in \{0, \ldots, N_m\} \tag{4.5b}$$

$$u_m^{(\upsilon)} \in \{0, \ldots, N_f\} \tag{4.5c}$$

$$u^{(\mu)} \tau_j r_j^{(\mu)} \geq \tau_{min} \qquad \forall j : r_j > 0 \tag{4.5d}$$

$$d_m^{(\upsilon)} \geq \tau_{min} \qquad \forall m \in \mathcal{M}^{(\upsilon)} \tag{4.5e}$$

$$\sum_{m=1}^{M^{(\upsilon)}} r_m^{(\upsilon)} \frac{u_m^{(\upsilon)}}{N_f} \leq N_{\text{PRB}}\left(1 - \frac{u^{(\mu)}}{N_f}\right) \tag{4.5f}$$

$$\sum_{j=1}^{\alpha_m} x_{m,j} = 1 \qquad \forall m \in \mathcal{M}^{(\mu)} \tag{4.5g}$$

$$x_{m,j} = 0 \qquad \forall j \geq \alpha_m + 1 \tag{4.5h}$$

$$\sum_{j=1}^{J} r_j \leq N_{\mathrm{PRB}} \tag{4.5i}$$

$$\frac{1}{M^{(\mu)}} \sum_{m=1}^{M^{(\mu)}} x_{m,j} \leq r_j \leq N_{\mathrm{PRB}} \sum_{m=1}^{M^{(\mu)}} x_{m,j} \qquad \forall j \in \mathcal{J} \tag{4.5j}$$

$$x_{m,j} \in \{0, 1\} \qquad \forall m \in \mathcal{M}^{(\mu)},\ \forall j \in \mathcal{J} \tag{4.5k}$$

$$r_j \in \{0, N_{\mathrm{PRB}}\} \qquad \forall j \in \mathcal{J} \tag{4.5l}$$

where (4.5a) to (4.5f) are the same constraints detailed in (4.3a) to (4.3f) for JMUS. Additionally, new constraints are included because of multicast subgrouping utilization, what implies that each user has to be assigned an MCS lower or equal than the reported CQI α_m (4.5g); and they can never be assigned an MCS higher than α_m (4.5h). The sum of PRBs allocated to each MCS is limited to the number of available PRBs (4.5i). Constraints (4.5j) to (4.5l) assure to assign PRBs only to subgroups with at least one user [30], that is

$$r_j = 0, \qquad \text{if} \sum_{m=1}^{M^{(\mu)}} x_{m,j} = 0$$

$$1 \leq r_j \leq N_{\mathrm{PRB}}, \quad \text{if} \sum_{m=1}^{M^{(\mu)}} x_{m,j} \geq 1.$$

4.3 Proposed algorithms for JMUS/JMSUT optimization problem

This section details the proposed algorithms to solve the JMUS and JMSUT optimization problems.

4.3.1 Exhaustive search scheme

An exhaustive search scheme (ESS) can be used to obtain the optimal values for JMUS and JMSUT maximization problems. Both SRA schemes use the knowledge of the MSs' CQI in the BS, and therefore the optimal MCS to each MS transmission. The ESS calculates the ADR of the multicast service using every MCS option and each one of different options in the number of subframes used for multicast transmissions, that is the complexity of an ESS for JMUS strategy is equal to $\mathcal{O}(JN_m)$. Each combination provides a different number of users that can be served by the multicast transmission (those users that can correctly decode the MCS used with the required BLER_0, and the remaining users require unicast transmissions for receiving the multicast service. When all the combinations are checked, the one which maximizes the ADR of the multicast service and fulfills the minimum data rate requirements for all the users is selected.

An ESS can also be used to achieve the optimal solution for the JMSUT scheme. The complexity of an ESS employed for a multicast subgrouping strategy is equal to $\mathcal{O}(N_{PRB}^J)$ [63]. Note that the computational cost required to carry out the SRA every TTI based on an ESS is prohibitive when the multicast service needs more than a few PRBs to be delivered. In the following chapters (see Chapters 5 and 6), we have proposed lower complexity solutions than the exhaustive search for the RRM when subgrouping strategies for multicast transmissions are employed.

4.3.2 Fast Search Scheme for JMUS

A fast search scheme (FSS) is proposed to reduce the computational complexity in the BS. This algorithm can achieve a suboptimal resource allocation for the JMUS strategy employing a much reduced number of iterations. First, an analysis of the problem feasibility has to be made to search a good starting point. This problem can be guaranteed to be feasible when the required τ_{min} is lower than the data rate that can be delivered using the most robust MCS, and the maximum number of available subframes for multi-

Algorithm 1 Fast search scheme: FSS

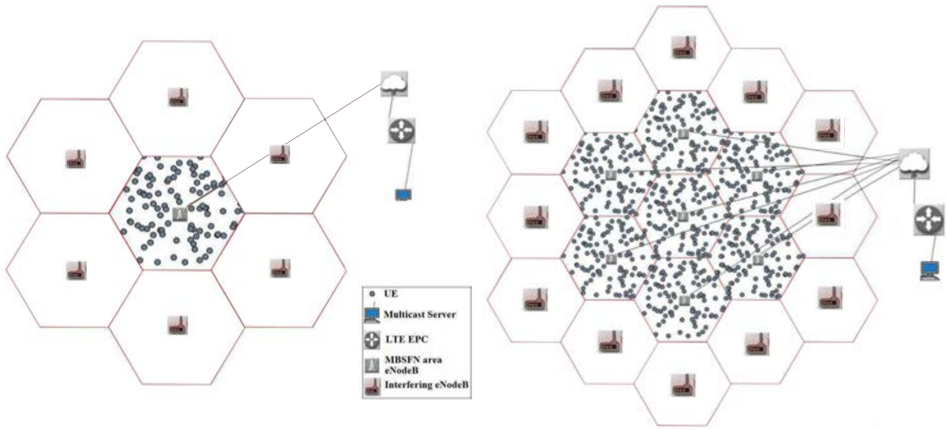
<u>**Inputs**</u>: $\boldsymbol{\alpha} = \{\alpha_m;\ m = 1, \ldots, M\}$, α_{min}, N_{PRB}
<u>**Initializations:**</u> $j = 0$, $N_m = 6$, $stop = false$, $ADR_{aux} = 0$
<u>**Search of $j^{(\mu)*}$ and $u^{(\mu)*}$ to maximize the ADR:**</u>
repeat
  repeat
    Compute $M^{(\mu)} : \alpha_m \geq j$
    Compute $M^{(\upsilon)} = M - M^{(\mu)}$
    for 1 to N_m **do**
      Compute $d^{(j^{(\mu)})} = \frac{u^{(\mu)}}{N_f}\tau_{j^{(\mu)}} N_{\text{PRB}}$ ▷ see (4.1)
    end for
    for N_m to N_f **do**
      Compute $d_m^{(\upsilon)} = \frac{u_m}{N_f}{}^{(\upsilon)} r_m^{(\upsilon)} \tau_{j(m)}$ ▷ see (4.2)
    end for
    for $m = 1, \ldots, M$ **do**
      if $\tau_m \leq \tau_{min}$ **then**
        $stop = true$
      end if
    end for
    Compute $ADR = M^{(\mu)} d^{(\mu)} + \sum_{m=1}^{M^{(\upsilon)}} d_m^{(\upsilon)}$ ▷ see (4.3)
    if $ADR \geq ADR_{aux}$ **then**
      $ADR_{aux} = ADR$
      $stop = false$
    else
      $stop = true$
    end if
    if $j < J$ **then**
      $j \leftarrow j + 1$
      $stop = false$
    else
      $stop = true$
    end if
  until $stop = true$
  if $N_m > 1$ **then**
    $N_m \leftarrow N_m - 1$
    $stop = false$
  else
    $stop = true$
  end if
until $stop = true$
$j^{(\mu)*} \leftarrow j - 1$
$u^{(\mu)*} \leftarrow \mu + 1$
<u>**Outputs**</u>: $j^{(\mu)*}$, $u^{(\mu)*}$

cast transmissions ($\tau_1 \times \frac{N_m}{N_f}$), i.e. all the MSs reporting a non-zero CQI can correctly decode the multicast transmission with the minimum required data rate. In order to prevent that the RRM problem could be unfeasible, the τ_{min} requirements must be selected according to the users' CQI distribution and the number of available resources N_{PRB}. Otherwise, the SRA algorithm may not satisfy the minimum requirements for each multicast user. After this consideration, the FSS algorithm should define a feasible starting point for the search of the suboptimal values [83]. Therefore, the most conservative MCS, $j^{(\mu)} = 1$, and the maximum number of multicast subframes, $u^{(\mu)} = N_m$, are selected as the starting point of the FSS. Next, the search algorithm looks for suboptimal values following these steps:

- Increase $j^{(\mu)}$ and check if the ADR is increased in the feasibility region. The feasibility region of this problem consists of the solutions that fulfill the τ_{min} requirements for all the MSs.
- Decrease $u^{(\mu)}$ and check if the ADR is increased in the feasibility region.
- If the ADR is indeed increased, increase $j^{(\mu)}$ and check if the ADR is increased in the feasibility region again. The new search starts from the $j^{(\mu)}$ value that maximizes the ADR for the $u^{(\mu)}$ value checked before.
- The search algorithm stops when $u^{(\mu)}$ is decreased and the ADR is not increased in the feasibility region.

4.4 Performance assessment

Let us consider that a multicast service is delivered in an LTE-like system. Two reference scenarios have been deployed to assess the SRA strategies employed. On the one hand, Figure 4.2a depicts a single-cell multicast scenario where one BS provides a multicast service to MSs which are uniformly distributed within the coverage area. Around this single-cell, a tier



(a) Single-cell multicast scenario. **(b)** Multi-cell multicast scenario.

Figure 4.2: Reference scenarios employed in JMUS/JMSUT assessment.

of six interfering BSs is deployed. On the other hand, Figure 4.2b illustrates a multi-cell multicast scenario, where seven BSs are coordinated in a single MBSFN area. The multicast service is provided in the 7-cell area to MSs which are uniformly distributed within each cell. Around the 7-cell MBSFN area, one tier of twelve BSs operating on the same frequency and transmission power as the seven BSs in the MBSFN area is deployed. The main parameters used in the simulations are illustrated in Table 4.1 and are based on the LTE standard. The available bandwidth is 3 MHz, thus 15 PRBs are available to deliver the service. One hundred multicast users are uniformly distributed in each cell, combining static and pedestrian (mobile users at 3 km/h) MSs. The SRA schemes have been employed with different QoS constraints, defined as τ_{min} requirements for all the multicast users[3].

4.4.1 Performance evaluation of JMUS scheme

We have compared the following SRA strategies in order to evaluate the performance of JMUS scheme:

[3] Note that this performance evaluation has employed a dedicated bandwidth of 3 MHz. The minimum data rate that is guaranteed for each user of the multicast service can be increased reserving a higher amount of resources (time and/or frequency).

Table 4.1: LTE system parameters to assess JMUS/JMSUT schemes.

Parameter	**Value**
Multi-cell system size	7 BSs
Interference model	1 tier of BSs
BSs geographical overlay	Hexagonal
Inter site distance	500 m
Transmission power	43 dBm
Antenna gain	11.5
Bandwidth	3 MHz
Number of PRBs	15
Downlink base frequency	2110 MHz
Pathloss model	3GPP Urban Macro cell
Multipath channel model	ITU Pedestrian B
BS transmission antennas	1
MSs per BS	100
MSs distribution	Uniform distribution
Pedestrian user speed	3 km/h

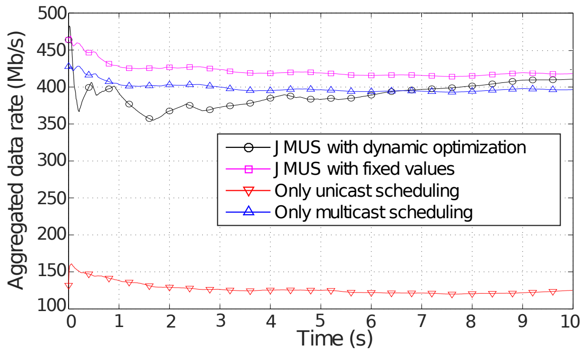
1. Pure unicast transmission with generic QoS-aware RRM as proposed in [81].

2. Pure multicast transmission with RRM using fixed values for $j^{(\mu)} = 7$ and $u^{(\mu)} = 6$.

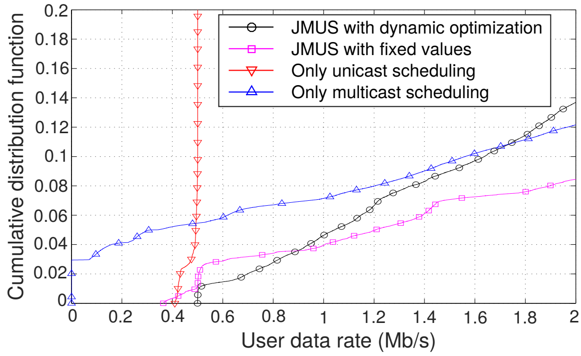
3. JMUS using fixed values for $j^{(\mu)} = 7$ and $u^{(\mu)} = 6$.

4. JMUS with dynamic optimization of the $j^{(\mu)}$ and $u^{(\mu)}$ values.

Figure 4.3a illustrates that the use of multicast transmissions clearly improves the achieved ADR using the pure unicast transmissions. Note that the use of the JMUS and the pure multicast RRM with fixed $j^{(\mu)}$ and $u^{(\mu)}$ values results in higher ADR performance than the JMUS with dynamic optimization. However, the QoS requirements imply the fulfillment of τ_{min}



(a) Aggregated data rate.



(b) CDF of the multicast users' data rate.

Figure 4.3: Performance assessment for different SRA strategies: unicast, multicast, fixed JMUS and dynamic JMUS, employing $\tau_{min} = 500$ kb/s

for each multicast user, what cannot be guaranteed employing non-dynamic SRA schemes as is depicted in Figure 4.3b. The cumulative distribution function (CDF) shows that the fulfillment of the QoS requirements, $\tau_{min} =$ 500 kb/s for each multicast user, can only be guaranteed using the JMUS with dynamic optimization. This goal cannot be ensured with the other SRA techniques employed in this performance assessment comparison.

Furthermore, an analysis of the influence of τ_{min} requirements has been assessed. Figure 4.4 depicts that increasing the τ_{min} requirements means that the feasibility of the maximization problem cannot be guaranteed, and consequently the search starting point may be not feasible. However, as it can be observed in Figure 4.4, using the JMUS with dynamic optimization

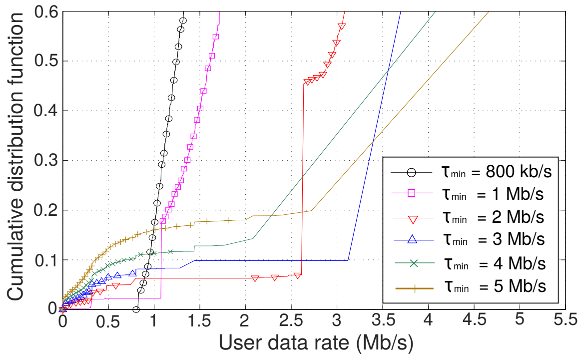


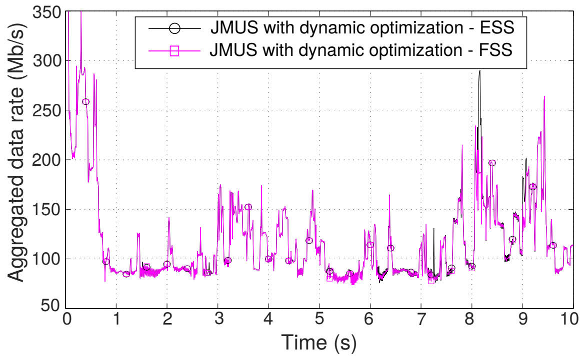
Figure 4.4: CDF of multicast users' data rate for different τ_{min} requirements.

allows us to guarantee 3 Mb/s for more than 90 percent of the multicast users.

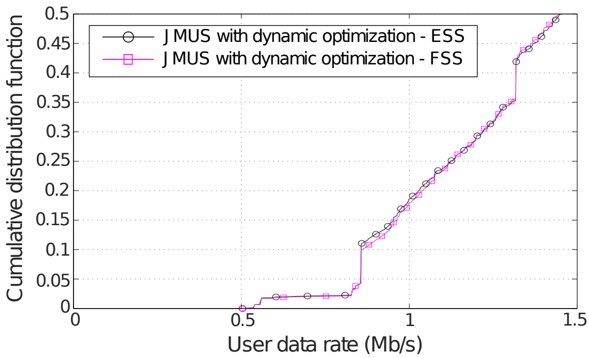
4.4.2 Comparison between ESS and FSS for JMUS

The comparison between FSS and ESS for the dynamic optimization is presented in Figure 4.5a where the ADR is shown for both searching algorithms. It can be observed that the use of both algorithms implies to achieve almost the same results each transmission frame. These results are confirmed in Figure 4.5b, where the CDF of the MSs' data rate is depicted using both ESS and FSS algorithms.

Additionally, the use of the proposed FSS algorithm highly reduces the number of iterations required to achieve the optimal values of the RRM every LTE frame. For example, considering the number of available MCSs in an LTE-like system, $J = 15$, and the maximum number of subframes that can be used for multicast transmissions, $N_m = 6$, the ESS employed for JMUS requires 90 (15×6) iterations each LTE frame to achieve the optimal values of the SRA scheme, whereas FSS requires an average of 10.16 iterations, that is the complexity of FSS is one order of magnitude lower than ESS.



(a) Aggregated data rate.



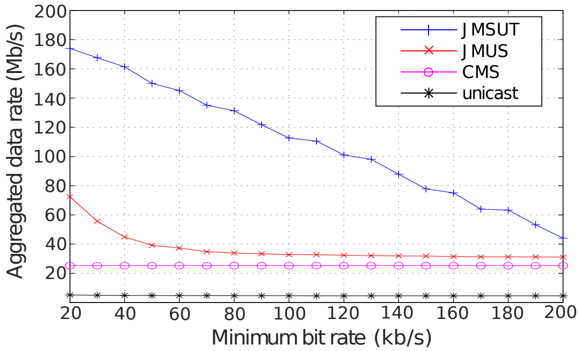
(b) CDF of the multicast users' data rate.

Figure 4.5: Performance comparison between ESS and FSS algorithms.

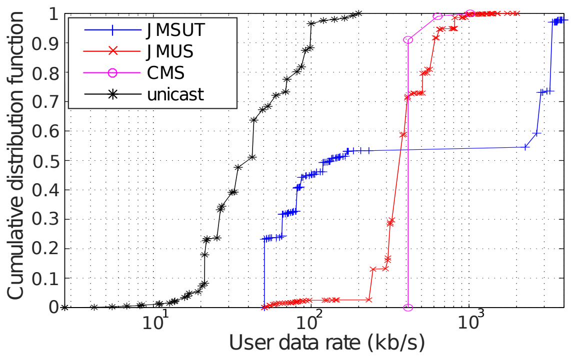
4.4.3 Performance comparison between JMUS and JMSUT

The performance comparison among the JMUS and the JMSUT has been performed in the multicast scenarios illustrated in Figure 4.2 using an ESS. The JMSUT, the JMUS, the CMS and the unicast SRA strategies are compared by means of two different assessments:

- The first evaluation consists in a comparison of different SRA strategies in a single-cell multicast scenario (Figure 4.2a). The results obtained using the proposed JMSUT and JMUS strategies are compared with the CMS and the use of only unicast transmissions.
- The second evaluation is based on the multi-cell multicast scenario



(a) Aggregated data rate.



(b) CDF of the multicast users' data rate, employing $\tau_{min} = 50$ kb/s.

Figure 4.6: Performance evaluation of JMSUT, JMUS, CMS and unicast transmissions in a scenario based on single-cell eMBMS using static and pedestrian users.

(Figure 4.2b). The results obtained in the central cell and a peripheral cell using both the JMSUT and JMUS strategies are compared. These results are also contrasted with the ones achieved in a single-cell multicast deployment (Figure 4.2a).

Comparison of SRA strategies in a single-cell multicast scenario

The first evaluation shows the results achieved using different SRA strategies in the single-cell multicast scenario with static and pedestrian users.

On the one hand, Figure 4.6a illustrates the ADR as a function of τ_{min}. It can be noticed that the use of multicast transmissions highly improves the

ADR of using only unicast transmissions. Nonetheless, the application of joint resource allocation techniques enhances the ADR results with respect of the most conservative multicast scheduling scheme (CMS), especially with the use of multicast subgrouping, since it can be observed how the JMSUT strategy results in important improvements in the service ADR over the use of the JMUS strategy. However, as τ_{min} is increased, this gain in the ADR is decreased. This is because the SRA strategy must ensure that the users with worst channel conditions reach these τ_{min} requirements, allocating more resources to the groups that are less efficient in terms of data rate.

On the other hand, Figure 4.6b shows the CDF of the achieved data rate per MS in this scenario, when τ_{min} is established to 50 kb/s. It can be noticed that the required τ_{min} for all the users cannot be guaranteed using only unicast transmissions. On the opposite side, the utilization of CMS guarantees the maximum fairness among all the users, but at the expense of the fact that the users with good channel conditions are not having benefit of it, and for that reason the ADR is low. On the other side, the JMSUT scheme not only allows all the users to achieve the τ_{min} requirements, but also the users that present good channel conditions can obtain higher data rates, and consequently the ADR of the service is greatly improved.

Note that the assessed strategies employ all the available resources in the LTE frame, either using only unicast transmissions or using a combination among multicast and unicast.

Comparison of SRA strategies in a multi-cell multicast scenario

The second evaluation illustrates the results obtained using the multi-cell multicast scenario with static and pedestrian users.

Figure 4.7 presents the service ADR achieved in the central cell and a peripheral one, using both the JMSUT and JMUS strategies. Furthermore, these results are compared with the ones achieved using the single-cell multicast scenario. It can be observed an important gain in the achieved ADR in the multi-cell scenario. The use of coordinated transmissions among 7-cells in an MBSFN area highly improves the channel conditions of the users

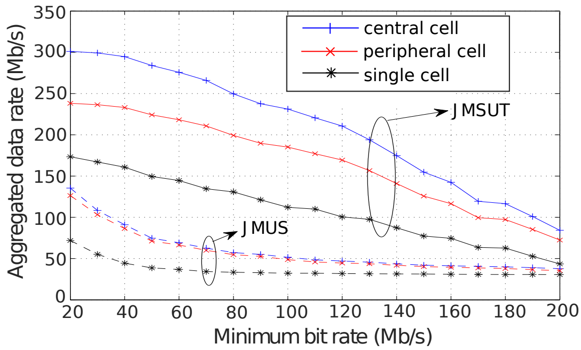


Figure 4.7: Performance evaluation of JMSUT and JMUS in single and multi cell scenario using static and pedestrian users.

in the cell-edge, especially in the central cell. In addition, this improvement in the channel conditions of the users leads to a higher gain using the JMSUT instead of the JMUS, since the τ_{min} required in the users with worst channel conditions can be fulfilled using a lower amount of resources.

