Uplink Grant-free Access for Ultra-Reliable Low-Latency Communications in 5G

Radio Access and Resource Management Solutions

Abreu, Renato Barbosa

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

UPLINK GRANT-FREE ACCESS FOR ULTRA-RELIABLE LOW-LATENCY COMMUNICATIONS IN 5G

RADIO ACCESS AND RESOURCE MANAGEMENT SOLUTIONS

BY
RENATO BARBOSA ABREU

DISSERTATION SUBMITTED 2019

AALBORG UNIVERSITY
DENMARK

Uplink Grant-free Access for Ultra-Reliable Low-Latency Communications in 5G

Radio Access and Resource Management Solutions

Ph.D. Dissertation

Renato Barbosa Abreu

Aalborg University
Department of Electronic Systems
Fredrik Bajers Vej 7
DK - 9220 Aalborg

Dissertation submitted: June 2019

PhD supervisor: Prof. Preben Mogensen
Aalborg University

Assistant PhD supervisors: Assoc. Prof. Gilberto Berardinelli
Aalborg University

Prof. Klaus Pedersen
Aalborg University

PhD committee: Professor Hans Peter Schwefel (chairman)
Aalborg University

Master Researcher Imadur Rahman
Ericsson

Dr. Ing., habil, Associate Professor Gerhard Wunder
Freie Universität Berlin

PhD Series: Technical Faculty of IT and Design, Aalborg University

Department: Department of Electronic Systems

ISSN (online): 2446-1628
ISBN (online): 978-87-7210-454-6

Published by:
Aalborg University Press
Langagervej 2
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2019

Curriculum Vitae

Renato Barbosa Abreu

Renato Barbosa Abreu obtained his degree in Electrical Engineering from Federal University of Viçosa, Brazil, in 2007. He obtained his master degree also in Electrical Engineering from Federal University of Amazonas, Brazil, in 2014. From 2008 to 2016, he worked as an engineer and researcher in the former Nokia Technology Institute INDT, Brazil, where he obtained experience in systems development and mobile technologies. In April 2016, Renato started his PhD studies in the Wireless Communication Networks (WCN) section at the Department of Electronic Systems at Aalborg University, Denmark. The research project was developed in collaboration with Nokia Bell Labs. His research focus is on ultra-reliable low latency communications in 5G networks.

Abstract

In the last years, cellular networks have evolved to provide ubiquitous broadband services with high data rates. Fifth generation (5G) systems, which should be introduced to the market in early 2020s, are also expected to support new types of services including mission-critical applications. To cover such applications, Ultra-Reliable Low-Latency Communications (URLLC) is defined in 5G, setting stringent requirements for transmitting a packet over the radio interface, such as 99.999% success probability within 1 ms.

Enabling URLLC has specific challenges in the uplink. The typical access procedure relies on scheduling request and *grant* for every packet transmission over dynamically allocated radio resources. This process causes excessive delays and demands reliable control signaling, which leads to an overhead in the communication link. To overcome these issues, *grant-free* solutions, in which the resources are pre-configured to each user, and eventually shared by a pool of users, come into place. However, grant-free access brings challenges in terms of resource utilization or increased interference levels. To deal with that, new techniques need to be developed and investigated.

This research focuses on the radio interface enhancements to efficiently support URLLC in the uplink. The first part of the thesis addresses the channel access solutions, with particular emphasis on the transmission and retransmission procedures. Schemes that rely on the preallocation of resources for transmissions as well as for retransmissions are initially studied. Their benefits are shown in terms of resource efficiency compared with conservative single-shot transmissions. Then, analyses are carried with focus on sporadic URLLC transmissions over grant-free resources shared by multiple users. Relevant retransmission and repetition schemes proposed for grant-free access are studied using detailed system level simulations. The conditions for efficiently employing the proposed approaches are identified.

In the second part of the thesis, radio resource management strategies for grant-free URLLC are investigated with the objective of improving the capacity for these services in the system. It is demonstrated that, by redefining the power control strategy with respect to traditional broadband settings and optimizing the parameters considering the URLLC requirements, the achievable

load can be greatly improved. Additionally, it is presented a resource allocation method. It encloses the configuration of multiple sub-bands with corresponding transmission parameters and an association scheme for URLLC users. The method reduces the probability of fully overlapping transmissions, and improves the URLLC performance. The influence of multi-antenna receivers is also considered. The spatial diversity and interference rejection capability show to be determinant for the performance of grant-free URLLC transmissions in shared resources. Lastly, multi-cell reception solutions are proposed to harvest combining gain and interference diversity by collecting soft information from assisting cells. High URLLC performance gains can be achieved with the cost of increased backhaul load.

The efficient support of heterogeneous services is also aimed in 5G systems. This motivates the third part of this thesis, which studies the problem of multiplexing grant-free URLLC and enhanced Mobile Broadband (eMBB) traffic. First, the impact of transmit power control settings on the performance of eMBB and URLLC using overlaying allocations is evaluated through system level simulations. Further insights are given on the configuration of the open loop power control for managing the performance of both services. Then, an analytical study of the supported load for each service is provided, comparing overlaying and separate bands allocation. The potential of overlaying allocation is revealed specially when employing advanced receivers with interference cancellation. On the other hand, separate bands show better performance, for instance, at low signal-to-noise ratio regimes or large payload size. Recommendations for 5G radio networks implementation are provided based on the presented results.

Resumé

I de seneste år har cellulære netværk udviklet sig til at supportere mobilt bredbånd med meget høje data rater og i et stort dækningsområde. Systemer at femte generation (5G), som forudsiges at komme på markedet i starten af 2020, forventes at supportere nye services såsom missions kritiske applikationer. For at understøtte disse applikationer, har 5G defineret en service kaldet Ultra-Reliable Low-Latency Communications (URLLC), som sætter yderst strikse krav til pakke transmissionen over radio interfacet såsom 99.999% sandsynlighed for maksimalt 1 ms forsinkelse.

Det er specielt vanskeligt at opnå understøttelsen af URLLC i uplink. Den typiske pakketransmissionsprocedure beror på en skeduleringsforespørgsel og skeduleringsgodkendelse til transmissionen af små pakker over dynamisk allokerede radioressourcer. Denne proces forårsager kritisk forsinkelse og kræver en pålidelig signalering af kontrolinformation, hvilket resultere i et overhead i radio forbindelsen. Grant-free løsninger, hvor brugerne er prækonfigureret med radio ressourcer og kan være delt mellem brugerne, kan benyttes til at undgå de førnævnte ulemper med de typiske pakketransmissionsprocedure. Dog er grant-free løsninger udfordret på enten radio ressource effektivitet eller forøget interferensniveauer. Derfor kalder grant-free løsninger på udvikling af nye undersøgelser og teknikker.

Forskningen præsenteret I denne afhandling, fokusere på forbedringer i radio interfacet med henblik på effektiv support af URLLC i uplink. Den første del af afhandlingen adressere radiokanal adgangsløsninger med særligt fokus på transmissions og retransmissions procedurer. Først, studeres protokoller der beror på præallokering af radio ressourcer til transmissioner og retransmissioner. Deres fordele er givet ved radio effektivitets forbedringer sammenlignet med konservative enkeltforsøgstransmissioner. Derefter præsenteres analyser af sporadisk URLLC transmissioner over grant-free radio ressourcer delt af flere brugere. Relevante protokoller med retransmissioner og repetitioner til grant-free transmissioner er studeret ved brug af detaljerede systemsimuleringer. Derudover identificeres betingelserne for effektiv udrulning af de foreslåede teknikker.

I den anden del af delle afhandling, undersøges radio ressource man-

agement teknikker til grant-free URLLC, med formålet at forøge den understøttede servicekapacitet. Det er demonstreret at med en omdefinering af power control strategien sammenlignet med traditionel bredbånd og en optimering af parametreud fra URLLC servicekravene, kan den opnåede serviceload forøges dramatisk. Derudover præsenteres en radio allokeringsteknik. Denne indbefatter en konfigurering af flere sub-bands med dertilhørende transmissionparaketre og en udvælgelsesalgorithme for URLLC brugerne. Denne teknik reducere sandsynligheden for fuldt overlappende transmissioner og forbedre URLLC performance. Påvirkningen af multiantenne radio modtagere er også inkluderet. Spatial diversitet og interference rejection viser sig at være afgørende for performance af grant-free URLLC transmissioner over delte radio resourcer. Til sidst foreslås anvendelsen af multi-cell reception til at opnå interference diversitet og forøget modtaget energi ved at modtage fra flere assisterende radio celler. En høj URLLC performance forøgelse kan opnås på bekostning af en forøget backhaul trafikbelastning.

Effektiv support af forskelligartede service er også en del af målet med 5G systemer. Dette motivere den tredje del af dette arbejde, hvori problemet med at multiplekse grant-free URLLC og enhanced Mobile Broadband (eMBB) trafik studeres. Først undersøges, med brug af systemsimuleringer, påvirkningen af transmission power control indstillinger på performance af eMBB or URLLC når overlappende radioresourceallokeringer anvendes. Yderligere indsigt gives ved brugen af fractional power control til håndtering af begge services performance. Derefter præsenteres et studie af den maksimalt understøttede traffikbelastning af for begge services, der anvendes til at sammenligne radioallokeringer på separate bånd eller i overlappende bånd. Potentialet af overlappende radioallokeringer er i særdeleshed størst når avancerede radiomodtagere med understøttelse af interference cancellation anvendes. På den anden side, radioallokeringer på separate bånd giver en bedre performance for eksempel ved forhold med lave signal-to-noise forhold eller ved transmissionen af store URLLC datapakker. Anbefalinger for 5G radio netværk implementeringer gives baseret på de præsenterede resultater.

Contents

List of Abbreviations

1G first generation

2G second generation

3D three dimensional

3G third generation

3GPP third generation partnership project

4G fourth generation

5G fifth generation

ACK positive acknowledgement

AMC adaptive modulation and coding

ARQ automatic repeat request

BLEP block error probability

BLER block error rate

BS base station

CCDF complementary cumulative distribution function

CCH control channel

CDF cumulative distribution function

CG configured-grant

CoMP coordinated multipoint

CQI channel quality indicator

CSI channel state information

DCI downlink control information

DL downlink

DMRS demodulation reference signal

eMBB enhanced Mobile Broadband

eMTC enhanced machine type communication

FDD frequency division duplex

GB grant-based

GF grant-free

gNB fifth generation NodeB

HARQ hybrid automatic repeat request

ICT Information Communication Technology

IIoT Industrial IoT

IMT International Mobile Telecommunications

IMT-2020 International Mobile Telecommunications for 2020 and beyond

IoT Internet of Things

IRC interference rejection combining

ITU International Telecommunication Union

KPI key performance indicator

LA link adaptation

LTE Long Term Evolution

LTE-A LTE-Advanced

MAC medium access control

MBB mobile broadband

MCC mission-critical communication

MCS modulation and coding scheme

MIMO multiple-input multiple-output

MMIB mean mutual information per coded bit

MMSE minimum mean square error

mMTC massive Machine Type Communication

MRC maximal-ratio combining

MT mobile terminal

MTC machine-type communication

MU multi user

MU-MIMO multi-user MIMO

NACK negative acknowledgement

NB-IoT narrowband IoT

NLOS non-line-of-sight

NOMA non-orthogonal multiple access

NR New Radio

OFDM orthogonal frequency-division multiplexing

OFDMA orthogonal frequency-division multiple access

OLLA outer-loop link adaptation

PDCCH physical downlink control channel

PDSCH physical downlink shared channel

PHY physical layer

PRB physical resource block

PSD power spectral density

PUCCH physical uplink control channel

PUSCH physical uplink shared channel

QoS quality of service

RACH random access channel

RAN radio access network

RB resource block

RE resource element

RLC radio link control

RRC radio resource control

RRM radio resource management

RS reference signal

RSRP reference signal received power

RTT round-trip time

SA Service and System Aspects

SC-FDMA single-carrier frequency division multiple access

SCS sub-carrier spacing

SDMA space-division multiple access

SIC successive interference cancellation

SINR signal to interference-and-noise ratio

SMS Short Message Service

SNR signal to noise ratio

SPS semi-persistent scheduling

SR scheduling request

SRS sounding reference signal

SU single user

SU-MIMO single-user MIMO

TB transport block

TCP transmission control protocol

TDD time division duplex

TSN time sensitive networking

TTI transmission time interval

UE user equipment

UL uplink

URLLC Ultra-Reliable Low-Latency Communications

Thesis Details

Thesis Title:	Uplink Grant-free Access for Ultra-Reliable Low-Latency Communications in 5G - Radio Access and Resource Management Solutions.
PhD Student:	Renato Barbosa Abreu.
Supervisors:	Prof. Preben Mogensen. Aalborg University. Prof. Gilberto Berardinelli. Aalborg University.

This PhD thesis is the result of three years of research at the Wireless Communication Networks (WCN) section (Department of Electronic Systems, Aalborg University, Denmark) and in collaboration with Nokia Bell Labs. In addition to the work presented herein, dissemination activities, external collaboration and mandatory courses were fulfilled, as part of the requirements for obtaining the PhD degree.

The main body of the thesis consists of the following articles:

Paper A: R. Abreu, P. Mogensen and K. I. Pedersen, "Pre-Scheduled Resources for Retransmissions in Ultra-Reliable and Low Latency Communications," *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, San Francisco, CA, 2017.

Paper B: R. Abreu, G. Berardinelli, T. Jacobsen, K. Pedersen and P. Mogensen, "A Blind Retransmission Scheme for Ultra-Reliable and Low Latency Communications," *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, Porto, 2018.

Paper C: T. Jacobsen, R. Abreu, G. Berardinelli, K. Pedersen, P. Mogensen, I. Z. Kovács and T. Kozlova, "System Level Analysis of Uplink Grant-Free Transmission for URLLC," 2017 IEEE Globecom Workshops (GC Wkshps), Singapore, 2017.

Paper D: R. Abreu, T. Jacobsen, G. Berardinelli, K. Pedersen, I. Z. Kovács and P. Mogensen, "Power control optimization for uplink grant-

free URLLC," 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, 2018.

Paper E: R. Abreu, T. Jacobsen, G. Berardinelli, K. Pedersen, I. Z. Kovács and P. Mogensen, "Efficient Resource Configuration for Grant-Free Ultra-Reliable Low Latency Communications," Submitted for peer-review in *IEEE Transactions on Vehicular Technology*, 2019.

Paper F: T. Jacobsen, R. Abreu, G. Berardinelli, K. Pedersen, I. Z. Kovács and P. Mogensen, "Multi-cell Reception for Uplink Grant-Free Ultra-Reliable Low-Latency Communications," *IEEE Access*, 2019.

Paper G: R. Abreu, T. Jacobsen, K. Pedersen, G. Berardinelli and P. Mogensen, "System Level Analysis of eMBB and Grant-Free URLLC Multiplexing in Uplink," *2019 IEEE 89th Vehicular Technology Conference (VTC Spring)*, Kuala Lumpur, 2019.

Paper H: R. Abreu, T. Jacobsen, G. Berardinelli, K. Pedersen, N. H. Mahmood, I. Z. Kovács and P. Mogensen, "On the Multiplexing of Broadband Traffic and Grant-Free Ultra-Reliable Communication in Uplink," *2019 IEEE 89th Vehicular Technology Conference (VTC Spring)*, Kuala Lumpur, 2019.

This thesis has been submitted for assessment in partial fulfillment of the PhD degree. The thesis is based on the submitted or published papers that are listed above. Parts of the papers are used directly or indirectly in the extended summary of the thesis. As part of the assessment, co-author statements have been made available to the assessment committee and also available at the Faculty.

Preface

This dissertation is the result of three years of research in a PhD project conducted at the Wireless Communication Networks section, Department of Electronic Systems, Aalborg University, Denmark. The research was carried in collaboration and with the support of Nokia Bell Labs, which has partly sponsored the project. Part of the research was also funded by the EU H2020-ICT-206-2 project ONE5G.

I want to start with a special thank to my supervisors Preben Mogensen, Gilberto Berardinelli, and Klaus Pedersen, who advised me all along these years. I am really glad and honored that you gave me the opportunity to learn from you. Many thanks also to Thomas Jacobsen, who joined me during this project. All the great discussions and your fundamental help, whenever we had to face the simulator beast, were determinant for the development of this research. And thanks to István Kovács and Nurul Mahmood for the valuable inputs to this work and for carefully reviewing our papers.

I would also like to express my gratitude to my colleagues from Aalborg University and from Nokia Bell Labs. Beginning with Mads Lauridsen and Ignacio Rodriguez, for inspiring me with the idea of doing a PhD in AAU, back in the time when you were visiting INDT in Brazil. To Dorthe Sparre and Linda Villadsen for the help with the administrative tasks during my period in AAU. To Zexian Li, Kimmo Valkealahti, Niko Kolehmainen, Martti Moisio and Panu Lahdekorpi for the support during the time we stayed in Nokia Espoo and after that as well. To all the colleagues that I met in the office, Ali E., Ali K., Beatriz, Benny, Claudio, Daniela, Dereje, Enric, Erika, Fernando, Frank, Guillermo, Huan, Jens, Jeroen, Lucas, Mads B., Marco, Melisa, Per Henrik, Rafhael, Rasmus, Roberto, Santiago, Tomasz, Troels, many thanks for the technical discussions, brainstorming, new ideas, paper reviews, help with the simulator, feedback and all the advises that you gave me. I am also grateful for all the unforgettable moments including the lunches together, cakes, chats in the sofa, danish lessons, Christmas lunches, summer events, sweets in the kitchen, Friday breakfasts, running events, biking trips, Elektronik Cups, among others (glad for this long list).

My greatest thanks is to my lovely wife, Flávia, for being on my side giv-

ing all the support and encouragement during the ups and downs moments along these years.

Finally, I dedicated this work specially to my parents, Zilda and Ronaldo, and to all my relatives and friends that even from far distances, have helped me with their best wishes.

All to the glory of God, in the name of Jesus Christ.

Renato Barbosa Abreu
Aalborg University, June 2019

Part I

Introduction

Background and Thesis Overview

The problem of reliably transmitting information over a radio interface remounts from the beginning of wireless communication. Back in the end of the 19th century, Roberto Landell de Moura, a Brazilian priest living in São Paulo, dreamed beyond the wireless telegraphy cumbersome methods for transmitting human to human information. Before continuing this story, the principles of radiotelegraphy are briefly recalled.

The radiotelegraph, invention attributed to Guglielmo Marconi in 1895, initiated the era of long distance communication without cables. In radiotelegraphy, a user sending a message should firstly encode each character, typically using Morse Code, i.e. a sequence of "dots" and "dashes". Using an on/off switch, the sequence is then converted to short and long duration electrical pulses which are transmitted via radio waves. The signal acquired by the receiver is reproduced as beeps, which should be interpreted by another user who decodes the message [1].

In this and all types of radio communication, the chance of successfully receiving the information is affected by equipment malfunction, radio propagation losses, interference and thermal noise. Besides, it is clear that both, the reliability and latency of the telegraphic communication depend directly on the "skills" of the two involved users. An untrained person would neither be able to timely type an urgent message, nor to make sense of a received beep sequence without loosing information.

Landell's objective was to directly transmit and reproduce the natural human voice over long distances without wires. Even without financial support and facing opposition due to his religious duties, he still persisted on the development of numerous experimental devices. The newspapers of the time mentioned his achievements on transmitting his words via "electrified air" through distances above 7 kilometers [2]. His firsts public demonstrations date from mid of 1899, what would be the earliest voice transmission over radio. Unfortunately, the inventor priest was far from the main axis of coun-

tries leading the technology revolution at that time. And the lack of consistent documentation and delayed registration of his inventions in North America, caused a low recognition of his work [3]. This, of course, could not prevent the advent of the radiotelephone and radio broadcast in the following years. With the development of the thermionic valve, radio equipment became viable for large scale production [4]. Voice signals could then be transmitted conveniently and reliably for human communication through long distances over the air. Such technologies have certainly brought great impacts to the society.

More than a hundred years have passed and radio communication has permeated the whole world. In the last decades, mobile radio communication became predominant and evolved from voice centric to data centric systems. The evolution of the mobile networks has been occurring in cycles of 10 years approximately. The first generation (1G) was an analog system that started to popularize the idea of mobile communication in the early eighties. In the beginning of the 90's emerged the second generation (2G) system which was fully digital, being more efficient and safer, and bringing new features like Short Message Service (SMS) and access to media content. The clear focus on broadband data services came with the third generation (3G) system, which had its first deployments in 2001. The peak data rate requirement for a 3G compliant system is set to 200 kbps. However, new data-enabled services such as social networks and media streaming have driven a rapid growth of mobile data traffic, scaled by the increasing number of smart devices. The fourth generation (4G) system was developed with clear targets for coping with the growing traffic demand in the current decade. By using packet switching principles for data and voice, among several other technical enhancements, 4G Long Term Evolution (LTE) in its latest versions supports peak data rates of 1 Gbps in the downlink and 0.5 Gbps in the uplink [5].

Looking toward the emergence of new demands from market and society for 2020 and beyond, the International Telecommunication Union (ITU), an United Nations agency which coordinates the shared use of the radio spectrum globally, defined new requirements for the next generation of mobile communication [6]. One of the main differences from the previous generations is the inclusion of support for mission critical applications. The native support for machine to machine communications is envisioned for enabling new applications with strict real-time constraints. Therefore, the shift from the human-centric to the machine-centric communication paradigm, calls for unprecedented levels of latency and reliability of the communication link.

1 Introduction to 5G

Though the previous generations had primary focus on increasing data rates, the fifth generation (5G) radio system should be different. A myriad of services are expected to be supported in 5G. The ITU recommendations report [7], which provides the vision for the next generation system defined as International Mobile Telecommunications for 2020 and beyond (IMT-2020), describes three main scenarios, which are illustrated in Figure I.1. These are summarized as:

- **enhanced Mobile Broadband (eMBB)**: As an extension of the current broadband services, this usage scenario regards to the evolved human-centric uses cases like multi-media streaming and high speed Internet services. An improved efficiency from the network is needed in order to provide the increasing demanded data rates, capacity and mobility. The main requirements for eMBB are peak data rates of 20 Gbps in downlink and 10 Gbps in uplink, respectively. The user experienced data rate should be 100 Mbps in downlink and 50 Mbps uplink.

- **massive Machine Type Communication (mMTC)**: Can also be seen as an extension of Internet of Things (IoT) technologies like enhanced machine type communication (eMTC) and narrowband IoT (NB-IoT) [8]. It addresses the use cases of large amount of IoT devices, e.g. in smart city applications, usually transmitting small data volumes sporadically. The network should comply with the extreme low cost and low power consumption required for these devices. The supported connection density should be at least 1 million devices per km^2. And the battery life should be above 10 years.

- **Ultra-Reliable Low-Latency Communications (URLLC)**: Focus on new use cases related to mission critical applications. Such domain has very stringent requirements of low end-to-end latency and high degree of reliability, to be applied for instance on, wireless industry automation, remote tactile control and teleprotection. The defined reliability requirement for URLLC radio interface is $1 - 10^{-5}$ success probability for transmitting a layer 2 packet of 32 bytes within 1 ms latency [6].

The third generation partnership project (3GPP), the organization developing the specifications for 5G defining the New Radio (NR) air interface, has envisioned a single and flexible technical framework addressing all these usage scenarios [9]. The established service requirements demand a significant improvement on key performance indicators like spectral efficiency, capacity, control and user plane latency, and reliability compared to current cellular

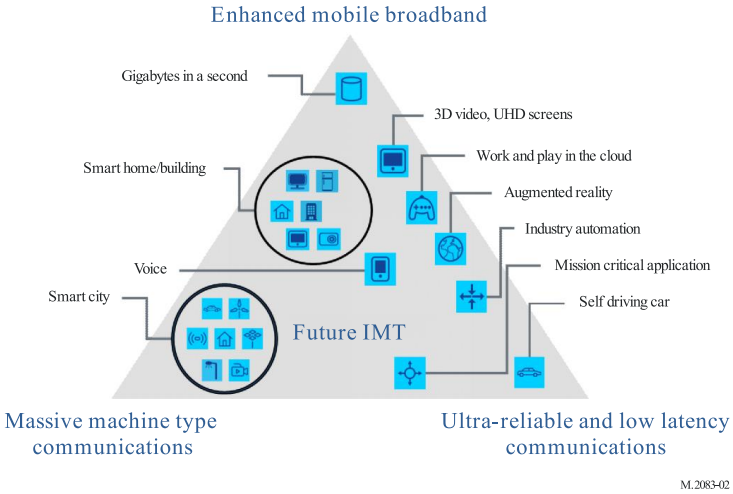


Fig. I.1: ITU envisaged usage scenarios for 2020 and beyond (from IMT-R report [7]).

systems as LTE. Therefore, new technical solutions have been developed for achieving the targets of IMT-2020.