4.5 Summary

The works presented in this chapter propose the use of a joint resource allocation strategy among the unicast and multicast subframes in LTE eMBMS service.

The JMUS strategy reaches the optimal values of the MCS and the number of subframes reserved for multicast transmissions in a dynamic way. The results achieved in the assessment show that JMUS can improve the performance of multicast services. This novel technique presents an important enhancement over pure unicast, pure multicast and non-dynamic SRA strategies to guarantee the QoS requirements of the multicast delivery. The pure unicast techniques can be used to guarantee a minimum data rate per user, however the achieved data rate is remarkably lower as compared to the use of the multicast techniques. On the other hand, the non-dynamic SRA techniques cannot guarantee a minimum data rate for all multicast users. We have demonstrated that the use of the JMUS with dynamic optimization allows the system to maximize the ADR taking into account the minimum data rate requirements.

We have proposed a fast search algorithm to implement the JMUS strategy. This algorithm has been evaluated achieving results very close to optimal values, and using one order of magnitude fewer iterations than an exhaustive search.

In order to enhance the JMUS performance, JMSUT proposes the creation of different multicast subgroups to allocate the available PRBs among them, what is employed in the multicast subframes. These multicast subgroups are combined with the transmissions in the reserved unicast subframes, which are used to serve the users experiencing worst channel conditions, in such a way this joint strategy maximizes the service ADR in the cell.

The performance assessment of the proposed algorithm shows how the JMSUT can greatly improve, in terms of service ADR, the results achieved using the CMS (i.e. one multicast group based on the user with the worst channel conditions), or the JMUS strategy (i.e. joint multicast and unicast transmissions using only one multicast group based on the joint optimization). At the same time, the JMSUT algorithm allows all the users to achieve the minimum required data rate. It is worth noting that the creation of different multicast subgroups causes that users with good channel conditions can be served using high data rates, and users with worse conditions can be served using the minimum data rate. Thus, the proposed algorithm allows the system to maximize the service ADR in the cell, at the time it guarantees a minimum service level for all the users.

Finally, the evaluation of the JMSUT in a multi-cell multicast scenario has presented significant higher service ADR than in single-cell multicast scenario. Note that this improvement is due to the fact that the channel conditions of the MSs are enhanced using a 7-cell MBSFN area, especially those users which are placed at cell-edge, leading to a higher gain in the service ADR when the JMSUT strategy is used.

This section has been based on the following published works:

- **Alejandro de la Fuente**, Ana García Armada, and Raquel Pérez Leal, "Joint Multicast/Unicast Scheduling with Dynamic Optimization for LTE Multicast Service", *20th European Wireless Conference (EW), May 2014.*
- **Alejandro de la Fuente**, Raquel Pérez Leal, and Ana García Armada, "Joint Strategy for LTE Resource Allocation: Multicast Subgrouping & Unicast Transmissions", *JITEL, Oct 2015.*

Chapter 5

Subband CQI feedback-based multicast resource allocation in MIMO-OFDMA networks

Most of the existing SRA algorithms for broadcast/multicast services are based on the use of either oversimplistic wideband CQI, or very conservative low complexity approaches using subband CQI. In order to go one step further with respect to the available solutions in the literature, in this chapter a new SRA framework that is based on a much more elaborate use of subband CQI is presented. The main contributions of this work can be summarized as follows:

- Based on the proposed physical layer abstraction in Section 3.2, a novel multicast SRA framework using channel-aware subgrouping is presented aiming at maximizing the ADR with constraints, first, on the minimum sustained data rate that has to be allocated to the whole set of users in the multicast group and, second, on the maximum BLER that the multicast data flow can support at any of the MSs. In order to solve such a complex optimization problem, a mechanism to derive an estimation of the so called effective SINRs from the subband CQIs is devised. Furthermore, the computational burden associated to the creation of an indeterminate number of multicast subgroups is avoided by constraining the proposal to the formation of only two subgroups. In a first stage, users in the first subgroup are guaranteed the minimum prescribed bit rate using as few resources as possible, whereas in the second stage the rest of users are allocated the remaining resources with the aim of maximizing the system ADR.

- Even this rather simplified two-step two-subgroup optimization problem is still combinatorial in nature and requires of a brute-force exhaustive search whose computational complexity is prohibitive even for modest values of the number of multicast users and PRBs. Consequently, a set of reduced-complexity algorithms providing different operating points on the performance versus complexity plane are proposed. Remarkably, the proposed novel subband CQI-based algorithms provide a substantial performance improvement over those available in the literature.

- A comprehensive performance evaluation of the proposed multicast SRA algorithms is conducted using a LTE-like MIMO-OFDMA multicellular system model. Numerical results clearly show the benefits that the use of subband CQI brings along when allocating the most adequate PRBs to each multicast subgroup under a wide range of setups, including the use of different system bandwidths, channel models, spatial correlation profiles and/or MIMO configurations.

5.1 Using subband CQIs to estimate the effective SINR

Recall that the CQI fed back by user m corresponding to subband b was denoted by $\alpha^{(m)}_{\mathcal{S}_b^{(m)}}$ (see Section 3.2.2). Let us assume that $\alpha^{(m)}_{\mathcal{S}_b^{(m)}} = j \in \mathcal{J}$, implying that MS m has selected MCS j as the most appropriate transmission mode on subband $\mathcal{S}_b^{(m)}$ in trying to guarantee a BLER less or equal than BLER_0 while maximizing the throughput. In fact, a $\alpha^{(m)}_{\mathcal{S}_b^{(m)}} = j$ ensures that ${\gamma_{\text{eff}}}_j^{(m)}\left(\mathcal{S}_b^{(m)}\right) \geq \Gamma_j({N_{\text{PRB}}}_S^{(m)})$ and thus, for $j \in \{1, \dots, J\}$, a suitable estimation for the SINR experienced by user m on any of the subcarriers in subband $\mathcal{S}_b^{(m)}$ can be obtained as (using linear units)

$$\hat{\gamma}_{s,j}^{(m)} = \Delta_j^{(m)} \Gamma_j \left({N_{\text{PRB}}}_S^{(m)} \right), \; \forall \, s \in \mathcal{S}_b^{(m)}, \tag{5.1}$$

where $\Delta_j^{(m)} \geq 1$ is a user- and MCS-dependent constant that can be used as a knob to regulate the amount of conservatism in the estimation process. The nearer the value of $\Delta_j^{(m)}$ is to one, the more conservative the estimation becomes. For the special case in which $\alpha_{\mathcal{S}_b^{(m)}}^{(m)} = 0$ we have that $\Gamma_0({N_{\text{PRB}}}_S^{(m)}) = 0$ and thus, instead of using (5.1) we propose

$$\hat{\gamma}_{s,0}^{(m)} = \Delta_0^{(m)} \Gamma_1 \left({N_{\text{PRB}}}_S^{(m)} \right), \; \forall \, s \in \mathcal{S}_b^{(m)}, \tag{5.2}$$

with $\Delta_0^{(m)} \leq 1$. In this particular case, the nearer the value of $\Delta_0^{(m)}$ is to one, the less conservative the estimation will be.

Now, using (5.1) and (5.2) in (3.14), it is quite obvious that ${\gamma_{\text{eff}}}_{j_g}^{(m)}\left(\mathcal{N}_{Bg}\right)$ can be estimated as

$$\widehat{\gamma_{\text{eff}}}_{j_g}^{(m)}\left(\mathcal{N}_{Bg}\right) = J_{\mathfrak{m}(j_g)}^{-1} \left(\frac{1}{N_d {N_{\text{PRB}}}_S^{(m)}} \sum_{\forall \, s \in \mathcal{S}_b^{(m)}} J_{\mathfrak{m}(j_g)} \left(\hat{\gamma}_{s,j_g}^{(m)} \right) \right). \tag{5.3}$$

5.2 Two-step two-subgroup optimization algorithm

A radical way of simplifying the problem of RRM for multicast service (see (3.19)) is to reduce the number of multicast subgroups to one and, hence, to set $M_G = 1$, $\mathcal{M}^{(1)} = \mathcal{M}$, $N_{\text{PRB}}^{(1)} = N_{\text{PRB}}$ and $\mathcal{N}_B^{(1)} = \mathcal{N}_B$. In this case, problem (3.19) reduces to selecting the optimal MCS to be allocated to the single multicast group, that is,

$$\begin{aligned}
&\underset{j^{(1)}}{\text{maximize}} \;\; d^{(1)} \sum_{\forall m \in \mathcal{M}} \left[1 - \psi_{N_{\text{PRB}}}^{(j^{(1)})} \left(\widehat{\gamma_{\text{eff}}}_{j^{(1)}}^{(u)} \left(\mathcal{N}_B \right) \right) \right] \\
&\text{subject to } d^{(1)} \geq \tau_{\min} \\
&\qquad\qquad \widehat{\gamma_{\text{eff}}}_{j^{(1)}}^{(m)} \left(\mathcal{N}_B^{(1)} \right) \geq \Gamma_{j^{(1)}} \left(N_{\text{PRB}} \right) \quad \forall\, m.
\end{aligned} \tag{5.4}$$

If this constrained optimization problem does not have a solution, a system outage is declared. Note that problem (5.4) can even be further simplified by assuming the use of the wideband CQIs reported by the MSs, denoted by $\alpha_B^{(m)}$ $\forall m \in \mathcal{M}$, because, in this case,

$$\begin{aligned}
&j^{(1)} = \min_{m \in \mathcal{M}} \alpha_B^{(m)} \\
&\text{subject to } d^{(1)} \geq \tau_{\min}.
\end{aligned} \tag{5.5}$$

This has been the solution usually implemented in the past to select the most appropriate MCS when multicasting information to a given group of MSs (see [58] and references therein). The main drawback of this approach, however, as will be shown later in the numerical results section, is that the global system performance is completely determined by the worst user in the system and, furthermore, the use of wideband CQI does not allow a proper exploitation of the channel frequency selectivity and/or the multiuser diversity. This simple SRA approach, usually termed in the literature as the CMS [84], will be used as one of the benchmark algorithms against which the performance of our proposed algorithms will be assessed.

Instead of adopting such a radical simplifying solution, we propose to capitalize on the availability of subband CQIs to implement a set of more powerful, yet still simple, two-stage two-subgroup optimization algorithms [30]. The fundamental intuition behind this proposal is that, in a first stage, an appropriate MCS $j^{(1)}$ should be determined with the aim of fulfilling the minimum data rate constraint of all multicast members in the system by using the minimum number of PRBs. This first stage problem can be simply stated as

$$\begin{aligned} &\underset{\mathcal{N}_B^{(1)},\, j^{(1)}}{\text{minimize}} \quad N_{\text{PRB}}^{(1)} \\ &\text{subject to } d^{(1)} \geq \tau_{\min} \\ &\qquad \widehat{\gamma_{\text{eff}}}_{j^{(1)}}^{(m)}\left(\mathcal{N}_B^{(1)}\right) \geq \Gamma_{j^{(1)}}\left(N_{\text{PRB}}^{(1)}\right) \quad \forall\, m. \end{aligned} \tag{5.6}$$

In those cases for which the first stage does not have a solution, a system outage is declared for this particular TTI. Otherwise, a second stage is carried out in which an MCS $j^{(2)}$ and the set $\mathcal{N}_B^{(2)} = \mathcal{N}_B \setminus \mathcal{N}_B^{(1)}$ of PRBs that have not been used in the first stage are allocated to the subgroup $\mathcal{M}^{(2)}$ of MSs making the most of them in terms of system ADR, while guarantying that they are allocated a data rate $d^{(2)} \geq \tau_{\min}$ and fulfill the target BLER constraint. That is,

$$\begin{aligned} &\underset{\mathcal{M}^{(2)},\, j^{(2)}}{\text{maximize}} \quad d^{(2)} \sum_{\forall m \in \mathcal{M}^{(2)}} \left[1 - \psi_{N_{\text{PRB}}^{(2)}}^{(j^{(2)})}\left(\widehat{\gamma_{\text{eff}}}_{j^{(2)}}^{(m)}\left(\mathcal{N}_B^{(2)}\right)\right)\right] \\ &\text{subject to } d^{(2)} \geq \tau_{\min} \\ &\qquad \widehat{\gamma_{\text{eff}}}_{j^{(2)}}^{(m)}\left(\mathcal{N}_B^{(2)}\right) \geq \Gamma_{j^{(2)}}\left(N_{\text{PRB}}^{(2)}\right) \quad \forall\, m. \end{aligned} \tag{5.7}$$

Note that the first multicast subgroup has then been reduced to $\mathcal{M}^{(1)} = \mathcal{M} \setminus \mathcal{M}^{(2)}$. On the occasions for which the second stage does not have a solution, a single multicast group is formed by solving problem (5.4).

Unfortunately, even the rather simplified constrained optimization problems defined in (5.6) and (5.7) are still combinatorial in nature. Hence, they require of a brute-force exhaustive search over all possible combinations of exclusive allocations of PRBs and subgroups of MSs, whose computational complexity is prohibitive even for modest values of the number of multicast users and PRBs. In the sequel we present a set of reduced-complexity algorithms that provide different operating points on the performance versus complexity plane.

Algorithm 2 First stage PRBs and MCS allocation: WCA

Inputs: $\alpha_B^{(m)}\ \forall\ m \in \mathcal{M}$, $\tau_{\min}$, $\mathcal{N}_B$
Allocation of MCS and PRBs to group $\mathcal{M}^{(1)}$:
$j^{(1)} = \min_{m\in\mathcal{M}} \left(\alpha_B^{(m)}\right)$
$N_{\text{PRB}}^{(1)} = \lceil \tau_{\min}/\tau_{j^{(1)}} \rceil$
if $N_{\text{PRB}}^{(1)} > N_{\text{PRB}}$ **then**
  $outage = true$
else
  $outage = false$, $d^{(1)} = \tau_{j^{(1)}} N_{\text{PRB}}^{(1)}$
  $\mathcal{N}_B^{(1)} \equiv$ Subset of $N_{\text{PRB}}^{(1)}$ PRBs of $\mathcal{N}_B$
end if
Outputs: $outage$, $j^{(1)}$, $\mathcal{N}_B^{(1)}$, $d^{(1)}$

5.2.1 First stage algorithms

Four different algorithms have been tested for the first stage of the two-step two-subgroup optimization approach, namely, the wideband CQI algorithm (WCA), the minimum subband CQI algorithm (MSCA), the sorted subband CQI algorithm (SSCA) and the greedy subband CQI algorithm (GSCA). The WCA and MSCA strategies are used to benchmark the SSCA and GSCA ones proposed in this work.

Wideband CQI algorithm (WCA) (see Algorithm 2)

The WCA-based approach was first proposed by Araniti *et al.* in [63, 30]. The RRM decisions taken by this algorithm are based on the wideband CQIs received from the MSs in the multicast group, that is, $\alpha_B^{(m)}\ \forall m \in \mathcal{M}$. The algorithm begins by determining which is the most spectral efficient MCS that can be supported by all the MSs in the system. Assuming the use of this MCS then the minimum number of PRBs it would require to fulfill the minimum data rate constraint is obtained. If the number of required PRBs is greater than the number of available PRBs, a system outage is declared. Otherwise, a subset of $N_{\text{PRB}}^{(1)}$ PRBs of $\mathcal{N}_B$ are allocated to multicast subgroup $\mathcal{M}^{(1)}$. Notice that due to the use of wideband CQIs the RRM procedure does not have any information about which are the most appropriate $N_{\text{PRB}}^{(1)}$ PRBs to be allocated to the multicast subgroup $\mathcal{M}^{(1)}$

and thus, the selection must be completely random. Furthermore, using the MCS corresponding to the received wideband CQI can only guarantee a BLER lower than BLER_0 when transmitting on the whole set of available PRBs. As only $N_{\mathrm{PRB}}^{(1)}$ PRBs are used, it can happen that some users in this multicast group experience a BLER larger than BLER_0 and, in this case, a BLER outage occurs.

Minimum subband CQI algorithm (MSCA) (see Algorithm 3)

This is a very conservative algorithm previously used by Condoluci *et al.* in [16] for LTE scenarios where multiple SVC streams are multicast to different groups of users. In this algorithm, RRM decisions are taken based on subband CQIs, that is, $\alpha_{\mathcal{S}_b^{(m)}}^{(m)}$ for all $m \in \mathcal{M}$ and for all $b \in \mathcal{S}_b^{(m)}$. Note that, on a given PRB $p \in \mathcal{N}_B$, the BS can reliably multicast data to any user $m \in \mathcal{M}$ by using the MCS associated to the CQI reported by the worst user on this PRB. Thus, the algorithm first obtains $\alpha_{\min}$ and $\alpha_{\max}$ defined, respectively, as the minimum and maximum of the minimum CQIs reported by the whole set of users over all PRBs in $\mathcal{N}_B$. Any MCS indexed by a CQI j such that $\alpha_{\min} \leq j \leq \alpha_{\max}$ can then be used by the BS to reliably multicast data on the set of PRBs $\mathcal{N}_{\mathrm{PRB}j} \triangleq \left\{ p : \min_{m \in \mathcal{M}} \alpha_{\mathcal{P}_p^{(m)}} \geq j \right\}$. The lower the index of the selected MCS the higher the number of PRBs necessary to multicast the data rate $\tau_{\min}$ will be. Thus, in order to minimize the number of PRBs used in the first stage, the algorithm determines the set of PRBs $\mathcal{N}_{\mathrm{PRB}j}$ that could support the use of the MCS indexed by CQI j, starting with $j = \alpha_{\max}$ and going down to $j = \alpha_{\min}$. If, eventually, the number of required PRBs is lower than the number of PRBs in $\mathcal{N}_{\mathrm{PRB}j}$, both MCS $j^{(1)} = j$ and a subset of $N_{\mathrm{PRB}}^{(1)}$ PRBs taken from $\mathcal{N}_{\mathrm{PRB}j}$ are allocated to multicast subgroup $\mathcal{M}^{(1)}$. Otherwise, a system outage is declared.

Sorted subband CQI algorithm (SSCA) (see Algorithm 4)

The MSCA-based approach is very conservative because the MCS used to multicast data on a given set of allocated PRBs is always the one that would

Algorithm 3 First stage PRBs and MCS allocation: MSCA

Inputs: $\alpha^{(m)}_{\mathcal{S}^{(m)}_b}$ $\forall\, m \in \mathcal{M}$ and $b \in \{1, \ldots, N^{(m)}_S\}$, $\tau_{\min}$, $\mathcal{N}_B$
Initializations: $outage = true$, $allocated = false$
$\alpha_{\min} \triangleq \min_{p \in \mathcal{N}_B} \left\{ \min_{m \in \mathcal{M}} \alpha_{\mathcal{P}^{(m)}_p} \right\}$
$\alpha_{\max} \triangleq \max_{p \in \mathcal{N}_B} \left\{ \min_{m \in \mathcal{M}} \alpha_{\mathcal{P}^{(m)}_p} \right\}$
Allocation of MCS and PRBs to group $\mathcal{M}^{(1)}$:
$j = \alpha_{\max}$
while ($j \geq \alpha_{\min}$) and ($allocated = false$) **do**
    $N^{(1)}_{\mathrm{PRB}} = \lceil \tau_{\min}/\tau_j \rceil$
    $\mathcal{N}_{\mathrm{PRB}\,j} \triangleq \left\{ p : \min_{m \in \mathcal{M}} \alpha_{\mathcal{P}^{(m)}_p} \geq j \right\}$
    if $N^{(1)}_{\mathrm{PRB}} \leq |\mathcal{N}_{\mathrm{PRB}\,j}|$ **then**
        $outage = false$, $allocated = true$, $j^{(1)} = j$, $d^{(1)} = \tau_{j^{(1)}} N^{(1)}_{\mathrm{PRB}}$
        $\mathcal{N}^{(1)}_B \equiv$ Subset of $N^{(1)}_{\mathrm{PRB}}$ PRBs of $\mathcal{N}_{\mathrm{PRB}\,j}$
    else
        $j = j - 1$
    end if
end while
Outputs: $outage$, $j^{(1)}$, $\mathcal{N}^{(1)}_B$, $d^{(1)}$

be allocated to reliably transmit to the worst user in the multicast group on the worst of the selected PRBs. Obviously, this strategy will never violate the target BLER constraint but this will be accomplished at the cost of a severe sacrifice in average ADR performance. In order to circumvent this drawback, the SSCA-based approach uses a less conservative strategy to select the set of PRBs to be used when transmitting using a given MCS. The rationale behind the proposed PRB selection strategy is that, although a given user may experience very bad channel conditions on some PRBs, it is not necessary to select the multicast MCS based solely on these worst PRBs. In fact, the same user may have other PRBs experiencing much better channel conditions that, when obtaining the joint effective SINR, can compensate for the bad behavior of the worst ones. Thus, one of the first steps of the SSCA consists of sorting the PRBs in descending order of the minimum expected spectral efficiency (i.e., max/min ordering). Certainly, there are other sorting strategies that can be used in this step but, among the ones we have tested, the max/min approach is the one providing the best

Algorithm 4 First stage PRBs and MCS allocation: SSCA

Inputs: $\alpha^{(m)}_{\mathcal{S}_b^{(m)}}$ $\forall\, m \in \mathcal{M}$ and $b \in \{1, \ldots, N_S^{(m)}\}$, $\tau_{\min}$, $\mathcal{N}_B$
Initializations: $outage = true$, $stop = false$
$\mathcal{P} \equiv$ Set of PRBs sorted in descending estimated minimum spectral efficiency order (max/min approach)
Allocation of MCS and PRBs to group $\mathcal{M}^{(1)}$:
$j = J$
while ($j \geq 1$) and ($stop = false$) **do**
  $N_{\text{PRB}}^{(1)} = \lceil \tau_{\min}/\tau_j \rceil$
  if $N_{\text{PRB}}^{(1)} \leq N_{\text{PRB}}$ **then**
    $\mathcal{N}_B^{(1)} \triangleq \left\{\mathcal{P}(1), \ldots, \mathcal{P}\left(N_{\text{PRB}}^{(1)}\right)\right\}$
    if $\widehat{\gamma_{\text{eff}}}_j^{(m)}\left(\mathcal{N}_B^{(1)}\right) \geq \Gamma_j\left(N_{\text{PRB}}^{(1)}\right)$ $\forall m \in \mathcal{M}$ **then**
      $outage = false$, $stop = true$, $j^{(1)} = j$, $d^{(1)} = \tau_{j^{(1)}} N_{\text{PRB}}^{(1)}$
    else
      $j = j - 1$
    end if
  else
    $outage = true$, $stop = true$
  end if
end while
Outputs: $outage$, $j^{(1)}$, $\mathcal{N}_B^{(1)}$, $d^{(1)}$

compromise between complexity and performance. In fact, as will be shown in the numerical results section, the SSCA-based approach provides performance metrics approaching those produced by the best proposed strategy at a much affordable computational complexity. Once the PRBs have been sorted, and in order to minimize the number of PRBs used in the first stage, the SSCA evaluates the different MCSs, starting with MCS $j = J$ and going down to MCS $j = 1$, in order to find the highest-order MCS able to fulfill both the minimum data rate and target BLER constraints. Note that the SSCA-based approach uses (5.1) to (5.3) to estimate the effective SINR experienced by the different users in the multicast group in order to check the (estimated) fulfillment of the BLER constraint. A system outage is declared either when the number of PRBs required by the selected MCS j to satisfy the minimum data rate constraint is greater than the number of available PRBs or when the algorithm is unable to fulfill the required constraints even when using the more robust MCS.

Greedy subband CQI algorithm (GSCA) (see Algorithm 5)

As it has been previously stated, the max/min sorting strategy used by the SSCA-based approach provides a good trade-off between performance and complexity. A way to increase the performance brought by the first stage of the multicast subgroup formation would be to exhaustively explore all the possible allocations of MCS and PRBs and select the one fulfilling the optimization constraints while using the minimum number of PRBs. The number of possible allocations to be explored when using such an exhaustive approach is upper-bounded by $J\left(2^{N_{\mathrm{PRB}}}-1\right)$, a value whose related complexity appears to be absolutely unaffordable for systems with any number of PRBs of practical interest. In these cases, a solution typically implemented to obtain near-optimal solutions at an acceptable complexity is the use of greedy allocation algorithms. The GSCA-based approach uses the effective SINR estimation procedure introduced in Section 5.1 to greedily select the minimum set of PRBs that can be used to reliably multicast data to the whole set of users in the multicast group while fulfilling the minimum data rate and BLER constraints. Note that, in this case, the selection of MCS and PRBs is based on a proper estimation of the system performance rather than on a pessimistic approach. In fact, assuming a subset $\mathcal{B}_p$ of PRBs, the corresponding CQIs for each user $m \in \mathcal{M}$ can easily be obtained from the reported subband CQIs and then, using (5.1) to (5.3) to estimate the effective SINR, an expression equivalent to that presented in (3.16) can be used to obtain the estimated CQI of user m on this set of PRBs, that is,

$$\begin{aligned}\widehat{\alpha}_{\mathcal{B}_p}^{(m)} = &\arg\max_{j\in\mathcal{J}} \tau_j \\ &\text{subject to } \widehat{\gamma_{\mathrm{eff}}}_j^{(m)}\left(\mathcal{B}_p\right) \geq \Gamma_j\left(\left|\mathcal{B}_p\right|\right).\end{aligned} \tag{5.8}$$

The algorithm begins with $\mathcal{N}_B^{(1)} = \emptyset$ and, in each iteration, it adds to this set the PRB that when combined with the previously selected ones, and after a proper selection of the corresponding MCS $j^{(1)}$, would provide the highest increase in the system ADR (i.e, greedy approach) while fulfilling the estimated BLER constraint. If, eventually, the data rate obtained in

Algorithm 5 First stage PRBs and MCS allocation: GSCA

Inputs: $\alpha^{(m)}_{\mathcal{S}^{(m)}_b}$ $\forall$ $m \in \mathcal{M}$ and $b \in \{1, \ldots, N^{(m)}_S\}$, $\tau_{\min}$, $\mathcal{N}_B$

Initializations: $stop = false$, $\mathcal{N}^{\text{free}}_B = \mathcal{N}_B$, $\mathcal{N}^{(1)}_B = \emptyset$, $d_{\text{aux}} = 0$

Allocation of MCS and PRBs to group $\mathcal{N}_1$:

while $(\mathcal{N}^{\text{free}}_B \neq \emptyset)$ and $(stop = false)$ **do**
  for $p \in \mathcal{N}^{(\text{free})}_B$ **do**
    $\mathcal{B}_p \triangleq \mathcal{N}^{(1)}_B \cup p, \quad \varpi_p \triangleq \min_{m \in \mathcal{M}} \widehat{\alpha}^{(m)}_{\mathcal{B}_p}$
  end for
  $p_{\text{sel}} = \arg\max_{p \in \mathcal{N}^{(\text{free})}_B} \varpi_p, \quad j^{(1)} = \max_{p \in \mathcal{N}^{(\text{free})}_B} \varpi_p$
  $d^{(1)} = \tau_{j^{(1)}} \left(\left| \mathcal{N}^{(1)}_B \right| + 1 \right)$
  if $d^{(1)} \geq d_{\text{aux}}$ **then**
    $\mathcal{N}^{(1)}_B = \mathcal{N}^{(1)}_B \cup p_{\text{sel}}$
    if $d^{(1)} \geq \tau_{\min}$ **then**
      $stop = true$, $outage = false$
    else
      $d_{\text{aux}} = d^{(1)}$, $\mathcal{N}^{(\text{free})}_B = \mathcal{N}_B \setminus \mathcal{N}^{(1)}_B$
    end if
  end if
end while

Outputs: $outage$, $j^{(1)}$, $\mathcal{N}^{(1)}_B$, $d^{(1)}$

one of these iterations happens to be higher than $\tau_{\min}$ then the algorithm allocates the corresponding selected MCS $j^{(1)}$ and the set of PRBs $\mathcal{N}^{(1)}_B$ to the first multicast group. The algorithm will stop and declare an outage when either the addition of a new PRB to the set of selected PRBs does not produce an increase of the system ADR or when the whole set of PRBs is exhausted without having fulfilled the minimum data rate constraint.

5.2.2 Second stage algorithms

The second stage algorithm is only executed when the first stage algorithm ends-up without having declared a system outage and with a non-empty set of non-allocated PRBs, that is with $\mathcal{N}^{(1)}_B \neq \mathcal{N}_B$. Furthermore, recall that if the second stage algorithm cannot provide a data rate greater or equal than $\tau_{\min}$ to the set of selected users then a single multicast group is formed by solving problem (5.4) or (5.5). In this subsection we propose two second stage algorithms, the first one is based on the use of wideband CQIs and the second one is based on the use of subband CQIs, namely, the second group wideband CQI algorithm (SG-WCA) and the second group subband CQI algorithm (SG-SCA), respectively.

Second group wideband CQI algorithm (SG-WCA) (see Algorithm 6)

Based on the algorithm proposed by Araniti *et al.* in [63, 30], the SG-WCA uses the wideband CQIs received from the MSs in the multicast group, that is, $\alpha_B^{(m)}$ $\forall m \in \mathcal{M}$, to select the MCS $j^{(2)}$ and the multicast subgroup $\mathcal{M}^{(2)}$ maximizing the estimated system ADR when transmitting on the set of PRBs $\mathcal{N}_B^{(2)} = \mathcal{N}_B \setminus \mathcal{N}_B^{(1)}$. The algorithm first determines which are both the MCSs $j_{\min}$ and $j_{\max}$ associated, respectively, to the minimum (except the no transmission case) and maximum wideband CQIs of the whole set of users in the multicast subgroup. Then, evaluates the utility ν_j (proportional to the estimated system ADR) provided by each available MCS $j \in \{j_{\min}, \ldots, j_{\max}\}$ and selects the MCS $j^{(2)}$ maximizing the utility. Once the MCS $j^{(2)}$ has been allocated to the second multicast subgroup, the data rate $d^{(2)}$ allocated to this subgroup can be calculated and compared to that obtained in the first stage. If $d^{(2)} \geq d^{(1)}$, the solution provided by the two-step two-subgroup optimization algorithm is accepted. Otherwise, a single multicast group is formed by solving problem (5.5).

Second group subband CQI algorithm (SG-SCA) (see Algorithm 7)

As previously stated, the wideband CQI metrics provide information on which is the best MCS to be used when transmitting on the whole system bandwidth. When the bandwidth over which the system has to transmit information is only a fraction of the system bandwidth, especially on those channels experiencing strong frequency selectivity, the wideband CQI is not a very reliable CSI metric. In these cases, subband CQI-based strategies can provide a clear advantage over wideband CQI-based ones. This is ba-

Algorithm 6 Second stage PRBs and MCS allocation: SG-WCA

Inputs: $\alpha_B^{(m)}\ \forall\, m \in \mathcal{M}$, $d^{(1)}$, $\mathcal{N}_B^{(2)} = \mathcal{N}_B \setminus \mathcal{N}_B^{(1)}$
Allocation of MCS and PRBs to group $\mathcal{N}_2$:
$j_{\min} = \max\left(\min_{m\in\mathcal{M}}\left(\alpha_B^{(m)}\right), 1\right)$
$j_{\max} = \max_{m\in\mathcal{M}}\left(\alpha_B^{(m)}\right)$
for $j = \{j_{\min}, \ldots, j_{\max}\}$ **do**
  $\mathcal{U}_j = \left\{m : \alpha_B^{(m)} \geq j\right\}, \quad \nu_j = |\mathcal{U}_j|\,\tau_j$
end for
$j^{(2)} = \arg\max_j \nu_j$, $\mathcal{N}_2 = \mathcal{U}_{j^{(2)}}$
$d^{(2)} = \left|\mathcal{N}_B^{(2)}\right| j^{(2)} \tau_{j^{(2)}}$
if $d^{(2)} \geq d^{(1)}$ **then**
  $Two-groups = true$
else
  $Two-groups = false$ (Solve problem (5.5))
end if
Outputs: $Two-groups$, $j^{(2)}$, $\mathcal{N}_2$

Algorithm 7 Second stage PRBs and MCS allocation: SG-SCA

Inputs: $\alpha_{\mathcal{S}_b^{(m)}}^{(m)}\ \forall\, m \in \mathcal{M}$ and $b \in \{1, \ldots, N_S^{(m)}\}$, $d^{(1)}$, $\mathcal{N}_B^{(2)} = \mathcal{N}_B \setminus \mathcal{N}_B^{(1)}$
Allocation of MCS and PRBs to group $\mathcal{N}_2$:
Calculate $\widehat{\alpha}_{\mathcal{N}_B^{(2)}}^{(m)}\ \forall\, m \in \mathcal{M}$
$j_{\min} = \max\left(\min_{j\in\mathcal{J}}\left(\widehat{\alpha}_{\mathcal{N}_B^{(2)}}^{(m)}\right), 1\right)$
$j_{\max} = \max_{m\in\mathcal{M}}\left(\widehat{\alpha}_{\mathcal{N}_B^{(2)}}^{(m)}\right)$
for $j = \{j_{\min}, \ldots, j_{\max}\}$ **do**
  $\mathcal{U}_j = \left\{m : \widehat{\alpha}_{\mathcal{N}_B^{(2)}}^{(m)} \geq j\right\}, \quad \nu_j = |\mathcal{U}_j|\,\tau_j$
end for
$j^{(2)} = \arg\max_j \nu_j$, $\mathcal{N}_2 = \mathcal{U}_{j^{(2)}}$
$d^{(2)} = \left|\mathcal{N}_B^{(2)}\right| j^{(2)} \tau_{j^{(2)}}$
if $d^{(2)} \geq d^{(1)}$ **then**
  $Two-groups = true$
else
  $Two-groups = false$ (Solve problem (5.5))
end if
Outputs: $Two-groups$, $j^{(2)}$, $\mathcal{N}_2$

sically the main difference between the SG-WCA and SG-SCA. The former is based on wideband CQI metrics while the latter is based on the use of subband CQIs. In fact, as it can be observed in Algorithm 7, the structure of SG-SCA is exactly the same as that of the SG-WCA except that, based on (5.8), it uses the subband CQIs to obtain estimations of the CQIs for all users in $\mathcal{N}_B$ assuming the transmission of data on the set of remaining PRBs after implementing the first stage allocation algorithm. Based on these metrics, the algorithm determines which are both the MCSs $j_{\min}$ and $j_{\max}$ associated, respectively, to the minimum (except the no transmission case) and maximum estimated CQIs. Then, it evaluates the utility provided by each available MCS $j \in \{j_{\min}, \ldots, j_{\max}\}$ and selects the MCS $j^{(2)}$ maximizing the utility. Again, if the data rate allocated to the second multicast group is greater than that allocated to the first multicast group, the solution provided by the two-step two-subgroup optimization algorithm is accepted. Otherwise, a single multicast group is formed by solving problem (5.5).

5.3 Numerical results

In this section, the downlink of an LTE-like MIMO-OFDMA wireless network is considered. The deployment scenario consists of nineteen omnidirectional cell sites located in the center of the corresponding cells and deployed in a hexagonal grid, where a central BS is surrounded by two concentric tiers of cells containing eighteen interfering BSs. As the performance analysis is restricted to the central cell, only the users served by the tagged BS are considered. In order to select the set of users in the multicast group, the following procedure is implemented: first, a large number of MSs is uniformly distributed on the coverage area of the BS of interest and then, in order to prevent an excessive and non-realistic number of multicast service delivery outages, a simple access control mechanism[1] is applied to select

[1]It is not our intention in this work to design access control mechanisms for broadcast/multicast schemes, we only use an access control strategy, common to all the proposed SRA algorithms, that helps preventing severe levels of multicast service delivery outages that could mask the desired system performance results under evaluation.

Table 5.1: Summary of simulation parameters used in two-step two-subgroup SRA algorithms.

Parameters	**Value**
BS transmit power (P_T)	47 dBm
BS/MS antenna gain (G_T/G_R)	6/3 dBi
MS noise figure (NF_{MS})	7 dB
Thermal noise (N_0)	-174 dBm/Hz
Cell radius (R)	500 m
Shadowing standard deviation (σ_s)	6 dB
Shadowing correlation (ρ)	0.5
Carrier frequency (f_0)	2.1 GHz
Default system bandwidth (B)	10 MHz
Subcarrier spacing (Δf)	15 KHz
Subcarrier per subband (N_{sc})	12
Multicarrier symbols per TTI (N_s)	14
Default number of orthogonal subbands $\left(N_B^{PHY}\right)$	50
TTI duration $\left(T_{TTI}^{PHY}\right)$	1 ms
Duration of long cyclic prefix $\left(T_{CP_{long}}\right)$	5.2 μs
Duration of short cyclic prefix $\left(T_{CP_{long}}\right)$	4.7 μs
Access control threshold $\left(\overline{\text{SINR}}_{\text{th}}\right)$	0 dB

the multicast MSs among those experiencing an average SINR higher than a prescribed access control threshold $\overline{\text{SINR}}_{\text{th}}$. Most of the system parameters have been defined based on current LTE/LTE-A specifications [74]. Table 6.1 summarizes the most relevant parameters considered in the simulations[2].

Recall that, on its way from the BS to the MSs, the transmitted signal experiences path-loss, large-scale shadow fading and small-scale frequency-, time- and space-selective fading.

The performance analysis is based on the following metrics:

- **ADR**: Defined as the sum of the useful data rates (measured in b/s)

[2]Note that the use of a normal CP has been assumed because the simulations have been performed in a single-cell multicast scenario. The use of an extended CP is only recommended in very large single-cell scenarios or in multicast scenarios using MBSFN. Nonetheless, it is worth stressing that the conclusions drawn using the proposed layout are also qualitatively valid for large single-cell and MBSFN scenarios.

offered to the whole set of users in the multicast group. The amount of data considered excludes protocol overhead data (i.e., CP, channel coding-related redundancy, control data) and only accounts for packets that are successfully delivered to the users.

- **Service outage probability**: Probability that the data rate the BS estimates that can be reliably transmitted (i.e., with a BLER less than or equal to the target BLER_0) to the users in the multicast group during a particular TTI is lower than the required minimum data rate $\tau_{\min}$.

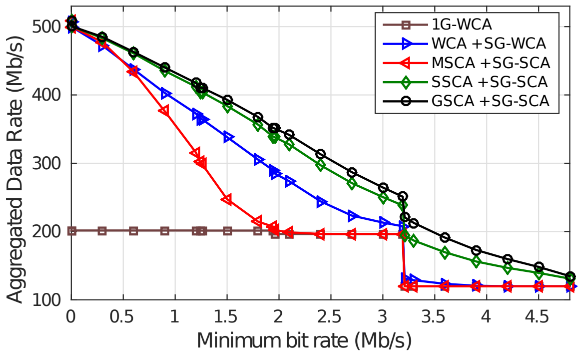
- **BLER outage probability**: Probability that, even though the BS predicts a reliable transmission, the measured BLER at any of the MSs in the multicast group is greater than the target BLER_0.

For the sake of accuracy, each simulation has been run over 1000 TTIs and has been repeated for 100 different random locations of the users.

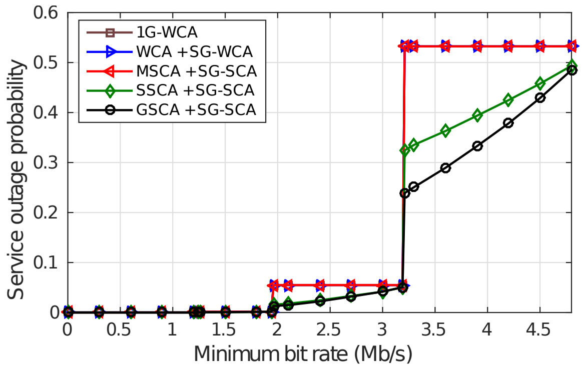
5.3.1 Benchmarking the proposed algorithms

Our aim in this subsection is to compare the proposed two-step two-subgroup subband CQI-based algorithms to the most representative algorithms proposed in the literature [30, 63, 16]. In particular, the GSCA+SG-SCA and SSCA+SG-SCA strategies are benchmarked against the MSCA+SG-SCA, WCA+SG-WCA and CMS schemes. All the numerical results presented in this subsection have been obtained using a BS with $N_T = 2$ transmit antennas multicasting data to a group of 50 MSs with $N_R = 2$ receive antennas on an 3GPP extended typical urban (ETU) channel with a low correlation profile using an LTE-like system with a bandwidth of $B = 10$ MHz split into $N_B^{\text{PHY}} = 50$ PRBs.

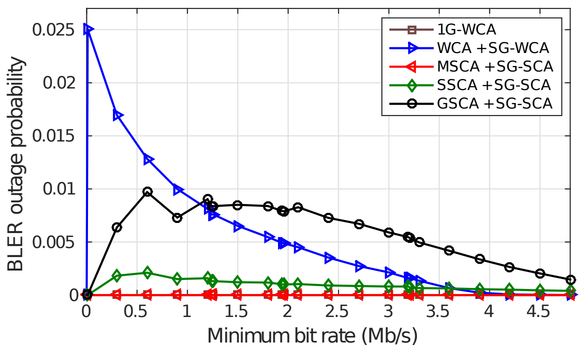
ADR, service outage probability and BLER outage probability are presented in Figures 5.1a to 5.1c, respectively, as a function of $\tau_{\min}$. Irrespective of the required $\tau_{\min}$, the subband CQI-based algorithms proposed in this chapter provide a clear ADR and service outage probability perfor-



(a) Aggregated data rate



(b) Service outage probability



(c) BLER outage probability

Figure 5.1: ADR, service outage probability and BLER outage probability as a function of $\tau_{\min}$ (ETU channel with a low correlation profile, 2×2 MIMO system, 50 users and 50 PRBs).

mance advantage in front of the benchmarking schemes, with the greedy-based one (GSCA+SG-SCA) outperforming the sorted subband CQI-based one (SSCA+SG-SCA). In particular, the single-group-WCA strategy is the one providing the poorest ADR and service outage probability performance among those algorithms considered in this work. This is because the same MCS is used to multicast data on the N_B^{PHY} available PRBs to the whole set of multicast users and, furthermore, the selection of this particular MCS is constrained by the channel conditions experienced by the worst MS in the multicast group. For low values of the required $\tau_{\min}$, the first stage of the conservative subband CQI-based MSCA+SG-SCA strategy is able to find a number of PRBs less than N_B^{PHY} allowing the fulfillment of the minimum data rate constraint using the most robust MCS among those reported by all the MSs on this set of PRBs. This still leaves room for the second stage to exploit the remaining PRBs in order to noticeably increase the ADR performance with respect to that provided by the CMS strategy. For high values of the required $\tau_{\min}$, however, the first stage of MSCA+SG-SCA tends to consume the whole set of available PRBs and transmit using an MCS that can be even more conservative than that required by the CMS strategy. In these cases, as it has been stated in Algorithm 3, instead of using the MCS corresponding to the minimum subband CQI, the system uses the MCS corresponding to the wideband CQI and, consequently, the MSCA+SG-WCA provides exactly the same performance as the CMS. The WCA+SG-WCA scheme, which is based on a less conservative strategy than that used by MSCA+SG-WCA, provides a much better ADR performance than the latter. However, even though the BS uses two disjoint subsets of PRBs to multicast data to the two subgroups of MSs generated in the first and second stages of the algorithm, all the decisions (i.e., selection of PRBs and selection of MCSs) are based on the wideband CQIs reported by the MSs. Thus, there is an obvious mismatch between the available CSI and the one that should ideally be managed by the algorithm, a mismatch that introduces a clear performance disadvantage of this benchmark algorithm when compared to that provided by the SSCA+SG-SCA and GSCA+SG-SCA schemes proposed in this work. Remarkably, as shown in Figure 5.1c, the ADR and service outage probability performance improvement shown by the proposed subband CQI-based algorithms is obtained at an almost negligible cost in terms of BLER outage probability that, for low values of the required minimum data rate, can even be lower than that produced when using the benchmark WCA+SG-WCA strategy. Obviously, given the fully conservative nature of CMS and MSCA+SG-SCA approaches, they are protected against BLER outages.

Note that in Figures 5.1a and 5.1b there are some values of $\tau_{\min}$ for which both the ADR and the service outage probability are discontinuous. The discontinuities occur at $\tau_{\min}$ values for which there are some TTIs and/or some of the random distributions of the multicast users for which a given MCS, when used to transmit on the whole set of available PRBs, is unable to support the required minimum data rate. When this happens, the system has to resort to a less robust MCS to fulfill the minimum data rate constraint with the consequent increase of the service outage probability and decrease of ADR. For the number of available PRBs considered in this subsection (50 PRBs), the MCS data rate values (measured in b/s/Hz) listed in Table 3.1 can be used in (3.18) to show that the discontinuities take place, as can be observed in Figure 5.1, at $\tau_{\min} = 1.25, 1.95, 3.20, 5.05, \ldots$ Mb/s.

Note that the proposed subband CQI-based algorithms are able to outperform the benchmarking ones because, thanks to a better exploitation of CSI, they are able to stick to the rule that the lower the number of PRBs used in the first stage the higher the ADR increase that can be generated in the second stage. In order to emphasize this fact, Figure 5.2 shows the average number of PRBs that the different algorithms require to allocate to the first multicast subgroup in order to fulfill the minimum data rate constraint. In accordance with the behaviour of the performance metrics presented in Figure 5.1, it can clearly be observed that the proposed subband CQI-based algorithms are able to fulfill the constraint on $\tau_{\min}$ using less PRBs than the benchmarking schemes. As an example, for a $\tau_{\min} = 1.5$ Mb/s, the MSCA, WCA, SSCA and GSCA schemes require an average number of PRBs equal to 36.6, 20.7, 16.1 and 15.2, respectively, to satisfy the minimum data rate

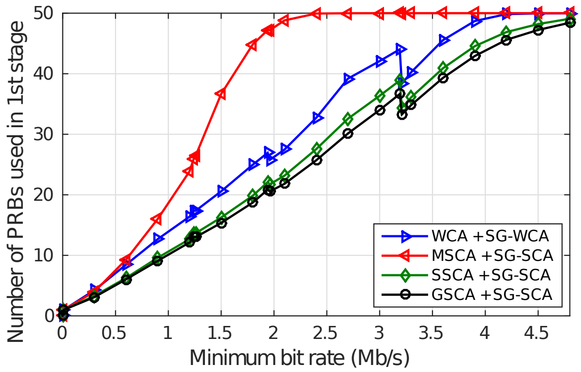


Figure 5.2: Average number of resources allocated in the first stage of the assessed algorithms (ETU channel with a low correlation profile, 2 × 2 MIMO system, 50 users and 50 PRBs).

constraint in the first stage. Using a small number of PRBs in the first stage, jointly with the fact that an algorithm using a better CSI produces less service outages thus increasing the number of times the RRM procedure can execute the second stage, frees a greater average number of resources to be used by the second stage in trying to increase the ADR.