A few technology components have a key role on shaping the 5G design as listed below:

- New spectrum and high bandwidth [10, 11]: compared to 4G, which is limited to operation in up to 3.5 GHz licensed spectrum and 5 GHz unlicensed, 5G exploits additional frequency ranges in the millimeter-wave spectrum up to 52.6 GHz. Transmission bandwidth of up to 400 MHz is supported, enabling very high capacity and peak data rates.

- Beamforming and massive-MIMO [12, 13]: beamforming allows focusing the radiated energy in a certain direction, improving coverage. And employing a large number of antennas allows to serve many users simultaneously, boosting the throughput and spectral efficiency.

- Scalable numerology and flexible frame structure [14, 15]: sub-carrier spacing from 15 kHz to 240 kHz allows different slot duration. And with a flexible frame structure allowing mini-slots of 1-13 OFDM symbols, very short transmission time intervals (TTIs) can be configured for low latency services. While the regular TTIs of 14 symbols, as in LTE, are used for high data rates.

- Advanced device capabilities [10, 13]: devices capable of storing local cache, exploiting device-to-device connectivity, and employing advanced receivers with interference suppression mechanisms further im-

prove the reliability and capacity of the system. In addition, fast processing capability reduces the communication latency.

- Network slicing and Edge computing [16, 17]: partitioning the network in different logical segments allows to support a variety of services using a common network infrastructure. And bringing storage and computational capability to the edge, i.e. closer to the users, reduces the end-to-end latency and improves the usage of the backhaul and core network resources.

The specification efforts for the NR in 3GPP started in the mid of 2016 with the definition of the requirements [9]. The initial target was to specify the functionalities for eMBB and enabling low latency for URLLC. By September of 2018 an important milestone was achieved with the finalization of the standalone version of 3GPP Release-15. The specification activities are still ongoing with approximately 25 Release-16 study items, covering different topics. The plan is to finalize the 3GPP Release-16 by the end of 2019. For enhanced URLLC, this should include higher spectral efficiency, and support for stricter latency and reliability requirements from different services as well as Industrial IoT (IIoT).

Among the features enabled in 5G, URLLC is of major importance for the Information Communication Technology (ICT) business. Despite the growth in data traffic and number of subscriptions in the recent years, the revenue of the ICT players has stagnated due to market competition and consumers demands, among other factors. However, URLLC should allow their expansion to newly emerging vertical markets in industry sectors like manufacturing, transportation and energy utilities. This will be possible by introducing the support for mission-critical applications and machine-type communications. Figure I.2 shows the trends of the current operator service revenues, compared with the trends of the revenue growth opportunity with industry digitalization enabled with 5G [18]. The forecast is based on market reports and interviews with 150 global leading market representatives [19]. It is clear that the current revenue is timid compared to what can be achieved with the inclusion of the new markets, which are not addressable with the deployed wireless technologies designed for human centric applications.

2 URLLC in 5G

URLLC should open the door for different kind of novel applications, including augmented/virtual reality, remote robotics, industry automation and intelligent transportation [7]. Traditionally, industrial control systems for example, have relied on wired networks since current wireless systems can not cope with the required reliability. Factory automation is one example

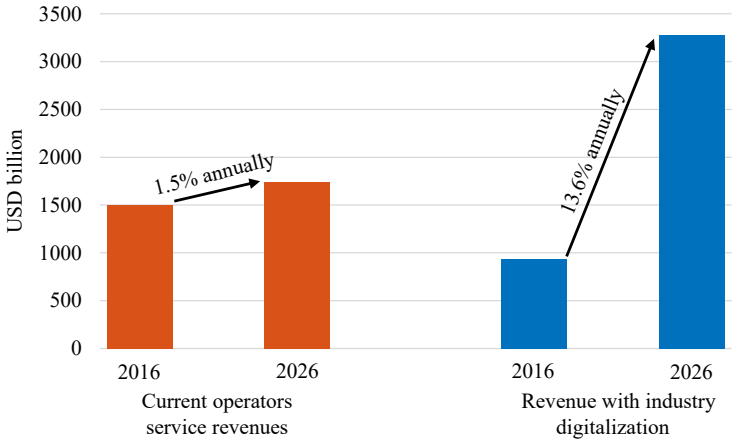


Fig. I.2: Revenue forecast for ICT with 5G (adapted from [18]).

which demands very reliable links for machine to machine communication and fast information exchanging between sensors and actuators in an automated process [20]. The use of reliable radio links powered by URLLC can bring substantial benefits in terms of flexibility, installation and maintenance costs. However, different application areas require distinct levels of reliability and latency. Table I.1 gives some examples identified by 3GPP within Service and System Aspects (SA) working group [21].

Table I.1: Reliability and latency for different applications [21, 22]

Scenario	Latency	Reliability	Traffic
Factory automation (motion control)	2 ms (end-to-end), 1 ms (radio interface)	99.9999%	periodic
Process automation (monitoring)	50 ms (end-to-end)	99.9%	aperiodic
Augmented reality and Virtual reality	1 ms (radio interface)	99.999%	aperiodic
Power distribution (protection)	15 ms (end-to-end), 7 ms (radio interface)	99.999%	periodic
Power distribution (outage management)	5 ms (end-to-end), 3 ms (radio interface)	99.9999%	aperiodic
Transport systems	5 ms (end-to-end), 3 ms (radio interface)	99.999%	periodic

From these examples, it can be noted that the requirements from different areas are quite heterogeneous. Some applications are characterized by periodic traffic in isochronous communication as in motion control. Others

consist of aperiodic transmissions triggered by events, for example, an alarm or an exceeding threshold in process automation. The latency requirements are set for the transmission from the user plane over the radio interface and from end-to-end. Figure I.3 shows a basic illustration of the elements of a mobile network and related links between the entities.

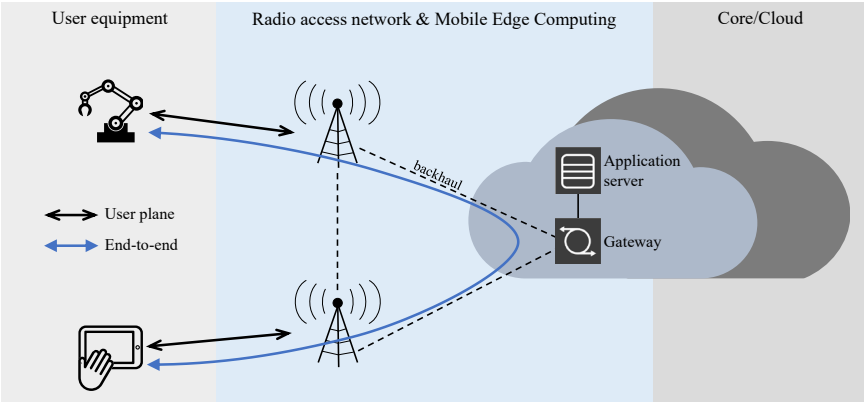


Fig. I.3: User plane and end-to-end relations in a 5G network.

The user plane latency is restricted to the time for transmitting in *one-way* via the radio interface of the user equipment (UE) and the base station in the radio access network (RAN). It includes queuing delays in the transmission buffer, processing time for preparing a payload for transmission through the physical layer, transmission over the wireless channel, and processing time for receiving the signal. In addition, the end-to-end latency includes delays in the backhaul communication and computing time in the core network. Low latency services, for example, benefit of more functionalities computed in the edge. On the other hand, the end-to-end latency increases when services depend on Cloud functionalities, such as computing, storage and networking, implemented far from the RAN. The factors that affect the user plane latency in the RAN are covered with more details in this thesis.

In order to establish a consistent definition and baseline capability target, each key performance indicator (KPI) for URLLC along with the minimum performance requirement for the RAN is presented in Table I.2, as proposed in [6, 9].

An outage event occurs when a URLLC packet cannot be successfully transmitted within the latency deadline, i.e. 1 ms. It is important to note that the presented requirements consider only the one-way latency. Other parts of the system which impact on the end-to-end performance, like core network and other external network interfaces, are not taken into account [23]. The network design should consider these aspects in order to meet the end-to-end latencies required by each service. One example is the use of mobile edge

Table I.2: Definition and baseline target for URLLC KPIs [6, 9]

KPI	Definition	Value
Control plane latency	The time to change from a battery efficient state to a continuous data transfer state, e.g. from idle to active	10 ms
User plane latency or just **latency**	The time for successfully deliver a packet from the radio protocol layer 2/3 ingress point to the radio protocol layer 2/3 egress point via radio interface in active mode	0.5 ms in average (not associated with a high reliability requirement)
Reliability	The success probability of transmitting X bytes within a certain user plane latency deadline, at a certain channel quality	$1 - 10^{-5}$ to transmit 32 bytes within 1 ms of user plane latency

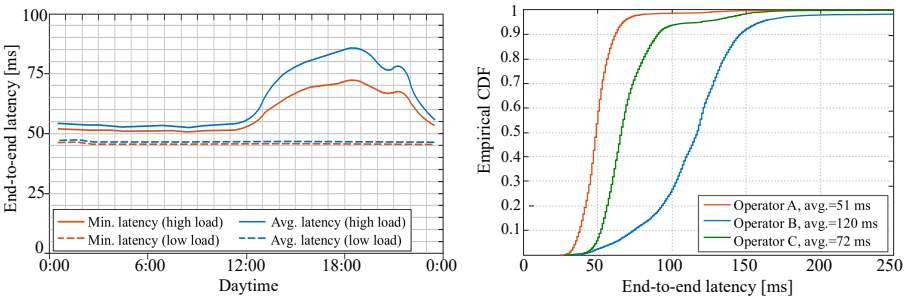
computing for distributing computation tasks closer to the end users [24]. Another KPI, not highlighted in Table I.2, is the RAN *availability*. It relates to the percentage of time that the cellular base station is available for communication. A common availability target for URLLC is not defined, varying depending on the service. In the scenarios described by [21] in Table I.2, for example, it corresponds to the same values as the reliability target.

The defined latency and reliability constraints impose together a design challenge for the RAN part, which has historically been optimized mainly for high network capacity. Fulfilling these requirements simultaneously, inevitably implies a reduced spectral efficiency, given the fundamental trade-off between the KPIs [25]. Therefore, radio resource management (RRM) strategies have also an important role for enabling a viable URLLC solution.

2.1 Latency and reliability in LTE

Here it is discussed the performance of currently deployed systems, showing their limitation for low latency communication. The current cellular technologies were designed with focus on human-centric applications. These applications are generally characterized by large volumes of data generated e.g. by audio/video streaming and Internet browsing. For achieving efficient use of the radio channel, long TTIs of 1 ms are typically utilized, obtaining high coding gains for transmitting data with large blocklength [15]. The cost of this is higher latency in the radio interface. In LTE networks, the radio interface contributes with at least 19 ms to the end-to-end latency, according to [26]. End-to-end latency measurements over LTE networks demonstrate the performance of current systems, as shown in Figure I.4 [27, 28].

Figure I.4a, from measurements in a German network, indicates how the load in the system impacts on the latency. It can be seen that the average latency increases from about 55 ms to 85 ms in the peak time. This is due to the increased number of active devices concurrently accessing the channel



(a) Latency variation in a German operator [27] **(b)** Latency of different Danish operators [28]

Fig. I.4: Measured end-to-end latency in LTE networks.

in a high load period. The results of a similar experiment executed in three different LTE networks in Denmark, are illustrated in Figure I.4b. The study shows that, despite the same radio interface technology is used, the performance is very different due to the distinct deployment of each operator. It should be noted that, while the average latency varies from 50 ms to 120 ms, the latency of up to 200 ms in the 99th percentile is way beyond the target for URLLC services. It is also relevant to note that, the end-to-end latency includes the delays in the backhaul and core network. So, deployments with application closer to the edge of the network should present lower latencies. Even though, the minimum latency, which is approximately 25 ms in the measurement campaigns, is bounded by the radio interface performance.

2.2 Uplink radio interface

Various aspects of a radio communication system impact on the latency and reliability of the transmissions. Here, the aspects related to the radio interface, which affect the performance of transmissions from the UE to the network, are discussed. Figure I.5 illustrates a generic multi-user multi-cell uplink communication system.

Each of the N UEs is connected and synchronized to at least one of the C base stations. Each UE has a transmitter with one or more antennas. And the base station is equipped with a receiver with one a more antennas.

When a packet is generated from the user application, it is transferred to the medium access control (MAC) layer of the radio protocol where it waits in a queue until the transmitter is ready to process it. The transmissions occur in a grid of time-frequency radio resources during a TTI. The UEs transmit using a set of orthogonal frequency-division multiplexing (OFDM) symbols following a specific radio resource configuration controlled by the base station. This may include the time-frequency resource allocation and

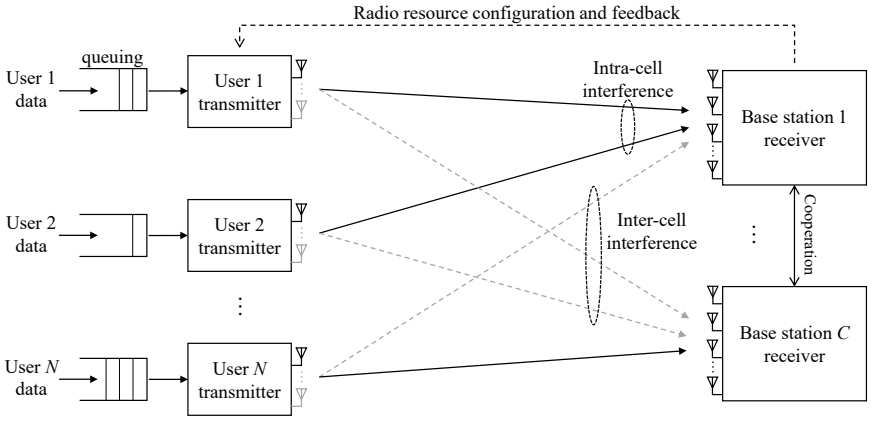


Fig. I.5: Uplink communication system.

transmission parameters like modulation and coding scheme (MCS), power control settings, demodulation reference signal (DMRS) to be used, and retransmission settings [29]. This configuration is dynamically indicated for each data packet transmission, or it is pre-configured for long term use.

UEs transmitting to the same cellular base station can interfere with each other if the transmissions overlap, i.e. intra-cell interference. Besides, overlapping transmission from neighbor cells generates inter-cell interference. Other factors affecting the user signal quality are time-variant and frequency-selective fading, limited power for compensating propagation losses, and noise in the receiver. The reliable decoding of a user's message depends directly on the resulting receive signal quality.

Different processing techniques can be employed in the receiver for suppressing interference and improve the signal quality. For example, the spatial diversity with multiple receive antennas can be exploited using linear combining techniques as minimum mean square error (MMSE) with interference rejection combining (IRC). This type of receiver can potentially suppress both intra-cell and inter-cell interference, as long as, the channel of users can be estimated. The receiver estimates the channel of each interferer signal and project the desired signal in a subspace which minimizes its mean square error [30, 31]. Successive interference cancellation (SIC) capable receivers, on the other hand, employ non-linear methods for successively decoding the signals from mutually interfering users. After decoding a user's signal, usually the strongest one, its interference is removed from the aggregate received signal before decoding the next user. Such iterative process can be computational heavy [32].

The base stations can also cooperate in many ways for improving the system performance. For example, they can apply interference management

techniques or exploit macro diversity with joint reception of the uplink signals. In case a user transmission fails due to poor signal quality, the base station can issue a negative acknowledgement (NACK) feedback or dynamically schedule a retransmission, for exploiting hybrid automatic repeat request (HARQ) mechanisms.

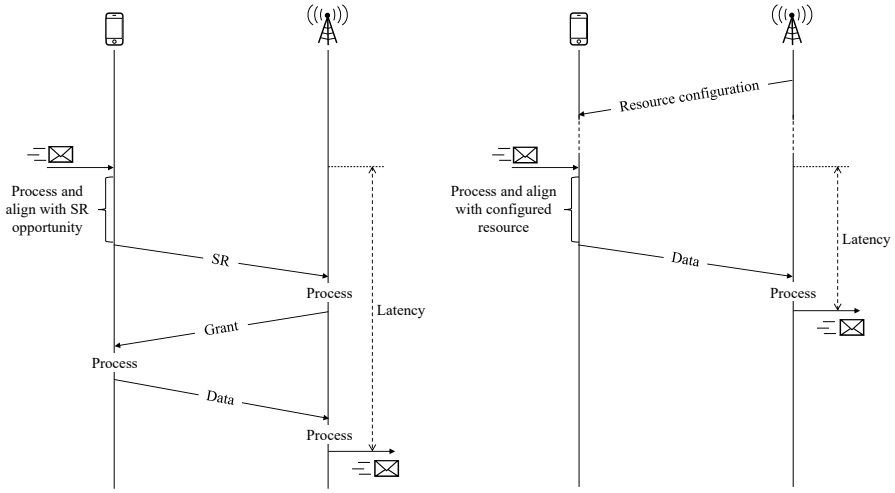
2.3 Resource allocation and channel access

Cellular systems typically employ a scheduling mechanism implemented in the MAC layer to dynamically allocate radio resources to the users, considering their quality of service (QoS) requirements [29]. The scheduler takes into account the channel state information (CSI) and a target block error rate (BLER) to allocate the channel resources for transmitting the buffered data. Link adaptation based on adaptive modulation and coding can be employed to improve the spectral efficiency [33]. The scheduler can multiplex the transmissions performing per-user allocation of the available time-frequency resources. Scheduling algorithms ensure efficient use of the channel and fairness for serving the multiple users while meeting their QoS requirement [34]. The basic access scheme is the orthogonal frequency-division multiple access (OFDMA). Depending on the availability of multiple antennas, multi-user MIMO (MU-MIMO) can be employed by scheduling multiple users over the same time-frequency resources.

For 5G NR, the support of URLLC in the downlink has been widely investigated in recent literature [35–37]. The scheduling prioritization for URLLC traffic, which is allocated in short TTIs, is among the main enablers for meeting the service requirements. In the downlink, the resource allocation and dynamic link adaptation can be promptly provided by a downlink control information (DCI) transmitted in the same TTI as the data. The UE can then quickly process the control information with the parameters and decode the data subsequently.

In the uplink, however, a *grant-based* procedure is usually performed, as illustrated in Figure I.6a. When a data arrives in the transmission buffer the user has to wait for a specific opportunity to transmit a scheduling request (SR) to the base station. The base station processes the SR signal and sends a scheduling grant to the user through a DCI, containing the necessary allocation and transmission parameters. The user processes the control signal and, finally, transmits the data using the granted resources.

An alternative access procedure is based on *grant-free* transmissions, as illustrated in Figure I.6b. In this case, the base station pre-configures the users with the resource allocation and transmission parameters. When a packet arrives, the user can perform the transmission using the preallocated resources, i.e. without needing a dynamic grant. This reduces the control channel overhead and the dependence on the control signaling, which is prone to errors



(a) grant-based **(b)** grant-free

Fig. I.6: Uplink transmission procedures.

and causes delays.

Semi-persistent scheduling (SPS) is a type of grant-free scheme, which is available in LTE since 3GPP Release-8. In SPS, the base station semi-statically allocates radio resources at minimum every 10 ms. It was meant for conveying periodic voice traffic [38]. For deterministic traffic as such, dedicated SPS allocation is clearly beneficial since the scheduling procedure for every transmission is avoided, reducing overhead and latency.

However, for aperiodic or sporadic traffic the preallocation of dedicated resources is inefficient, since the resources would be wasted when the user has no data to transmit. In order to improve the resource efficiency, a group of users can share the same resource allocation [39]. But this creates another problem. Transmissions from these users are susceptible to collisions, causing intra-cell interference, which can jeopardize the reliability. That is the trade-off between latency, reliability and spectral efficiency taking place. Whenever two of these KPIs is improved, the third one is degraded.

2.4 Latency and reliability enhancements from LTE to 5G

In order to reduce the latency for different kinds of traffic, various enhancements were proposed in 3GPP Release-14 [39]. The SPS framework, previously designed focusing on voice traffic, was further extended. The enhancements bring the support for reduced allocation periodicity as short as 1 TTI, and uplink skipping when the user has no data to transmit in the physi-

cal uplink shared channel (PUSCH). This allows a user to transmit in a TTI immediately after the arrival of a packet in the transmitter buffer, or avoid using the channel when the buffer is empty. Multiple users can share the same PUSCH resource allocation, and user specific DMRS can be used for identifying the users transmissions. With the default 1 ms TTI, such fast uplink access allows reaching uplink user plane latency of about 4.5 ms [27].

A fundamental component for achieving lower latencies in 5G NR systems is the adoption of a flexible frame structure which allows for short TTIs. Broadband traffic should be scheduled with usual longer TTI, with default slot duration of 14 OFDM symbols. While latency critical traffic is multiplexed using TTI of up to 0.25 ms [15]. In NR, this is possible by using a frame structure which supports transmissions in mini-slots of 1 to 13 symbols. Additionally, a new and scalable numerology, with sub-carrier spacing of $2^{\mu} \times 15$ kHz with $\mu = \{0, 1, 3, 4\}$, allows the use of shorter symbol duration and different frequency bands [14, 40]. Table I.3 shows the supported 5G NR numerology and some examples of mini-slot duration. Lastly, efficient pipeline processing should guarantee fast processing times of about 2.5 - 6 symbols for URLLC. This is permitted by adopting front loaded reference signal and avoiding interleaving across OFDM symbols, which allow the device to immediately start the data processing before buffering the whole slot [10, 41].

Table I.3: NR numerology and mini-slots duration in microseconds (μs) [42]

Sub-carrier spacing	**Symbol duration [μs]**	**Cyclic prefix [μs]**	**Slot [μs] (14-symb)**	**Mini-slot [μs] (7-symb)**	**(4-symb)**	**(2-symb)**
15 kHz	66.67	4.69	1000	500	286	143
30 kHz	33.33	2.34	500	250	143	71.5
60 kHz	16.67	1.17	250	125	71.5	36
120 kHz	8.33	0.57	125	62.5	36	18
240 kHz	4.17	0.29	62.5	31.3	18	9

In parallel with the development of this work, new 5G NR features were specified. 3GPP has further extended the support for grant-free transmissions in Release-15, referred as configured grant operations [42, 43]. Two types of operation are defined. In Type 1, the transmission parameters and allocation are directly provided and activated via radio resource control (RRC) signaling. In Type 2, the activation is provided via physical downlink control channel (PDCCH). Various physical layer settings are configurable like periodicity and resource allocation, MCS, DMRS, frequency hopping, open loop power control settings, HARQ and repetitions settings. The configured grant operation, together with mini-slots and aggressive processing capability, are the basic NR components for enabling URLLC in uplink. The recent advances in the NR mostly concerns to reaching low latency and sufficient reliability in low load conditions. Although, achieving higher reliability lev-

els and improving the resource utilization for heterogeneous traffic is still actively researched [44].

3 Scope and Objectives of the Thesis

The research described in this dissertation addresses the problem of URLLC in 5G NR, focusing on uplink solutions for the radio interface. It is well understood that the system capacity is very limited when high reliability and low latency requirements are imposed [25]. Due to the fundamental trade-off between capacity and reliability, if the traffic load is increased, the reliability is compromised. The limited capacity implies then a high cost for utilizing URLLC services. Therefore, solutions that allow to improve the reliability or the outage capacity, while meeting the URLLC requirements, are desirable.

The objective of the research is to design and evaluate methods for achieving the URLLC requirements with efficient usage of the available radio spectrum resources. This means that the strict latency and reliability constraints of the service should be met without draining the network capacity. Thereby, the network can make the most of its available resources, e.g. for supporting higher traffic loads or multiplexing other services traffic, while satisfying the required QoS. Analytical and system level simulation tools are used to assess the proposed concepts. As a baseline target, this work adopts the stringent reliability requirement of 99.999% success probability for transiting a small data packet within 1 ms of user plane latency. The scope of the work is illustrated in Figure I.7. Control plane and core network solutions are not included in the scope of this work.

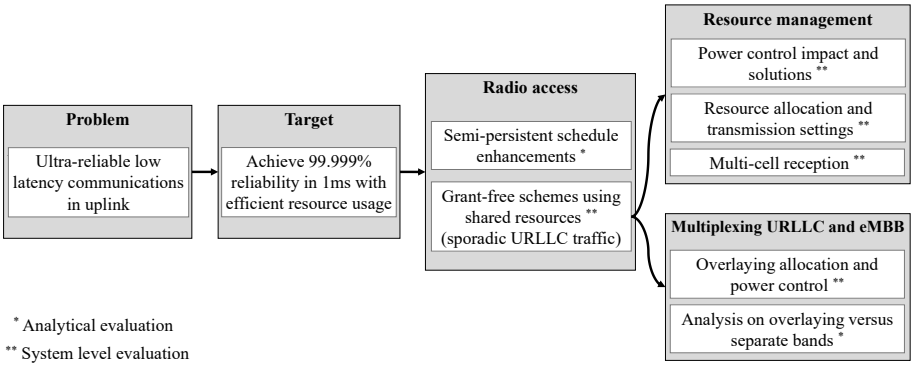


Fig. I.7: Scope of the thesis and evaluation approach

The support for URLLC in the uplink has further issues in comparison with the downlink, as mentioned earlier. The usual uplink radio access and resource management solutions are based on dynamic allocations, tra-

ditionally optimized for high throughput. These solutions are not directly applicable when targeting stringent latency and reliability, due to the signaling issues. Radio access procedures based on pre-configuration of the resources are fundamental components for the solution. In this work, hybrid schemes exploiting pre-configuration of dedicated and shared resources are evaluated, being potentially applicable for deterministic URLLC traffic. For aperiodic traffic, typical for process automation and electric distribution systems as shown previously in Table I.1, grant-free access is considered using shared resources for multiple users. Further enhancements on transmission procedures, resource management, reception schemes and multiplexing are necessary for fulfilling the requirements with efficient usage of resources.