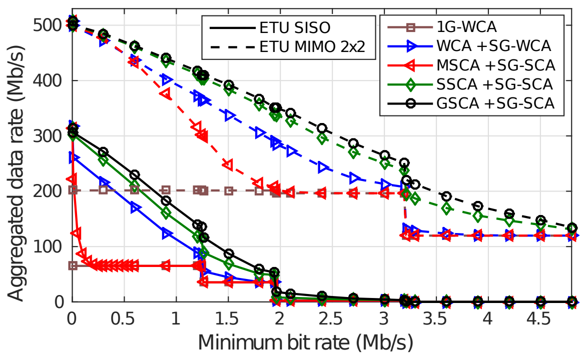
Comparing the results for the two-step two-subgroup algorithms presented in Figures 5.1a and 5.2, it can be observed that for very low and very high values of $\tau_{\min}$ all the algorithms tend to provide similar ADR performance levels and the greater performance differences are obtained, for the default scenario considered in this subsection, for required minimum data rates around 2.5 Mb/s. This is because for very low (or very high) values of $\tau_{\min}$, the average number of PRBs used in the second stage (or the first stage) approaches N_B^{PHY} and thus, the SRA strategies based on both the wideband and subband CQI tend to use the same amount of information to estimate/predict the channel conditions and accordingly tend to provide the same performance results. In contrast, as the average number of PRBs used in either the first or the second stages approximate $N_B^{\mathrm{PHY}}/2$, the RRM-related estimations/predictions that are based on the use of subband CQI are clearly better than those based on wideband CQI thus yielding an explicit performance advantage to the proposed subband CQI-based algorithms in terms of both ADR and service outage probability.

To sum up, the numerical results presented in this subsection show that, irrespective of the $\tau_{\min}$ requirements and thanks to a clever use of the available subband CQI, the novel algorithms proposed in this work allow increasing the ADR and decreasing the number of service outages when compared to well-known benchmarking schemes and this is done at a negligible cost in terms of BLER outage probability. Among the proposed two-step two-subgroup subband CQI-based algorithms, the GSCA+SG-SCA clearly outperforms the SSCA+SG-SCA due to a better exploitation of subband CQI but at the cost of a significant complexity increase[3].

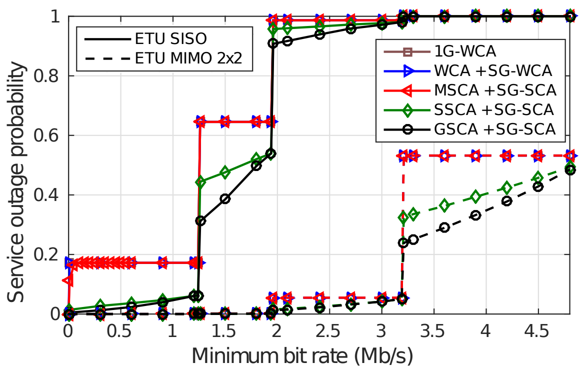
5.3.2 Channel model and MIMO configuration effects

Let us now turn our attention to the effects that the channel models and MIMO configurations have on the quantitative and qualitative performance of both the proposed and benchmarking algorithms. Figures 5.3a and 5.3b show that, as expected, the lower SINR values provided by the use of single input single output (SISO) systems result in considerably lower ADRs and higher service outage probabilities, respectively, when compared to TD-based MIMO arrangements. Remarkably, the qualitative behaviour of all the considered SRA algorithms is exactly the same irrespective of the system configuration (either SISO or MIMO) under consideration, with the proposed greedy-based approach (GSCA+SG-SCA) outperforming the sorted subband CQI-based one (SSCA+SG-SCA) and both of them, in turn, outperforming all the benchmarking schemes. Needless to say, the minimum data rates at which the behaviour of all the considered two-step two-subgroup SRA algorithms collapses to the behaviour of the single-group WCA strategy are considerably lower for SISO systems than for TD-based

[3]Note that both SSCA+SG-SCA and GSCA+SG-SCA have been proposed to overcome the complexity limitations of an exhaustive search based on subband CQI that implies a complexity equal to $\mathcal{O}(MN_{PRB}^{J})$. As an example, for $M = 50$ users, $N_{PRB} = 50$ PRBs and $J = 15$ MCSs, the number of operations required in an exhaustive search is higher than 10^{27}. However, for this configuration, SSCA+SG-SCA and GSCA+SG-SCA require less than 10^{3} and 10^{5} operations, respectively.



(a) Aggregated data rate

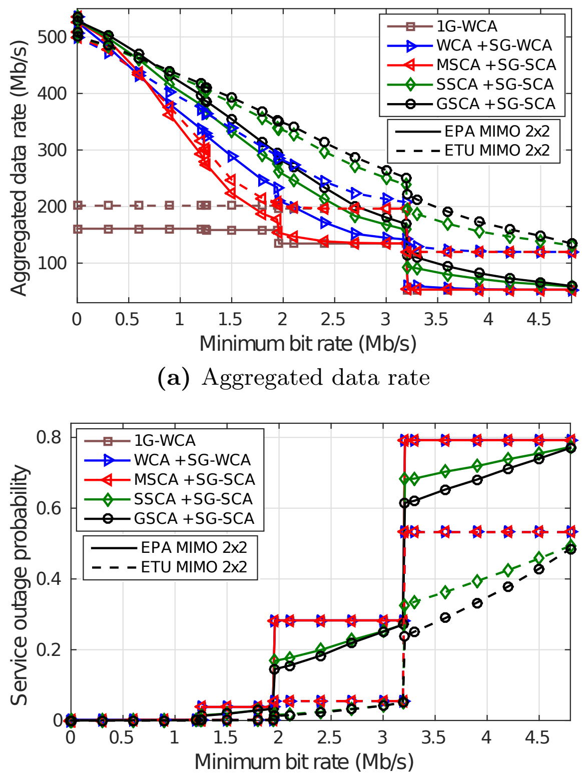


(b) Service outage probability

Figure 5.3: ADR and service outage probability *versus* minimum required data rate for both 2×2 MIMO (dashed lines) and SISO (solid lines) systems (ETU channel with low correlation profile, 50 users and 50 PRBs).

MIMO schemes.

Figure 5.4 compares the performance of the assessed algorithms assuming that the BS multicasts information on either an ETU or a 3GPP extended pedestrian A (EPA) channel model. When compared to ETU, the EPA channel model is characterized by a much lower frequency selectivity. Hence, as the requirements of $\tau_{\min}$ rise it becomes increasingly difficult to find a sufficient number of PRBs that allow compliance with this restriction and consequently, as shown in Figure 5.4b, the number of service outages grows far beyond the number of outages experienced by the system when transmitting on an ETU channel model. Moreover, as clearly unveiled by



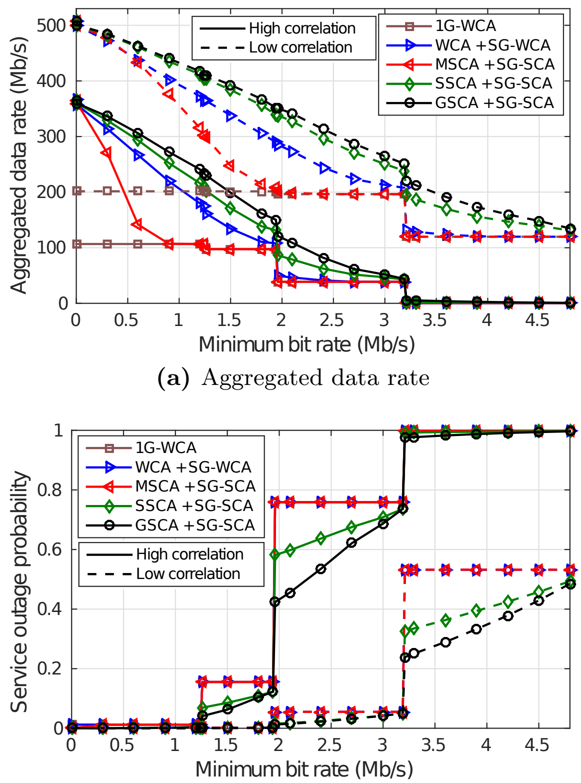
(a) Aggregated data rate

(b) Service outage probability

Figure 5.4: ADR and service outage probability *versus* minimum required data rate for both ETU (dashed lines) and EPA (solid lines) channels with low correlation profile (2×2 MIMO system, 50 users and 50 PRBs).

results presented in Figure 5.4a, the performance advantages of the SRA algorithms proposed in this work over the benchmarking WCA+SG-WCA scheme are more evident when transmitting on the ETU channel than when transmitting on the EPA channel. This is due to the fact that the advantages provided by the availability of subband-CQIs can be better exploited when there is a lot of frequency selectivity. Indeed, in a frequency flat channel the subband CQI-based algorithms would not provide any advantage with respect to wideband CQI-based schemes.

Figures 5.5a and 5.5b illustrate the fact that increasing the correlation among MIMO transmit and/or receive antennas produces a quantitative



(a) Aggregated data rate

(b) Service outage probability

Figure 5.5: ADR and service outage probability *versus* minimum required data rate for ETU channel models with both low (dashed lines) and high (solid lines) correlation profiles (2 × 2 MIMO system, 50 users and 50 PRBs).

decrease (increase) of the ADR (service outage probability) provided by any of the SRA strategies assessed in this work. However, they also confirm that the qualitative behaviour of all the algorithms is essentially the same under different antenna correlation conditions and that, irrespective of the $r_{\min}$ requirements, a correct use of the available subband CQI provides the novel algorithms proposed in this work with the capacity of outperforming the benchmarking schemes even under heavy spatial correlation conditions.

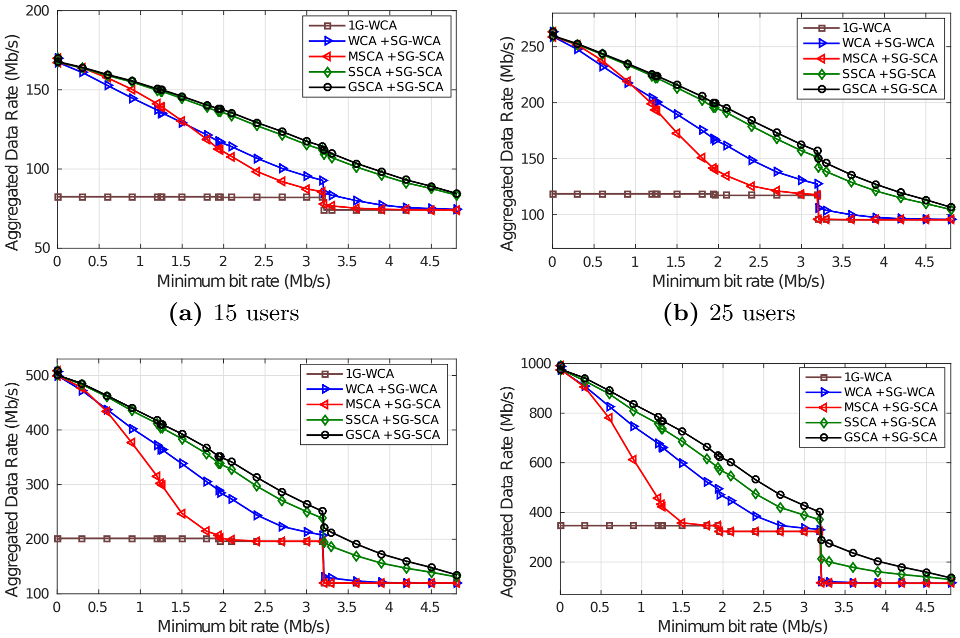
In summary, results presented in Figures 5.3 to 5.5 allow us to guarantee that the two-step two-subgroup subband CQI-based SRA solutions that we have proposed (SSCA+SG-SCA and especially GSCA+SG-SCA) outperform the benchmarking strategies regardless of the SISO/MIMO configuration, the channel model and/or the transmit/receive antenna correlation profile.

5.3.3 Effects of the number of users, PRBs and PRBs/ subband

Our main aim in this subsection is to investigate the effects that the number of users in the multicast subgroup, the number of available PRBs and the number of PRBs per subband have on the performance of the proposed SRA algorithms when compared to the benchmarking strategies. Numerical results presented in the following figures have all been obtained using BSs with $N_T = 2$ transmit antennas multicasting data to MSs with $N_R = 2$ receive antennas on an ETU channel with a low correlation profile.

Figures 5.6a to 5.6d present the ADR as a function of $\tau_{\min}$ for a number of users in the multicast group equal to 15, 25, 50 and 100, respectively. For the considered scenarios, increasing the number of users in the multicast group results in an improved ADR even for the 1G-WCA scheme. However, the most important conclusion that can be drawn from the comparative examination of these figures is that the qualitative behaviour of all the algorithms under study is approximately the same irrespective of the number of users in the multicast group, with the novel two-step two-subgroup subband CQI-based algorithms proposed in this work (i.e., SSCA+SG-SCA and GSCA+SG-SCA) clearly outperforming those previously proposed in the literature.

Increasing the system bandwidth or, equivalently, the number of PRBs per TTI, produces a similar effect on the performance of the different algorithms as that observed when increasing the number of users in the multicast group. That is, as it can be observed in Figures 5.7a to 5.7d, using a larger system bandwidth allows multicasting larger quantities of data to the set of users in the multicast group. Furthermore, it is worth noting that the values of $\tau_{\min}$ for which the ADR (and the service outage probability) show discontinuities are proportional to the number of available PRBs in the sys-

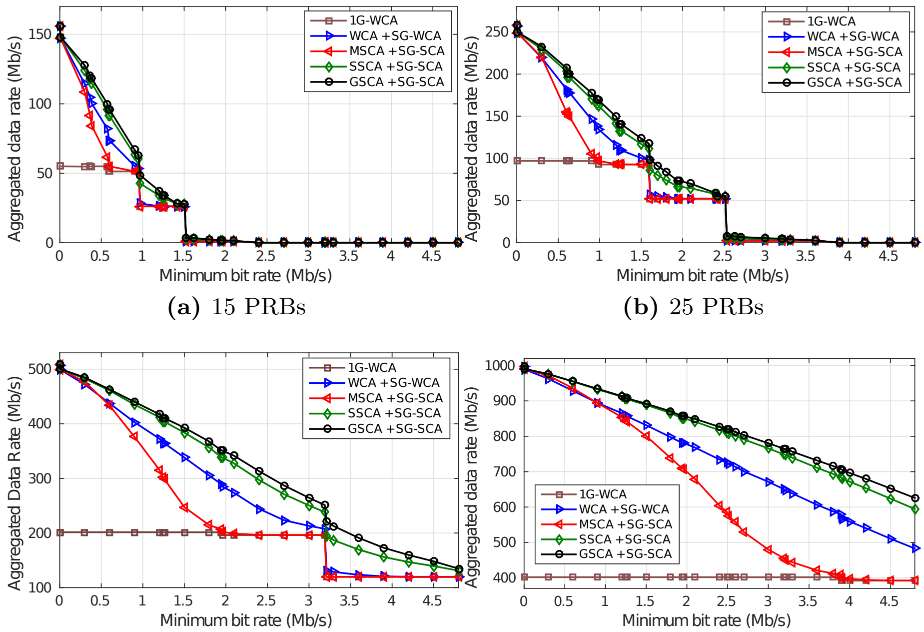


(a) 15 users
(b) 25 users
(c) 50 users
(d) 100 users

Figure 5.6: ADR as a function of $\tau_{\min}$ and with the number of users in the multicast group as parameter (ETU channel with a low correlation profile, 2×2 MIMO system, and 50 PRBs).

tem. Recalling that for a system with 50 PRBs the discontinuities occurred at $\tau_{\min} = 1.25, 1.95, 3.20, 5.05, \ldots$ Mb/s, it is quite easy to show that the discontinuities take place at $\tau_{\min} = 0.375, 0.585, 0.960, 1.515, \ldots$ Mb/s for a system with 15 PRBs, at $\tau_{\min} = 0.625, 0.975, 1.600, 2.525, \ldots$ Mb/s for a system with 25 PRBs, and at $\tau_{\min} = 2.5, 3.9, 6.4, 10.1, \ldots$ Mb/s for a system with 100 PRBs.

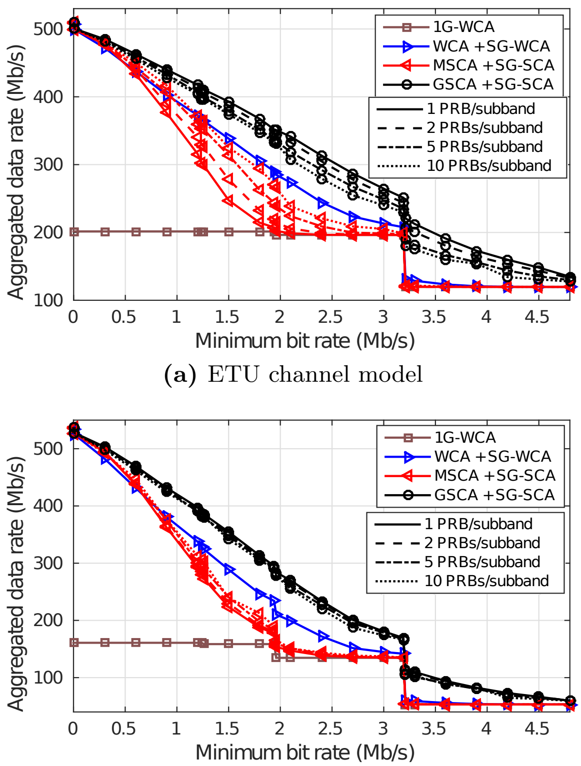
Finally, Figures 5.8a and 5.8b present the ADR as a function of $\tau_{\min}$ with the number of PRBs per subband as parameter. The former considers a BS multicasting on an ETU channel whereas the latter considers the same BS multicasting on an EPA channel. Obviously, SRA strategies based on the use of wideband CQI (i.e., 1G-WCA and WCA+SG-WCA) provide exactly the same performance irrespective of the number of PRBs per subband and hence, their performance is just shown for the sake of comparison. Note



(a) 15 PRBs (b) 25 PRBs (c) 50 PRBs (d) 100 PRBs

Figure 5.7: ADR as a function of $\tau_{\min}$ and with the number of available PRBs in the system as parameter (ETU channel with a low correlation profile, 2×2 MIMO system, and 50 users).

also that, in account of clarity, among the three strategies that are based on subband CQI (i.e., MSCA+SG-SCA, SSCA+SG-SCA and GSCA+SG-SCA) only the ADRs provided by the GSCA+SG-SCA and MSCA+SG-SCA schemes are compared. Indeed, the performance results obtained using the SSCA+SG-SCA approach are only slightly worse than those provided by the GSCA+SG-SCA scheme and follow the same trends as these ones when varying the number of PRBs per subband. The first important fact worth stressing is that any subband CQI-based strategy using a single subband containing all the PRBs in the system is equivalent to a WCA+SG-WCA scheme. Thus, increasing the number of PRBs per subband (or equivalently, reducing the number of subbands) leads to performance metrics increasingly similar to those provided by the WCA+SG-WCA scheme. Indeed, as the number of subbands decreases, the ADR obtained using a WCA+SG-WCA



(a) ETU channel model

(b) EPA channel model

Figure 5.8: ADR as a function of $\tau_{\min}$ and with the number of PRBs per subband as parameter (channels with low correlation profiles, 2×2 MIMO system, and 50 users).

algorithm acts as either a lower or an upper bound for the ADR provided by the GSCA+SG-SCA and MSCA+SG-SCA algorithms, respectively. Remarkably, the performance results provided by the proposed GSCA+SG-SCA algorithm when reducing the number of subbands from 50 to 5 (i.e., 10 PRBs/subband) are still substantially better than those obtained using the benchmarking WCA+SG-WCA scheme, especially when transmitting on the EPA channel model whose frequency selectivity is much lower than that of the ETU channel.

Summarizing the results presented in this subsection, it is worth mentioning that, again, the two-step two-subgroup subband CQI-based SRA solutions proposed in this work outperform the benchmarking strategies regardless of the number of users in the multicast subgroup, the number of available PRBs and/or the number of PRBs per subband. Moreover, as expected, increasing the granularity when feeding back the CQI from the MSs to the BS (i.e., increasing the number of subbands) is especially advantageous when transmitting on channels showing strong frequency selectivity.

5.4 Summary

The work presented in this chapter has explored the efficient utilization of subband CQI in order to optimize multicast transmissions in the context of MIMO-OFDMA networks. Making use of the physical abstraction developed in the system model in Chapter 3, we have modelled the CQI fed back by the users to the BS under different subband partitions and for various MIMO configurations. This physical abstraction allows the formulation of a multicast throughput constrained optimization problem whose restrictions establish minimum QoS requirements to be fulfilled by all multicast users. Since the derivation of the optimal solution is, in general, computationally infeasible, suboptimal solutions have been sought. In particular, it has been established that an effective strategy consists of greedily splitting the multicast users into two subgroups with the aim of, first, satisfying a minimum data rate request (users in the first subgroup) and second, maximizing the system ADR (users in the second subgroup), all being subject to target average BLER constraints. Results obtained in an LTE-like context reveal that the two-step two-subgroup approach leads to substantial benefits in terms of ADR irrespective of the MIMO configuration in use. The exploitation of the granularity that subband CQI brings along becomes especially significant in channels with strong frequency selectivity, where wideband CQI becomes very imprecise regarding the true potential capacity of the users' channels. Remarkably, the two subband CQI-based techniques introduced in this chapter, namely, the GSCA+SG-SCA and the SSCA+SG-SCA, are shown to significantly outperform state-of-the-art proposals with regard to ADR and service outage probability for minimum data rate requirements of practical interest and under most of the system and channel configurations that can be found in real scenarios.

This chapter has been based on the following submitted work:

- **Alejandro de la Fuente**, Guillem Femenias, Felip Riera-Palou, and Ana García Armada, "Subband CQI feedback-based Multicast Resource Allocation in MIMO-OFDMA Networks", *Recommended for publication in IEEE Transactions on Broadcasting.*

Chapter 6

Radio resource allocation for multicast services based on multiple video layers

The fact that the search space of the optimal solution for multicast SRA strategies can be extremely large depending on the number of multicast members, available resources, and the number of multicast groups [71] requires particular attention. This results in the need of focusing the development of multicast SRA algorithms on reducing the search space of possible solutions, which facilitates their implementation in real systems.