Resource management solutions which are able to reduce the outage probability for URLLC transmissions in a certain load are investigated. These solutions allow, conversely, to increase the load in the system while meeting the target reliability requirement. The considered RRM technical components include power control, resource allocation with multiple MCS, and multi-cell reception mechanisms. The impact of the receiver capability for interference rejection is also studied.

For efficient usage of the radio interface resources, it is desirable to serve multiple traffic types using a single pool of resources. Therefore, the implications of co-scheduling traffic with distinct characteristics, such as eMBB over URLLC resources, are investigated. The receiver configuration is taken into account for determining the performance of multiplexing, either in overlaying resources or in separate bands for each service.

The research is conducted in alignment with the agreements for 5G NR specification in 3GPP. The focus of the analysis is on the performance in frequency band below 6 GHz, which is beneficial for URLLC due to the better propagation characteristics, i.e. lower signal blockage probability compared with centimeter and millimeter-waves bands [45]. Short TTI, using mini-slots, and fast processing times are considered. This enables reduced round-trip time (RTT) and allows enhancements based on retransmissions, or usage of multiple blind repetitions. Baseline receiver types, such as MMSE-IRC with multiple receive antennas [46], are employed for obtaining spatial diversity and multi-user detection in the collision prone scenarios. The multiple access physical resources comprise time and frequency dimensions. Further non-orthogonal multiple access (NOMA) mechanisms were not considered, since no specific scheme was agreed for 5G NR until the moment of this research, being therefore left for future work.

The following list summarizes the research questions and hypotheses addressed in this work:

Q1 What transmission/retransmission schemes should be utilized for achieving the best URLLC performance in uplink?

H1 Grant-free transmission schemes achieve better performance then grant-based, since the delay and error prone scheduling procedure is avoided. Resource sharing should improve the resource efficiency of the system, when the channel is used sporadically. With the use of short TTIs, the target reliability should be achieved aided by repetitions or reactive re-transmissions schemes. And the use of robust MCS, and multi-user detection receiver with interference rejection capability should permit to achieve the stringent requirements.

Q2 How to improve the resource efficiency of URLLC for supporting higher achievable loads in the system?

H2 RRM enhancements are necessary for improving the resource efficiency for grant-free URLLC. Power control, for instance, when optimized with focus on reliability instead of throughput, should improve the outage capacity for URLLC. Besides, using multiple MCS configurations, which can be assigned to the users depending on the channel condition, should further improve the achievable URLLC load. The use of multi-antenna receivers as well as multi-cell reception should increase the level of diversity combining for the benefit of URLLC in uplink.

Q3 How to multiplex eMBB and grant-free URLLC to support both services with improved resource utilization?

H3 In uplink, eMBB cannot be preempted by grant-free URLLC. Thus, it should be possible to multiplex both traffic over the same pool of resources, as long as the power density of eMBB is reduced to a certain extent. Employing overlaying allocations should translate in a significant improvement in terms of efficient usage of resources, compared with using separate bands for URLLC and eMBB. The benefit however should depend on the operation condition, as well as their target load and the receiver type.

4 Research Methodology

For pursuing the objectives of the research, a classical scientific approach is taken. The adopted methodology is summarized as follows:

1. **Identification of the problem and research questions:** An extended survey is conducted for acquiring the background knowledge about the state-of-art. Based on that, the open problems, for instance, which limit the performance of URLLC in uplink, are identified. Research questions are elaborated for triggering the formulation of possible hypotheses and problem solutions.

2. **Formulation of hypothesis along with a potential solution:** An hypothesis is formulated as a tentative answer for the research questions, e.g. by outlining the potential of different transmission mechanisms in the new context of URLLC. Besides, the expected benefits of a proposed solution can be described. Predictions can be drawn in respect of possible outcomes as consequence of the hypothesis.

3. **Modeling the system and propose solutions:** The system is modeled using analytical methods and/or implemented in a Monte Carlo system level simulator. The analytical approach is utilized to provide insights about the trade-offs among the main variables affecting the KPIs. Many simplifications are typically necessary for obtaining a tractable analytical model. System level simulations are used for capturing most of the complex effects caused by the different elements of a realistic radio network. Adopted assumptions are stated, for allowing the study to be comparable and reproducible.

4. **Collection of results and analysis:** Numerical evaluations are carried out for collecting the performance results of the involved mechanisms and test the formulated hypothesis. For the evaluations conducted using system level simulations, the amount of samples collected should be high enough for obtaining statistical relevant results, which account very rare events that impacts URLLC performance. The numerical results are then analyzed and conclusion related to the validity of the hypothesis are drawn. Insights about the meaning of the results for practical applications can also be provided. If further issues and potential enhancements are identified, new hypotheses and possible future work directions can be suggested.

5. **Dissemination of the findings:** In the end of each part of the study, the proposed ideas and learnings are disseminated through the publication of scientific papers, seminar presentations and contribution to specification forums. In addition, patent applications are disclosed in the cases where novel concepts are identified.

5 Contributions

The main contributions of this work are listed below:

1. **New schemes which exploit preallocation of resources for reducing the dependence of control signaling for URLLC retransmissions.**

   The proposed schemes provide efficient HARQ retransmission opportunity without requiring a dynamic scheduling signaling for the retransmission. The idea is to preallocate shared retransmission resources

for a group of users. Two approaches for recovering the signal in the shared resources are presented. One in which only non-acknowledged users utilize the resource. Other which relies on SIC to remove interference from early decoded replicas. Probabilistic models are derived and used to evaluate the schemes, which show better resource efficiency than single-shot transmissions.

2. **Detailed system level analysis and comparison of uplink transmission/retransmission schemes for sporadic URLLC traffic.**

   A study of grant-free access is carried out for different retransmission schemes, listed as, K-repetitions, reactive HARQ, and proactive (repetitions with early termination). The performance with grant-based scheduling is also presented for comparison. Detailed latency statistics are shown and the recommendations regarding the use of each scheme are discussed. The reactive HARQ is highlighted as the most efficient scheme upon a short round trip time.

3. **Sensitivity analysis of the impact of power control settings on the performance of grant-free URLLC, along with a retransmission boosting strategy.**

   The outage probabilities for URLLC are shown for different open loop power control settings. The achievable URLLC load shows a considerable improvement when optimized settings, focused on reliability, are utilized. Recommendations for the power control configuration using full path loss compensation are provided. It is also discussed the applicability and limitations of power boosted retransmission in the macro scenario.

4. **A grant-free design for sporadic traffic transmission, including radio resource management solutions for improving the achievable URLLC load.**

   The proposal includes the usage of HARQ retransmissions, and resource allocation with MCS and power control settings assignment according to the average channel condition of the user. Users in high coupling gain condition can use the power budget to transmit using higher MCS in smaller sub-bands, reducing the overlapping issues. In addition, it is shown that multi-antenna receivers with interference rejection capability and multi-packet reception are determinant for achieving higher URLLC loads.

5. **A comprehensive study on the potential of multi-cell reception for grant-free URLLC, accompanied by RRM enhancements.**

   The relevant settings for a multi-cell reception system are demonstrated. It is shown, for instance, that it is more important to have more users

connected to a few assisting cells than to have a high number of assisting cells per user. The backhaul load is accounted for soft combining, selection combining and a hybrid combining scheme. Multi-cell reception greatly improves the URLLC outage capacity, but shows lower impact when other diversity mechanisms are in place, like multi-antenna receiver and HARQ. In addition, multi-cell aware RRM mechanisms allow to reduce the resource usage by users assisted by many cells.

6. **System level evaluation and insights about the effect of overlaying allocations for high volume eMBB data and grant-free URLLC.**

   Allocating eMBB over resources used for grant-free URLLC has the potential to improve resource efficiency. Power control should be distinguished for both services, and used for managing the trade-off between their achievable load. It is shown as well that not only URLLC but also eMBB should use full path loss compensation and reduced power density for prioritizing URLLC.

7. **An analytical framework for determining the outage probability and the achievable load for URLLC and eMBB using overlaid or separate resources, along with numerical results based on 5G NR assumptions.**

   The method considers the outage probability for MMSE receivers. The performance of the multiplexing considers also the case when ideal SIC is utilized. Based on that, it can be estimated whether using overlaying allocation or using separate frequency bands should be preferred, depending on the operation regime, such as TTI duration, reliability requirement, number of antennas, among other settings. Overlaying allocations show benefits mainly for low URLLC load and with the use of MMSE and SIC in medium to high SNR conditions.

Based on this research, the following scientific publications were authored, forming the main content of this thesis.

Paper A: R. Abreu, P. Mogensen and K. Pedersen, "Pre-scheduled Resources for Retransmissions in Ultra-Reliable and Low Latency Communications", *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, March, 2017.

Paper B: R. Abreu, G. Berardinelli, T. Jacobsen, K. Pedersen and P. Mogensen, "A Blind Retransmission Scheme for Ultra-Reliable and Low Latency Communications", *IEEE 87th Vehicular Technology Conference (VTC Spring)*, July, 2018.

Paper C: T. Jacobsen, R. Abreu, G. Berardinelli, K. Pedersen, P. Mogensen, I. Z. Kovács and T. K. Madsen, "System Level Analysis of Uplink

Grant-Free Transmission for URLLC", *IEEE 2017 GlobeCom Workshops*, December, 2017.

Paper D: R. Abreu, T. Jacobsen, G. Berardinelli, K. Pedersen, I. Z. Kovács and P. Mogensen, "Power Control Optimization for Uplink Grant-Free URLLC", *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, April, 2018.

Paper E: R. Abreu, T. Jacobsen, G. Berardinelli, K. Pedersen, I. Z. Kovács and P. Mogensen, "Efficient Resource Configuration for Grant-Free Ultra-Reliable Low Latency Communications", *IEEE Transactions of Vehicular Technology*, 2019. **Submitted for publication**.

Paper F: T. Jacobsen, R. Abreu, G. Berardinelli, K. Pedersen, I. Z. Kovács and P. Mogensen, "Multi-cell Reception for Uplink Grant-Free Ultra-Reliable Low-Latency Communications", *IEEE Access*, 2019.

Paper G: R. Abreu, T. Jacobsen, K. Pedersen, G. Berardinelli and P. Mogensen, "System Level Analysis of eMBB and Grant-Free URLLC Multiplexing in Uplink", *2019 IEEE 89th Vehicular Technology Conference (VTC Spring)*, April, 2019.

Paper H: R. Abreu, T. Jacobsen, G. Berardinelli, N. H. Mahmood, K. Pedersen, I. Z. Kovács and P. Mogensen, "On the Multiplexing of Broadband Traffic and Grant-Free Ultra-Reliable Communication in Uplink", *2019 IEEE 89th Vehicular Technology Conference (VTC Spring)*, April, 2019. **Received the Best Student Paper Award**.

Additionally, supplementary publications were co-authored during collaboration in other researches closely related to this work. These papers are included in the appendix and will be also refereed in the relevant parts of this thesis in the following order:

Paper I: T. Jacobsen, R. Abreu, G. Berardinelli, K. Pedersen, I. Z. Kovács and P. Mogensen, "System Level Analysis of K-Repetition for Uplink Grant-Free URLLC in 5G NR", *European Wireless*, May, 2019.

Paper J: T. Jacobsen, R. Abreu, G. Berardinelli, K. Pedersen, I. Z. Kovács and P. Mogensen, "Joint Resource Configuration and MCS Selection Scheme for Uplink Grant-Free URLLC", *IEEE 2018 GlobeCom Workshops*, December, 2018.

Paper K: G. Berardinelli, R. Abreu, T. Jacobsen, N. H. Mahmood, K. Pedersen, I. Z. Kovács and P. Mogensen, "On the Achievable Rates over Collision-Prone Radio Resources with Linear Receivers", *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, September, 2018.

The novelties identified during the research were filed as patent applications in co-operation with Nokia Bell Labs and in accordance with the frame agreement with Aalborg University. The titles of the filed invention reports are listed as follows:

Patent Application 1: Semi-Persistent Scheduling of Contention Based channel for Transmission Repetitions.

Patent Application 2: Blind Retransmissions over Shared Resources.

Patent Application 3: Delayed robust side information.

Patent Application 4: Resource and MCS configuration and dynamic adjustment for grant-free uplink transmission.

Patent Application 5: Efficient configured grant operations for data transfer for RRC inactive UEs.

The observations and proposals from this work were disseminated, as part of the collaboration in projects with Nokia Bell Labs, and utilized as input for 5G NR specification in 3GPP. Some of the related contributions for RAN1 working group are refereed in [47], [48], [49], [50], [51], [52], [53].

A considerable part of this work was dedicated to the development of system level simulator functionalities. The proprietary simulator has been developed in collaboration with Nokia Bell Labs, and includes detailed modeling of LTE and 5G network functionalities. The development is based on C++ object-oriented programming. The list below includes a short description of the main implemented features which were necessary for the studies:

- **Grant-free transmission on shared resources**: overall framework for multiple grant-free access schemes and simultaneous transmissions over configured time-frequency resources.
- **Detailed statistics for uplink**: logging of transport block transmission details and latency measurements in different layers for statistics calculations.
- **Grant-free sub-bands allocation and hopping**: support multiple grant-free sub-bands and hopping between transmission repetitions.
- **Collision probability statistics**: collect statistics for transmissions overlapping in time-frequency resources.
- **Power control enhancements**: implementation of power control scheme and statistics per transmission, power headroom and power boosting.
- **Joint MCS and power control association**: transmit parameters association according channel/coupling condition and additional statistics.

- **Grant-based conservative scheduler**: scheduling aiming on reduce queuing, minimize block error rate and avoid segmentation.
- **Even load for heterogeneous traffic**: configurable deployment of users in scenario according service category.
- **Multiplexing of eMBB and grant-free URLLC**: power control differentiation depending on the service class of the data.
- **Co-scheduling heterogeneous traffic**: support full buffer traffic scheduled together with grant-free sporadic traffic.
- **Simulation campaigns**: scripts to launch parallel executions, and consolidate a huge amount of data from multiple simulation drops.
- **Post-processing and statistics representation**: implementation of complementary Matlab scripts to post process detailed logs information, calculate relevant statistics and generate plots.
- **Modeling and testing**: verification of the simulation models comparing against reference results and calibration curves.

6 Thesis Outline

This dissertation consists of an introductory chapter and a collection of papers organized in three main parts complemented by an appendix. Each part includes a general overview with the problem description, objectives and main findings together with associated recommendations, followed by the respective papers. An overview of the thesis structure is illustrated in Figure I.8.

The thesis outline is as follows:

Part I - An introductory chapter that presents the overview about the research topic. One section is dedicated to the aspects related to detection and control signaling issues for URLLC in uplink. General considerations for this work are then delineated.

Part II - In this part, radio access procedures for URLLC relying on semi-static resource configuration, as grant-free, and grant-based procedures are studied. The impact of retransmission schemes in terms of latency, reliability, and resource efficiency are discussed. Papers A, B and C form the main body of this part, and Paper I is related in appendix.

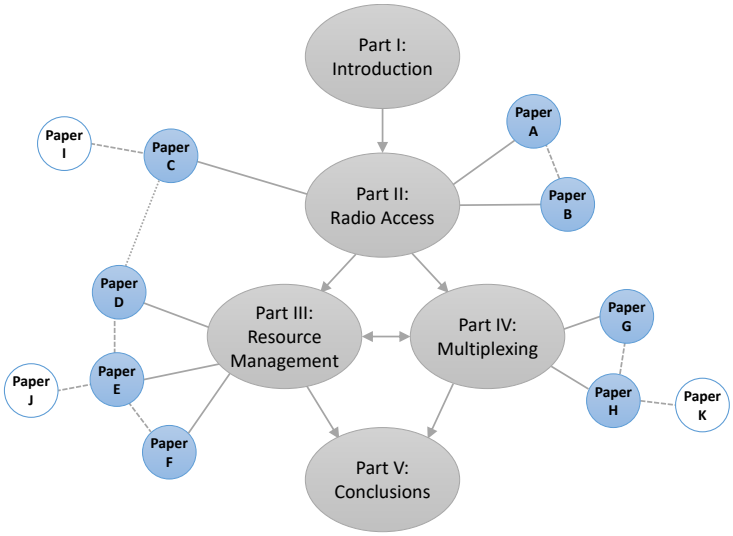


Fig. I.8: Thesis structure.

Part III - Radio resource management solutions are presented with the focus on improving the achievable URLLC load. Resource allocation, power control, modulation and coding scheme selection, as well as reception mechanisms are studied in a system level perspective. Papers D, E and F are included in this part which also refers to Paper J in appendix.

Part IV - This part presents the aspects related to the multiplexing of eMBB and URLLC services, which have different characteristics in terms of traffic and requirements. The use of shared resources for eMBB and grant-free URLLC is analyzed through simulations and with a theoretical model. Papers G and H form the main body of this part, and the related work in Paper K is included for reference in appendix.

Part V - A summary of the main conclusions and final remarks of the work are presented in this part. It is finalized with potential future paths for the research.

Part VI - The appendix includes additional information and co-authored papers that relates to the core of this research.

References

[1] S. Hong, *Wireless: from Marconi's black-box to the audion*. Cambridge, MA and London: MIT Press, 2001.

[2] Jornal do Commercio (Rio de Janeiro), "O Teléforo," Jun. 1899, page 1, sixth column, reprinted from (São Paulo) Diario Hespanhol.

[3] New York Herald, Fifth section, "Talking Over A Gap Of Miles Along A Ray Of Light - Brazilian Priest's Invention," Oct. 1902, page 9.

[4] "The invention of radio," accessed on 24-April-2019. [Online]. Available: http://www.makingthemodernworld.org.uk/stories/the_age_of_the_mass/07.ST.04/?scene=4

[5] H. Holma and A. Toskala, *LTE Advanced: 3GPP Solution for IMT-Advanced*. Wiley, 2012.

[6] ITU-R, "Report ITU-R M.2410-0 - Minimum requirements related to technical performance for IMT-2020 radio interface(s)," International Telecommunication Union (ITU), Tech. Rep., Nov. 2017.

[7] International Telecommunication Union (ITU), "IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond," ITU Radiocommunication Sector, Tech. Rep., Sep. 2015.

[8] R. Ratasuk, N. Mangalvedhe, J. Kaikkonen, and M. Robert, "Data Channel Design and Performance for LTE Narrowband IoT," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, Sep. 2016.

[9] 3GPP TR 38.913 v14.1.0, "Study on Scenarios and Requirements for Next Generation Access Technologies," Mar. 2017.

[10] S. Parkvall, E. Dahlman, A. Furuskar, and M. Frenne, "NR: The New 5G Radio Access Technology," *IEEE Communications Standards Magazine*, vol. 1, no. 4, pp. 24–30, Dec. 2017.

[11] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of Millimeter Wave Communications for Fifth-Generation (5G) Wireless Networks - With a Focus on Propagation Models," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 12, pp. 6213–6230, Dec 2017.

[12] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[13] 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, Feb. 2014.

[14] A. A. Zaidi, R. Baldemair, H. Tullberg, H. Bjorkegren, L. Sundstrom, J. Medbo, C. Kilinc, and I. D. Silva, "Waveform and Numerology to Support 5G Services and Requirements," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 90–98, Nov. 2016.

[15] K. I. Pedersen, G. Berardinelli, F. Frederiksen, P. Mogensen, and A. Szufarska, "A Flexible 5G Frame Structure Design for Frequency-Division Duplex Cases," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 53–59, Mar. 2016.

[16] K. Samdanis, X. Costa-Perez, and V. Sciancalepore, "From network sharing to multi-tenancy: The 5G network slice broker," *IEEE Communications Magazine*, vol. 54, no. 7, pp. 32–39, Jul. 2016.

[17] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration," *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1657–1681, May 2017.

[18] Telefonaktiebolaget LM Ericsson - Tech Report, "The 5G Business Potential," 5G Americas, Tech. Rep., Sep. 2017.

[19] B. Virag and K. Taga and V. Dimitrov and G. Peres and T. Gaar, "Major strategic choices ahead of TelCos: Reconfiguring for value," Arthur D. Little, Tech. Rep., Jan. 2017.

[20] B. Holfeld, D. Wieruch, T. Wirth, L. Thiele, S. A. Ashraf, J. Huschke, I. Aktas, and J. Ansari, "Wireless Communication for Factory Automation: an opportunity for LTE and 5G systems," *IEEE Communications Magazine*, vol. 54, no. 6, pp. 36–43, Jun. 2016.

[21] 3GPP TS 22.261 v16.5.0, "Service requirements for the 5G system," Sep. 2018.

[22] 3GPP TR 38.824 v1.0.0, "Study on physical layer enhancements for NR ultra-reliable and low latency case (URLLC)," Nov. 2018.

[23] J. Sachs, G. Wikstrom, T. Dudda, R. Baldemair, and K. Kittichokechai, "5G Radio Network Design for Ultra-Reliable Low-Latency Communication," *IEEE Network*, vol. 32, no. 2, pp. 24–31, Mar. 2018.

[24] M. S. Elbamby, M. Bennis, and W. Saad, "Proactive edge computing in latency-constrained fog networks," in *2017 European Conference on Networks and Communications (EuCNC)*, Jun. 2017.

[25] B. Soret, P. Mogensen, K. I. Pedersen, and M. C. Aguayo-Torres, "Fundamental Tradeoffs among Reliability, Latency and Throughput in Cellular Networks," in *2014 IEEE Globecom Workshops*, Dec. 2014.

[26] H. Holma and A. Toskala, *LTE for UMTS: Evolution to LTE-Advanced*, 2nd ed. Wiley Publishing, 2011.

[27] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, I. Riedel, A. Puschmann, A. Mitschele-Thiel, M. Muller, T. Elste, and M. Windisch, "Latency Critical IoT Applications in 5G: Perspective on the Design of Radio Interface and Network Architecture," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 70–78, Feb. 2017.

[28] M. Lauridsen, L. C. Gimenez, I. Rodriguez, T. B. Sorensen, and P. Mogensen, "From LTE to 5G for Connected Mobility," *IEEE Communications Magazine*, vol. 55, no. 3, pp. 156–162, Mar. 2017.

[29] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*, 2nd ed. Wiley, 2011, pp. 108–120.

[30] J. H. Winters, "Optimum combining in digital mobile radio with cochannel interference," *IEEE Transactions on Vehicular Technology*, vol. 33, no. 3, pp. 144–155, Aug. 1984.

[31] F. M. L. Tavares, G. Berardinelli, N. H. Mahmood, T. B. Sørensen, and P. Mogensen, "On the Potential of Interference Rejection Combining in B4G Networks," in *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*, Sep. 2013.

[32] N. I. Miridakis and D. D. Vergados, "A Survey on the Successive Interference Cancellation Performance for Single-Antenna and Multiple-Antenna OFDM Systems," *IEEE Communications Surveys Tutorials*, vol. 15, no. 1, pp. 312–335, Apr. 2013.

[33] C. Rosa, D. L. Villa, C. U. Castellanos, F. D. Calabrese, P. H. Michaelsen, K. I. Pedersen, and P. Skov, "Performance of Fast AMC in E-UTRAN Uplink," in *IEEE ICC*, May 2008, pp. 4973–4977.

[34] H. Fattah and C. Leung, "An overview of scheduling algorithms in wireless multimedia networks," *IEEE Wireless Communications*, vol. 9, no. 5, pp. 76–83, Oct. 2002.

[35] G. Gerardino, "Radio Resource Management for Ultra-Reliable Low-Latency Communications in 5G," Ph.D. dissertation, Aalborg University, 2017.

[36] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, "Ultra-Reliable and Low-Latency Communications in 5G Downlink: Physical Layer Aspects," *IEEE Wireless Communications*, vol. 25, no. 3, pp. 124–130, Jun. 2018.

[37] A. Anand and G. de Veciana, "Resource Allocation and HARQ Optimization for URLLC Traffic in 5G Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 11, pp. 2411–2421, Nov. 2018.

[38] D. Jiang, H. Wang, E. Malkamaki, and E. Tuomaala, "Principle and Performance of Semi-Persistent Scheduling for VoIP in LTE System," in *2007 International Conference on Wireless Communications, Networking and Mobile Computing*, Sep. 2007.

[39] 3GPP TR 36.881 v14.0.0, "Study on latency reduction techniques for LTE," Jul. 2016.

[40] 3GPP TS 38.300 V15.2.0, "NR; NR and NG-RAN Overall Description," Jun. 2018.

[41] R1-1807825, "Summary of Maintenance for DL/UL Scheduling," May 2018.

[42] 3GPP TS 38.214 v15.4.0, "NR; Physical layer procedures for data," Jan. 2019.

[43] 3GPP TS 38.331 V15.4.0, "NR; Radio Resource Control (RRC) protocol specification," Jan. 2019.

[44] RP-182089, "Study on physical layer enhancements for NR ultra-reliable and low latency communication (URLLC)," Sep. 2018.

[45] O. Semiari, W. Saad, M. Bennis, and M. Debbah, "Integrated Millimeter Wave and Sub-6 GHz Wireless Networks: A Roadmap for Joint Mobile Broadband and Ultra-Reliable Low-Latency Communications," *ArXiv e-prints, arXiv:1802.03837 [eess.SP]*, Oct. 2018.

[46] 3GPP TR 38.802 v14.0.0, "Study on New Radio Access Technology," Mar. 2017.

[47] R1-1609748, "Semi-persistent scheduling for 5G new radio URLLC," Oct. 2016.

[48] R1-1612251, "Enhanced semi-persistent scheduling for 5G URLLC," Nov. 2016.

[49] R1-1703329, "UL grant-free transmission for URLLC," Feb. 2017.

[50] R1-1808570, "On Configured Grant enhancements for NR URLLC," Aug. 2018.

[51] R1-1813117, "Solution for UL inter-UE multiplexing between eMBB and URLLC," Nov. 2018.

[52] R1-1813118, "On Configured Grant enhancements for NR URLLC," Nov. 2018.

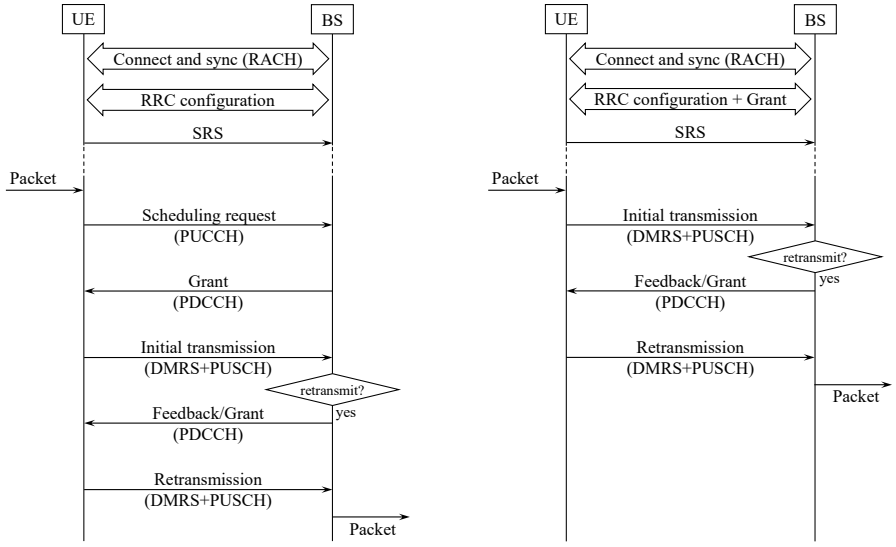
[53] R1-1900931, "Solution for UL inter-UE multiplexing between eMBB and URLLC ," Jan. 2019.

Challenges and Research Assumptions

The first part of this section discusses the impact of control and data channel errors on the overall transmission reliability. Understanding this aspect is important, since the specific signaling of the procedures used in uplink affects the communication performance. The analyses presented here determine the reliability constraints of the channels, which should be taken into account in the design for supporting URLLC. In the last part of the section, the general assumptions adopted in the research are discussed.

1 Reliability for data and control channels

Figure I.9 shows the signaling related to the uplink procedures in 5G NR cellular networks, as well as the used channels in each step, for grant-based and grant-free transmissions. Considering the user plane URLLC performance, the UEs are not restricted by a battery efficient state, as stated in [1]. So, it is assumed that the UEs have performed the random access channel (RACH) procedure for establishing the connection and synchronization with the base station. Maintaining the time-frequency synchronization is important for avoiding inter-carrier and inter-symbol interference in OFDM, and is required prior to the signal equalization and coherent detection in the receiver. After the RACH procedure, the UE can establish the RRC connection. Session management, security functions, QoS management, mobility functions, measurement and reporting configurations and a sort of other functions can be configured by the RRC signaling during the connection setup. Some physical layer parameters related to power control, DMRS and HARQ can also be configured for both, grant-based and grant-free procedures [2]. The RRC signaling is typically robust, being protected by automatic repeat request (ARQ).



(a) grant-based **(b)** grant-free

Fig. I.9: Signaling for uplink transmission.

1.1 Transmission success probability

After the connection is settled in the control plane, the UE can send sounding reference signal (SRS) periodically or eventually triggered from the base station, so that the latter can estimate the channel quality. In the grant-based procedure, upon a data packet arrival, the UE should send a SR using the configured physical uplink control channel (PUCCH). This procedure is very flexible and resource efficient, because the base station can assign the transmission parameters and resources accurately, based on the UE buffer status report and the estimated channel quality. However, the availability and periodicity of the SR resources have a direct impact on the latency, since the UE needs to wait for the SR opportunity. And in case that SR resources are not available in the PUCCH, the UE needs to initiate a RACH procedure, causing even higher delays. Besides, for URLLC, the amount of SR resources in the PUCCH should be sufficiently high for a reliable detection, leading to a high control channel overhead. By detecting the SR, the base station should reply with an uplink grant using an also robust DCI transmitted through the PDCCH. The DCI contains the remaining parameters needed for the uplink transmission, including at least the MCS, resource allocation and HARQ configuration [3]. After the UE decodes the DCI, it can finally transmit the data in the allocated resource blocks (RBs) within the PUSCH. The RBs carrying the data are accompanied with a DMRS that enables channel estimation.

HARQ is known for improving the spectral efficiency and reliability of the communication through the wireless channel [4]. However, the support of HARQ depends on the control channel for conveying a feedback or a retransmission grant. Moreover, the latency budget should afford for the HARQ RTT. The RTT is the defined by the duration from the beginning of the initial transmission until the moment when the UE has received the feedback and is ready for performing a retransmission. In case a HARQ retransmission cannot be issued within the latency budget or if a reliable control channel is unavailable, the URLLC transmission should count on a single-shot attempt. Without HARQ, the success probability of the grant-based transmission is expressed as

$$P_s = (1 - \xi_{uc})(1 - \xi_{dc})(1 - \xi_1). \tag{1.1}$$

where ξ_{uc} is the error probability of the SR signal detection in the uplink control channel, ξ_{dc} is the error probability of the grant in the downlink control channel, and ξ_1 is the error probability of the initial transmission in the PUSCH.

With short TTI and fast processing time, one HARQ retransmission can be possibly utilized without violating the URLLC latency deadline. In that case, the UE should monitor the PDCCH carrying the feedback and potential grant for a retransmission in case the initial one has failed. One possible option is to utilize a single-bit NACK feedback signal for indicating to the UE to retransmit using the same resources as in the initial attempt. Another option is to use a DCI signal to dynamically grant the time-frequency resources for the retransmission. In either option, the error probability of the control signal is represented here by ξ_{dc}. Though in practice the achievable error probability values can be different due to the distinct message formats. Considering that the UE can perform one HARQ retransmission within the latency budget, the success probability is given by

$$P_s = (1 - \xi_{uc})(1 - \xi_{dc}) \left[(1 - \xi_1) + \xi_1(1 - \xi_{dc})(1 - \xi_2)\right], \tag{1.2}$$

where ξ_2 is the error probability of the retransmission.

For grant-free transmissions, the base station should have prior knowledge about the traffic flow, which can be based on the QoS characteristics defined during the session establishment [5]. The base station configures all transmission parameters, through RRC signaling, and grant the resources in advance to the UE. Therefore, the dependence on the uplink SR signaling and dynamic grant for every packet transmission is eliminated. Then, the base station just needs to decode the RBs configured for the UE transmissions. The transmission success probability in this case, without HARQ retransmissions, can be expressed as

$$P_s = (1 - \xi_1). \tag{1.3}$$

And if one HARQ retransmission is supported, the success probability is given by

$$P_s = (1 - \xi_1) + \xi_1(1 - \xi_{dc})(1 - \xi_2), \tag{1.4}$$

which of course depends on the downlink control signaling carrying the feedback and granting the retransmission.

For deterministic traffic, the base station can assign the resources to the UE in accordance with its traffic pattern. In this case, the base station knows exaclty when a transmission occurs. While for sporadic traffic, the base station should configure the resources allowing the UE to transmit whenever a packet arrives in the transmission buffer. And if the UE has no data to transmit it will not use the configured resource. This can incur in a high blind detection effort from the base station. Besides, unnecessary feedback signaling would be issued whenever the base station misinterpret the lack of transmission as a failure. An alternative is to use the DMRS prepended to the data as pilots for transmission detection [6, 7]. Based on that, the transmission success probability is given by

$$P_s = (1 - \xi_{rs})(1 - \xi_1), \tag{1.5}$$

which depends on the error probability ξ_{rs} for detecting the prepended reference signal. And with one HARQ retransmission supported, the success probability is

$$P_s = (1 - \xi_{rs}) \left[(1 - \xi_1) + \xi_1(1 - \xi_{dc})(1 - \xi_2)\right]. \tag{1.6}$$

Note that the expression considers that the reference signal of the initial transmission attempt should be detected successfully, so that the base station can issue a grant for retransmission, in case the initial transmission fails.

1.2 Reliability constraints

The data and control channels exhibit different impacts on the overall performance depending on the transmission procedure [8]. Knowing the performance of the data transmissions in the uplink channel, which is determined by the error probabilities ξ_1 and ξ_2, permits to derive the reliability required for the control channels, depending on the target communication reliability. The value of ξ_1 depends on the MCS and the post-processing signal to interference-and-noise ratio (SINR) of the received signal. The base station is in charge of allocating the PUSCH resources and assigning the transmission parameters for meeting a target BLER. The lower is this target value, the

lower should be the MCS and higher amount of resources are needed for the initial transmission.

In LTE systems with focus on high spectral efficiency, the BLER target for the initial transmission is in the order of 10% with the use of HARQ. However, for URLLC this value should be lower, i.e. $\xi_1 \leq 10^{-2}$ [9]. In case a retransmission is issued with the same MCS and in the same channel conditions as the initial transmission, a combining gain of 3 dB can be achieved, by soft combining both transmissions. Otherwise, the base station can adapt the retransmission parameters to ensure that it is received with a high probability. Therefore, the retransmission can achieve a very low error probability. In this analysis, it is assumed that ξ_2 reaches an error floor of 10^{-5}, as in [10].

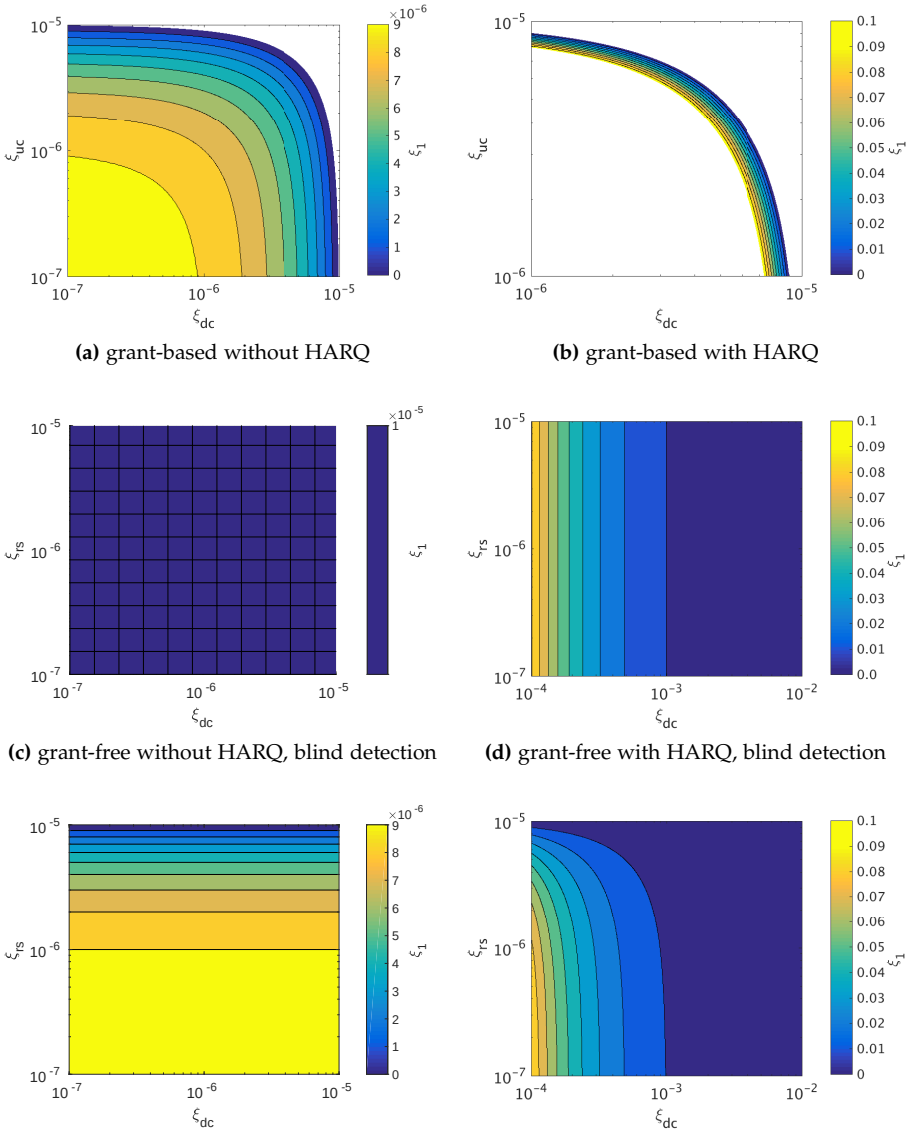
Figure I.10 shows the error probability constraints for the different channels in order to meet a success probability $P_s = 1 - 10^{-5}$, with the different transmission procedures. Each plot from (a) to (f) is based on the results according the success probability expressions from 1.1 to 1.6, respectively. It can be observed that grant-based schemes impose a stringent requirement for the uplink and downlink control channels, since the error probability of SR and grant signals should be lower than 10^{-5}. With HARQ, the reliability requirement of the control channels should be still high, but the error target for the initial transmission can be relaxed, counting that much higher reliability is achieved after soft combining with a retransmission. The lower is the reliability of the control channels, the higher should be the reliability of the data channel, hence, the lower is the spectral efficiency.

For grant-free transmissions with base station performing blind detection over the allocated resources, there is no dependence on uplink control channel or on the reference signal for activity detection purpose. When HARQ is not enabled, the one-shot transmission should be sufficiently robust to meet the reliability target. While with HARQ, the reliability requirement of the initial transmission can be relaxed as long as the downlink control channel used for the HARQ feedback is sufficiently reliable. For instance, with an initial BLER of 10^{-2}, the reliability of the PDCCH should be in the order of 10^{-3}. Without blind detection, the base station needs to reliably detect the presence of the reference signal, i.e. $\xi_{rs} < 10^{-5}$, in order to decode the initial transmission or to issue a retransmission grant.

1.3 Signaling impact

Achieving the mentioned reliability levels is very challenging, specially in low SINR conditions, requiring diversity and a high amount of radio resources. The reliability of the control signaling can be improved by using repetitions and higher aggregation levels, i.e. using more resource elements for the control channels. However, this reduces the capacity of the system.

Another important aspect, as will be further discussed in this work, is



(a) grant-based without HARQ

(b) grant-based with HARQ

(c) grant-free without HARQ, blind detection

(d) grant-free with HARQ, blind detection

(e) grant-free without HARQ, uplink detection

(f) grant-free with HARQ, uplink detection

Fig. I.10: Error probability limits for control and data channel.

that each of the steps of the procedures illustrated in Figure I.9 impacts on the latency. The duration of the steps depends on multiple factors such as TTI size, numerology and symbol duration, UE and base station processing times, resource availability, etc. The handshake procedure, required by the

grant-based scheme for every transmission, leads to an overhead which can jeopardize the communication latency.

Grant-free procedures, have lower dependence on fast and reliable control signaling, though they are less flexible in terms of resource allocation and link adaptation due to the semi-static RRC configuration. This calls for solutions to improve the radio resource utilization of the data channel, which are addressed in this work.

2 General research assumptions

In order to focus on the performance trends of the grant-free access and resource management solutions, a few assumptions were considered throughout the research. In this section, some of the general assumptions are discussed, while specific assumptions for each part of the work are detailed in the embodied papers.

Spectrum

Low frequency bands, below 6 GHz, are assumed for the URLLC deployments, given the lower attenuation and better support for non-line-of-sight (NLOS) wide-area environments [11, 12]. Besides, large bandwidths should be available for commercial deployments of 5G in the of 3.3 to 4.2 GHz range [13].

Frame structure

Transmissions occur in a time-frequency grid using an air interface based on OFDM frame structure. Frequency division duplex (FDD) is assumed, preventing potential delays caused by discontinuities in the link direction. Short sub-carrier spacing, such as 15 kHz and 30 kHz, used typically for macro cellular scenarios with higher delay spreads, are considered throughout the study. This implies longer slot durations compared to higher sub-carrier spacings. The latter are mainly applicable for indoor scenarios and at high carrier frequencies, given the robustness to phase noise [14]. Mini-slots with 2 to 7 symbols are then assumed for enabling short TTIs.

Timing

Due to the low latency requirement each transmission is localized in the short TTIs. Packet segmentation over multiple TTIs is avoided. The base station and the UE have fast processing capability allowing to achieve low round trip time. Propagation delays are considered negligible. The packet data spans over the available bandwidth employing robust MCS and exploiting

diversity in frequency domain. The potential for diversity in time domain is very limited, since the channel coherence time is usually longer then the latency constraint [15]. Hence, retransmissions are mainly employed to obtain combining gain and interference diversity.

Traffic and channel variability

The grant-free configuration is semi-static, i.e. it can only be changed on a slow basis, with the assumption that the traffic and channel characteristics have low variability in the URLLC usage scenarios. That should be the case of machine-to-machine communication without high mobility requirements, for instance in electricity distribution, process automation and monitoring applications [16]. The users traffic are assumed to be uncorrelated. Reducing queuing delays has a high relevance for scheduler design. However, solutions for queuing are not directly addressed here, since the grant-free procedure is assumed to make the best effort in terms of scheduling by immediately transmitting the packet in the upcoming transmission opportunity.

Control channel and detection

For the part of the work in which system level simulations are utilized for the evaluations, control channel errors and lack of control channel resources are not taken into account. This tends to favor grant-based schemes which are used as reference for comparison with the studied grant-free schemes. For grant-free, it is assumed that the base station is able to ideally detect the presence of the UEs transmissions over the preallocated resources. In case of deterministic traffic, the base station can know exactly when the UEs transmissions occur. However, in case of sporadic traffic, the practical implication of the assumption is that the base station needs to perform a blind detection in every allocated resources, for detecting the occurrence of a UE transmission. Otherwise, reliable reference sequences should prepend the data allowing the base station to perform fast detection, e.g. by cross-correlation, before decoding the data. And in case the grant-free resources are shared by multiple UEs, orthogonal reference sequences need to be used by the transmitting UEs, to allow the base station to differentiate them and estimate their channel. It is worth to mention that the detection performance based on DMRS is a topic that has being recently studied considering 5G NR design. A misdetection probability depends on various factors including DMRS configuration, false alarm target, resource allocation, and intra/inter-cell interference [17].

Reception

The UEs and all the base stations are fully synchronized. This is important, specially when interference rejection and combining techniques are applied

in the receiver, which can take into account intra- and inter-cell interference signals for computing the interference covariance matrix. It is assumed that the base station can perform accurate channel estimation for the uplink signals from the users cell and neighbors cells. Based on that, the receiver can project the desired signal in a subspace which minimizes the minimum-mean square error upon the presence of multiple interferers [18]. Accurate channel estimation can be achieved in quasi-static scenarios where the channel conditions have low variability, which is the case for the considered URLLC deployments. Network synchronization is obtained, for example, through backhaul connection between the base stations [19]. While the UE synchronization relies on primary and secondary synchronization signals [20], and updated time alignment received from the base station.

References

[1] 3GPP TR 38.913 v14.1.0, "Study on Scenarios and Requirements for Next Generation Access Technologies," Mar. 2017.

[2] 3GPP TS 38.331 V15.4.0, "NR; Radio Resource Control (RRC) protocol specification," Jan. 2019.

[3] 3GPP TS 38.300 V15.4.0, "NR; NR and NR-RAN Overall Description," Dec. 2018.

[4] P. Wu and N. Jindal, "Coding versus ARQ in Fading Channels: How Reliable Should the PHY Be?" *IEEE Transactions on Communications*, vol. 59, no. 12, pp. 3363–3374, Dec 2011.

[5] 3GPP TS 23.501 V15.4.0, "System Architecture for the 5G System," Dec. 2018.

[6] 3GPP TR 36.881 v14.0.0, "Study on latency reduction techniques for LTE," Jul. 2016.

[7] P. Popovski, J. J. Nielsen, C. Stefanovic, E. d. Carvalho, E. Strom, K. F. Trillingsgaard, A. S. Bana, D. M. Kim, R. Kotaba, J. Park, and R. B. Sørensen, "Wireless Access for Ultra-Reliable Low-Latency Communication: Principles and Building Blocks," *IEEE Network*, vol. 32, no. 2, pp. 16–23, Mar. 2018.

[8] H. Shariatmadari and S. Iraji and R. Jäntti and P. Popovski and Z. Li and M. A. Uusitalo, "Fifth-Generation Control Channel Design: Achieving Ultrareliable Low-Latency Communications," *IEEE Vehicular Technology Magazine*, vol. 13, no. 2, pp. 84–93, Jun. 2018.

[9] G. Pocovi, H. Shariatmadari, G. Berardinelli, K. Pedersen, J. Steiner, and Z. Li, "Achieving Ultra-Reliable Low-Latency Communications: Challenges and Envisioned System Enhancements," *IEEE Network*, vol. 32, no. 2, pp. 8–15, Mar. 2018.

[10] H. Shariatmadari, Z. Li, S. Iraji, M. A. Uusitalo, and R. Jäntti, "Control Channel Enhancements for Ultra-Reliable Low-Latency Communications," in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2017.

[11] 3GPP TR 38.802 v14.0.0, "Study on New Radio Access Technology," Mar. 2017.

[12] I. Parvez, A. Rahmati, I. Guvenc, A. I. Sarwat, and H. Dai, "A Survey on Low Latency Towards 5G: RAN, Core Network and Caching Solutions," *IEEE Communications Surveys Tutorials*, vol. 20, no. 4, pp. 3098–3130, May 2018.

[13] J. Lee, E. Tejedor, K. Ranta-aho, H. Wang, K. Lee, E. Semaan, E. Mohyeldin, J. Song, C. Bergljung, and S. Jung, "Spectrum for 5G: Global Status, Challenges, and Enabling Technologies," *IEEE Communications Magazine*, vol. 56, no. 3, pp. 12–18, Mar. 2018.

[14] A. A. Zaidi, R. Baldemair, H. Tullberg, H. Bjorkegren, L. Sundstrom, J. Medbo, C. Kilinc, and I. D. Silva, "Waveform and Numerology to Support 5G Services and Requirements," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 90–98, Nov. 2016.

[15] J. Sachs, G. Wikstrom, T. Dudda, R. Baldemair, and K. Kittichokechai, "5G Radio Network Design for Ultra-Reliable Low-Latency Communication," *IEEE Network*, vol. 32, no. 2, pp. 24–31, Mar. 2018.

[16] 3GPP TR 22.804 v2.0.0, "Study on Communication for Automation in Vertical Domains," May 2018.

[17] R1-1901330, "Summary of 7.2.6.3 Enhanced configured grant PUSCH transmissions," Jan. 2019.

[18] F. M. L. Tavares, G. Berardinelli, N. H. Mahmood, T. B. Sørensen, and P. Mogensen, "On the Potential of Interference Rejection Combining in B4G Networks," in *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*, Sep. 2013.

[19] D. Bladsjo, M. Hogan, and S. Ruffini, "Synchronization aspects in lte small cells," *IEEE Communications Magazine*, vol. 51, no. 9, pp. 70–77, Sep. 2013.

[20] P. Wang and F. Berggren, "Secondary Synchronization Signal in 5G New Radio," in *2018 IEEE International Conference on Communications (ICC)*, May 2018.