In order to go one step further with respect to the available solutions in the literature, this chapter complements the works of [16, 66, 30], incorporating the multicast video layers to the optimization framework, and presents a novel SRA strategy to deliver multicast video services, that focuses on maximizing the ADR while minimizing the solution space. This reduction of the solution space signifies low complexity procedures. The main contributions of the work presented in this chapter can be summarized as follows:

- We propose a new formulation to optimize the ADR. Differently from [66, 30], the multicast video layers are explicitly incorporated to the optimization framework.

- By analyzing the structure of the formulated problem, we propose a novel scheme, named as MAMVL, that significantly reduces the complexity of the problem. This characteristic makes MAMVL adequate for real time implementations.

- The proposed algorithm to solve the optimization problem enhances the spectral efficiency, as the MAMVL scheme strictly delivers the data rates required by each layer, freeing resources for other layers and/or services.

6.1 Scalable video coding for broadcast/multicast transmissions in LTE

SVC has become a very attractive solution to transmit video with a given quality in the current wireless systems. The H.264/SVC standard allows the system to transmit a variety of different quality layers for a video sequence [85]. This layered approach allows users to choose the layers they can correctly decode according to their channel conditions.

In order to support scalability, H.264/SVC systems are able to use more than one layer of a single video stream, which implies that different layers of the same content are transmitted. Scalability can be achieved along three orthogonal dimensions, namely spatial, quality and temporal, that refer to scalability in terms of resolution, compression level and number of frames per second, respectively. These scalable dimensions can be used either jointly or independently to generate a H.264/SVC stream.

There is one mandatory layer that consists of the basic representation in each of the temporal, spatial and quality dimensions. Then, a variable number of enhancement layers can be encoded. Each enhancement layer is an improvement in terms of one or more of the 3 dimensions. In order to correctly decode an enhancement layer, the correct reception and decoding of all the lower layers is required. Using this approach the quality of a particular sequence of video can be adapted to the device that is used to visualize the video content and to the varying conditions of the wireless channel [86].

Applying H.264/SVC to broadcast/multicast video transmission over eMBMS LTE-A networks, the service can be seen as a PtM delivery of a data stream that consists of L different layers [71]. In particular, the basic layer provides a basic reconstruction quality, which is gradually improved by the remaining $L-1$ layers, called enhancement layers. In agreement with the layered structure, the higher the number of layers that can successfully be decoded the better the level of the user QoS.

3GPP contemplates the use of SVC employing multiple layers in eMBMS [87]. 3GPP proposes the allocation of multiple multicast radio bearers using different MCS for different SVC layers. The base layer can be transmitted using robust MCS in order to guarantee high-priority, so that the base layer is delivered using low data rate. On the other hand, the enhancement layers can be transmitted using less robust MCSs to provide high data rate. The effect that results of allocating multi-level MCS channels for SVC is that the users experiencing good signal strength may receive all base and enhancement layers, whereas the users located in an area of poor signal strength may only receive base layer data. Therefore the multi-level MCS allocation for SVC is adaptive to channel condition and the quality is degraded based on the number of layers the user can correctly receive. For example, the radio resources can be divided to carry SVC layers in different MCS channels.

6.2 Optimization problem

Let us consider the downlink of an LTE-like OFDMA single-cell system consisting of one BS which is providing a multicast service to M users distributed over the corresponding coverage area as depicted in Figure 6.1. The set $\mathcal{M} = \{1, \dots, M\}$ will be used to index the multicast users. Both SISO and MIMO set-ups will be considered. Interference signal is considered

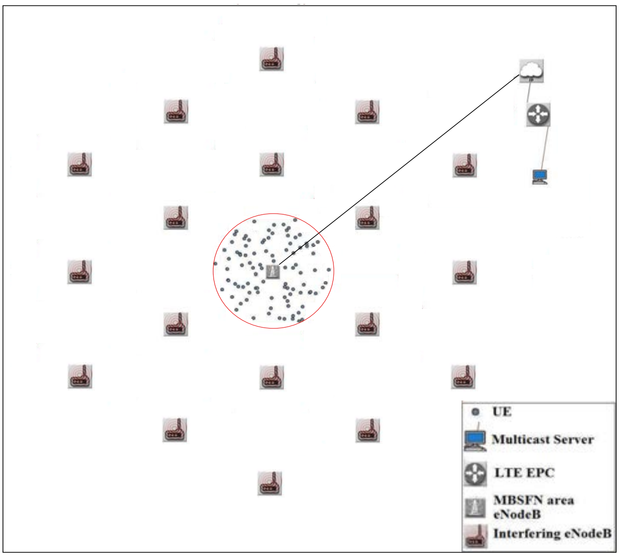


Figure 6.1: Single-cell multicast scenario employed in MAMVL assessment

from the neighbouring BSs to be added, jointly with thermal noise, to the desired signal received by the multicast user in order to model the SINR measured by the MS.

Let us consider a full-buffer traffic model in which, at any given time, the BS has some video traffic to deliver. This video traffic is encoded in L different video layers (hereafter simply referred to as layers) using SVC, and the set of available layers is denoted by $\mathcal{L} = \{1, \ldots, L\}$. The BS may select the MCS $j \in \mathcal{J} = \{1, \ldots, J\}$ suitable for each layer $l \in \mathcal{L}$. The main characteristics of the MCSs are those given for LTE/LTE-A systems and listed in Table 3.1. The total number of PRBs is denoted by N_{PRB} and depends on the bandwidth reserved for that purpose.

The wideband CQI values reported by each multicast user are denoted by $\boldsymbol{\alpha} = \{\alpha_m;\ m = 1, \ldots, M\}$. These CQIs determine the MCS so that each user can correctly decode the transmission. The binary matrix $\boldsymbol{X} = \{x_{m,j};\ m = 1, \ldots, M;\ j = 1, \ldots, J\}$ describes which MCS j user m is assigned to, from those MCSs which the user is able to decode (see (3.22)).

Each user must be assigned to one and only one MCS depending on the reported wideband CQIs and the allocation of PRBs among the MCSs. Hence $x_{m,j}$ denotes the MCS (with allocated PRBs) assigned to each user. Note that this selection is made from the MCSs which can be successfully decoded given the CQI reported by the user. Next subsection describes (P1), that is an optimization problem developed to allocate the PRBs among the MCSs.

6.2.1 MT problem

The authors of [30] formulate the MT problem (P1) (see (3.20)) to maximize the ADR of the service delivered to all the multicast members per transmission frame. The results of this maximization problem are: *(i)* the optimal allocation vector $\boldsymbol{r} = \{r_j;\ j = 1, \dots, J\}$ that denotes the number of PRBs allocated to each MCS, and *(ii)* the binary matrix $\boldsymbol{X}$ defined in (3.22). This problem is formulated as

$$(\text{P1}) \quad \underset{r_j,\, x_{m,j}}{\text{maximize}} \quad \sum_{j=1}^{J} \tau_j r_j \sum_{m=1}^{M} x_{m,j} \tag{6.1}$$

$$\text{subject to} \quad \sum_{j=1}^{\alpha_m} x_{m,j} = 1 \qquad \forall m \in \mathcal{M} \tag{6.1a}$$

$$x_{m,j} = 0 \qquad \forall j \geq \alpha_m + 1 \tag{6.1b}$$

$$\sum_{j=1}^{J} r_j \leq N_{\text{PRB}} \tag{6.1c}$$

$$\frac{1}{M} \sum_{m=1}^{M} x_{m,j} \leq r_j \leq N_{\text{PRB}} \sum_{m=1}^{M} x_{m,j} \qquad \forall j \in \mathcal{J} \tag{6.1d}$$

$$x_{m,j} \in \{0, 1\} \qquad \forall m \in \mathcal{M}, \quad \forall j \in \mathcal{J} \tag{6.1e}$$

$$r_j \in \{0, \dots, N_{\text{PRB}}\} \qquad \forall j \in \mathcal{J} \tag{6.1f}$$

where the ADR is calculated as (6.1) (see (3.20) where MT cost function is defined).

In (P1), constraints (6.1a) to (6.1f) are the same constraints detailed in (4.5g) to (4.5l) for JMSUT. However, (P1) maximizes the ADR without considering SVC.

6.2.2 MT with multiple video layers

We now reformulate (P1) to both incorporate SVC to multicast video transmission and to reduce the complexity of the original problem (P1).

We first reduce the complexity of the new problem by realizing that variables $x_{m,j}$ can be explicitly determined, what considerably simplifies (P1) as r_j are the only variables of the new problem. Indeed, as $x_{m,j} = 1$ represents that MCS j will be assigned to user m, the value of $x_{m,j}$ can be calculated from the reported CQI α_m, the variable r_j and the number of available resources for CQIs lower than α_m from the following conditions:

- The number of resources allocated to MCS j must be not null, what implies that $r_j > 0$.
- The CQI of user m is larger or equal to j, what implies $\alpha_m \geqslant j$.
- No resource block has been assigned for CQIs lower than α_m, implying that $\sum_{\rho=j+1}^{\alpha_m} r_\rho = 0$.

These conditions lead to the new constraint (6.4a).

Now, we introduce auxiliary variables $\mathbf{W}$ and $\mathbf{Z}$ to incorporate multiple layers to (P1). As a discrete number of SVC layers with different data rates is used, the binary matrix $\mathbf{W} = \{w_{j,l};\ j = 1, \ldots, J;\ l = 1, \ldots, L\}$ will denote the allocation of MCS j to SVC layer l, whose elements $w_{j,l}$ are defined as

$$w_{j,l} = \begin{cases} 1, & \text{if MCS } j \text{ is assigned to video layer } l \\ 0, & \text{otherwise.} \end{cases} \tag{6.2}$$

The elements of **W** can be calculated from the variables r_j, given that the number of resources allocated to MCS j, must provide a data rate, represented by $r_j \tau_j$, equal or larger than the required data rate of layer l d_l. Then, $w_{j,l} = 1$ if $r_j \tau_j \geq d_l$.

Next, we define variable **Z**, which determines if user m can correctly decode layer l. Hence, the elements of **Z** are defined as

$$z_{m,l} = \begin{cases} 1, & \text{if user } m \text{ can correctly receive layer } l \\ 0, & \text{otherwise.} \end{cases} \tag{6.3}$$

As a result, taking into consideration the definitions of $x_{m,j}$ and $w_{j,l}$, we have that $\mathbf{Z} = \mathbf{XW}$.

Considering the data rate that corresponds to each available video layer l (note that this information is given by the service provider), which are denoted by the vector $\boldsymbol{d} = \{d_l;\ l = 1, \ldots, L\}$, we now formulate the MT problem explicitly considering multiple layers:

$$\text{(P2)} \quad \underset{r_j}{\text{maximize}} \quad \sum_{l=1}^{L} d_l \sum_{m=1}^{M} z_{m,l} \tag{6.4}$$

$$\text{subject to} \quad x_{m,j} = (r_j > 0) \wedge (\alpha_m \geq j) \wedge \left(\sum_{\rho=j+1}^{\alpha_m} r_\rho = 0 \right) \quad \forall m \in \mathcal{M}, \quad \forall j \in \mathcal{J} \tag{6.4a}$$

$$w_{j,l} = (r_j \tau_j \geq d_l) \qquad \forall j \in \mathcal{J}, \quad \forall l \in \mathcal{L} \tag{6.4b}$$

$$\boldsymbol{Z} = \boldsymbol{XW} \tag{6.4c}$$

$$x_{m,j} \in \{0, 1\} \qquad \forall m \in \mathcal{M}, \ \forall j \in \mathcal{J} \tag{6.4d}$$

$$r_j \in \{0, \ldots, N_{\text{PRB}}\} \qquad \forall j \in \mathcal{J} \tag{6.4e}$$

$$\sum_{j=1}^{J} r_j \leq N_{\text{PRB}} \tag{6.4f}$$

where the ADR is calculated in (6.4). Note that three new constraints have been included in (P2) with respect to (P1). Constraint (6.4a) refers to the calculation of matrix $\mathbf{X}$ according to the conditions given at the begining of this section. Constraint (6.4b) refers to the calculation of matrix $\mathbf{W}$ as stated above, and constraint (6.4c) refers to the calculation of matrix $\mathbf{Z}$.

6.3 The low complexity MAMVL scheme

In this section, we propose the multicast resource allocation based on multiple video layers (MAMVL) scheme to solve (P2) with reduced complexity. MAMVL takes advantage of the nature of multicast transmissions, providing a solution that guarantees high spectral efficiency, since it maximizes the data rate of a large number of users using the minimum number of PRBs.

Let us examine the general case of a combinatorial problem with J variables (MCSs) with N_{PRB} possible values (allocated PRBs). The number of available solutions for an ESS is given by

$$\mathcal{C}_{ESS} = N_{\mathrm{PRB}}^{J}. \tag{6.5}$$

In the case of (P1), constraint (6.1c) leads to a reduction of the search space according to the optimized search scheme (OSS) of [63]. The number of possible solutions to (P1) results in a version of the classic stars-and-bars problem in combinatorics [88]. For any pair of natural numbers J and N_{PRB}, the number of distinct J-tuples of non-negative integers whose sum is less or equal than N_{PRB} is given by the number of multisets of cardinality J taken from a set of size $N_{\mathrm{PRB}}+1$, or equivalent by the binomial coefficient. Thus the solution space is given by

$$\mathcal{C}_{OSS} = \binom{N_{\mathrm{PRB}}+J}{N_{\mathrm{PRB}}} = \frac{(N_{\mathrm{PRB}}+J)!}{J!N_{\mathrm{PRB}}!}. \tag{6.6}$$

Note that the number of possible solutions depends on the number of available PRBs when the number of MCSs is fixed (15 MCSs are used in LTE/LTE-A). For instance, for $N_{\mathrm{PRB}} = 25$ PRBs (a dedicated bandwidth of 5 MHz), the number of possible solutions with OSS is 4.0225×10^{10}, what implies an unaffordable computational cost for practical implementation.

The MAMVL scheme proposes a search space reduction that brings a significant decrease of complexity with respect to the solution space of (P1), without any degradation in the optimality of the resource allocation scheme, as the next section details. Algorithm 8, presented in Section 6.3.2, implements MAMVL.

6.3.1 Search space reduction

Due to the nature of the maximization problem (P2), an important reduction of the search space can be achieved. This process consists of three steps, as we detail next.

The first step considers the number of layers $L < J$. The MAMVL scheme requires the correct reception of at least the basic layer, namely $l = 1$, and not more than L layers. The total number of permutations of a maximum of N_{PRB} elements (the option of not using all the PRBs is considered) into J groups is given by (6.5). However, the introduction of a limited number of layers results in a number of possible solutions given by the permutations of N_{PRB} resource elements into L possible layers as many times as J MCS elements can be permuted into L possible layers. Hence, the number of possible solutions is given by

$$\mathcal{C}_{MAMVL_1} = \binom{N_{\text{PRB}}}{L}\binom{J}{L} = \frac{N_{\text{PRB}}!}{(N_{\text{PRB}} - L)!L!}\frac{J!}{(J-L)!L!}\,, \tag{6.7}$$

which is much less than $\mathcal{C}_{OSS}$.

For the second step, let us denote the MCS assigned to each video layer by the vector $\boldsymbol{\mu} = \{\mu_l;\ l = 1, \ldots, L\}$. Considering that the basic video layer $l = 1$ must be delivered to all the users guaranteeing a correct decoding, the MCS used to deliver $l = 1$ must be

$$\mu_1 = \min(\alpha_m) \qquad \forall m \in \mathcal{M}. \tag{6.8}$$

The number of PRBs required depends on the used MCS. Let us consider the matrix $\boldsymbol{U} = \{u_{j,l};\ j = 1, \ldots, J;\ l = 1, \ldots, L\}$, where the element $u_{j,l}$ denotes the number of PRBs required to deliver the layer l using the MCS j. The basic layer encoded with the lower data rate is transmitted using μ_1 (6.8), and the number of PRBs allocated to the first layer is calculated as

$$u_{\mu_1,1} = \lceil d_1/\tau_{\mu_1} \rceil. \tag{6.9}$$

The remaining columns of matrix $\boldsymbol{U}$ denote the PRBs required to deliver the enhancement layers, and are calculated as

$$u_{j,l} = \lceil d_l/\tau_j \rceil; \qquad j = [\mu_1, J] \in \mathcal{J},\ \ l = [2, L] \in \mathcal{L} \tag{6.10}$$

and the following positions of matrix $\boldsymbol{U}$ remain at '0', since they are not considered possible solutions for (P2):

$$u_{j,l} = 0 \begin{cases} \text{if } \ l = 1 \ \ j \neq \mu_1 \ \ \text{, or} \\ \text{if } \ l > 1 \ \ j < \mu_1. \end{cases} \tag{6.11}$$

Consequently, after the allocation of the basic layer $l = 1$, the number of available MCSs J', PRBs N'_{PRB}, and remaining layers L' decrease and are given by

$$\begin{aligned} J' &= J - \mu_1, \\ N'_{PRB} &= N_{\text{PRB}} - r_{j,1}, \\ L' &= L - 1. \end{aligned} \tag{6.12}$$

For instance, 8 PRBs (see Table 3.1) are required to deliver a basic video layer with a data rate of 200 kb/s using the most robust MCS, $j = 1$. Thus, applying the same reasoning that was used to obtain (6.7), the number of possible solutions of the second step for different number of layers is given by

$$\mathcal{C}_{MAMVL_2} = \binom{N'_{PRB}}{L'}\binom{J'}{L'} = \frac{N'_{PRB}!}{(N'_{PRB}-L')!L'!}\frac{J'!}{(J'-L')!L'!}. \tag{6.13}$$

Now, some considerations allow us to introduce the third search space reduction for the matrix $\boldsymbol{U}$:

- For each video layer, there are different options to allocate the PRBs using different MCSs. However, if the same number of PRBs is required to deliver one layer data rate using different MCSs, the most robust MCS is used, as the more robust the MCS, the higher the number of users that can decode the transmission. For example, if the MCSs $j = 2$ and $j = 3$ require the same number of PRBs to deliver the data rate of layer $l = 2$, only $j = 2$ is considered. Then, the following condition must be added

$$u_{j+1,l} < u_{j,l}; \quad \forall j \in \mathcal{J},\ \ l = [l_2, L] \in \mathcal{L}. \tag{6.14}$$

- The sum of PRBs allocated to deliver the complete set of layers cannot exceed the number of PRBs N_{PRB}. It can be expressed as

$$\sum_{j=1}^{J}\sum_{l=1}^{L}(u_{j,l}\, w_{j,l}) \leq N_{\text{PRB}}. \tag{6.15}$$

- Using SVC transmission implies that the enhancement layers will be decoded only by the users experiencing better channel conditions, whereas the basic layer must be decoded by every user. This means that the higher the video layer, the less robust the MCS used to encode it. This can be expressed as

$$w_{j,l+1} = 0 \qquad \forall j \in (1,\rho] : w_{\rho,l} = 1 \tag{6.16}$$

  where ρ denotes the MCS used to deliver the layer l. Therefore, only MCSs higher than ρ are possible solutions to deliver the layer $l+1$.

Note that the elements of matrix $\boldsymbol{W}$ are defined in (6.2).

After the three steps described above, we achieve an important reduction in the dimension of the solution space of 30 orders of magnitude with respect to ESS, as we show in Section 6.4.

6.3.2 Algorithm to solve the MAMVL scheme

The proposed algorithm for the MAMVL scheme solves (P2) and takes advantage of the search space reduction developed in the previous subsection. The corresponding pseudocode is given in Algorithm 8.

Algorithm 8 uses as inputs the wideband CQIs received from the multicast members $\boldsymbol{\alpha}$, the data rate per PRB achieved using each MCS $\boldsymbol{\tau}$, the data rate required to deliver the video content of each layer $\boldsymbol{d}$, and the number of available PRBs N_{PRB}. First, the MCS μ_1 required to satisfy the user that reports the worst CQI is calculated (line 7). After that, the algorithm applies the first reduction of the search space, following (6.9) to (6.10), where $u_{j,l}$ values are calculated (lines 8-10).

We now update the binary matrix $\boldsymbol{W}$ defined in (6.2) according to the prior considerations. Recall that $\boldsymbol{W}$ denotes the assignment matrix after the second and third search space reduction. Therefore, the element of $\boldsymbol{W}$ in row 1 corresponding to μ_1 is set to '1', indicating the only possible solution for layer $l = 1$, i.e. $w_{\mu_1,1} = 1$, the remaining elements in row 1 are set to '0' (line 11), and the remaining PRBs N'_{PRB} are calculated (line 12).

The third search space reduction (lines 13-24) calculates the values of $\boldsymbol{W}$ for layers $l = 2, ..., L$ as follows: $w_{j,l}$ is set to '1' to denote that using MCS j requires a number of PRBs less or equal than N'_{PRB} and fulfilling (6.14) (lines 13-19). Next, the binary matrix $\boldsymbol{W}$ is updated to reduce the possible solutions according to (6.15) and (6.16) (lines 20-24). For that purpose, $w_{j,l}$ is set to '0' if one of the following conditions holds: *(i)* the sum of PRBs required to transmit all layers is larger than the available number of PRBs N'_{PRB}, or *(ii)* j_{l+1} is less robust than j_l, i.e. $j_{l+1} < j_l$.

As a result, every multicast user receives the video with a quality (data

Algorithm 8 Multicast resource allocation based on multiple video layers: MAMVL

1: **Inputs**:
2: Vector of reported CQIs: $\boldsymbol{\alpha}$
3: Vector of MCS' data rate per PRB: $\boldsymbol{\tau}$
4: Vector of video layer's data rate: $\boldsymbol{d}$
5: Number of available PRBs: N_{PRB}
6: **Search space reduction** ▷ Obtain all possible solutions:
7: $\mu_1 \leftarrow min_{\alpha_m}(\boldsymbol{\alpha})$ ▷ Search worst CQI reported (6.8)
8: **for all** $(l \in \mathcal{L})$ *and* $(j \in \mathcal{J})$ **do** ▷ Matrix $\boldsymbol{U}$ building
9: $\quad u_{j,l} \leftarrow \lceil d_l/\tau_j \rceil$ ▷ (6.10)
10: **end for**
11: $w_{j,1} = \begin{cases} 1, & \text{if } j = \mu_1 \\ 0, & \text{otherwise} \end{cases}$ ▷ Update matrix $\boldsymbol{W}$ for layer 1
12: $N'_{PRB} \leftarrow N_{\mathrm{PRB}} - r_{j,1}$ ▷ (6.12)
13: **for all** $l \in [2, L], \quad j \in [\mu_1, J]$ **do**
14: $\quad$ **if** $(u_{j,l} \leq N'_{PRB})$ *and* $(u_{j+1,l} < u_{j,l})$ **then** ▷ (6.14)
15: $\quad\quad w_{j,l} \leftarrow 1$ ▷ Update matrix $\boldsymbol{W}$ for layers 2,..,L
16: $\quad$ **else**
17: $\quad\quad w_{j,l} \leftarrow 0$
18: $\quad$ **end if**
19: **end for**
20: **for all** $(j, l)/\{w_{j,l} = 1\}$ **do**
21: $\quad$ **if** $\left(\sum_{l\in[2,L]} u_{j,l} > N'_{PRB}\right)$ *or* $(w_{j,l+1} = 0, \forall j \in (1, \rho] : w_{\rho,l} = 1)$ **then** ▷ (6.15) and (6.16)
22: $\quad\quad w_{j,l} \leftarrow 0$
23: $\quad$ **end if**
24: **end for**
25: **Search for MT maximization** ▷ The search space is denoted by $\boldsymbol{W}$
26: Select $r_j^* \leftarrow \arg\max \sum_{l=1}^{L} d_l \sum_{m=1}^{M} z_{m,l}$ ▷ (6.4)
27: **Outputs**:
28: Optimal allocation vector: $\boldsymbol{r^*}$

rate) according to the number of layers that the user can correctly decode. Finally, the sum of all users data rate is evaluated according to (6.4) in order to determine the configuration that ensures to maximize the ADR (line 26).