Part II

Radio Access for Uplink URLLC

Radio Access for Uplink URLLC

1 Problem Description

As discussed in the previous chapter, achieving reliable communication in uplink has specific challenges, not present in downlink. This part of the thesis focuses on the issues related to the radio access procedures used by the user equipments (UEs) for transmitting a packet with low latency and high reliability in uplink. In line with the Ultra-Reliable Low-Latency Communications (URLLC) service requirements, the adopted target is delivering a 32 bytes payload over the radio interface with 99.999% reliability within 1 ms [1].

The usual grant-based procedure, though is very flexible and efficient for high data rate use cases, requires very reliable control channels and scheduling opportunities available at any instant, which lead to a large overhead. Semi-persistent scheduling (SPS) has been presented as a grant-free solution for low latency communication given its potential for reducing the signaling overhead and delays caused by request/grant procedures. The base station can preallocate either dedicated resources per-UE or shared resources per group of UEs. The principles of SPS are revisited here, taking into account recent enhancements for latency reduction, such as low periodicity and short transmission time interval (TTI) as described in [2]. The standard hybrid automatic repeat request (HARQ) mechanism for SPS relies on grant signaling for re-scheduling failed transmissions. Hence, retransmissions still depend on fast and reliable control signaling, which might not be available.

In this work the preallocation of resources is considered for initial transmissions, retransmissions or both. The first part of this chapter, including Paper A and Paper B, studies the option of preallocating resources to a group of UEs for retransmissions, in order to reduce the dependence of control signaling. And the second part of the chapter considers the case in which shared resources are configured for the initial transmission. The latter case is partic-

ularly relevant for 5G NR standardization for the support of sporadic URLLC traffic.

The traffic type is determinant for the choice of resource allocation and transmission scheme. For deterministic/periodic traffic, the application of SPS mechanisms using dedicated resources for initial transmissions is beneficial. On the other hand, for sporadic/aperiodic traffic, resources should be shared for improving the resource utilization, leading however to potential collisions. The pure slotted ALOHA access procedure is not suitable for URLLC due to its low reliability under collision channel and delays for collision resolution [3]. Robust modulation and coding using large bandwidth, and receiver with multi-user detection capabilities should be employed for coping with the degradation on users signals caused by mutual interference. And also in this case, the potential of retransmissions mechanisms has to be considered for guaranteeing a reliable reception.

The feasibility of grant-free transmissions over shared resources requires detailed evaluation. In Paper C, the performance of different retransmission mechanisms for grant-free URLLC are addressed, namely K-repetitions, reactive HARQ and proactive repetitions with early termination. This part of the study is conducted using detailed system level simulations, to capture the dynamics of a realistic multi-user multi-cell network. An urban macro scenario is utilized. This scenario is important, considering that the first URLLC deployments should occur on sites where the cellular operators have existing infra-structure. Further evaluation on K-repetitions with frequency hopping is given in the appendix Paper I.

2 Objectives

This part of the work has, in summary, the following objectives:

- Study radio access solutions for uplink URLLC, investigating the potential of grant-free as an alternative to grant-based procedures.
- Investigate mechanisms for reducing the dependence of control signaling for transmission/retransmissions and improve the performance of URLLC in uplink.
- Evaluate the feasibility of grant-free transmissions of sporadic traffic over shared resources, and compare different retransmission mechanisms in terms of outage performance.
- Design and assess the performance of the uplink URLLC solutions at system level considering New Radio (NR) evaluation assumptions.

3 Included Articles

The following papers form this part of the thesis:

Paper A. Pre-scheduled Resources for Retransmissions in Ultra-Reliable and Low Latency Communications

This paper considers the usage of SPS, in order to reduce the dependence on control signaling for URLLC, and presents a scheme which relies on preallocated resources for retransmissions. The initial transmissions use dedicated resources, as in standard SPS. While the retransmission resources are also preallocated and shared by a group of UEs to avoid resource wastage in case the retransmission occurrences are very rare. Only non-acknowledged UEs contend for the retransmission resources. The scheme is compared with a conservative allocation method for avoiding retransmissions, counting on robust encoding for achieving the reliability target in a single-shot. The analyses are based on a semi-analytical approach using probabilistic models and link level performance curves.

Paper B. A Blind Retransmission Scheme for Ultra-Reliable and Low Latency Communications

This paper addresses the problem of reducing the impact of signaling and retransmission delays for URLLC using SPS. A scheme using preallocated resources for blind retransmissions (repetitions) is proposed. As in paper A, the resources used for blind retransmissions are shared by group of UEs. Here, however, the scheme counts with successive interference cancellation (SIC) for removing earlier decoded replicas. Since the initial transmissions occur in reliable dedicated resources, most of the interference can be canceled from the shared pool. The performance is also compared analytically with single-shot transmissions, and with feedback based retransmissions.

Paper C. System Level Analysis of Uplink Grant-Free Transmission for URLLC

In this paper, grant-free procedures for URLLC users transmitting sporadic traffic are studied. The grant-free resources are shared by multiple UEs. Different proposed retransmission schemes are evaluated through detailed system level simulations. A multi-cell synchronous network is considered using a 3D urban macro scenario. A minimum mean square error (MMSE) receiver with two antennas and interference rejection combining (IRC) capability is assumed in the base station. The utilized mathematical models and simulation methodology are based on third generation partnership project (3GPP) evaluation assumptions [4].

4 Main Findings and Recommendations

Preallocation mechanisms for transmissions and retransmissions

Employing pre-scheduling mechanisms for the initial transmissions, as in SPS, is naturally beneficial for URLLC, specially for deterministic traffic. Besides, preallocating resources also for retransmissions further reduces the dependence on control signaling. The resource efficiency can be improved by sharing the reserved retransmission resources between groups of UEs.

The analyses in Paper A show that this method is up to 28% more resource efficient than using conservative transmissions targeting 10^{-5} without retransmissions. Instead, an initial BLER in the order of 10^{-3} is targeted and the shared retransmission resource is used for reducing the failure probability to 10^{-5}. However, reasonable gains can only be achieved when more than 8 UEs can be grouped, or when the retransmission resources can be reallocated when no initial transmission fails. These restrictions can limit the applicability of the scheme.

The scheme proposed in Paper B, has the advantage of not depending on a feedback signaling for determining the usage of the retransmission resources. The transmission latency is therefore lower, not being affected by the round-trip time (RTT). Also in this case, resource efficiency gains higher than 20% are achieved only for groupings of 10 UEs or more. Besides, the usage of SIC receiver in the latter scheme adds more complexity to the system design. The preallocation of dedicated resources for initial transmissions subsumes deterministic traffic. However, the analyses can be extended for aperiodic traffic by using the initial BLER value as a packet arrival probability times the failure probability of the initial transmission.

Grant-free transmissions for sporadic traffic

The usage of grant-based scheduling procedures for sporadic URLLC transmissions is unfavorable not only due to the need of robust control signaling, as discussed in Part I.1, but also due to the delays caused by the scheduling request and grant processing. Considering the assumptions utilized in Paper C, with short TTI of 0.143 ms and processing time taking this same duration, even if an error free scheduling request opportunity is assumed to be available at every TTI, only one transmission can be issued within the 1 ms latency deadline. This leaves no room for potential retransmissions, which are important for resource efficiency and reliability. In addition, the scheduling process incurs higher queuing delays, while the base station coordinates each UE to transmit in orthogonal resources.

Grant-free transmission in uplink are facilitated by the use of low modulation and coding scheme (MCS) orders and linear receivers with multi-

user detection and interference rejection capabilities, such as MMSE-IRC. With grant-free procedures, the initial transmission can be received before 0.5 ms with the considered assumptions. This allows employing retransmission mechanisms for improving the reliability, making grant-free a more attractive option for use cases with tight latency constraints.

The results from Paper C show that, in comparison with K-repetitions and proactive repetitions with early termination, grant-free with reactive HARQ retransmissions achieves higher URLLC loads (approximately 400 packets per second per cell in average). This is due to the lower usage of the data channel, similarly to grant-based procedures. Efficient reactive HARQ is possible due to the short RTT duration with mini-slots. K-repetitions on its hand does not depend on the feedback, being able to achieve the reliability requirement with lower latency under low load conditions (approximately 100 packets per second per cell in average). The potential of the proactive scheme is limited by the RTT since it cannot avoid unnecessary repetitions before receiving the feedback. A drawback of repetition schemes is the increased queuing, since the effective load through the UE transmission buffer scales with the number of configured repetitions.

It is important to point out that, for use cases in which the user plane latency requirement can be relaxed to e.g. 2 ms, and under the assumption that a reliable control channel is available, the grant-based procedure should be preferable. This is because it can achieve the reliability requirement with less usage of the data channel, being more resource efficient.

System level performance in multi-user multi-cell network

The simulations conducted for the urban macro scenario show that, with MCS QPSK1/8 and 2-antenna MMSE-IRC receiver, the URLLC requirement can be achieved for the outdoor deployment. Even though, retransmissions should be enabled for obtaining combining gain and the required level of diversity to reach the 10^{-5} outage probability. UEs in the cell edge or suffering high path loss tend to be power limited, not being able to achieve the required SINR for reliable decoding. Thus, the obtained URLLC performance is achievable when the UEs are not subject to outdoor to indoor penetration losses.

For grant-free, the multi-user detection capability has an important role. The MMSE-IRC can suppress an interfering signal up to its degree of freedom, for decoding simultaneous transmissions. And the low MCS adds a tier of protection for cases in which the desired signal post-processing SINR is very degraded by collisions and fading (e.g. transport blocks using QPSK1/8 can be reliably decoded in until -5 dB SINR). Accurate channel estimation and full network synchronization are required for reaching the maximum performance with the MMSE-IRC receiver. Further improvements for

the URLLC system performance through radio resource management (RRM) mechanisms, as power control and MCS selection methods, should be investigated.

Main recommendations

Based on the findings, the following recommendations are drawn:

- For deterministic traffic, pre-scheduling dedicated resources for initial transmissions, thus avoiding control overhead for URLLC, should be preferable.
- Retransmission resources should be preallocated and shared by groups of a least 8 UEs for improving the resource efficiency, while further reducing the dependency of control signaling for retransmission.
- For sporadic traffic, grant-free access using shared resource should be allocated using robust MCS (e.g. QPSK 1/8) and with multi-user detection receiver with interference rejection capability, such as MMSE-IRC.
- Grant-free with reactive HARQ retransmissions should be employed for achieving higher URLLC loads.
- Grant-free with K-repetitions can be employed in case of low URLLC loads for lower latency requirements (<1 ms).
- Grant-based procedure should be preferable in case the latency requirement is relaxed, e.g. to 2 ms, if reliable control channel is available.

References

[1] ITU-R, "Report ITU-R M.2410-0 - Minimum requirements related to technical performance for IMT-2020 radio interface(s)," International Telecommunication Union (ITU), Tech. Rep., Nov. 2017.

[2] 3GPP TR 36.881 v14.0.0, "Study on latency reduction techniques for LTE," Jul. 2016.

[3] P. Popovski, J. J. Nielsen, C. Stefanovic, E. d. Carvalho, E. Strom, K. F. Trillingsgaard, A. S. Bana, D. M. Kim, R. Kotaba, J. Park, and R. B. Sørensen, "Wireless Access for Ultra-Reliable Low-Latency Communication: Principles and Building Blocks," *IEEE Network*, vol. 32, no. 2, pp. 16–23, Mar. 2018.

[4] 3GPP TR 38.802 v14.0.0, "Study on New Radio Access Technology," Mar. 2017.

Paper A

Pre-scheduled Resources for Retransmissions in Ultra-Reliable and Low Latency Communications

Renato Abreu, Preben Mogensen, Klaus Pedersen

The paper has been published in the
2017 IEEE Wireless Communications and Networking Conference (WCNC)

Abstract

The fifth generation (5G) cellular network demands new solutions to meet, in an efficient way, the stringent targets for ultra-reliable and low latency communication (URLLC), such as 1-10^{-5} reliability within 1 ms. In a wireless system, the control signaling of the scheduling process is also a source of errors and delays. Semi-persistent scheduling (SPS) is an option to reduce the signaling, leading to lower latency and improved transmission reliability. However, conventional SPS still applies grant signaling to schedule the retransmission. In this work it is proposed an alternative scheme in which a group of users shares a pre-scheduled resource for retransmission. The benefit is that it provides a retransmission opportunity without needing a scheduling control information. Besides that, if the pre-scheduled resource can not be reallocated, the sharing mechanism avoids excessive capacity loss. It is demonstrated through a simple analytical model that, for right grouping sizes and initial transmission error rates, the target error probability e.g. 10^{-5} can be achieved. It is also shown that the suggested scheme can provide improved resource efficiency compared to a single conservative transmission which also avoids re-scheduling.

1 Introduction

The possibilities opened for the mission critical communication with ultra-reliable and low latency communication (URLLC) in fifth generation (5G) networks, may bring a big amount of novel applications for new markets. Some examples are wireless industry automation, vehicle-to-everything communication (V2X) and remote tactile control [1]. At the same time, big challenges emerge to achieve the stringent requirements needed in these contexts, e.g. 1-10^{-5} reliability within 1 ms and average user plane latency of 0.5 ms [2].

Many applications demand low latency and reliable transmissions of predictable traffic. For instance, machines remotely controlled via Tactile Internet with real-time, synchronous and haptic feedback [3]; and V2X, with broadcast of periodic awareness information in form of Cooperative Awareness Messages [4]. Such machine type communication can generate a significant amount of small packets by a large number of user equipments (UEs). Dynamically scheduling this kind of data at each transmission time interval (TTI) would cause an excessive control signaling overhead. And this, besides being a bottleneck in terms of capacity, is also a source of errors and delays.

Semi-persistent scheduling (SPS) was introduced in LTE standard to support VoIP services, solving the problem of the tight delay requirement for small periodic traffics and the scarcity of control channel resources [5]. In SPS, resources are pre-scheduled with a certain periodicity, to avoid the overhead caused by multiple assignment/grant messages. Recently, SPS has gained more attention in the context of latency reduction considering short-

ened TTIs and periodicities. It can specially benefit the uplink, as the scheduling request and grant process can be skipped [6]. For URLLC, errors in the data and in the control channels should be strictly avoided in order to meet the tight requirements. In that sense, SPS can bring extra benefits, not only by reducing latency but also the role of the control channel as an error source [7].

The drawback of pre-scheduling is that, typically, the reserved resources can not be used by other UEs, limiting the resource utilization. For URLLC, which requires a very robust transmission, the cost in terms of resources can be very high, specially in bad coverage conditions. So, employ a data retransmission scheme like hybrid automatic repeat request (HARQ) is important to enhance the resource efficiency [8]. Otherwise, a large amount of resources needs to be reserved for each pre-scheduled cycle, for a conservative transmission.

The conventional SPS includes a persistent scheduling for the initial (first) transmission and a dynamic scheduling for the retransmissions (re-scheduling) [9]. For URLLC it may be desired to avoid also the signaling for the re-scheduling due to the possible errors in the control channel. Besides that, extra-latency can be caused by the late re-scheduling in high loaded scenarios and by the grant processing itself.

This paper presents an alternative scheme to provide HARQ retransmission opportunity for URLLC. The basic idea is to have a pre-scheduled resource for retransmission which is shared by a group of UEs. This way, the control signaling used to re-schedule the transmission when it does not succeed, can be suppressed. At the same time, with the sharing of the reserved resource, excessive capacity loss can be avoided. A model for the system is presented to show how the transmission success probability varies depending on the dimensioning of the group and on the initial transmission error rate. The resource efficiency of the system is finally compared with a conservative method that uses a robust modulation and coding scheme (MCS), targeting 10^{-5} error probability in a single transmission (which also avoids re-scheduling).

The rest of the paper is organized as follows: Section 2 describes the concept of the proposed scheme. Section 3 presents the system model and the main assumptions. Section 4 shows the numeric evaluation regarding the reliability and resource efficiency. Section 5 finalizes with the main conclusions of this work.

2 Shared Retransmission Scheme

The basic principle of the shared retransmission opportunity for a group of UEs is illustrated in Fig. A.1.

In the proposed scheme the base station (BS) should group and coordinate

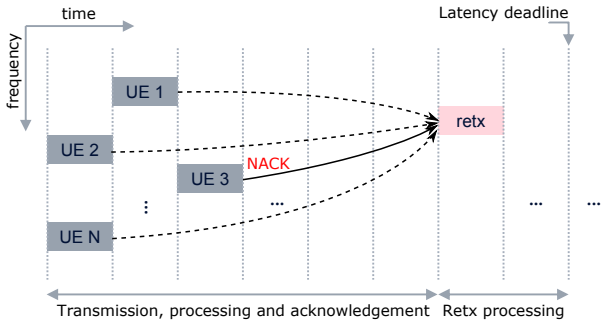


Fig. A.1: Pre-scheduled retransmission opportunity shared by UE 1 to UE N.

the UEs with similar traffic characteristics, and configure them to contend for a shared retransmission resource if the initial transmission fails. The grouping and the allocation should aim at a better resource utilization than a conservative transmission. At the same time, it should have a low probability of contention for the retransmission opportunity in order to achieve the target success probability.

The time location of the retransmission resource should allow that the transmission and decoding of the packet is concluded within the latency deadline (the maximum time for a packet to be delivered successfully in the receiver side). It is worth to notice that the initial transmissions of all UEs may not necessarily be aligned in time, as long as the processing and acknowledgment of all transmissions finishes before the reserved retransmission moment. Furthermore, transmitting in different TTIs can permit to accommodate the data packets of UEs in poor channel conditions in the available band during a TTI. Another advantage is to uncorrelate possible errors caused by sudden interference on the grouped UEs.

Both the dedicated resources for the initial transmission and the retransmission resources are pre-scheduled including a certain periodicity according to the traffic pattern. So, retransmissions occur as a synchronous HARQ, at fixed time-intervals. The pre-scheduling configuration can be made through radio resource control (RRC) signaling protected by automatic repeat request (ARQ), like in SPS, so the potential errors on the control channel can be neglected.

The main idea is that, if the initial transmission in the dedicated resource is not decoded, the shared resource can be used for one of the UEs in the group, e.g. UE3 in Fig. A.1. A possible implementation in the downlink case is, if more than one UE does not acknowledge on initial transmission, the BS decides to which one it will retransmit on the reserved resource. Only the selected UE can decode the data, while the others will not be able to decode that retransmission resource. In the uplink, the BS can solve the contention by

issuing a simple 1-bit signal, or a NACK, only to the UE that should use the retransmission resource. So the collision is avoided in case the retransmission is demanded for more than one UE. This procedure is not susceptible to the granting errors of dynamic re-scheduling because the selected UE knows, from the initial configuration, the time-frequency allocation for the retransmission. Here it is considered that, if the initial transmission fails and the UE does not get the retransmission, the packet is dropped. This is the worst case, considering that there is no available resource, reliable control or time budget for a re-scheduling. The remaining issue is to know how the contention based access to the retransmission resource can provide sufficient reliability.

3 System Model

In this section it is presented a model to estimate the success probability according to the number of UEs in a group and their transmission error probabilities. A formulation for the inherent boundaries of the system is also shown.

A single retransmission opportunity for the group of N UEs during each transmission cycle is considered. This is a reasonable assumption in the context of URLLC since the tight latency requirement may not allow multiple retransmissions. The initial transmission of each UE can randomly fail, then requiring the retransmission. This can be modeled like a Slotted ALOHA process [10] in which the probability of each UE to contend for the retransmission resources, i.e. contention based retransmission, is the probability of failing in the initial transmission P_1. Here, it is assumed that all UEs in the same group have the same error probability target. The probability of the reserved retransmission resources to be idle is given by

$$P_{idle} = (1 - P_1)^N, \tag{A.1}$$

while the probability of the resource to be required for a single UE is written

$$P_{single} = \binom{N}{1} P_1 (1 - P_1)^{N-1}. \tag{A.2}$$

Finally, the probability that the retransmission resource is required for more than one UE is simply obtained as

$$P_{collision} = 1 - P_{single} - P_{idle}. \tag{A.3}$$

In case the retransmission is demanded for more than one UE, the BS can decide which of them gets the reserved resource (the "winner"). So, assuming that each UE has an equal chance to win, the probability of having the packet

successfully decoded is then given by

$$P_{success} = (1 - P_1) + P_1(1 - P_2) \sum_{n=1}^{N} \binom{N-1}{n-1} (P_1)^{n-1}(1 - P_1)^{N-n}(1/n), \quad \text{(A.4)}$$

where P_2 is the error probability in the retransmission. It is worth noting that the probability of a grant/assignment error, typical of a dynamic re-scheduling scheme, does not appear in equation (A.4). That is basically replaced by another term that considers the contention for use the retransmission resource, which is the summation term in (A.4). This term depends mainly on the error probability of the first transmission and on the grouping size N. It sets boundaries on the success probability, independent of the error probability of the retransmission (i.e. $0 \leq P_2 \leq 1$), which are written

$$(1 - P_1) \leq P_{success} \leq (1 - P_1) + P_1 \sum_{n=1}^{N} \binom{N-1}{n-1} (P_1)^{n-1}(1 - P_1)^{N-n}(1/n). \quad \text{(A.5)}$$

So, there is a clear trade-off between the number of UEs in the group and the maximum success probability. It is important to point out that, for the sake of simplicity to present the main idea, the feedback errors were omitted in the model. However such errors impacts the final success probability of the system, requiring a lower error target on transmissions or smaller groupings, to be compensated.

4 Performance Analysis

In order to achieve a certain final success probability with the described scheme, the objective is to find the number of UEs that can be grouped and the required success probability for the initial transmission. After that, it is important to quantify the resource efficiency when applying the proposed procedure. A fair comparison can be made with a single conservative transmission, which also does not require a re-schedule signaling, but spends a large amount of resources aiming to succeed with one transmission.

4.1 Grouping and reliability evaluation

For finding the number of UEs that can be grouped under a certain initial block error rate (BLER, taken as the transmission error probability), the BLER on the retransmission (after the soft combining) is fixed to 10^{-5}, to match with the baseline reliability of the 5G access technologies [2]. Fig. A.2 shows the

final error probability $(1 - P_{success})$ according to the first BLER for different number of UEs grouped to share the retransmission opportunity. It can be

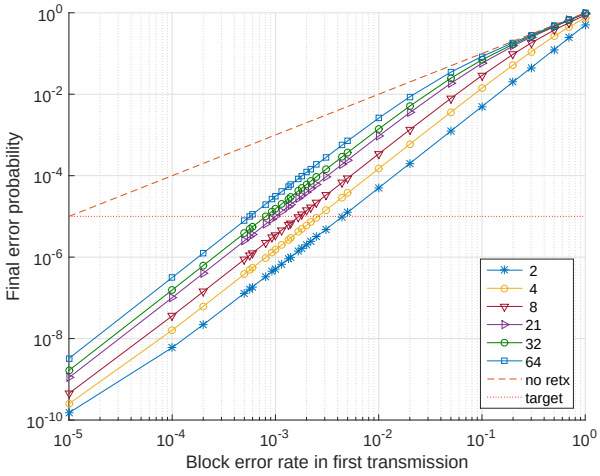


Fig. A.2: Reliability according to the first BLER for N UEs.

seen that, for instance 21 UEs can be grouped to share one retransmission opportunity when the initial BLER is 10^{-3}. That UEs can still achieve the final target error probability of 10^{-5}, without needing a control signal to re-schedule eventual retransmissions. It can be noticed also that, the higher the number of UEs is a group, the lower should be the BLER on the initial transmission to achieve the target error probability. Since the minimum grouping size is 2, the maximum BLER allowed for the initial transmission to achieve the final error probability of 10^{-5}, is 4.4×10^{-3}. As stated before, instead of a granting error probability in equation (A.4), there is a summation term which accounts for the probability of winning the retransmission opportunity in case of contention. The complement of that, which is the probability of not getting the retransmission opportunity, is given by

$$P_{notwin} = 1 - \sum_{n=1}^{N} \binom{N-1}{n-1} (P_1)^{n-1} (1 - P_1)^{N-n} (1/n). \tag{A.6}$$

These probabilities are shown for different number of UEs in Fig. A.3. The dashed line (limit) represents the maximum value for P_{notwin} in order to achieve less than 10^{-5} final error probability. That is equivalent to the maximum error probability required for the granting in a dynamic re-scheduling scheme. The proposed scheme can operate within the target reliability if the number of UEs in the group and the initial BLER are in the region below the limit line. Taking the intersections with the limit line, the maximum number of UEs at each initial BLER condition can be extracted as shown on Fig. A.4.

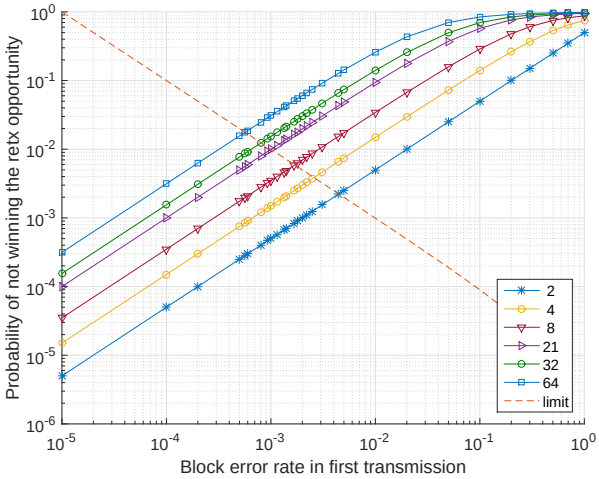


Fig. A.3: Probability of not winning on contention for retransmission.

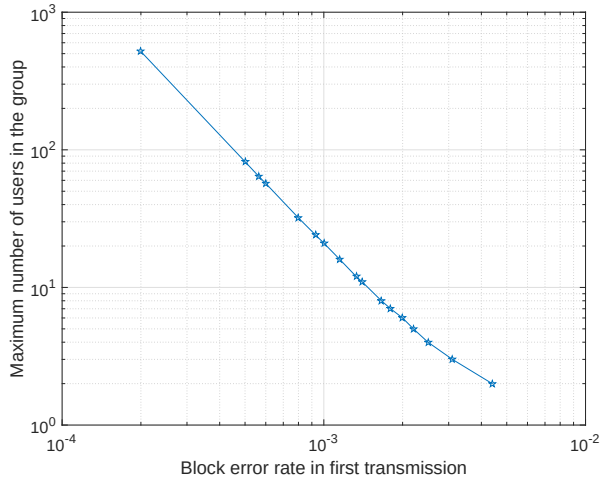


Fig. A.4: Maximum grouping according to the first BLER for 10^{-5} error target.

4.2 Resource efficiency evaluation

This section shows an estimation of the resource efficiency gain, when comparing the scheme with shared retransmission opportunity against a conservative transmission.

A link abstraction model was used to derive the coding rate needed to achieve each required BLER, when transmitting a packet of 256 bits at a certain signal-to-noise ratio (SNR). Typical modulation orders were assigned to each SNR interval like: QPSK from -10 to 0 dB, 16QAM from 0 to 5 dB,

64QAM from 5 to 10 dB and 256QAM from 10 dB onwards. The model was obtained considering turbo codes, which is one of the coding schemes proposed for URLLC that has presented better performance for block sizes of 200 bits onwards [11]. Fig. A.5 shows some example performance curves of the model for an additive white Gaussian noise (AWGN) channel. It can be

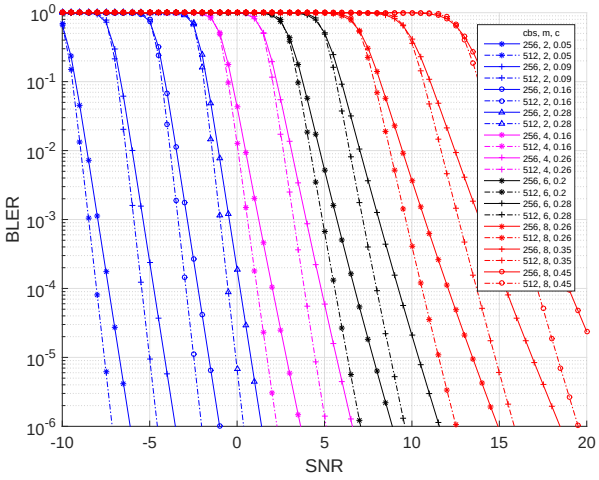


Fig. A.5: Example of performance curves from the link abstraction model for different code block sizes (cbs), modulation orders (m) and coding rates (c).

noticed in Fig. A.5 that, for small packets like 256 bits (baseline packet size for URLLC evaluation [2]), the curves are not as steep as for larger packets, so the modulation and coding rate requirements are more sensible to changes on the BLER target.

To account for the resource utilization, the number of used resource elements per information bit is considered. For a conservative transmission, i.e. without a retransmission opportunity, it is written

$$\phi_c = \frac{1}{r_c(1-P_c)}\,, \tag{A.7}$$

where r_c is the transmission rate utilizing a conservative modulation order (m) and coding rate (c) to achieve the required success probability, i.e. $r_c = m \times c$; and P_c is the error probability, which should be the target BLER itself, considering ideal link adaptation.

For the proposed scheme, the required resources per bit can be simply given by the resources on the first transmission ϕ_1, which is less conservative, and the shared resources divided by N UEs ϕ_2, so

$$\phi_s = \phi_1 + \phi_2 = \frac{1}{r_1(1-P_1)} + \frac{1}{r_2(1-P_2)N}\,, \tag{A.8}$$

where r_1 and r_2 are the transmission rates for the initial and for the retransmission, respectively. For simplicity of the analysis, it is assumed that the grouped UEs have similar channel conditions, requiring the same MCS. It is also assumed that the MCS for the retransmission is equal to the initial transmissions (i.e. $r_1 = r_2$). With this, it was verified using the link model (from -10 to 10 dB SNR) that, with the soft combining providing 3 dB gain, the retransmission error probability is lower than the target, in this case 10^{-5}.

Efficiency gain without resource reallocation

Fig. A.6 shows the gains in resource efficiency when comparing the scheme with shared retransmission opportunity against the conservative single initial transmission, that is ϕ_c/ϕ_s. Here it is first considered that, if all the initial transmissions are acknowledged, the reserved retransmission resource is wasted. It can be seen that, as expected, the efficiency is higher when more

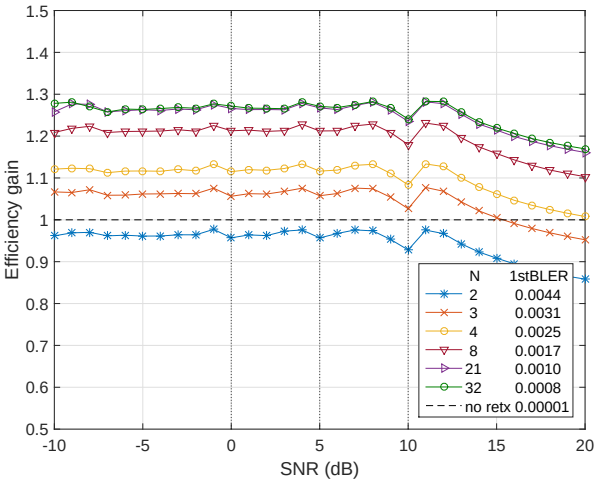


Fig. A.6: Efficiency gain for different groupings of UEs (256-bit packet).

UEs share the retransmission resources. Taking the case with initial BLER at 10^{-3}, which permits groupings of up to 21 UEs achieving the 10^{-5} reliability, it can be noticed that the shared retransmission scheme brings gains of up to 28% on resource efficiency compared to a conservative transmission. However, as shown in the previous section, larger groups demand lower BLER on initial transmission, which can be more challenging to accommodate in a TTI due to the larger amount of resources needed. It can also be observed that larger groups, e.g. greater than 21 UEs, do not provide better efficiency, since the required initial BLER become as low as for a conservative transmission.

For small groups of UEs, the gain drops since the wasting for having the

reserved retransmission resource is higher than the gain given by the relaxed initial BLER target.

The slight variations in each curve is due to the discrete changes of MCS at each SNR. On higher SNRs the efficiency gain reduces, since the MCS and success rate of the conservative transmissions become high as in the proposed scheme.

Efficiency gain considering resource reallocation

In Fig. A.7, similar resource efficiency evaluation was made, but now considering that the reserved retransmission resource can be re-allocated to a non-URLLC UE. These type of UEs, are normal mobile broadband users that do not have stringent latency and reliability requirements, so they can deal with possible errors and delays in granting procedures. In this case, since it is considered that the retransmission resource is not wasted when all the URLLC UEs succeed in initial transmission, the resources per bit is given by

$$\phi_{s'} = \phi_1 + \phi_2(1 - P_{idle}) = \frac{1}{r_1(1 - P_1)} + \frac{1 - P_{idle}}{r_2(1 - P_2)N}. \tag{A.9}$$

The re-allocation permits a better resource utilization in general since the wasting is avoided. It can be observed that, in this case, smaller groupings outperforms the bigger groupings. However, to consider that all the reserved resources of smaller groups can be reallocated, it is necessary sufficient demand from non-URLLC UEs in the network.

If there is a high traffic demand of non-URLLC UEs and low load of URLLC UEs in the network, it can be even worthy to reserve retransmission resources to each single URLLC UEs. For that case, a link adaptation scheme like in [12] could be applied for finding an efficient MCS.

It is important to note that, to apply the reallocation, there should be sufficient time budget for the base station, after the acknowledgments of the URLLC UEs, to grant the reserved resource to a non-URLLC UE.

5 Conclusion

In this paper it was proposed a scheme that employs pre-scheduling of resources shared by a group of URLLC UEs, for retransmissions. The analysis shows that, with the right dimensioning of groups and BLER target, the probability of contention for the shared retransmission can be sufficiently low. This means that the final error probability can be achieved without rescheduling procedures. The resource efficiency of the method was compared against a single conservative transmission aiming at 10^{-5} of error probability. Considering that the reserved resources are wasted when all URLLC UEs initially succeed, it can be seen that the efficiency gain is higher (up to 28% for

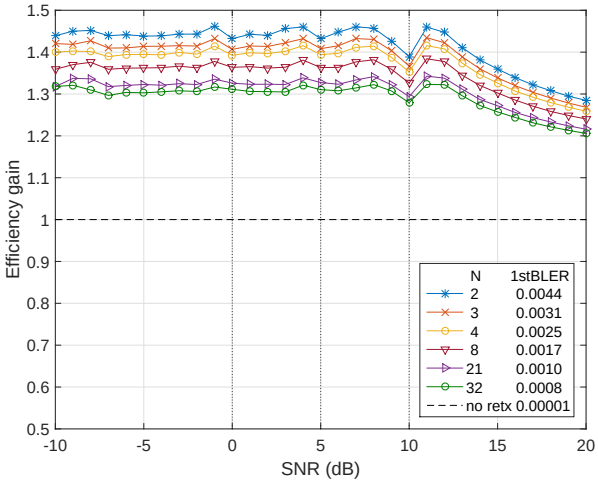


Fig. A.7: Efficiency gain considering reallocation of the retransmission resource.

256-bit packet) when more UEs are grouped. However, this requires lower initial BLER. For small groups (e.g.: 2), the wasting for having the reserved retransmission resource is higher than the gain of the relaxed initial transmission. On the other hand, when the reserved resources can be reallocated (e.g. to a non-URLLC UE), the efficiency of the proposed scheme is generally higher since the waste is avoided. Future work can consider enhancements for unpredictable traffic and simulations considering non-ideal link adaptation.

Acknowledgment

The authors would like to thank Krzysztof Bąkowski for the work on the link abstraction model.

References

[1] ITU-R M.2083-0, "IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond," Sept. 2015

[2] 3GPP TR 38.913, "Study on Scenarios and Requirements for Next Generation Access Technologies," 3GPP Tech. Rep., V14.0.0, Oct. 2016.

[3] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, G. Fettweis, "5G-Enabled Tactile Internet," IEEE Journal on Selected Areas in Communications, vol. 34, no. 3, Mar. 2016.

[4] The 5G Infrustructure Public Private Partnership. *5G PPP White Papers on Energy, Automotive, Factories, and eHealth Vertical Sectors.* [Online]. Available: https://5g-ppp.eu/white-papers/. Accessed: Aug. 20, 2016.

[5] 3GPP TS 36.321, "Medium Access Control (MAC) protocol specification," 3GPP Tech. Spec., V13.2.0, Jun. 2016.

[6] 3GPP TR 36.881, "Study on latency reduction techniques for LTE," 3GPP Tech. Rep., V14.0.0, Jun. 2016.

[7] R1-167309, "Semi-persistent scheduling for 5G new radio URLLC," 3GPP TSG-RAN WG1 #86, Aug. 2016.

[8] H. Shariatmadari, S. Iraji, R. Jäntti, "Analysis of Transmission Methods for Ultra-Reliable Communications," IEEE 26th PIMRC, Sept. 2015.

[9] D. Jiang, H. Wang, E. Malkamaki, E. Tuomaala," Principle and Performance of Semi-persistent Scheduling for VoIP in LTE System," IEEE WiCom, Sept. 2007.

[10] L. G. Roberts, "Aloha packet system with and without slots and capture," SIGCOMM Comput. Commun. Rev., Apr. 1975.

[11] M. Sybis, K. Wesołowski, K. Jayasinghe, V. Venkatasubramanian, V. Vukadinovic, "Channel coding for ultra-reliable low-latency communication in 5G systems," IEEE VTC Fall, Sept. 2016.

[12] H. Shariatmadari, Z. Li, M. A. Uusitalo, S. Iraji, R. Jäntti, "Link adaptation design for ultra-reliable communications," IEEE ICC, May 2016.

Paper B

A Blind Retransmission Scheme for Ultra-Reliable and Low Latency Communications

Renato Abreu, Gilberto Berardinelli, Thomas Jacobsen, Klaus Pedersen, Preben Mogensen

The paper has been published in the
IEEE 87th Vehicular Technology Conference (VTC Spring)

Abstract

This work is related to 5G new radio concept design, with focus on ultra-reliable and low latency communication (URLLC) use cases. We mainly target to achieve the stringent latency and reliability requirements for transmissions over the air interface, such as 99.999% success probability within 1 ms. Meeting these requirements in an efficient way, that is, without draining the network capacity is one of the main challenges for the new radio standardization. In this work, we propose a scheme to perform blind retransmissions on shared radio resources together with the application of successive interference cancellation to receive remaining non-decoded data with low delay penalty. The method avoids control errors and extra delays existent on feedback-based retransmission schemes. The investigations also show that blind retransmission on shared resources is more resource efficient than a conservative single shot transmission, depending on the number of users sharing the resources.

1 Introduction

The advent of ultra-reliable and low latency communication (URLLC) for mission critical applications in cellular networks brings new challenges due to specific characteristics of these systems, such as tight delay and reliability tolerances (e.g. $1 - 10^{-5}$ within 1 ms) and in some cases, infrequent small data traffic [1]. URLLC requires a careful redesign of technology components such as radio numerology, frame structure, scheduling and transmission protocols [2]. Acknowledged transmission mechanisms suffer from inherent delays due to the round trip time (RTT) of the feedback signaling, impacting negatively the latency distribution and potentially jeopardizing the possibility of coping with the URLLC target. Besides that, errors can occur either in the decoding of the feedback or grant signaling messages, affecting the reliability of system [3].

Semi-persistent scheduling (SPS) was extended in LTE Release-14 for faster uplink (UL) access reducing the overhead caused by the request/grant procedures. For unpredictable data traffic, pre-scheduled allocation could result in wasting of radio resources in case user equipment (UE) has no data available for transmission. So, it was proposed that SPS resources could be shared by multiple UEs [4]. In the case that more than one UE transmit at the same time in the shared resources, a collision happens and the base station (BS) may not decode the data. So, the collision should be detected in order to arrange a retransmission of the data of each UE. This can result in an extended latency and compromise the application in URLLC use cases. It should be noted that retransmissions in shared resources are not supported for SPS in LTE, meaning that they should only be scheduled in dedicated resources.

The usage of a shared channel for retransmissions was considered in [5].

In that case, a shared retransmission resource is pre-scheduled to a group of UEs. If more than one UE fails on their initial transmissions, they need to content for the pre-scheduled resource. The procedure relies on a feedback signal to solve the contention for the resource.

Different multi-user detection (MUD) approaches exist to combat interference at the receiver. In conventional successive interference cancellation (SIC), the signal with large signal-to-interference-plus-noise-ratio (SINR) is decoded, reconstructed and subtracted from the aggregated signal. Subsequently, the signal with low SINR is decoded from the other signal [6]. Recently, coded random access schemes using SIC receivers have being proposed in [7]. Such techniques have the potential of boosting cell throughput but the increased average delay does not cope with URLLC requirements.

In 5G New Radio (NR), it is expected that URLLC exploits the usage of blind repetitions, in order to increase the success probability of transmitting a message with low delay penalty [8, 9]. The node just proactively retransmits for a predetermined number of attempts or until a positive acknowledgment is received, rather than stop and wait for a feedback upon each transmission. However, this method can also lead to poor resource utilization and excessive interference, since further retransmissions might not be needed if the message is already detected on the initial transmission.

In summary, retransmissions are beneficial to improve reliability but the problems are:

- blind retransmissions can drain capacity
- stop-and-wait protocols lead to a delay penalty

Hence, in this work we evaluate a scheme that permits the nodes to perform blind retransmissions with low delay penalty and improved resource utilization. A receiver that performs the cancellation of initially decoded transmissions is considered for recovering retransmissions on a shared resource pool. We describe a simple analytical model used to evaluate its success probability achieved with different configurations. We also compare its performance in terms of resource utilization and latency with other schemes.

The paper is organized as follows. Section 2 describes the proposed scheme. Section 3 formulates the system model. The performance evaluation is presented in Section 4, and the conclusions are drawn in Section 5.

2 Blind Retransmission over Shared Resources

Fig. B.1 illustrates a group of N UEs performing the initial transmission on dedicated resources, and the principle of sharing M resources to perform blind retransmissions in a total of T transmission attempts.

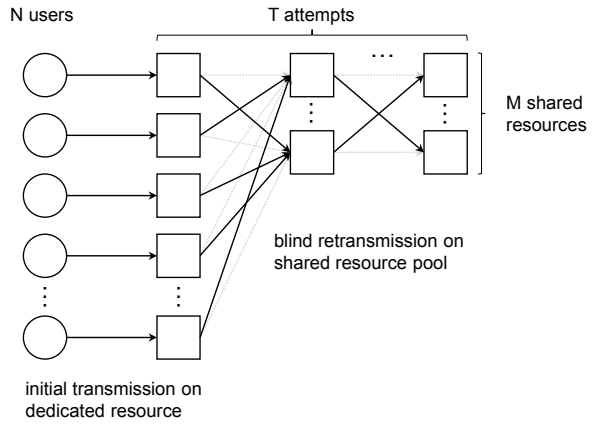


Fig. B.1: Example of M shared resources and T transmissions by N UEs.

Thought the principle could be used for both downlink and uplink, it should be more relevant in uplink, where each transmitter node might not be interested on the data decoding of the other nodes. It is important to mention that the UEs should be time and frequency synchronized in the uplink. In the proposed scheme a group of UEs perform their initial transmission on dedicated resources that can be granted or semi-statically assigned. Subsequently, the devices transmit again the same information $T-1$ times without waiting for a feedback, aiming low latency and reliability. However, instead using dedicated resources, the UEs in the group perform their repetitions using a shared resource pool, for better resource utilization. The shared resource pool can be pre-reserved and its size should be smaller than the amount of resources utilized for the dedicated transmissions ($M < N$). If the pool contains multiple resources, the one to be used for each retransmission can be predefined or randomly selected to avoid extra control signaling.

Since the UEs in the group can perform the same procedure, collision will occur during the retransmissions. Then, a successive interference cancellation (SIC) receiver is used to recover a payload that was possibly not decoded on the initial transmission. Since the initial transmission occurs in "safer" resources, most of them should be early decoded for a low initial block error rate (BLER) target. The already decoded signals can be then subtracted from the received signal in the shared resources, therefore increasing the chances of correctly retrieving the payloads whose detection had failed earlier.

Fig. B.2 illustrates the reception process. The received signal $y_{m,j}$ on a shared resource $j \in \{1, ..., M\}$ at a certain retransmission attempt is a combination of the signals from all the UEs retransmitting in there, considering

also the channel effect over each transmission stream. This can be written as

$$y_{m,j} = \sum_{i \in \Psi} h_{i,j} x_i + \sum_{i \in \Omega} h_{i,j} x_i + w \ , \tag{B.1}$$

where x_i are the signals transmitted by the UEs, $h_{i,j}$ are the channel fading coefficients of the i-th UE transmitting over the j-th resource, w denotes the Gaussian noise, Ψ is the set of indexes of the UEs whose payload was not yet decoded, and Ω is the set of the ones whose payload was decoded, and are being retransmitted over the same shared resource j. So, the receiver should be able to detect the UEs and estimate their channel responses (for instance, by assuming orthogonal reference sequences used by the different UEs) and reconstruct the signal from the previously decoded ones. After that, it cancels their interference over the non-decoded signals. That is part of the SIC decoding process. Ideally, each successfully decoded replica should permit to remove its interference in the other replicas, at each retransmission. Therefore, the successive decoding process on the shared channel resolves fast the remaining non-decoded transmissions.

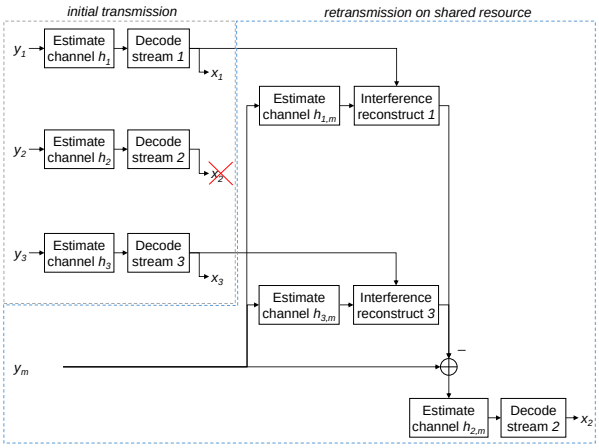


Fig. B.2: Example of reception process with shared retransmission resource.

The scheme can be summarized as follows (e.g. for an uplink transmission implementation):

1. The BS configures semi-persistent or dynamically granted resources for the UEs initial UL transmission.

2. The BS also configures the UEs to perform blind retransmissions on shared retransmission channels.

3. The UE performs the initial transmission in dedicated channel and

blind retransmissions on a shared channel according to the configuration from steps 1 and 2.

4. The BS attempts to decode the initial transmissions from the UEs in their dedicated resources and store the successfully decoded signals.
5. The BS attempts decoding the shared channel after subtracting the already decoded signals from the combined received signal.

3 Success Probability Model

To investigate the reliability achieved with the described transmission procedure we model the probability to successfully deliver a data packet. The following assumptions are considered in this study:

- Same error probability on the initial transmission P_1 for the grouped UEs.
- The decoded transmissions can be fully canceled from the shared resource.
- For a predefined pool with $M > 1$, the retransmission occurs in one randomly selected resource from the pool.
- A transmission can be decoded on shared resource in case it does not collide with other non-decoded transmission.

The probability of u UEs to fail on the initial transmission and contending on the shared resource pool with a UE of interest is given by

$$P_f(u) = \binom{N-1}{u-1}(P_1)^{u-1}(1-P_1)^{N-u}. \tag{B.2}$$

For u UEs failing on the first transmission, the probability of a UE of interest to be the only failing UE transmitting in a certain resource from the pool is

$$P_g(u) = \left(\frac{M-1}{M}\right)^{u-1}. \tag{B.3}$$

For one retransmission attempt ($T = 2$), the probability of UE transmission to be singleton, that is, the only transmission that was not yet decoded in a certain shared resource is given by

$$P_s = \sum_{n=1}^{N} P_f(n)P_g(n) = \\ = \sum_{n=1}^{N} \binom{N-1}{n-1}(P_1)^{n-1}(1-P_1)^{N-n}\left(\frac{M-1}{M}\right)^{n-1}. \tag{B.4}$$

And the final probability that a packet transmission to be successfully received can be given by

$$P_r = (1 - P_1) + P_1 P_s (1 - P_2), \tag{B.5}$$

where P_2 is the error probability in the retransmission.

In a typical feedback-based retransmission scheme, the error probability of the control signaling should be taken into account [3]. However, in the studied scheme the signaling errors do not appear. Instead, equation (B.5) considers the contention when using the shared retransmission resources, which is the probability of being singleton P_s.

4 Performance Evaluation

In this section, we present first the reliability and resource utilization analysis, and then a case study with latency evaluation. We compare the described scheme with an aggressive single shot transmission. We also consider for the sake of comparison, the feedback-based scheme in which an UL grant is needed for the retransmissions, as was recently agreed for NR in 3GPP [10].

4.1 Reliability and resource efficiency

Employing the model presented in the previous section, we first analyze the resulting failure probability for different number of UEs grouped to share the retransmission resource pool. Fig. B.3 shows the final failure probability $(1 - P_r)$ achieved. As in [3] and [5] the failure probability for any retransmission (which should be singleton in our case) is assumed to be 10^{-5} after the detection and soft-combining with the initial transmission. It is obvious that the failure probability reduces with the lower block error rate on the initial transmission. In any case, for the assumed failure probability on the retransmission, the final failure probability is lower than for a baseline case without retransmission. The initial BLER for achieving the target success probability of $1 - 10^{-5}$ is in the order of $\sim 10^{-3}$. For instance, for 16 UEs sharing 2 resources and for 8 UEs sharing 1 resource the initial BLER should be at most 1.2×10^{-3} to meet the target.