6.4 Performance assessment

In this section, a circular single cell downlink LTE-like MIMO-OFDMA wireless system is considered (see Figure 6.1). A BS delivers video content to a set of multicast users uniformly distributed over the whole coverage area, in which the BS is equipped with one or two transmit antennas, and the multicast users are also equipped with one or two receive antennas, respectively. Eighteen interfering BSs have been considered around the coverage area in order to model the interference signal that the multicast users are receiving. The MIESM [78] has been used as the link abstraction technique employed to obtain a scalar value of effective SINR that can be mapped onto a CQI value for a required target $\text{BLER}_0 = 0.1$. System parameters have been defined based on current LTE/LTE-A specifications [74], and Table 6.1 summarizes the most relevant parameters considered in the simulations[1].

The transmitted signal experiences path-loss, large-scale shadow fading and small-scale fading where the frequency, time and space characteristics of the channel are modelled. Path-loss and shadowing have been simulated using the macro cell propagation model for urban area described in [89]. Small-scale fading has been generated using channel power delay profiles conforming to either the ETU or the EPA channel models defined within LTE/LTE-A [90]. The ETU and EPA channel models have been simulated using maximum Doppler frequencies of $f_d = 70$ Hz and $f_d = 5$ Hz, respectively. That is, the ETU channel with $f_d = 70$ Hz represents an environment showing high time variability and frequency selectivity, whereas the EPA channel with $f_d = 5$ Hz is a good example of a low-mobility scenario with low frequency selectivity.

Simulations using both ETU and EPA channel profile, as well as SISO and MIMO configurations, have been performed. For the sake of accuracy, each simulation has been run over 100 TTIs and has been repeated for 100 random positions of the users. The path-loss and large-scale shadow fading are fixed during each simulation of 100 TTIs, whereas these fixed positions are random for the 100 different simulations. As small-scale fading varies every TTI according to ETU or EPA channel models, this allows us to take

[1]Note that we assume the use of normal cyclic prefix, since we perform our simulations in a single cell multicast scenario. Extended cyclic prefix is recommended in large cell scenarios, or multicast scenarios using MBSFN among multiple cells.

Table 6.1: Summary of simulation parameters in MAMVL assessment.

Parameters	**Value**
BS transmit power	47 dBm
BS/MS antenna gain	6/3 dBi
MS noise figure	7 dB
Thermal noise	-174 dBm/Hz
Number of interfering BSs	18
Cell radius	500 m
Shadowing deviation	6 dB
Shadowing correlation	0.5
Channel model	ETU & EPA
Carrier frequency	2.1 GHz
Subcarrier spacing	15 KHz
Subcarrier per subband	12
Multicarrier symbols per TTI	14
TTI duration	1 ms
Duration of long cyclic prefix	5.2 μs
Duration of short cyclic prefix	4.7 μs
Multicast user distribution	Uniform
Number of TTIs in the simulation	100
Number of experiments in the simulation	100

into account the effects of user mobility in the multicast resource allocation, since the large-scale effects are almost constant during a 100 TTIs simulation even when the user is moving with a high speed.

We have simulated a multicast system with 2, 3 and 4 video layers to obtain a comparison in terms of ADR, fairness and spectral efficiency. SISO and MIMO[2] 2×2 (hereinafter referred to as MIMO) systems have been considered using, for the second system, transmit and receive correlation matrices conforming to the low correlation profile defined in [90].

This section first presents a complexity evaluation of the search space

[2]In a multicast service scenario, MIMO TD schemes are the most suitable ones. In this work, only TD schemes for two transmit antennas and one codeword will be considered, and will be compared with the single transmit antenna scheme.

reduction proposed for the MAMVL scheme. Next, an analysis of the multi-layer multicast system in terms of the ADR, fairness and spectral efficiency with different set-ups is presented. After that, specific services are considered to evaluate the influence of the type of service. Finally, a comparison among the proposed MAMVL scheme and other strategies presented in the literature is shown.

6.4.1 Complexity analysis considering the search space reduction

In this section, we compare our MAMVL with different approaches existing in the literature in terms of complexity. With this purpose, Figure 6.2 illustrates the cardinality of the solution space when different schemes and number of layers are used.

We observe that the reduction of the solution space using the MAMVL scheme is very significant with respect to ESS and OSS, which implement the computational burden reduction for multicast subgrouping in [63], and with respect to the multicast resource allocation (MRA)/multicast subgrouping for multilayer video applications (MSML) SVC-based schemes proposed in [16]. Note that this difference is several orders of magnitude. At the same time, Figure 6.2 shows that the complexity of MAMVL using 2 video layers is similar to the optimal solution for multicast subgrouping (MT) proposed in [30], adding the benefits of taking into account the SVC technique to improve the ADR performance, as it can be observed in Section 6.4.4. Furthermore, the number of possible solutions only increases slightly with the number of PRBs with MAMVL. This leads to low complexity regardless of the number of resources and makes it possible to implement MAMVL in real systems with a large number of resources.

It can be noticed that the number of operations depends on the number of layers used, i.e. an average of 15, 40 and 122 possible solutions must be evaluated for 100 PRBs using MAMVL with 2, 3 and 4 video layers, respectively. These results are relevant in comparison with the cardinality of the solution space for MRA/MSML SVC-based schemes [16], that is

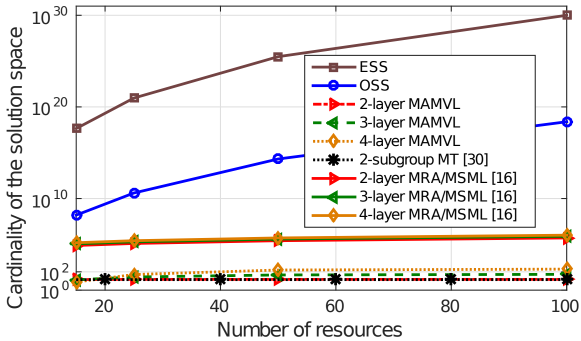


Figure 6.2: Comparison of the search space for different approaches: ESS, OSS, the optimal solution for multicast subgrouping [30], and SVC schemes using 2, 3 and 4 layers (MRA/MSML and the proposed MAMVL) to split the multicast content.

Table 6.2: Data rates available for each video layer in MAMVL assessment.

System	Layer	Data rate options (Kb/s)
2-layer	1	25-50-100-200
	2	500-1000-2000-4000-8000
3-layer	1	25-50-100-200
	2	500-1000-2000-4000
	3	1000-2000-4000-8000
4-layer	1	25-50-100-200
	2	500-1000-2000
	3	1000-2000-4000
	4	2000-4000-8000

5×10^5, 7.5×10^5 and 10^6 using 2, 3, and 4 video layers, respectively. This enhancement is even more relevant in comparison with ESS and OSS [63] that require 10^{30} and 2.4×10^{18} operations every TTI, respectively.

6.4.2 MAMVL scheme performance evaluation

In this subsection we present the results achieved by means of the MAMVL scheme using different number of layers. The results have been obtained using a wide set of available data rates (Table 6.2) to deliver a service and selecting the one that maximizes the ADR in each configuration[3]. The performance analysis will be based on the following metrics:

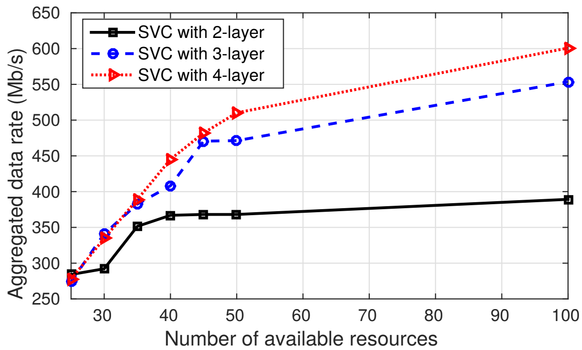
- **ADR** $[b/s]$: Defined as the sum of the useful data rates (measured in b/s) offered to the whole set of users in the multicast group (note that ADR was previously defined in the numerical results of Chapter 5).
- **Fairness**: Jain's fairness index (FI) [91] has been used to evaluate how fair is the solution among all the multicast users. Jain's FI is expressed as

$$FI = \frac{\left\{ \sum_{m=1}^{M} \sum_{l=1}^{L} d_l z_{m,l} \right\}^2}{M \left\{ \sum_{m=1}^{M} \left(\sum_{l=1}^{L} d_l z_{m,l} \right)^2 \right\}} \tag{6.17}$$

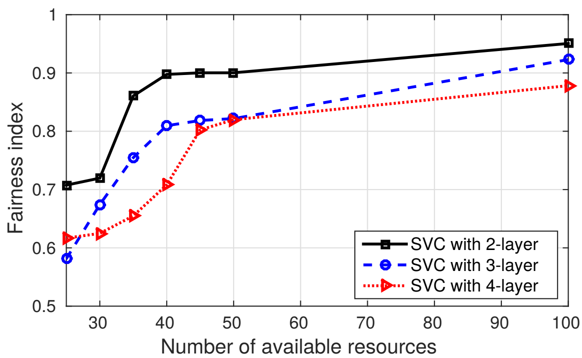
  where $\sum_{l=1}^{L} d_l z_{m,l}$ corresponds to the data rate of the video content delivered to each multicast user in a certain TTI.
- **Spectral efficiency**: $[b/s/Hz]$: Defined as the ratio between the ADR and the bandwidth, i.e PRBs, exploited by the BS.

Figures 6.3 to 6.5 show the MAMVL performance in terms of the three metrics defined above with respect to the number of available PRBs, for 50 users uniformly distributed over the coverage area. Figures 6.3a, 6.4a and 6.5a show that regardless of the channel model and MIMO configuration, the ADR increases with the number of PRBs and layers. Note that for every simulation we select the bit rate configuration that maximizes the ADR from the whole set (see Table 6.2), which depends on the reported CQIs by the multicast users and the number of PRBs. This saturates the linear increase of the ADR when the system uses a high number of PRBs.

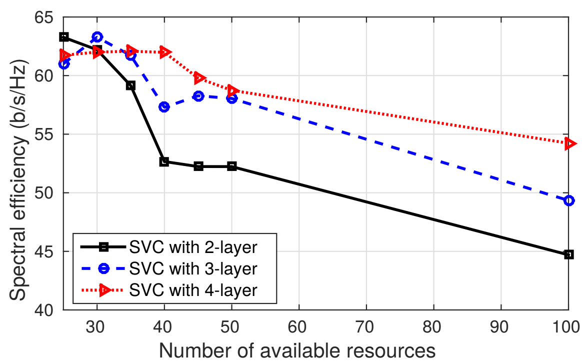
[3]Multiple combinations of data rates for each video layer have been tested to select the configuration that maximizes the ADR in every scenario. In order to achieve consistent results, this complete test is performed since the number of operations and the performance, i.e ADR, fairness or spectral efficiency, depends on the set of data rates used. In contrast, video delivery in a real system consists of a given set of video layers with fixed data rates, as it is analyzed in the following subsection.



(a) Aggregated Data Rate (ETU 1x1)



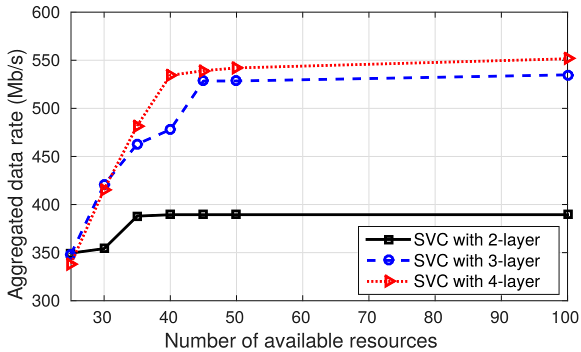
(b) Fairness Index (ETU 1x1)



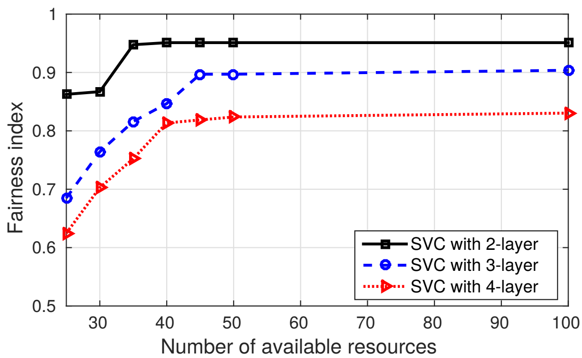
(c) Spectral Efficiency (ETU 1x1)

Figure 6.3: Performance evaluation of MAMVL scheme with 2, 3 and 4 SVC layers to deliver a multicast video service to 50 users in an ETU channel, SISO system.

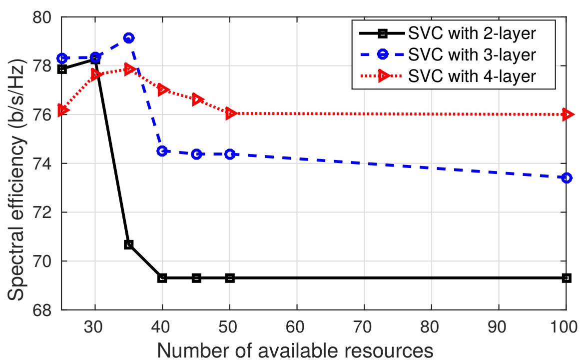
Comparing the results for 2, 3 and 4 layers, it is observed that a lower layer system reaches this saturation using lower number of PRBs, what indicates



(a) Aggregated Data Rate (ETU 2×2)



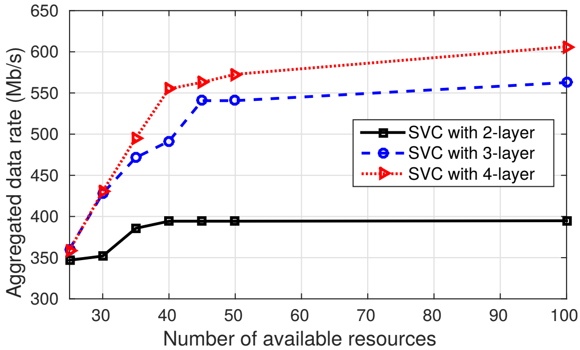
(b) Fairness Index (ETU 2×2)



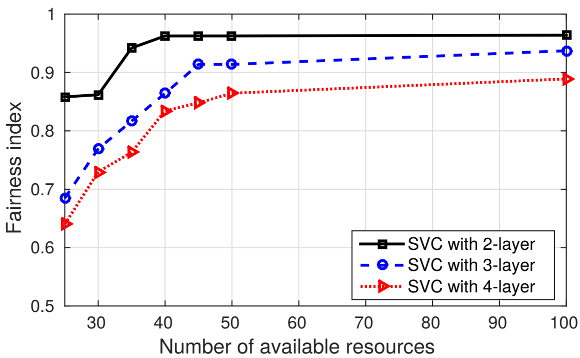
(c) Spectral Efficiency (ETU 2×2)

Figure 6.4: Performance evaluation of MAMVL scheme with 2, 3 and 4 SVC layers to deliver a multicast video service to 50 users in an ETU channel, 2×2 MIMO system.

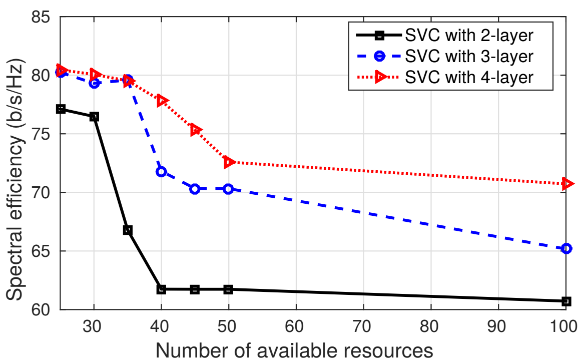
that using a higher layer increases the ADR at the expense of an excessive number of PRBs to deliver high data rates. Furthermore, such an ADR



(a) Aggregated Data Rate (EPA 2×2)



(b) Fairness Index (EPA 2×2)



(c) Spectral Efficiency (EPA 2×2)

Figure 6.5: Performance evaluation of MAMVL scheme with 2, 3 and 4 SVC layers to deliver a multicast video service to 50 users in an EPA channel, 2×2 MIMO system.

can be achieved in both SISO channel (Figure 6.3a) and MIMO channel (Figure 6.4a), but at the expense of wasting resources, since the channel conditions reported by the users in a SISO channel are worse. Observing the ADR achieved in an EPA channel (Figure 6.5a), we can conclude that it is slightly higher than in an ETU channel (Figure 6.4a), due to its lower time variability and frequency selectivity.

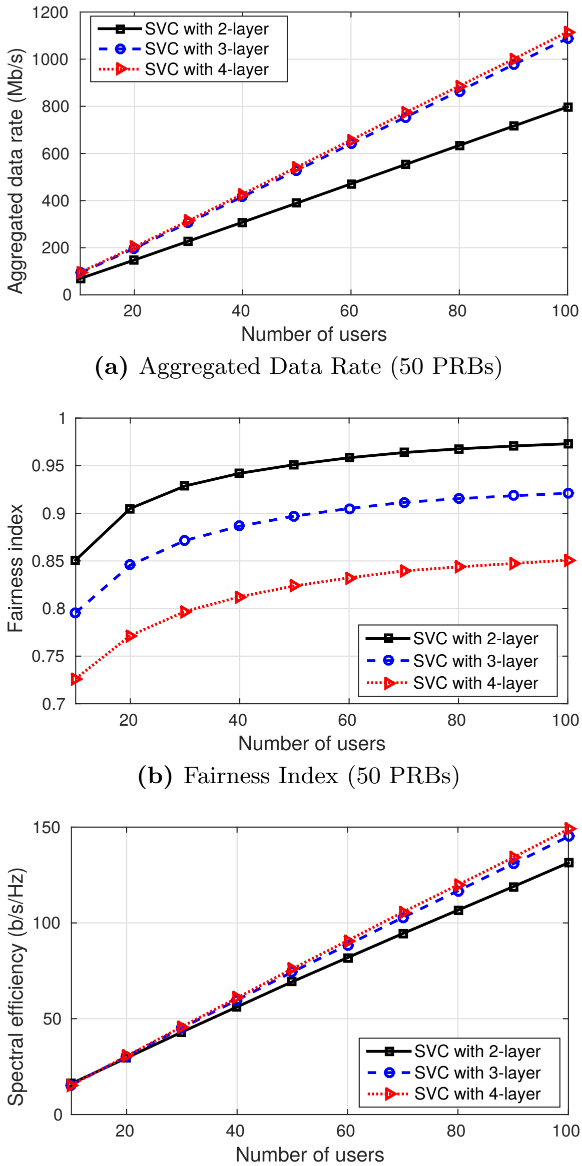
Figures 6.3b, 6.4b and 6.5b present the results in terms of fairness among users. These results show that the higher the number of layers the lower the fairness among the users. In contrast, Figures 6.3c, 6.4c and 6.5c illustrate the benefits of using higher number of layers in terms of spectral efficiency. This enhancement is more noticeable when the number of available resources is higher.

Regarding channel characteristics, Figures 6.3 to 6.5 show that MAMVL presents similar performance independently of the channel profile. Although MAMVL benefits of the lower time variability and frequency selectivity of the EPA channel in order to achieve better performance, both in terms of the ADR and fairness, the same general conclusions can be derived from using the MAMVL over ETU and EPA channel profiles.

SISO configuration (Figures 6.3a to 6.3c) results in a significant degradation of the performance with respect to MIMO configuration (Figures 6.4a to 6.4c). As it is expected, the lower SINR experienced by the users in the SISO channel set-up the lower the ADR achieved and consequently the spectral efficiency.

To study how the number of multicast users affects the performance of MAMVL, Figure 6.6 presents the ADR, fairness and spectral efficiency results when 50 PRBs are available for different number of users in an ETU channel with MIMO configuration. Both the ADR (Figure 6.6a) and the spectral efficiency (Figure 6.6c) proportionally increase with the number of multicast users, resulting in a better performance using higher number of layers. Note that a high number of PRBs (50) results in quite similar performance of using 3 and 4 layers in terms of ADR and spectral efficiency. In contrast, 2-layer system performance is gradually degrading in comparison with higher layer systems when the number of multicast users is increased.

It should be noted that the results in terms of fairness among users



(a) Aggregated Data Rate (50 PRBs)

(b) Fairness Index (50 PRBs)

(c) Spectral Efficiency (50 PRBs)

Figure 6.6: Performance evaluation of MAMVL scheme with 2, 3 and 4 SVC layers to deliver the multicast video service to a variable number of users (ETU channel, 2×2 MIMO system, and 50 PRBs).

(Figure 6.6b) improve sharply with an increase of users when the number of multicast users is low, and fairness gradually increases for large number of users. It can be observed that the distance among the curves corresponding to 2-, 3- and 4-layer system is almost constant. Although the proposed scheme is not designed to maximize the fairness, the use of SVC multi-layer system results in a high performance in terms of fairness, and using 2-layer system achieves the highest fairness performance.

6.4.3 Performance analysis of MAMVL scheme for different types of video service

For this analysis, three different video services are selected with fixed data rates per layer (see Table 6.3), where one low rate service (NEWS), one medium rate service (ICE) and one high rate service (CITY) are specified.

The numerical results in terms of ADR, fairness, spectral efficiency and the average number of resources are presented when a multicast video service is delivered to 50 users and MAMVL is employed for ETU channel model with SISO and MIMO configurations (Tables 6.4 and 6.5) and EPA channel model with MIMO configuration (Table 6.6). The number of PRBs (100) is sufficiently large to deliver all services without any constraint.

Regardless the channel model and the MIMO configuration, the high data rate service (CITY) presents the better performance both in terms of ADR/spectral efficiency and fairness among users, at the expense of using a higher number of resources. The ADR is largely higher for services that require high data rate, as expected, whereas fairness and spectral efficiency results are quite similar regardless the data rate the service requires.

Regarding the number of layers, the results obtained in this subsection agree with the prior subsection. The higher the number of layers the better the performance in terms of ADR and spectral efficiency. In contrast, the higher the number of layers, the lower the fairness and the higher the number of required resources.

Comparing the results for SISO and MIMO channels, we observe how the ADR achieved can be similar, or even higher in the SISO channel depending on the service and the number of layers. For instance, Table 6.4 shows the

Table 6.3: Data rates (Kb/s) per layer for different services [92] employed in MAMVL assessment.