The relation between the maximum number of UEs that can be grouped and the initial BLER to achieve the target success probability for different sharing settings is illustrated in Fig. B.4. The curve for $T = 3$ transmission attempts was derived through simulation. It is obvious that the higher the number of resources in the shared pool, the higher is the number of UEs that can be supported in the group for the same initial BLER. For instance, from $M = 1$ to $M = 3$ and initial BLER of $\sim 10^{-3}$, the number of UEs sharing the pool can increase from 10 to 30.

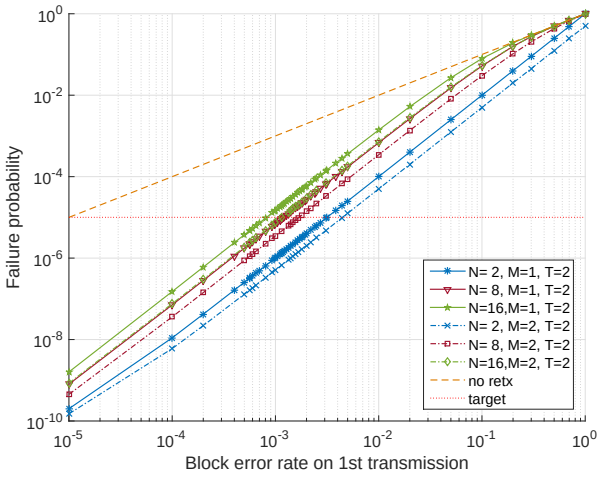


Fig. B.3: Reliability for N UEs sharing M retransmission resources and $T = 2$.

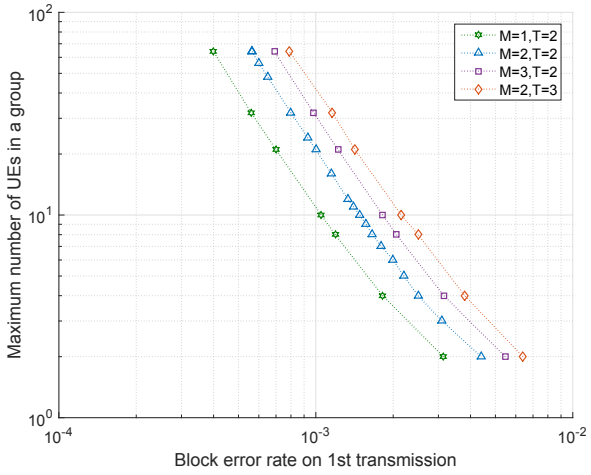


Fig. B.4: Maximum number of UEs in the group versus initial BLER.

To account for the resource utilization, it was applied the same procedure as in [5]. For the single shot transmission the used resources per bit can be calculated as

$$\phi_c = \frac{1}{r_c(1 - P_c)}\,, \tag{B.6}$$

where r_c is the rate of a robust modulation and coding scheme, considering ideal link adaptation, which gives a failure probability P_c that in this case

should be equal to 10^{-5}. The rates are obtained considering the link performance with turbo codes for the transmission of a small packet of 32 bytes.

For blind retransmissions over shared resources we calculate resource utilization as

$$\phi_s = \frac{1}{r_1(1-P_1)} + \frac{M}{r_2(1-P_2)N}\,, \tag{B.7}$$

including the resources occupied for the dedicated initial transmission and the M resources shared by N grouped UEs for the case of one retransmission attempt. The rate for the initial transmission r_1 and for the retransmission r_2 are assumed equal here.

We can calculate the resources utilized in the case of a feedback-based retransmission scheme with the following equation

$$\phi_f = \frac{1}{r_1(1-P_1)} + \frac{P_1}{r_2(1-P_2)(1-\xi)}\,, \tag{B.8}$$

where ξ is the failure probability of the feedback signal which carries the retransmission grant.

The resource efficiency of two shared retransmission configurations ($M = 1$ and $M = 2$, for $T = 2$) is compared against a transmission that targets 10^{-5} BLER in a single shot. To compare with a feedback-based retransmission scheme we assume a fixed failure probability of $\xi = 10^{-3}$ for the feedback signal. The used resources for the initial transmission and its failure probability is set to be the same as for the blind retransmission scheme with $M = 1$ for the shared pool.

Fig. B.5 shows the obtained gain in terms of bits per symbol as function of the number of grouped UEs sharing the resource pool. It can be observed that the gain for $M = 1$ is generally higher, though it requires a lower initial BLER as shown on previous figures. Also for $M = 1$, in case there are only 2 UEs sharing the resource, no gain is achieved. For $M = 2$, a gain on resource efficiency is achieved when the number of UEs sharing the pool is higher than 5. In both cases, the gain saturates at $\sim 23\%$, since a high number of UEs sharing the pool requires higher initial BLER targets which translates in lower code rates. In practice, such groups with high number of UEs can be formed, for instance, by machine-type communications devices with similar traffic characteristics and located in the same area. In a high mobility scenario, the grouping may require a more complex coordination. It can be also noticed in Fig. B.5 that the feedback-based retransmission scheme has, in general, a better resource efficiency. The difference tends to decrease when comparing to the cases where more UEs can be grouped to share the retransmission resources. And, as mentioned previously, the feedback-based scheme comes with the cost of the extra signaling. This can translate to higher latencies as will be discussed next.

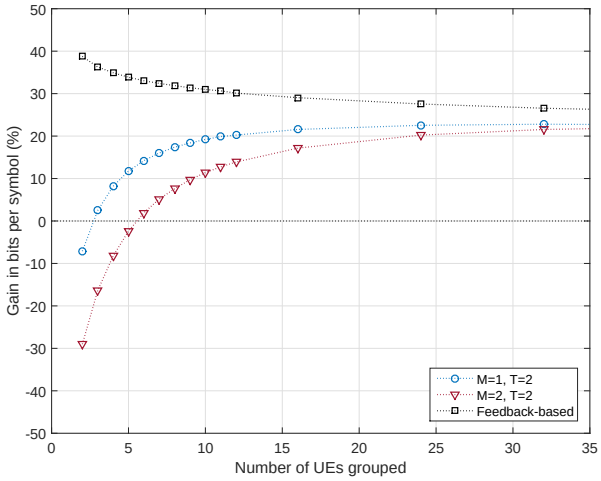


Fig. B.5: Gain on resource efficiency compared to a single shot aggressive transmission.

4.2 Case study

Here we consider a frame-based system alike LTE where the resources are arranged in a time-frequency grid and the transmissions occur in mini-slots of a few OFDM symbols (2 to 7) as considered for NR [9]. For uplink transmissions without grant, like in SPS, the latency of a packet transmission is composed by the frame alignment time, transmitter processing, propagation and receiver processing time. The alignment time is a random value from the moment a packet arrives in the transmission buffer until the beginning of the next transmission time interval (TTI). As in [11] and [12] we assume a fast processing time of 1 TTI for transmitting/receiving and also 1 TTI for processing, both in the UE and in the BS side. For the feedback-based retransmission, the HARQ round trip time should also be accounted. The value of it is scaled with the TTI duration and is considered to take 4 TTIs, matching with the time between the beginning of a transmission attempt until the end of its feedback processing. Queuing delays in the UEs transmission buffers are not considered.

Fig. B.6 shows the complementary cumulative distribution function (CCDF) of the achievable latencies in terms of TTIs, from the time a packet arrives in the transmission buffer until it is received and decoded. We can observe that the single shot transmission obviously achieves the lower latency of 3 TTIs at the 10^{-5} percentile, with the cost of low resource efficiency as discussed previously. The blind retransmissions using shared pools with $M = 1$ and $M = 2$ for 2 and 3 transmission attempts respectively, take 4 to 5 TTIs. While the feedback-based option takes 7 TTIs due to the impact of the RTT on the retransmissions.

Table B.1: Achievable latency at 10^{-5} failure probability

Example of numerology configuration	TTI size (ms)	Single shot	Feedback -based	Shared pool
60 KHz SCS, 7 symbols	0.125	0.375	0.875	0.5
15 KHz SCS, 2 symbols	0.143	0.429	1.0	0.572
30 KHz SCS, 7 symbols	0.25	0.75	1.75	1.0
15 KHz SCS, 7 symbols	0.5	1.5	3.5	2.0

Considering in particular the baseline URLLC target of $1 - 10^{-5}$ success probability within 1 ms, we show some cases on Table B.1 for different mini-slot configurations, highlighting the options that do not meet the requirement. Mini-slot durations will depend on the subcarrier spacing (SCS) and on the number of OFDM symbols for a given SCS, adopted according to the type of deployment and carrier frequency [13]. It is important to note that the assumed processing times and RTT can be optimistic for the practical NR implementation. If the RTT takes longer time (for instance 8 TTIs like is typically in LTE), then the feedback-based option would not to meet the latency constraints even for very short TTIs.

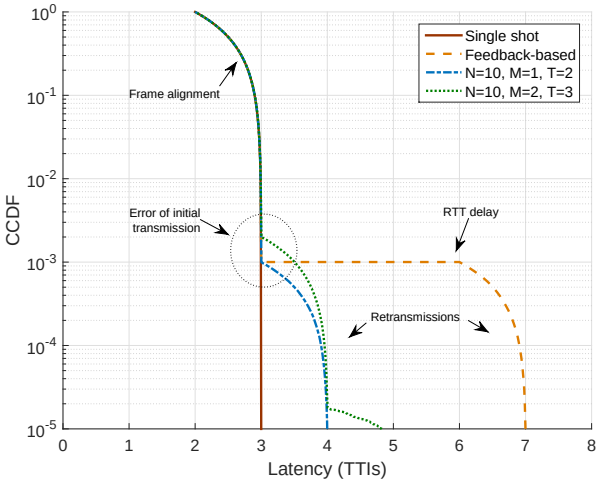


Fig. B.6: Example of latency CCDFs for different settings.

5 Conclusions

In this paper we have proposed a scheme for URLLC in which groups of UEs can use a shared resource pool to perform blind retransmissions. The scheme avoids possible errors and delays caused by feedback signaling and re-scheduling procedures for retransmission. One or more retransmission opportunities can be provided on the shared resources.

The scheme can be more resource efficient than single shot transmissions, especially when more UEs share the retransmission resources. While if the number of UEs is too large the efficiency gain saturates since the BLER for the initial transmission needs to be low. Feedback-based retransmissions have generally better resource utilization than the studied scheme, but might not be able to achieve strict URLLC targets, depending on the numerology and processing times.

The studied solution does not require extra control signaling to allow the UE to perform retransmissions. It counts with an interference cancellation receiver that should be able to reconstruct retransmissions that were previously decoded and subtract them from the received signal in the shared resources. Further, it can be beneficial to consider the performance with multi-user detection receivers which have the potential to capture multiple non-decoded retransmissions on shared resources.

Acknowledgment

This research is partially supported by the EU H2020-ICT-2016-2 project ONE5G. The views expressed in this paper are those of the authors and do not necessarily represent the project views.

References

[1] 3GPP TR 38.913, "Study on Scenarios and Requirements for Next Generation Access Technologies," 3GPP Tech. Rep., V14.0.0, Oct., 2016.

[2] K. I. Pedersen, G. Berardinelli, F. Frederiksen and A. Szufarska, "A flexible 5G frame structure design for frequency-division duplex cases", IEEE Communications Magazine, vol. 54, no. 3, pp. 53-59, March, 2016.

[3] H. Shariatmadari, Z. Li, S. Iraji, M. A. Uusitalo, R. Jäntti, "Control channel enhancements for ultra-reliable low-latency communications", IEEE ICC Workshops, May, 2017.

[4] 3GPP TR 36.881, "Study on latency reduction techniques for LTE," 3GPP Tech. Rep., V14.0.0, Jun., 2016.

[5] R. Abreu, P. Mogensen and K. I. Pedersen,"Pre-Scheduled Resources for Retransmissions in Ultra-Reliable and Low Latency Communications", IEEE Wireless Communications and Networking Conference (WCNC), March, 2017.

[6] Miridakis et. al., "A Survey on the Successive Interference Cancellation Performance for Single-Antenna and Multiple-Antenna OFDM Systems", vol. 15, no. 1, IEEE Comm. Surveys & Tutorials, 2013

[7] E. Paolini, C. Stefanovic, G. Liva, P. Popovski, "Coded Random Access: Applying Codes on Graphs to Design Random Access Protocols", IEEE Comm. Magazine, June, 2015.

[8] 3GPP TR R1-1612246, "Discussion on HARQ support for URLLC", RAN1 #87, Reno, Nevada, 2016.

[9] 3GPP TR 38.802, "Study on New Radio Access Technology," 3GPP Tech. Rep., v14.2.0, Sept., 2017.

[10] 3GPP TSG RAN WG1 Meeting #90, "RAN1 Chairmans Notes", Prague, Czech Republic, Aug., 2017.

[11] G. Pocovi, B. Soret, K. I. Pedersen, and P. Mogensen, "MAC Layer Enhancements for Ultra-Reliable Low-Latency Communications in Cellular Networks," in IEEE International Conference on Communications (ICC) 2017 Workshop, May, 2017.

[12] T. Jacobsen, R. B. Abreu, G. Berardinelli, K. I. Pedersen, P. E. Mogensen, I. Kovcs, and T. Kozlova, "System level analysis of uplink grant-free transmission for URLLC," in IEEE Globecom Workshops, Dec., 2017.

[13] A. A. Zaidi et al., "Waveform and Numerology to Support 5G Services and Requirements," in IEEE Communications Magazine, vol. 54, no. 11, pp. 90-98, Nov., 2016.

Paper C

System Level Analysis of Uplink Grant-Free Transmission for URLLC

Thomas Jacobsen, Renato Abreu, Gilberto Berardinelli, Klaus Pedersen, Preben Mogensen, István Z. Kovács, Tatiana K. Madsen

The paper has been published in the
IEEE 2017 GlobeCom Workshops

Abstract

In the context of 5'th Generation (5G) New Radio (NR), new transmission procedures are currently studied for supporting the challenging requirements of Ultra-Reliable Low-Latency Communication (URLLC) use cases. In particular, grant free (GF) transmissions have the potential of reducing the latency with respect to traditional grant-based (GB) approaches as adopted in Long Term Evolution (LTE) radio standard. However, in case a shared channel is assigned to multiple users for GF transmissions, the occurrence of collisions may jeopardize the GF potential. In this paper, we perform a system analysis in a large urban macro network of several transmission procedures for uplink GF transmission presented in recent literature. Specifically, we study K-Repetitions and Proactive schemes along with the conventional HARQ scheme referred to as Reactive. We evaluated their performance against the baseline GB transmission as a function of the load using extensive and detailed system level simulations. Our findings show that GF procedures are capable of providing significant lower latency than GB at the reliability level of $1 - 10^{-5}$*, even at considerable network loads. In particular, the GF Reactive scheme is shown to achieve the latency target while supporting at least 400 packets per second per cell.*

1 Introduction

Ultra-Reliable Low-Latency Communication (URLLC) represents the most challenging set of services/use cases [1] for upcoming 5th Generation (5G) New Radio (NR), with ambitious latency and reliability targets (1 ms with $1 - 10^{-5}$ reliability) for small packet transmissions [2]. A number of technology components including spatial diversity [3], frame structure [4, 5], resource allocation [6] including link adaptation and transmission schemes, all need to be redesigned when dealing with requirements that are beyond current Long Term Evolution (LTE) capabilities [7].

In particular, the transmission procedures, including Hybrid Automatic Repeat Request (HARQ) retransmissions, play a major role in achieving the URLLC requirements [8]. LTE utilizes dynamic scheduling as a basic transmission mode, which is referred to as Grant Based (GB) scheduling (specified in [9]). A traditional GB transmission requires the User Equipment (UE) to be scheduled by the base station (BS). The scheduling procedure is initiated by the UE with a scheduling request which the BS can respond by issuing a scheduling grant.

Grant-Free (GF) transmission schemes are also well known solutions that are meant for fast uplink access, by removing the phases of scheduling request and grant issuing [10]. With Semi-Persistent-Scheduling (SPS), the BS can configure the UE to have pre-allocated periodic radio resources available for transmissions [11, 12]. For periodic traffic, SPS is expected to be a

valid solution to meet the URLLC requirements. However, in case of aperiodic (sporadic) traffic, pre-allocating dedicated resources may lead to a large waste and will scale poorly with the number of URLLC users. A possible solution to overcome this limitation, is to pre-schedule shared resources for contention-based transmissions [4].

Conventional HARQ operations in LTE allows for retransmissions only upon reception of a negative acknowledgement. This requires the BS to have first received the payload, processed it and issued the feedback. Such HARQ scheme is often referred to as *Reactive* since retransmissions are triggered based on the knowledge about the previous transmission.

However, the reactive HARQ scheme can only support a limited number of retransmissions before the URLLC requirements is no longer met. Therefore different HARQ strategies to further reduce latency and improve reliability have been recently studied. One technique that has been considered for 5G, is to run a number of blind transmissions of the same payload. The BS can then perform soft combining of the transmissions to improve the decoding reliability [13]. Such kind of solution is already part of the recent 3GPP agreements for NR and are referred to as *K-Repetitions* (K-Rep) [14].

In a proactive version of the HARQ scheme mentioned above, the UE can still transmit in consecutive frames (like K-Rep), but it will stop when it has received and decoded a positive feedback from the BS. Such scheme is known as repetition scheme with early termination, and is mentioned in [15] and [16]. This scheme is more computational heavy for the UE, which needs to monitor the feedback. However, it is also likely to be more resource efficient than K-Rep if the number of blind repetitions is overestimated and more reliable if the number is underestimated.

The theoretical foundation of the transmission procedures mentioned above is already well established. However, to the best of our knowledge their suitability for URLLC has been so far evaluated in simplified scenarios, such as single cell (and therefore no inter-cell interference impact), basic abstraction models for contention-based transmissions and throughput mapping. In this paper, we perform a detailed system level evaluation of the identified transmission procedures in an outdoor 3GPP urban macro setup with 21 cells, including realistic traffic and radio propagation models, receiver types and open loop power control. GB with conventional HARQ scheme is used as performance baseline. The transmission schemes are then evaluated in terms of the latency and reliability and as a function of the load imposed by URLLC devices in the network. Our aim is to assess the effective system benefits of the identified techniques and their potential in a network of URLLC devices.

The paper is structured as follows. The considered URLLC UL transmission schemes are described in section 2. The simulation assumptions are outlined in section 3, while the results are presented in section 4. The work is discussed in section 5 and concluded in section 6.

2 URLLC UL Transmission Schemes

This section provides a general description of the transmission schemes considered in this paper. A frame-based system alike LTE is assumed, meaning that transmissions can start on a frame basis. The transmissions occur when the UE is already synchronized and in connected state. We consider both GB and GF solutions.

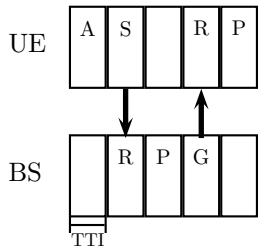


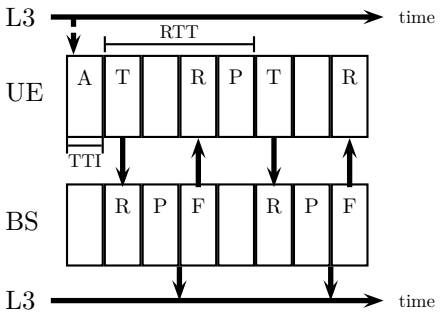
Fig. C.1: Scheduling request model used for Grant-Base access. Legend: A = Frame alignment, S = Scheduling Request, R = Reception, P = Processing, G = Scheduling Grant.

The GB approach is the common method to perform an UL transmission in cellular networks, and is evaluated with the usual LTE scheduling grant procedure as illustrated in illustrated in Fig. C.1 and with the conventional HARQ scheme (reactive Fig. C.2(a)).

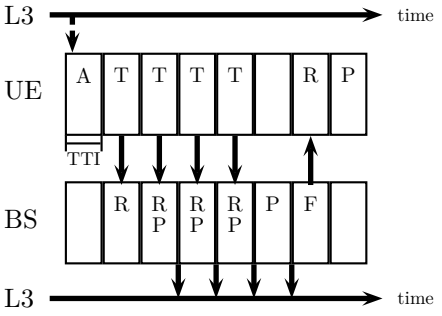
When using the GB approach, each UL transmission is coordinated by the base-station (BS). Upon a packet arrival on layer 3 (L3), a UE waits for the next subframe occurrence for transmitting a scheduling request (SR) signal (S). After processing the SR, the BS transmits a scheduling grant (G) which indicates the time-frequency resources among other settings that the UE should use for its uplink data transmission (T). Only after receiving (R) and processing (P) the scheduling grant, the UE can perform the data transmission. This procedure allows the BS to assign resources in a very flexible manner, leading to a high spectral efficiency. Further, the transmissions are collision-free.

The scheduling process comes with a number of drawbacks; it is time consuming, which makes it harder to make the URLLC requirements, it introduces a large signalling overhead for small packets which might be a limiting factor for scalability and the signalling is error prone. The cost is that the transmissions becomes prone to collisions and intra-cell interference.

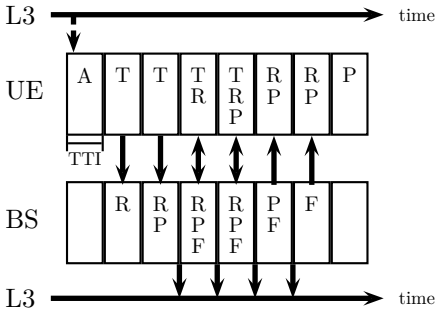
Three HARQ schemes are considered for GF transmissions, namely a Reactive, K-Rep and Proactive scheme. The Reactive scheme is illustrated in Fig. C.2(a). When the UE has finalized its initial uplink data transmissions, its signal is processed at the BS, which will transmit a positive or negative acknowledgement. Upon processing the feedback, the UE can transmit a new payload or retransmit the same payload again. The time duration of the cy-



(a) Reactive



(b) K-Repetitions (K-Rep) with $K = 4$ repetitions



(c) Proactive with maximum 4 repetitions

Fig. C.2: The considered Uplink HARQ Schemes for URLLC. Shown for Grant-Free transmissions. Legend: A = Frame alignment, T = transmission, R = Reception, P = Processing, F = Feedback.

cle from the beginning of a transmission until the processing of its feedback is called the HARQ Round-Trip-Time (RTT). In the illustration it is assumed that the BS spends 1 transmission time interval (TTI) for processing and 1 TTI for transmitting the feedback. These assumptions are similar to the ones used by the authors in [8].

The K-Rep scheme is illustrated in Fig. C.2(b). The UE is configured to autonomously transmit the same packet K times before waiting for feedback from the BS. Each repetition can be identical, or be a different redundancy versions of the encoded data. This method can reduce the delay in the HARQ process, with a potential waste of resources if the number of repetitions is overestimated.

The last HARQ scheme considered for GF transmissions is the Proactive scheme which is illustrated in Fig. C.2(c). Similarly to the K-Rep scheme, the UE aims at repeating the initial transmission for a number of times, however, it will receive a feedback at every repetition. This allows the UE to stop the chain of repetitions earlier in case of a positive feedback. A reduction of the overall transmission resources can be obtained compared to the K-Rep scheme in case the time spent for the K'th transmission is higher than the HARQ RTT. Further it might enhance the reliability compared to the K-Rep, in case K is underestimated.

Note that both GB and GF transmissions can be subject to queuing delays. This occurs due to the limit that a UE can only transmit one packet per TTI or if the UE runs out of Stop-And-Wait (SAW) channels. A SAW channel is occupied throughout the entire transmission, meaning from the initial transmission until the stopping criteria determined by the HARQ RTT from the last transmission.

3 Simulation assumptions

The simulation assumptions and parameters used for this study are in line with the guidelines for NR performance evaluations presented in [17] and are summarized in Table C.1.

The system level simulation of the multi-cell synchronous network includes inter-cell interference, realistic propagation models, link-to-system mapping and modeling of major radio resource management (RRM) functionalities in accordance with the evaluation methodology of recent 3GPP standardization agreements.