Name	Layer 1	Layer 2	Layer 3	Layer 4
CITY	448	923	1288	1943
NEWS	121	259	372	564
ICE	277	548	767	1123

Table 6.4: Numerical results for the delivery of CITY, ICE and NEWS services to 50 users with a multi-layer multicast system in an ETU channel model with a SISO system.

Name	Number of layers	ADR (Mb/s)	Fairness	Spec. eff. (b/s/Hz)	Resources (avg.)
CITY	2	114.500	0.9658	39.0407	16.3269
	3	169.440	0.9430	42.0002	22.4462
	4	200.740	0.9135	46.3277	24.1134
ICE	2	67.084	0.9667	37.1198	10.0598
	3	99.899	0.9442	41.2610	13.4702
	4	118.539	0.9147	43.9372	15.0124
NEWS	2	32.768	0.9645	34.4013	5.3039
	3	46.904	0.9257	37.3742	6.9893
	4	48.757	0.8407	40.4270	6.7237

ADR = 169.440 Mb/s for CITY service in the SISO channel using 3 layers, whereas Table 6.5 presents this ADR = 164.880 Mb/s for CITY service in the MIMO channel using 3 layers, too. However, it is at the expense of using much higher number of PRBs in the SISO channel (an average of 22.4462) than in the MIMO channel (an average of 13.9008), since users report worse channel conditions. It results in a much higher spectral efficiency when the multicast service is delivered in a MIMO channel (66.0450 b/s/Hz) than in a SISO channel (42.0002 b/s/Hz). Note that the spectral efficiency achieved is extremely high, both in a MIMO and SISO channel, because of the use of multicast transmission to deliver the same content to 50 users. Table 3.1 shows spectral efficiency for unicast transmissions. It can be observed that

Table 6.5: Numerical results for the delivery of CITY, ICE and NEWS services to 50 users with a multi-layer multicast system in an ETU channel model with a 2×2 MIMO system.

Name	Number of layers	ADR (Mb/s)	Fairness	Spec. eff. (b/s/Hz)	Resources (avg.)
CITY	2	114.560	0.9662	60.6209	10.5260
	3	164.880	0.9312	66.0450	13.9008
	4	187.883	0.8843	67.3983	15.5252
ICE	2	67.114	0.9671	55.2731	6.8758
	3	93.011	0.9029	60.1471	8.6087
	4	96.400	0.8144	62.4989	8.5800
NEWS	2	30.864	0.9191	54.5516	3.1457
	3	34.014	0.7504	59.1827	3.1988
	4	34.680	0.6292	62.6915	3.0888

spectral efficiency using multicast is one order of magnitude higher than using unicast with the less robust MCS.

Furthermore, we compare the performance results in EPA (Table 6.6) and ETU (Table 6.5) channels. Looking at the average resource consumption we can observe how EPA channel requires higher number of resources than ETU channel what affects the spectral efficiency, i.e. delivering a service in an EPA channel would present slightly worse spectral efficiency than delivering the same service in an ETU channel.

Finally, Figure 6.7 presents the ADR achieved when the CITY service is delivered to 50 users and the number of available resources is constrained. It can be noticed that the lower the number of available PRBs the more convenient the use of lower number of layers. Figure 6.7 shows that the availability of only 5 PRBs is required to ensure the delivery of the CITY service using SVC with 2 layers. However, 7 and 9 available PRBs are required to guarantee the delivery of the video CITY service using SVC with 3 and 4 layers, respectively. The benefits in terms of ADR of using 3 layers to deliver CITY service occur when the number of available PRBs is

Table 6.6: Numerical results for the delivery of CITY, ICE and NEWS services to 50 users with a multi-layer multicast system in an EPA channel model with a 2×2 MIMO system.

Name	Number of layers	ADR (Mb/s)	Fairness	Spec. eff. (b/s/Hz)	Resources (avg.)
CITY	2	115.860	0.9750	53.1822	12.1370
	3	171.850	0.9542	57.4219	16.6745
	4	202.480	0.9234	62.1916	18.1296
ICE	2	67.868	0.9756	49.5424	7.6228
	3	100.240	0.9481	54.4479	10.2507
	4	112.120	0.8885	58.5884	10.6631
NEWS	2	32.234	0.9518	47.7691	3.7595
	3	40.541	0.8616	53.6739	4.2173
	4	41.329	0.7839	60.4067	4.1329

higher than 9, and SVC with 4 layers requires more than 12 PRBs available to deliver the CITY service resulting in better ADR results.

6.4.4 Performance comparison between the MAMVL scheme and the MT strategy

This subsection presents a comparison between the performance achieved to deliver the three services of Table 6.3 to 50 multicast users employing the MAMVL scheme and the 2-subgroups MT strategy presented in [30]. Recall that 2-subgroups MT uses subgrouping strategy without the knowledge of the available video data rate to be transmitted. Note that the MRA and MSML and strategies [16] that employ SVC have also been compared with MAMVL. However, these results are not included in this work as all these schemes provide optimal solutions according to the utility function they optimize and the use of the same utility function in all of them implies the same performance results. The utility functions employed imply that the MAMVL scheme achieves higher ADR than the MRA and MSML schemes,

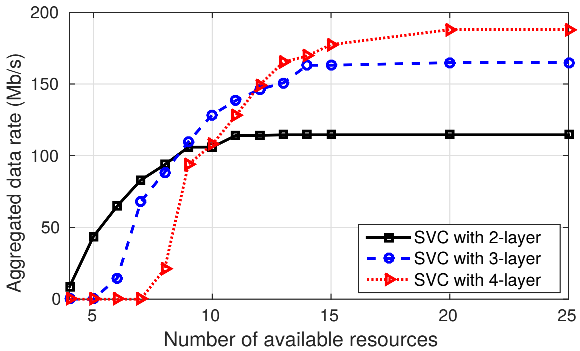


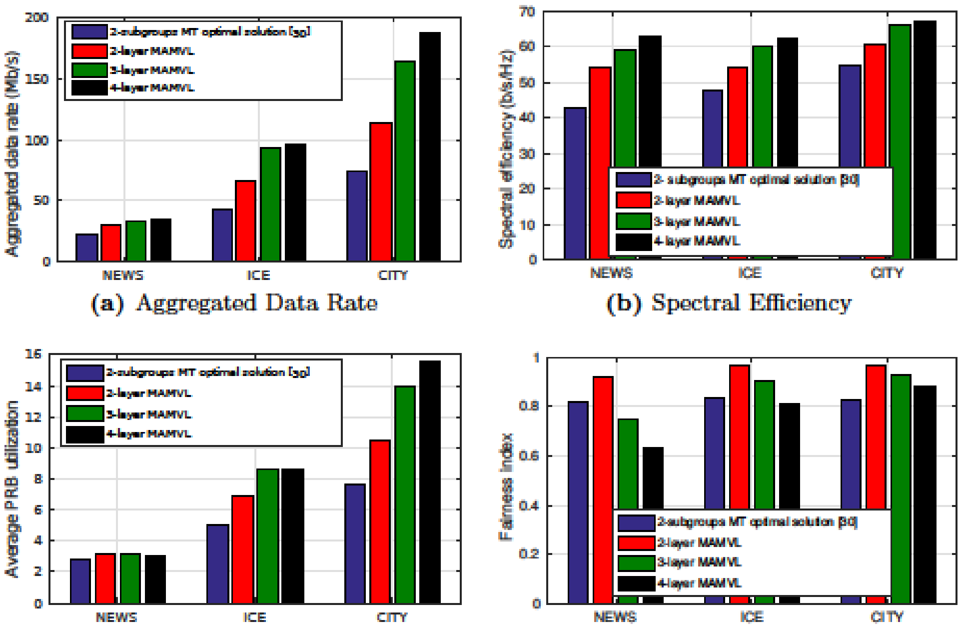
Figure 6.7: ADR as a function of the number of available PRBs using MAMVL scheme with 2, 3 and 4 SVC layers to deliver the CITY video service (ETU channel, 2×2 MIMO system, and 50 users).

because ADR is the utility function that is directly maximized in MAMVL.

Figure 6.8a shows the achieved ADR. It can be observed that the MAMVL scheme improves the ADR achieved with MT regardless of the number of layers. The enhancement of using MAMVL can be also appreciated in terms of spectral efficiency (Figure 6.8b), although MT uses a lower number of PRBs, as Figure 6.8c illustrates, the MAMVL scheme achieves better performance of the PRBs employed.

Figure 6.8d shows how the fairness decreases with the number of layers. Note that this reduction is sharper for low rate services (NEWS) than for high rate services (CITY). That is, MAMVL scheme presents better fairness performance regardless of the number of layers than MT strategy for high data rate services. However, the fairness performance degrades with high number of layers for services that require lower data rate. E.g. MAMVL with 4 layers for ICE service presents slightly worse fairness than MT, and MAMVL with 3 or 4 layers for NEWS service results in clearly worse fairness than MT.

Figures 6.8a to 6.8c show that the MAMVL scheme allocates PRBs more efficiently than the MT strategy, as it improves the performance in terms of ADR, fairness (for high data rates) and specially spectral efficiency. Consequently, by using MAMVL it is possible to maximize the ADR depending on the service delivered, using a reduced number of PRBs. For example, using



(a) Aggregated Data Rate

(b) Spectral Efficiency

(c) Average resource utilization

(d) Fairness Index

Figure 6.8: Comparison of the performance achieved by MAMVL and MT strategies to deliver a multicast service (ETU channel, 2×2 MIMO system, 50 users, and 50 PRBs).

a 4-layer MAMVL scheme, the delivery of the CITY service is optimized in terms of ADR maximization using an average number of 15.5 PRBs, and only an average number of 3.1 PRBs is used to deliver the NEWS service. This obviously leaves a big number of unused resources for other services.

6.5 Summary

In this chapter, a new resource allocation scheme is developed to deliver multicast video services that makes use of the SVC multiple layers. The proposed MAMVL scheme is focused on reducing the complexity of the solution space to develop an algorithm easy to implement in real systems and maximizing the ADR.

Numerical results obtained for 2, 3 and 4 video layers reveal the bene ts of MAMVL both in terms of the ADR and the fairness among users, but specially in terms of the spectral efficiency achieved, because of a better allocation of the available PRBs among the different layers. The performance evaluation indicates that the spectral efficiency achieved by using MAMVL improves MT strategies regardless the number of layers and the kind of service delivered.

This work presents the proposed MAMVL algorithm and the results of the resource allocation performed individually for different kinds of multicast service. This strategy can be extended to deliver more than one multicast service with the introduction of new policies to manage the resource allocation among different multicast groups, e.g. intergroup fairness, ADR or resource consumption.

A future MAMVL scheme for joint allocation of several multicast services and unicast deliveries will provide a more complete tool for real systems, where the intra-group resource allocation scheme is in charge of maximizing the ADR, whereas new policies must be taken into account in order to manage the different QoS requirements for each unicast and multicast delivery.

This chapter has been based on the following submitted (revised manuscript) work:

- **Alejandro de la Fuente**, José Joaquín Escudero Garzás, and Ana García Armada, "Radio resource allocation for multicast services based on multiple video layers", *Submitted the revised version to IEEE Transactions on Broadcasting.*

Conclusions

Conclusions

The *tsunami* of new mobile applications, which implies an extraordinary increase in video transmissions, brings stringent requirements in the future mobile networks. Such demand requires the provision of broadband services which enable high data rates, high spectrum efficiency, low latency and, in general, high system capacity. Broadcast/multicast service is a viable and effective solution to provide those applications which achieves an efficient use of the spectrum. The evolution of the multicast service will make it more dynamic, useful in the vast majority of the applications and more scalable depending on the number of users. Nevertheless, emerging technologies need to be continuously enhanced to achieve the goals of next generation broadcasting.

Radio resource management (RRM) for multicast transmissions is one of the main challenges that this technology has to face in order to provide the enhancements that the 5th generation of mobile communications (5G) networks demand. In recent years, the investigations on dynamic multicast scheduling and resource allocation (SRA) strategies have multiplied in order to maximize the benefits of using multicast transmissions in heterogeneous networks, where the service is delivered to users that experience different signal to interference plus noise ratio (SINR). This thesis includes novel works that we have investigated with the aim of improving the literature in the field of SRA solutions for multicast services in mobile networks. Towards the development and validation of the contributions presented, we have implemented a physical abstraction described in the system model. Such abstraction is based on the effective SINR, which allows us to model the channel quality indicator (CQI) fed back by the users to the base station (BS).

The first contribution we have proposed in this thesis consists in the use of joint SRA strategies among the complete set of subframes in Long Term Evolution (LTE) evolved multimedia broadcast and multicast service (eMBMS). The joint multicast/unicast scheduling (JMUS) strategy obtains the optimal values of the modulation and coding scheme (MCS) and the number of subframes reserved for multicast transmissions in a dynamic way, at the time it employs the unicast subframes to serve the users experiencing poor channel conditions. We have implemented a fast search algorithm through JMUS strategy, that achieves very close to optimal results with a reduction in the computation complexity. In addition, the joint multicast subgrouping and unicast transmissions (JMSUT) scheme have been developed with the aim of enhancing the JMUS performance. The JMSUT proposes the creation of different multicast subgroups and splits the allocation of the available physical resource blocks (PRBs) among them. These multicast subgroups are combined with the transmissions in the reserved unicast subframes, which are used to serve the users with worst conditions. The JMSUT can greatly improve, in terms of service aggregated data rate (ADR), the results achieved using the conventional multicast scheme (CMS) or the JMUS strategy. At the same time, both the JMUS and the JMSUT allow all the users to achieve the minimum required data rate.

Another one of the main contributions made explores the efficient utilization of subband CQI in order to optimize multicast transmissions in the context of current mobile networks. Since the derivation of the optimal solution to the ADR maximization problem is, when subband CQI is employed, computationally infeasible, suboptimal solutions have been proposed. In particular, it has been established that an effective strategy consists of greedily splitting the multicast users into two subgroups with the aim of, first, satisfying a minimum data rate request and second, maximizing the multicast service ADR, all being subject to block error rate constraints. The two-step two-subgroup approach based on subband CQI leads to substantial benefits in terms of the achieved ADR. The exploitation of the granularity that subband CQI brings along becomes especially significant in channels with strong frequency selectivity, where wideband CQI becomes very imprecise regarding the true potential capacity of the users' channels. We have implemented two solutions based on subband CQI, namely, the greedy subband CQI algorithm (GSCA) used together with second group subband CQI algorithm (SG-SCA) and the sorted subband CQI algorithm (SSCA) used together with SG-SCA, which are shown to significantly outperform state-of-the-art proposals with regard to the ADR and the service outage probability, employing minimum data rate requirements of practical interest and under most of the system and channel configurations that can be found in real scenarios.

Finally, in the last main contribution presented, we have proposed a new resource allocation scheme to deliver multicast video services, that makes use of the scalable video coding (SVC) multiple layers. The proposed multicast resource allocation based on multiple video layers (MAMVL) scheme is mainly focused on reducing the complexity of the solution space and, at the same time, maximizing the ADR. In such a way, it is possible to develop a low complexity algorithm which is easy to implement in real systems. The MAMVL using 2, 3 and 4 video layers presents substantial benefits both in terms of the ADR and the fairness among users, but especially in terms of the spectral efficiency, because of a better allocation of the available PRBs among the different layers than those algorithms that do not use the a priori knowledge of the content data rate to be delivered.

As a conclusion, it is worth mentioning that all the contributions made in this thesis are becoming part of the state of the art in the RRM for multicast service proposals, providing enhancements in the spectral effciency and mechanisms for developing low complexity solutions that could be implemented in real systems. Regarding the first topic, both the splitting of the available resources into multicast subgroups and the use of subband CQI information have shown to be excellent options for developing algorithms that provide high spectral efficiency. On the other hand, the utilization of systems based on multiple layers, where the data rates of the content available to be delivered in each layer is previously known, allows us to implement low complexity solutions, which are essential for the use of these schemes in real systems.

Future lines of research

The different contributions made in this thesis are intended to stimulate and serve as a reference for the ongoing research on the promising field of the next generation broadcast/multicast services, especially in RRM, the most critical part for their evolution. Further research work will pursue the design of novel SRA algorithms taking into consideration the following challenging issues:

- The development of SRA algorithms, based on a clever use of subband-CQI, that are in charge of jointly optimizing the allocation of resources to both multicast and unicast services. Note that the works presented in this thesis allocate resources for the delivery of only one multicast service. These strategies can be extended to deliver more than one multicast service with the introduction of new policies to manage the resource allocation among different multicast and/or unicast services, e.g. intergroup fairness, ADR, or resource consumption. Therefore, a future SRA scheme for joint allocation of several multicast and unicast services will provide a more complete tool for real system analysis. Employing multi-objective utility functions, the intra-group resource allocation scheme is in charge of maximizing the ADR, whereas new policies must be taken into account in order to manage the different quality of service (QoS) requirements for each unicast and multicast deliveries.

- A very demanding challenge consists in adapting the proposed SRA strategies developed for cellular links to a new heterogeneous environment. Small cells and device-to-device (D2D) communications are considered an interesting means to extend coverage and improve the channel conditions of the users located at the macro cell edge. Cellular links can be offloaded with the assistant of both small cells and D2D short-range links. The multicast subgrouping strategy can be extended to allocate not only the macro cell resources but also the short-range links ones.

- Machine-type communication (MTC) is the main enabler of a wide range of novel applications and services expected in 5G systems. The MTC characteristics and its specific requirements, such as short connection time, low-energy data transmission, discontinuous reception (DRX), reduction of signalling, control traffic and high reliability to the thousands of devices which are foreseen to be camped in each cell, make RRM for MTC a challenging issue. Furthermore, we should take into consideration that SRA strategies that efficiently support group-oriented MTC traffic are still needed to be designed. These schemes require to cut delays and allow scalability when the number of receivers is huge and the MTC data traffic is usually characterized by small packets.

Conclusiones

Conclusiones

El *tsunami* de nuevas aplicaciones móviles, que implica un extraordinario crecimiento en la transmisión de vídeo, trae consigo requisitos muy exigentes en las redes de próxima generación. Dicha demanda require la provisión de servicios móviles de banda ancha que proporcionen altas tasas de transmisión, alta eficiencia espectral, baja latencia y, en general, un sistema de altas prestaciones. El servicio *broadcast/multicast* resulta una solución efectiva y viable para la provisión de dichas aplicaciones, proporcionando una utilización eficiente del espectro. La continua evolución de los servicios *multicast* hace que sean actualmente más dinámicos, utilizables en una gran mayoría de aplicaciones y escalables en función del número de usuarios. No obstante, para alcanzar los retos que demandan los servicios *multicast* de nueva generación, resulta preciso seguir incorporando mejoras a las nuevas tecnologías.

La gestión de recursos radio (RRM) en las transmisiones *multicast* es uno de los retos más importantes que se deben encarar para proporcionar las mejoras que demandan las redes de la quinta generación de comunicaciones móviles (5G). En los últimos años ha proliferado la investigación en estrategias de asignación dinámica de recursos que permitan maximizar los beneficios de emplear transmisiones *multicast* en redes heterogéneas, donde los usuarios que demandan el servicio experimentan diferente relación señal a interferencia más ruido (SINR). En esta tesis hemos incluido trabajos novedosos llevados a cabo con el objetivo de mejorar el estado del arte en el campo de las soluciones de asignación de recursos (SRA) para servicios *multicast* en redes móviles. Para el desarrollo y validación de las diferentes contribuciones expuestas, hemos implementado una abstracción de la capa física descrita en el modelo del sistema. Dicha abstracción se basa en el uso de la SINR efectiva, la cual nos permite modelar el indicador de calidad del canal (CQI) reportado por los usuarios a la estación base (BS).

La primera contribución propuesta en esta tesis consiste en la utilización de estrategias de SRA de forma conjunta para todas las subtramas del servicio *broadcast* y *multicast* multimedia evolucionado (eMBMS) en *Long Term Evolution* (LTE). La estrategia de asignación conjunta *multicast/unicast* (JMUS) obtiene dinámicamente el valor óptimo del esquema de modulación y codificación (MCS) y del número de subtramas reservadas para las transmisiones *multicast*, a la vez que utiliza las subtramas *unicast* para dar servicio a los usuarios que sufren peores condiciones de canal. Hemos implementado un algoritmo de búsqueda rápida mediante la estrategia JMUS, el cual consigue resultados muy próximos al óptimo con una importante reducción en la complejidad computacional. Además, se ha desarrollado el esquema conjunto de subgrupos *multicast* y transmisiones *unicast* (JMSUT) con el objetivo de mejorar las prestaciones alcanzadas con JMUS. El esquema JMSUT propone, para las subtramas *multicast*, la creación de diferentes subgrupos *multicast* y dividir la asignación de los bloques de recursos físicos (PRBs) disponibles entre todos ellos. Estos subgrupos *multicast* se combinan con transmisiones *unicast* en las subtramas reservadas para tal finalidad, y que son las que se encargan de servir a los usuarios con peores condiciones. JMSUT mejora de forma sustancial los resultados, en términos de tasa total de datos (ADR), que se obtienen con el esquema *multicast* convencional (CMS) o con la estrategia JMUS. Además, tanto JMUS como JMSUT permiten que todos los usuarios alcancen la tasa de datos mínima requerida.

Otra de las principales contribuciones que aportamos explora la utilización eficiente del CQI por subbanda con el objetivo de optimizar las transmisiones *multicast* en el contexto general de las actuales redes móviles. Dado que la obtención de la solución óptima al problema de maximización del ADR utilizando la información del CQI por subbanda resulta inviable computacionalmente, hemos propuesto y estudiado diferentes soluciones subóptimas. En particular, se observa que una solución eficaz consiste en una división *greedy* de los usuarios *multicast* en dos subgrupos con el objetivo, en primer lugar, de satisfacer la tasa de datos mínima requerida y, a continuación, maximizar el ADR del servicio *multicast*, todo ello sujeto a las restricciones que implican los requisitos de tasa de error de bloque. El enfoque basado en el CQI por subbanda utilizando dos etapas con dos subgrupos nos proporciona importantes mejoras en el ADR obtenido. Al aprovechar la granularidad que el CQI por subbanda nos brinda se obtienen mejoras significativas, especialmente en aquellos canales que experimentan una alta selectividad frecuencial, ya que el CQI de banda ancha se muestra más impreciso para aprovechar la capacidad potencial que ofrecen los canales de los usuarios. Hemos implementado dos técnicas basadas en el CQI por subbanda, el algoritmo *greedy* con CQI por subbanda (GSCA) combinado con un algoritmo basado en CQI por subbanda para el segundo grupo (SG-SCA) y el algoritmo de ordenación con CQI por subbanda (SSCA) combinado con un SG-SCA, que muestran una mejora significativa sobre propuestas del estado del arte en términos de ADR y probabilidad de entrega del servicio, utilizando unos requisitos de tasa mínima y configuraciones de sistema y canal que se corresponden con los escenarios de interés para los sistemas reales.

Finalmente, en otra de las contribuciones principales de esta tesis, hemos propuesto un nuevo esquema de asignación de recursos para la provisión de servicios *multicast* con transmisión de vídeo, para la cual se emplea la codificación de vídeo escalabe (SVC) en múltiples capas. El esquema propuesto para la asignación de recursos *multicast* basado en múltiples capas (MAMVL) está enfocado principalmente en reducir la complejidad del espacio de posibles soluciones y, a su vez, maximizar el ADR del servicio. De esta manera es posible desarrollar un algoritmo con baja complejidad de fácil implementación en sistemas reales. El esquema MAMVL utilizando 2, 3 y 4 capas de vídeo presenta importantes ventajas en términos de ADR y de equidad entre usuarios, pero especialmente en la eficienca espectral alcanzada debido a una mejor asignación de los PRBs disponibles entre las diferentes capas que aquellos algoritmos que no emplean el conocimiento a priori de la tasa de datos del contenido a transmitir.

Como conclusión, merece resaltarse que todas las contribuciones aportadas en esta tesis se están convirtiendo en parte del estado del arte de las propuestas para RRM en servicios *multicast*, proporcionando mejoras en la eficiencia espectral y mecanismos para el desarrollo de soluciones con baja complejidad computacional que pueden ser implementables en sistemas reales. Respecto al primer aspecto, tanto la división de los recursos disponibles en subgrupos *multicast*, como la utilización de información del CQI por subbanda muestran amplias posibilidades para el desarrollo de algoritmos que proporcionen alta eficiencia espectral. Por otro lado, la utilización de sistemas basados en múltiples capas, con conocimiento a priori de las tasas de datos del contenido a transmitir, permite la implementación de soluciones de baja complejidad, imprescindibles para la aplicación real de dichos esquemas.

Líneas de investigación futura

Las diferentes contribuciones aportadas en esta tesis pretenden servir de estímulo y referencia para la continuación de las investigaciones relativas al prometedor campo de los servicios *broadcast/multicast* de próxima generación, especialmente en RRM, la parte más crítica para su evolución. Los próximos trabajos de investigación deben enfocarse en el diseño de nuevos algoritmos para SRA que tengan en cuanta los siguientes desafíos:

- La implementación de algoritmos para SRA, basándose en una adecuada utilización del CQI por subbanda, que se encarguen de forma conjunta de optimizar la asignación de recursos tanto a los servicios *multicast* como *unicast*. Hay que tener presente que los trabajos incluidos en esta tesis se encargan de la asignación de recursos para la entrega de un único servicio *multicast*. Estas estrategias se pueden ampliar para dar cobertura a la entrega de más de un servicio *multicast* mediante la inclusión de nuevas políticas para manejar la asignación de recursos entre los diferentes servicios *multicast* y/o *unicast*, e.g. equidad entre grupos, ADR, o consumo de recursos. Por lo tanto, un esquema futuro de SRA para la asignación conjunta de recursos a varios servicios *multicast* y *unicast* nos proveerá de una herramienta más completa para el análisis de sistemas reales. Mediante la utilización de funciones de utilidad multi-objetivo, el esquema de asignación de recursos intra-grupo se encargaría de la maximización del ADR, mientras que es necesaria la inclusión de nuevas políticas para manejar los diferentes requisitos de QoS que tenga cada servicio *unicast* o *multicast*.

- Un importante desafío consiste en adaptar las estrategias propuestas de SRA desarrolladas para enlaces celulares a nuevos entornos heterogéneos. Las celdas pequeñas y las comunicaciones directas entre dispositivos (D2D) se consideran un medio interesante para extender la cobertura y mejorar las condiciones del canal de aquellos usuarios situados en el borde de la macro celda. Los enlaces celulares se pueden descargar de tráfico con la ayuda tanto de celdas pequeñas como de enlaces D2D de corto alcance. La técnica *multicast* basada en subgrupos se puede adaptar para encargarse de la asignación de recursos no solo en la macro celda sino también en los enlaces de corto alcance.

- La comunicación entre máquinas (MTC) resulta fundamental para una amplia gama de nuevas aplicaciones y servicios cuyo lanzamiento está planificado con los sistemas 5G. Las características de MTC junto con sus requerimientos específicos, como el tiempo de conexión corto, la transmisión de datos con bajo consumo energético, la recepción discontinua (DRX), la reducción de la señalización, el control del tráfico y la alta fiabilidad a los miles de dispositivos que se prevé estén ubicados en cada celda, hacen que el desarrollo de estrategias de SRA para

MTC se convierta en un desafío importante Además, actualmente se precisa el diseño de nuevas estrategias de SRA que soporten de manera eficiente el tráfico MTC orientado a grupos. Dichas estrategias se han de enfocar en minimizar retardos y permitir escalabilidad cuando el número de receptores sea muy elevado, teniendo en cuenta que el tráfico de datos en MTC se suele caracterizar por paquetes pequeños.

Bibliography

[1] "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021", Cisco, Tech. Rep., 2017, [Online] Available at: www.cisco.com.

[2] "Ericsson mobility report", Ericsson AB, Tech. Rep., 2017, [Online] Available at: www.ericsson.com.

[3] D. Lecompte and F. Gabin, "Evolved multimedia broadcast/multicast service (eMBMS) in LTE-advanced: Overview and rel-11 enhancements", *IEEE Communications Magazine*, vol. 50, no. 11, pp. 68–74, 2012.

[4] G. Araniti, M. Condoluci, P. Scopelliti, A. Molinaro, and A. Iera, "Multicasting over emerging 5G networks: Challenges and perspectives", *IEEE Network*, vol. PP, no. 99, pp. 2–11, 2017.

[5] "LTE - Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 - (3GPP TS 36.300 version 14.3.0 Release 14)", 3GPP, Tech. Rep., 2017.

[6] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?", *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[7] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G", *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.

[8] "5G RAN architecture and functional design (white paper)", METIS II, Tech. Rep., 2016, [Online] Available at: www.metis-ii.5g-ppp.eu.

[9] "View on 5G architecture (white paper)", 5G-PPP Architecture Working Group, Tech. Rep., 2016, [Online] Available at: www.5g-ppp.eu.

[10] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. D. Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice", *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, 2017.

[11] C. Cox, *An introduction to LTE: LTE, LTE-Advanced, SAE and 4G mobile communications*. Wiley, 2012.

[12] E. Dahlman, S. Parkvall, and J. Skold, *4G, LTE-Advanced Pro and the road to 5G*, 3rd. Academic Press, 2016.

[13] J. Lee, Y. Kim, Y. Kwak, J. Zhang, A. Papasakellariou, T. Novlan, C. Sun, and Y. Li, "LTE-Advanced in 3GPP rel-13/14: An evolution toward 5G", *IEEE Communications Magazine*, vol. 54, no. 3, pp. 36–42, 2016.

[14] A. de la Fuente, R. P. Leal, and A. G. Armada, "New technologies and trends for next generation mobile broadcasting services", *IEEE Communications Magazine*, vol. 54, no. 11, pp. 217–223, 2016.

[15] L. Zhang, Y. Wu, G. K. Walker, W. Li, K. Salehian, and A. Florea, "Improving LTE eMBMS with extended OFDM parameters and layered-division-multiplexing", *IEEE Transactions on Broadcasting*, vol. 63, no. 1, pp. 32–47, 2017.

[16] M. Condoluci, G. Araniti, A. Molinaro, and A. Iera, "Multicast resource allocation enhanced by channel state feedbacks for multiple scalable video coding streams in LTE networks", *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 2907–2921, 2016.

[17] A. de la Fuente, R. P. Leal, and A. G. Armada, "Performance analysis of eMBMS in LTE: Dynamic MBSFN areas", in *Annual Technology Conference OPNETWORK*, 2013, pp. 1–6.

[18] ——, “Análisis de prestaciones de eMBMS en LTE: Redes de frecuencia única”, in *XXVIII Simposium Nacional URSI*, 2013, pp. 1–4.

[19] J. Calabuig, J. F. Monserrat, and D. Gómez-Barquero, “5th generation mobile networks: A new opportunity for the convergence of mobile broadband and broadcast services”, *IEEE Communications Magazine*, vol. 53, no. 2, pp. 198–205, 2015.

[20] “LTE - Multimedia Broadcast/Multicast Service (MBMS); Stage 1 - (3GPP TS 22.146 version 13.0.0 Release 13)”, 3GPP, Tech. Rep., 2015.

[21] “LTE - Introduction of the multimedia broadcast/multicast service (MBMS) in the radio access networks (RAN)-Stage 2 - (3GPP TS 25.346 version 14.0.0 Release 14)”, 3GPP, Tech. Rep., 2017.

[22] “LTE - Multimedia Broadcast/Multicast Service (MBMS); Protocols and codecs (3GPP TS 26.346 version 14.3.0 Release 14)”, 3GPP, Tech. Rep., 2017.

[23] “Information technology - dynamic adaptive streaming over HTTP (DASH) - part 1: Media presentation description and segments formats”, ISO-IEC 23009-1, Tech. Rep., 2014.

[24] T Stockhammer, A. Shokrollahi, M. Watson, M. Luby, and T. Gasiba, “Application layer forward error correction for mobile multimedia broadcasting”, *Handbook of Mobile Broadcasting: DVB-H, DMB, ISDB-T and Media Flo, CRC Press*, pp. 239–280, 2008.

[25] A. de la Fuente, C. Lentisco, L. Bellido, R. Perez Leal, E. Pastor, and A. Garcia Armada, “Analysis of the impact of FEC techniques on a multicast video streaming service over LTE”, in *European Conference on Networks and Communications (EuCNC)*, 2015, pp. 219–223.

[26] C. M. Lentisco, L. Bellido, A. de la Fuente, E. Pastor, R. P. Leal, and A. G. Armada, “A model to evaluate MBSFN and AL-FEC techniques in a multicast video streaming service”, in *IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2014, pp. 691–696.

[27] A. de la Fuente, C. M. Lentisco, L. Bellido, R. P. Leal, A. G. Armada, E. Pastor, and A. G. Bolívar, "End to end measurements of multimedia streaming over LTE", in *European Conference on Networks and Communications (EuCNC)*, 2016, pp. 57–61.

[28] "LTE - Evolved Universal Terrestrial Radio Access (E-UTRA); Study on single-cell point-to-multipoint transmission for E-UTRA (3GPP TR 36.890 version 13.0.0 Release 13)", 3GPP, Tech. Rep., 2015.

[29] A. Awada, D. Navrátil, and M. Säily, "A study on single-cell point-to-multipoint transmission for public safety communications with eMBMS LTE networks", in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2016, pp. 1–6.

[30] G. Araniti, M. Condoluci, M. Cotronei, A. Iera, and A. Molinaro, "A solution to the multicast subgroup formation problem in LTE systems", *IEEE Wireless Communications Letters*, vol. 4, no. 2, pp. 149–152, 2015.

[31] C.-H. Ko, C.-C. Chou, H.-Y. Meng, and H.-Y. Wei, "Strategy-proof resource allocation mechanism for multi-flow wireless multicast", *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3143–3156, 2015.

[32] "LTE - Study on RAN Improvements for Machine-type Communications - (3GPP TR 37.868 version 11.0.0 Release 11)", 3GPP, Tech. Rep., 2011.

[33] M. Condoluci, G. Araniti, T. Mahmoodi, and M. Dohler, "Enabling the IoT machine age with 5G: Machine-type multicast services for innovative real-time applications", *IEEE Access*, vol. 4, pp. 5555–5569, 2016.

[34] L. D. Xu, W. He, and S. Li, "Internet of things in industries: A survey", *IEEE Transactions on Industrial Informatics*, vol. 10, no. 4, pp. 2233–2243, 2014.

[35] "Narrowband IoT (NB-IoT); RP-151621 (3GPP TSG RAN meeting #69 - Qualcomm, Incorporated)", 3GPP, Tech. Rep., 2015.

[36] C. H. Wei, R. G. Cheng, and S. L. Tsao, "Performance analysis of group paging for machine-type communications in LTE networks", *IEEE Transactions on Vehicular Technology*, vol. 62, no. 7, pp. 3371–3382, 2013.

[37] G. Wunder, "5GNOW: Non-orthogonal, asynchronous waveforms for future mobile applications", *IEEE Communications Magazine*, vol. 52, no. 2, pp. 97–105, 2014.

[38] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results", *IEEE Communications Magazine*, vol. 49, no. 2, pp. 102–111, 2011.

[39] E. Dahlman, G. Mildh, S. Parkvall, J. Peisa, J. Sachs, Y. Selén, and J. Sköld, "5G wireless access: Requirements and realization", *IEEE Communications Magazine*, vol. 52, no. 12, pp. 42–47, 2014.

[40] "5G systems (white paper)", Ericsson AB, Tech. Rep., 2015, [Online] Available at: www.ericsson.com.

[41] J. Abdoli, M. Jia, and J. Ma, "Filtered OFDM: A new waveform for future wireless systems", in *IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2015, pp. 66–70.

[42] A. Kliks, O. Holland, A. Basaure, and M. Matinmikko, "Spectrum and license flexibility for 5G networks", *IEEE Communications Magazine*, vol. 53, no. 7, pp. 42–49, 2015.

[43] S. Ping, A. Aijaz, O. Holland, and A. Aghvami, "SACRP: A spectrum aggregation-based cooperative routing protocol for cognitive radio ad-hoc networks", *IEEE Transactions on Communications*, vol. 63, no. 6, pp. 2015–2030, 2015.

[44] S. Yrjölä et al., "Strategic choices for mobile network operators in future flexible UHF spectrum concepts?", in *Cognitive Radio Oriented Wireless Networks*, Springer, 2015, pp. 573–584.

[45] O. Holland, A. Aijaz, S. Ping, S. Wong, J. Mack, L. Lam, and A. de la Fuente, "Aggregation in TV white space and assessment of an aggregation-capable IEEE 802.11 white space device", in *IEEE International Conference on Communications (ICC)*, 2016, pp. 1–5.

[46] "LTE broadcast: Evolving and going beyond mobile", Qualcomm Technologies, Inc, Tech. Rep., 2014, [Online] Available at: www.qualcomm.com.

[47] M. Iwamura, "NGMN view on 5G architecture", in *IEEE 81st Vehicular Technology Conference (VTC Spring)*, 2015, pp. 1–5.

[48] Y. Cui, D. Jiang, and Y. Wu, "Analysis and optimization of caching and multicasting in large-scale cache-enabled wireless networks", *IEEE Transactions on Wireless Communications*, vol. PP, no. 99, pp. 1–1, 2016.

[49] B. Zhou, H. Hu, S. Q. Huang, and H. H. Chen, "Intracluster device-to-device relay algorithm with optimal resource utilization", *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 2315–2326, 2013.

[50] J Alonso and M Dohler, "Machine-to-machine technologies & markets-shift of industries", *IEEE WCNC Tutorial*, vol. 6, 2014.

[51] K. Zheng, F. Hu, W. Wang, W. Xiang, and M. Dohler, "Radio resource allocation in LTE-advanced cellular networks with M2M communications", *IEEE Communications Magazine*, vol. 50, no. 7, pp. 184–192, 2012.

[52] "LTE - Study on Enhancements to Machine-Type Communications (MTC) and other Mobile Data Applications; Radio Access Network (RAN) aspects (3GPP TR 37.869 version 12.0.0 Release 12)", 3GPP, Tech. Rep., 2013.

[53] F. Ghavimi and H. H. Chen, "M2M communications in 3GPP LTE/LTE-A networks: Architectures, service requirements, challenges, and applications", *IEEE Communications Surveys Tutorials*, vol. 17, no. 2, pp. 525–549, 2015.

[54] O. Niamut, G. Thomas, E Thomas, R van Brandenburg, L D'Acunto, and R Gregory-Clarke, "Live event experiences-interactive UHDTV on mobile devices", in *IET IBC2014 Conference*, 2014, p. 11.1.

[55] "5G use cases and requirements (white paper)", Nokia Networks, Tech. Rep., 2014.

[56] Y. G. Kim and W. J. Kim, "Implementation of augmented reality system for smartphone advertisements", *International Journal of Multimedia and Ubiquitous Engineering*, vol. 9, no. 2, pp. 385–392, 2014.

[57] "Scenarios, requirements and KPIs for 5G mobile and wireless system (METIS deliverable D1.5)", METIS, Tech. Rep., 2015, [Online] Available at: www.metis2020.com.

[58] R. O. Afolabi, A. Dadlani, and K. Kim, "Multicast scheduling and resource allocation algorithms for OFDMA-based systems: A survey", *IEEE Communications Surveys Tutorials*, vol. 15, no. 1, pp. 240–254, 2013.

[59] J. Xu, S. J. Lee, W. S. Kang, and J. S. Seo, "Adaptive resource allocation for MIMO-OFDM based wireless multicast systems", *IEEE Transactions on Broadcasting*, vol. 56, no. 1, pp. 98–102, 2010.

[60] T. P. Low and M. O. Pun and Y. W. P. Hong and C. C. J. Kuo, "Optimized opportunistic multicast scheduling (OMS) over wireless cellular networks", *IEEE Transactions on Wireless Communications*, vol. 9, no. 2, pp. 791–801, 2010.

[61] P. K. Gopala and H. El Gamal, "On the throughput-delay tradeoff in cellular multicast", in *International Conference on Wireless Networks, Communications and Mobile Computing (WIRELESSCOM)*, vol. 2, 2005, pp. 1401–1406.

[62] C. Suh and J. Mo, "Resource allocation for multicast services in multicarrier wireless communications", *IEEE Transactions on Wireless Communications*, vol. 7, no. 1, pp. 27–31, 2008.

[63] G. Araniti, M. Condoluci, L. Militano, and A. Iera, “Adaptive resource allocation to multicast services in LTE systems”, *IEEE Transactions on Broadcasting*, vol. 59, no. 4, pp. 658–664, 2013.

[64] P. Eusiebio and A. Correia, “Two QoS regions packet scheduling for MBMS”, in *2nd International Symposium on Wireless Communication Systems (ISWCS)*, 2005, pp. 777–781.

[65] G. Araniti, V. Scordamaglia, M. Condoluci, A. Molinaro, and A. Iera, “Efficient frequency domain packet scheduler for point-to-multipoint transmissions in LTE networks”, in *IEEE International Conference on Communications (ICC)*, 2012, pp. 4405–4409.

[66] A. Orsino, P. Scopelliti, and M. Condoluci, “A multi-criteria approach for multicast resource allocation over LTE and beyond cellular systems”, in *22th European Wireless Conference (EW)*, 2016, pp. 1–6.

[67] M. Condoluci, L. Militano, G. Araniti, A. Molinaro, and A. Iera, “Multicasting in LTE-A networks enhanced by device-to-device communications”, in *IEEE Globecom Workshops (GC Wkshps)*, 2013, pp. 567–572.

[68] S. Pizzi, M. Condoluci, G. Araniti, A. Molinaro, A. Iera, and G. M. Muntean, “A unified approach for efficient delivery of unicast and multicast wireless video services”, *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 8063–8076, 2016.

[69] L. Christodoulou, O. Abdul-Hameed, A. M. Kondoz, and J. Calic, “Adaptive subframe allocation for next generation multimedia delivery over hybrid LTE unicast broadcast”, *IEEE Transactions on Broadcasting*, vol. 62, no. 3, pp. 540–551, 2016.

[70] J. Chen, M. Chiang, J. Erman, G. Li, K. K. Ramakrishnan, and R. K. Sinha, “Fair and optimal resource allocation for LTE multicast (eMBMS): Group partitioning and dynamics”, in *IEEE Conference on Computer Communications (INFOCOM)*, 2015, pp. 1266–1274.

[71] A. Tassi, I. Chatzigeorgiou, and D. Vukobratović, “Resource-allocation frameworks for network-coded layered multimedia multicast services”, *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 2, pp. 141–155, 2015.

[72] P. Chau, Y. Lee, T. D. Bui, J. Shin, and J. P. Jeong, “An efficient resource allocation scheme for scalable video multicast in LTE-advanced networks”, in *11th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, 2017, p. 76.

[73] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, “An overview of MIMO communications-a key to gigabit wireless”, *Proceedings of the IEEE*, vol. 92, no. 2, pp. 198–218, 2004.

[74] “LTE - Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation (3GPP TS 36.211 v14.3.0 Release 14)”, 3GPP, Tech. Rep., 2017.

[75] A. G. Fabregas and G. Caire, “Coded modulation in the block-fading channel: Coding theorems and code construction”, *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 91–114, 2006.

[76] A. Serra-Pagès, “Link level performance evaluation and link abstraction for LTE/ LTE-Advanced downlink”, PhD Thesis, Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, 2015.

[77] J. Colom-Ikuno, “System level modeling and optimization of the LTE downlink”, PhD Thesis, Institute of Telecommunications - Technische Universität Wien, 2012.

[78] K. Brueninghaus, D. Astely, T. Salzer, S. Visuri, A. Alexiou, S. Karger, and G. A. Seraji, “Link performance models for system level simulations of broadband radio access systems”, in *IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, vol. 4, 2005, pp. 2306–2311.

[79] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation", *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 927–946, 1998.

[80] A. de la Fuente, A. G. Armada, and R. P. Leal, "Joint multicast/unicast scheduling with dynamic optimization for LTE multicast service", in *20th European Wireless Conference (EW)*, 2014, pp. 1–6.

[81] G. Monghal, K. Pedersen, I. Kovacs, and P. Mogensen, "QoS oriented time and frequency domain packet schedulers for the UTRAN long term evolution", in *IEEE Vehicular Technology Conference, (VTC Spring)*, 2008, pp. 2532–2536.

[82] A. de la Fuente, R. P. Leal, and A. G. Armada, "Joint strategy for LTE resource allocation: Multicast subgrouping & unicast transmissions", in *XII Jornadas de Ingeniería Telemática (JITEL)*, 2015, pp. 9–15.

[83] E. Castañeda, R. Samano-Robles, and A. Gameiro, "Sum rate maximization via joint scheduling and link adaption for interference-coupled wireless systems", *EURA-SIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, pp. 1–17, 2013.

[84] J. Liu, W. Chen, Z. Cao, and K. B. Letaief, "Dynamic power and sub-carrier allocation for OFDMA-based wireless multicast systems", in *IEEE International Conference on Communications (ICC)*, 2008, pp. 2607–2611.

[85] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, 2007.

[86] P. McDonagh, C. Vallati, A. Pande, P. Mohapatra, P. Perry, and E. Mingozzi, "Investigation of scalable video delivery using H.264 SVC on an LTE network", in *14th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, 2011, pp. 1–5.

[87] "LTE - Improved Video Coding Support - (3GPP TR 26.904 version 13.0.0 Release 13)", 3GPP, Tech. Rep., 2015.

[88] W. Feller, *An introduction to probability theory and its applications: Volume I.* John Wiley & Sons, 1968.

[89] "LTE - Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) system scenarios - (3GPP TR 36.942 version 14.0.0 Release 14)", 3GPP, Tech. Rep., 2017.

[90] "LTE - Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception - (3GPP TS 36.104 version 14.4.0 Release 14)", 3GPP, Tech. Rep., 2017.

[91] R. Jain, *The art of computer systems performance analysis: Techniques for experimental design, measurement, simulation, and modeling.* John Wiley & Sons, 1990.

[92] S. Sharangi, R. Krishnamurti, and M. Hefeeda, "Energy-efficient multicasting of scalable video streams over WiMAX networks", *IEEE Transactions on Multimedia*, vol. 13, no. 1, pp. 102–115, 2011.