In this work we compare the GF schemes with a baseline GB scheme. As in [8], we assume here 1 TTI for transmitter and receiver processing time. It is worth mentioning that a higher processing time directly translates to a higher delay on the scheduling procedure and HARQ schemes. To ensure a fair comparison between GF and GB schemes we use the same amount

Table C.1: Simulation assumptions

Parameter	Value
Network layout	3GPP Urban Macro (UMa) [17] with 21 cells, 500 m inter-site distance
UE deployment	Uniformly distributed outdoor, speed of 3 km h^{-1}, without handover
Carrier and Bandwidth	10 MHz at 4 GHz
PHY numerology	2 OFDM symbols per TTI, subcarrier spacing of 15 kHz, 12 subcarriers/PRB
Uplink receiver	MMSE-IRC
Uplink antenna	1x2 antenna configuration
Channel model	3D UMa propagation model, noise density of -174 dBm Hz^{-1}
HARQ configuration	4 TTI RTT and 1 TTI processing (for both UE and BS), 4 SAW channels
Frame alignment model	Uniform random variable up to 1 TTI
Traffic model	FTPModel3 with 32 B packet size and Poisson arrival of 10 packets per second (PPS) per UE
Link-Adaptation	Conservative modulation and coding scheme fixed to QPSK 1/8
Power control	Open Loop Power Control (OLPC) with $\alpha = 0.8$ and $P_0 = -85$ dBm
SR configuration	SR periodicity of 1 TTI
Shared channel configuration	48 RB contention based channel, all UEs can transmit in any TTI

of resources for the uplink shared channel used by GF and GB. Uplink and downlink is separated in frequency (FDD), where the uplink shared channel has 48 resource blocks (RBs) in the 10 MHz bandwidth. The shared channel is assumed to be available in all subframes for GF transmission. For the GB procedure, the configured SR periodicity of 1 TTI permits the UE to ask to be scheduled at every TTI. No additional control overhead is assumed. In this work, we assume the control signalling to be error free, meaning that particular the GB results can be optimistic.

The scenario used in our study is slightly deviating from the one specified in [17], since here all UEs are deployed outdoor. Indoor users showed an tendency to get power limited and were hence unable reach URLLC reliabilities.

Open loop power control is used in this study by the UE to compensate the coupling loss and is configured with $\alpha = 0.8$ and $P_0 = -85$ dBm. In the considered deployment this configuration permits the UEs to operate mostly

below the maximum transmit power (23 dBm).

It is assumed that the URLLC UEs are pre-configured with 48 RB for contention based uplink transmissions. The modulation and coding scheme (MCS) is also pre-configured as very conservative (QPSK with coding rate 1/8), which permits the UE to transmit the 32 B packet (in accordance with baseline in [2]) in 1 TTI using the full band.

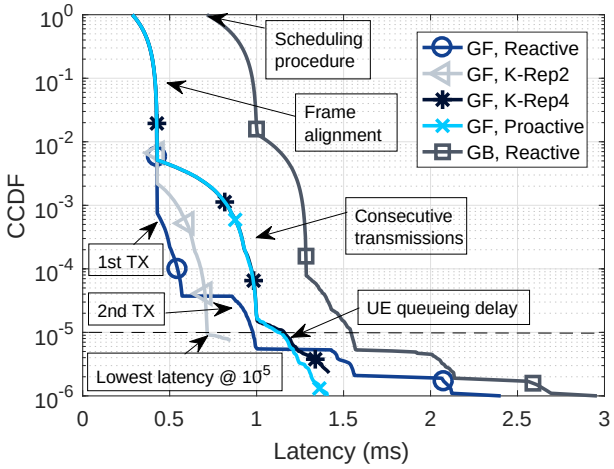
The adopted Minimum Mean Square Error Interference Rejection Combining (MMSE-IRC) receiver is assumed to be able to ideally estimate the interference covariance matrix for suppressing intra-cell and inter-cell interference. Given the 2 receive antennas, up to one interfering stream can be suppressed. This also means the decoding of two simultaneously transmitting UEs in the same cell is still possible and depends on the post-detection Signal-to-Noise Plus Interference Ratio (SINR) and the selected MCS.

We focus on the user plane latency and reliability for small packet transmissions assuming the UE is in connected mode. The latency is measured as a one-way latency from when the packet leaves the L3 buffer at the UE until it enters L3 layer at the BS. Throughout the study it has been observed that the packet generation rate per UE impacts the queuing delay and hence forces an upper bound of the load. In order to circumvent this limitation, a variable cell load is simulated by varying the number of UEs per cell, while their packet generation rate is maintained constant. However this comes at the penalty of increased computational complexity of the simulation when more UEs are added. In order to have an acceptable simulation time for different number of UEs, we chose a mean packet generation rate of 10 packets per second giving a theoretical lower bound probability (depending on the HARQ scheme) of a packet being queued at $\approx 10^{-6}$.

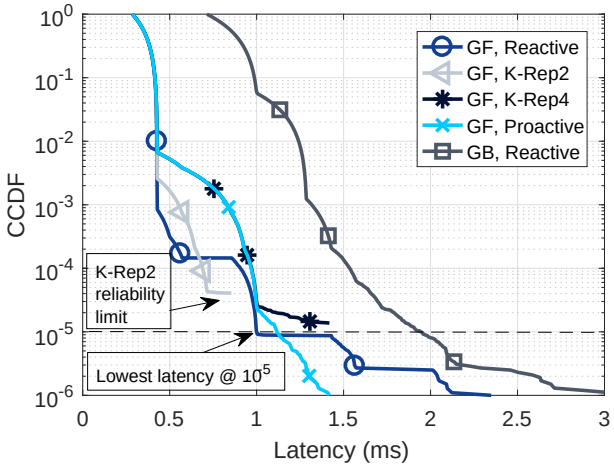
4 Results

The evaluation of the UL transmission schemes is carried out with Monte Carlo simulations. More than 5×10^6 samples per simulations are acquired to ensure sufficient statistical confidence in the 10^{-5} quantile [8]. The transmission schemes are evaluated at different loads, determined by URLLC densities. Results are presented in terms of one-way latency for a packet transmission, as well as number of transmissions per packet. Unsuccessful packets are represented as void samples and are used to reflect the achievable reliability.

In Fig. C.3(a) the emperical Complementary Cumulative Distribution Function (CCDF) of the one-way latency for the different GF HARQ transmission schemes is shown along with the GB baseline with low load (10 UEs / cell). On the horizontal axes the latency is shown in ms and on the vertical axes the outage probability quantiles are shown. The GF schemes clearly provide



(a) Low load (10 UE / cell)



(b) High load (40 UE / cell)

Fig. C.3: CCDF of the latency for GF and GB baseline for low (a) and high (b) load.

a better latency for the same reliability compared to the GB reference. One of the main differences between these are the unavoidable delay offsets from the scheduling procedure. The first slope from ≈ 0.3 ms to ≈ 0.4 ms corresponds to the uniformly distributed frame alignment delay.

The Reactive HARQ scheme is the one providing the best reliability for the the first transmission. The stair behaviour is caused by the HARQ RTT. K-Rep scheme with 2 repetitions follows the initial transmission with a similar

slope for the second consecutive transmission, and is capable of providing 1 ms latency with the target $1 - 10^{-5}$ reliability. The curve has a tail caused by low probability events corresponding the probability of packet buffering at the UE.

The K-Rep scheme with 4 repetitions and Proactive scheme have a similar latency and reliability performance until 1 ms. This can be explained from the fact that the Proactive scheme earliest determination time depends on the HARQ RTT which here it is assumed to be 4 TTIs. Since more than 4 repetitions is rarely needed in this scenario, K-Rep4 and Proactive perform almost identically. The schemes shows different tail tendencies, where the Proactive scheme is better on handling the low probability events where more than $K = 4$ repetitions is needed.

Comparing the HARQ Reactive transmission scheme for GF and GB transmission, they show a similar stair behaviour, with the initial step occurring at different latency and reliability combinations (e.g. 0.6 ms and 1.6 ms for GF and GB respectively). The reason for the reliability difference for the initial transmission is due to the impact of intra-cell interference. Further the GB curve shows tendencies for higher packet queuing probability due to the longer pre-transmission time caused by the scheduling procedure.

Performance at a higher load (40 UE / cell) is shown in Fig. C.3(b). The impact of a higher load is clearly visible for the Reactive HARQ schemes. The CCDF of the Reactive HARQ scheme shows an increase in the probability of needing multiple retransmissions and causing its tail to be longer compared with the low load. The GF K-Rep schemes reach a reliability floor around $\approx 1 - 4 \times 10^{-5}$ instead of $\approx 1 - 10^{-5}$. With this load, only the Proactive and Reactive HARQ schemes for GF transmissions are able to achieve the $1 - 10^{-5}$ reliability and only the Reactive HARQ scheme is capable of doing within the 1 ms latency target.

Figure C.4 illustrates the impact of the load on the achievable latency with $1 - 10^{-5}$ reliability. At low load, the Reactive scheme and the K-Rep scheme with 2 repetitions meet the URLLC performance target, where the latter has the lowest latency. For more than 40 UEs / cell no GF or GB scheme is capable of achieving the URLLC target. However, at high load the GF Proactive scheme leads to the lowest latency.

Figure C.5 shows the empirical Cumulative Distribution Function (CDF) of the average SINR per RB for the case of 40 UE / cell. Here it is possible to see that the GB transmissions presents the best SINR condition since intra-cell interference is avoided in this procedure. GF with the K-Repetitions and Proactive scheme on the other hand presents the worst SINR due to the extra intra-cell interference caused by the blind repetitions. The GF Reactive scheme presents a better SINR then the other GF schemes given that it avoids unnecessary retransmissions. This explains why each transmission of the Reactive scheme presents a higher reliability, compared to the cases with

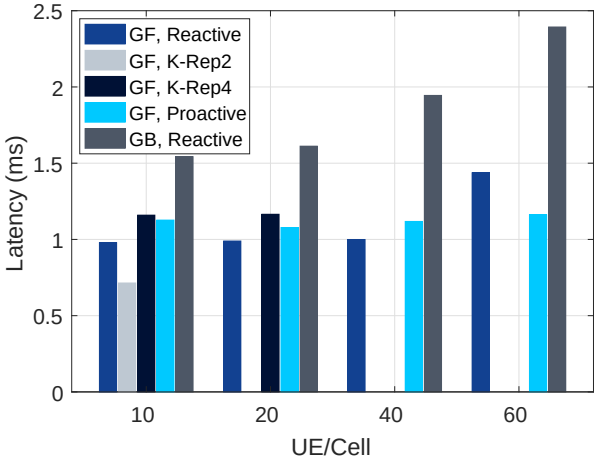


Fig. C.4: Achieved latency at $1 - 10^{-5}$ reliability as a function of load.

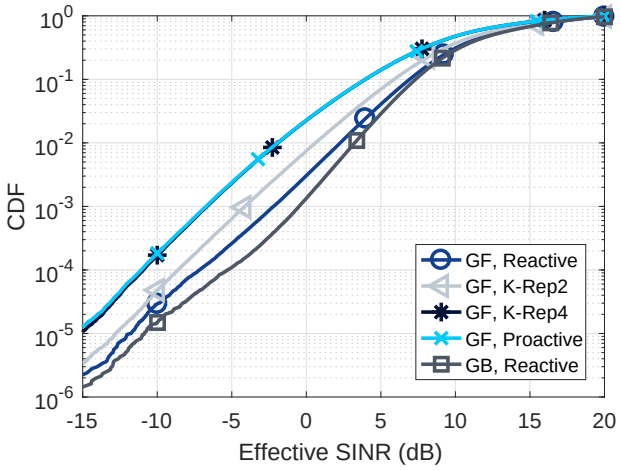


Fig. C.5: Average effective SINR per RB for GF and GB (40 UE / cell).

blind repetitions. In this case, for GF Reactive, a $1 - 10^{-5}$ reliability can be achieved with 2 transmission attempts. While, for instance, in the Proactive or K-Rep after 4 attempts the achieved reliability is even lower.

As showed in [7], achieving low latency and high reliability has a cost in terms of resource utilization and therefore spectral efficiency. Figure C.6 shows the empirical CCDF of the number of transmissions used for successfully delivering a packet for the different schemes, assuming a load of 40 UEs / cell. The GB scheme presents, not surprisingly, the lower probability of re-

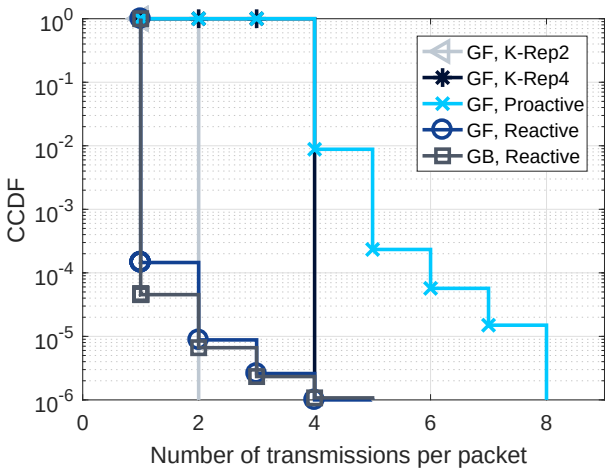


Fig. C.6: CCDF of the number of transmissions per packet (40 UE / cell).

quiring multiple channel accesses for transmitting a packet. The curve for the GF Reactive scheme is slightly higher compared to the GB Reactive. This is likely due to the presence of collisions. The K-Rep schemes are very deterministic in terms of channel usage, while GF Proactive occupies the channel at least during the RTT. The two former schemes, besides not meeting the baseline requirement, also presents the lowest spectral efficiency at this scenario and with this load.

5 Discussion

The evaluated GF solutions clearly show better latency performance than GB transmission at $1 - 10^{-5}$ reliability, despite the impact of collisions. Our results also showed that GF schemes are not outperformed by GB even in the case of 40 devices per cell. This section discussions the dominating factors impacting our results.

GB avoids intra-cell interference by ensuring a single transmit UE per TTI, but also causes a latency increase by waiting for the channel to become available. The GF schemes have no such limitation, but are instead affected by the intra-cell interference from competing UEs. Therefore GB has the potential to achieve the $1 - 10^{-5}$ reliability when the latency requirement is relaxed, to e.g. 2 ms for the referred loads, causing a lower interference in the network.

The reasoning behind the usage of GF K-Rep schemes, is to cope with tight time constraints by allowing a number of consecutive transmissions in

a short time. Our findings show, however, that the additional intra-cell interference due to the multiple transmissions is the major impacting factor and surpasses the benefits of the combining gain. One way to lower the average intra-cell interference with K-Rep schemes is to use a faster reconfiguration cycle that sets higher number of repetitions only for the UEs in worse channel condition, though requiring additional RRC signalling.

In the studied scenario with GF, the use of a robust MCS (QPSK 1/8) ensures a high decoding probability even under a potentially high intra-cell interference. Another aspect is the benefit of HARQ which adds combining gain and diversity, given also that a packet has lower probability of colliding.

As mentioned in Section 3, results are obtained with a MMSE-IRC receiver with 2 antennas, which is able to resolve two simultaneous transmissions from two different UEs. It is left for future analysis to investigate the impact of other receiver types and antenna configurations, whose capabilities of resolving the interference may affect the trade-off between GB and GF transmissions. The use of a successive Interference Cancellation (SIC) receiver is also considered.

With GF transmissions the BS has to conduct blind decoding as every connected UE has the possibility to transmit in every TTI. The BS should be able to identify a UE before attempting to decode it. This assumes a system design where the UE identity is mapped over e.g. preambles and header at each transmission [18]. The impact on the preamble and header design on the GF performance is left for future work.

Moreover, in this work the control channel is assumed to be ideal and not introducing any overhead. While the control signalling is typically designed to be very robust, the potential errors may not be negligible for the range of reliability expected for URLLC. Errors in control signalling can significant impact the schemes relying on feedback, such as the Proactive and particular the Reactive schemes, as well as the scheduling procedure for GB. These are also the scheme relying on the most DL resources due to the signalling. The impact of error-prone control signalling is left for further analysis.

The GF analysis can also be extended with the adoption of other enhancements, as a Non-Orthogonal Coded Access scheme like proposed in [19], that increases the capacity and reduce collisions with additional spreading codes.

6 Conclusion

In this paper, we studied the performance of uplink GF schemes in a large outdoor urban macro scenario and compared its performance with a traditional GB scheme. In particular, the schemes referred to as GF Reactive, K-Rep and Proactive, are evaluated. The results are obtained using extensive system level simulations to include the complexity of the receiver, inter-cell

interference, power control and HARQ operations including soft combining. The main findings of this work together with the recommendations for a 5G NR design are:

- GF in general outperforms GB transmission procedures in terms of latency at the target reliability ($1 - 10^{-5}$). This makes them valuable candidates for achieving the baseline URLLC requirements in an outdoor scenario.
- The GF Reactive scheme is strongly recommended as it is capable of supporting the largest load among the GF schemes. The maximum achieved load is found to be 400 packets per second per cell (40 UEs per cell generating 10 packets per second on average). This scheme is also the most uplink resource efficient next to the GB baseline.
- The GF Proactive scheme gives the smallest latency performance degradation for loads higher than 400 packets per second.
- GB transmissions can achieve the target reliability if the latency requirements is relaxed to e.g. 2 ms.

The presented results are obtained by relying on a robust MCS (QPSK 1/8) for packet transmission, interference suppression by IRC receiver and HARQ combining gain from repetitions and retransmissions. Future work will investigate the impact on the GF performance of factors such as dynamic link adaptation, power boosting, multiple receiver types and antenna configurations.

Acknowledgment

This research is partially supported by the EU H2020-ICT-2016-2 project ONE5G. The views expressed in this paper are those of the authors and do not necessarily represent the project views.

References

[1] International Telecommunication Union (ITU), "IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond," ITU Radiocommunication Sector, Tech. Rep., Sep. 2015.

[2] 3GPP TR 38.913 v14.1.0, "Study on Scenarios and Requirements for Next Generation Access Technologies," Mar. 2017.

[3] G. Pocovi, B. Soret, M. Lauridsen, K. I. Pedersen, and P. Mogensen, "Signal Quality Outage Analysis for Ultra-Reliable Communications in Cellular Networks," in *2015 IEEE Globecom Workshops*, Dec. 2015.

[4] 3GPP TR 36.881 v14.0.0, "Study on latency reduction techniques for LTE," Jul. 2016.

[5] K. I. Pedersen, G. Berardinelli, F. Frederiksen, P. Mogensen, and A. Szufarska, "A Flexible 5G Frame Structure Design for Frequency-Division Duplex Cases," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 53–59, Mar. 2016.

[6] C. She, C. Yang, and T. Q. S. Quek, "Radio Resource Management for Ultra-Reliable and Low-Latency Communications," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 72–78, Jun. 2017.

[7] B. Soret, P. Mogensen, K. I. Pedersen, and M. C. Aguayo-Torres, "Fundamental Tradeoffs among Reliability, Latency and Throughput in Cellular Networks," in *2014 IEEE Globecom Workshops*, Dec. 2014.

[8] G. Pocovi, B. Soret, K. I. Pedersen, and P. Mogensen, "MAC Layer Enhancements for Ultra-Reliable Low-Latency Communications in Cellular Networks," in *2017 IEEE International Conference on Communications Workshops*, May 2017.

[9] 3GPP TS 36.213 V14.2.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," Mar. 2017.

[10] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, I. Riedel, A. Puschmann, A. Mitschele-Thiel, M. Muller, T. Elste, and M. Windisch, "Latency Critical IoT Applications in 5G: Perspective on the Design of Radio Interface and Network Architecture," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 70–78, Feb. 2017.

[11] D. Jiang, H. Wang, E. Malkamaki, and E. Tuomaala, "Principle and Performance of Semi-Persistent Scheduling for VoIP in LTE System," in *2007 International Conference on Wireless Communications, Networking and Mobile Computing*, Sep. 2007.

[12] RP-161788, "V2V Work Item Completion," Sep. 2016.

[13] R1-1705246, "UL grant-free transmission for URLLC," Apr. 2017.

[14] 3GPP TSG RAN WG1, "RAN1 Chairman's Notes," Jan. 2017.

[15] R1-1612246, "Discussion on HARQ support for URLLC," Nov. 2016.

[16] 3GPP TSG RAN WG1, "RAN1 Chairman's Notes," Feb. 2017.

[17] 3GPP TR 38.802 v14.0.0, "Study on New Radio Access Technology," Mar. 2017.

[18] K. M. Rege, M. A. Kocak, K. Balachandran, J. H. Kang, and M. K. Karakayali, "On the Design of Preamble for Autonomous Communications with Extended Coverage," in *VTC Spring 2017 - IEEE 85th Vehicular Technology Conference*, Jun. 2017.

[19] Z. Zhao, D. Miao, Y. Zhang, J. Sun, H. Li, and K. Pedersen, "Uplink Contention Based Transmission with Non-Orthogonal Spreading," in *VTC Fall 2016 - IEEE 84th Vehicular Technology Conference*, Sep. 2016.

Part III

Radio Resource Management for Grant-free URLLC

Radio Resource Management for Grant-free URLLC

1 Problem Description

Although high spectral efficiency is not a major requirement for Ultra-Reliable Low-Latency Communications (URLLC) in 5G New Radio (NR), the radio-frequency spectrum is limited asset, specially below 6 GHz. Therefore avoiding the waste of radio resources is important for guaranteeing viable solutions, which support higher URLLC loads in the system as discussed in the first section. In the previous part of the study, the transmission schemes for URLLC where evaluated without further exploiting radio resource management (RRM) techniques like power control, link adaptation and macro-diversity. Those are generally relevant for satisfying quality of service (QoS) requirements while ensuring an efficient utilization of the network resources.

Transmit power control is a crucial RRM functionality in cellular networks. It aims at achieving a desirable signal level at the receiver, and at the same time, limiting the user equipment (UE) power consumption and generated intra- and inter-cell interference. In Long Term Evolution (LTE) systems, open loop power control with fractional path loss compensation is utilized for favoring cell capacity in detriment of cell-edge bit rate [1]. For URLLC, however, the power control configuration needs to be rethought. Latency and reliability should be taken into account as main key performance indicators (KPIs), and the aim should be on minimizing the outage probability for the URLLC transmissions in the system. With perfect knowledge of the channel state information, the optimum power control can be given by truncated channel inversion as described in [2]. However, for grant-free procedures in uplink, only a large scale fading gain estimation is available based on path loss measurements from the downlink reference signals. To deal with the potential signal quality degradation in the presence of fast fading and collisions, more power should be devoted to the transmissions than what would be necessary in a collision-free case [3]. The extra margin should

cope with subtle channel variations and at the same time an excessive generated interference should be avoided. A customized power control strategy is therefore necessary for supporting grant-free URLLC.

As discussed in [4], stringent latency and reliability requirements imply low data rate. The usage of low modulation order and robust coding in order to ensure a low block error rate (BLER) naturally reduces the spectral efficiency. Another factor which penalizes the spectral efficiency is that conventional channel coding is less efficient for small URLLC payloads, which need to be encoded with a lower rate compared with transmissions of large blocklengths [5]. These are fundamental limitations for efficient URLLC. In cellular systems, like LTE, the scheduler dynamically allocates the radio resources and sets the modulation and coding scheme (MCS) for every requested transmission based on buffer status report provided by the UE and on the estimated channel state information (CSI). LTE systems also count on fast adaptive modulation and coding including an outer-loop link adaptation (OLLA) scheme for meeting a target BLER [6]. Such scheme corrects the CSI estimation and adjusts the MCS selection algorithm according to the feedback from initial transmissions. However, the convergence of such method is very slow for URLLC, due to the low failure rate targeted for the transmissions and strict latency requirement [7]. Moreover, for grant-free, this procedure is not applicable since the resource configuration is semi-static, i.e. not changing per transmission basis as mentioned in Part I. The preallocated resources and transmission parameters should be sufficient to cope with signal quality variation margin. This variation is caused by fast fading and sudden interference in the uplink channel. For coping with that, the MCS selection needs to be redefined together with the use of wideband allocations to offer frequency diversity. Besides, the network can make use of spatial diversity and receiver combining mechanisms.

The reception scheme is of great importance for enabling efficient URLLC. Multi-antenna receivers are able to provide spatial degrees of freedom, which can either provide combining gain with micro-diversity, or enable spatial multiplexing for receiving signals from different sources [8]. A more advanced scheme exploits macro-diversity, in which the signal received by multiple base stations is combined for lowering the packet error probability. The problem for utilizing such advanced reception schemes is the cost and augmented complexity due to demanding receiver processing, precise synchronization and backhaul load.

2 Objectives

The goal of this part of the thesis is to improve the achievable URLLC load in the system by mean of RRM solutions. For that, the following objectives are

settled:

- Revise the power control strategy and transmission parameters settings with emphasis on URLLC KPIs, instead of throughput oriented settings.
- Propose alternatives for link adaptation applicable for grant-free URLLC transmissions.
- Study diversity techniques which exploit multi-antenna receivers and multi-cell reception schemes for grant-free URLLC.

3 Included Articles

The content of this part of the thesis is formed by the following papers:

Paper D. Power Control Optimization for Uplink Grant-Free URLLC

In this paper open loop power control is investigated with focus on the performance of grant-free URLLC. The objective is to quantify the impact of power control on URLLC KPIs values, hence obtaining insights about configuration strategies which lead to higher URLLC loads. In addition, a simple method encompassing power boosting steps for reducing the outage probability of hybrid automatic repeat request (HARQ) retransmissions is presented. Recommendations are provided regarding the configuration of open loop power control and power boosted retransmissions. The study is based on system level simulations considering an urban macro scenario, following the methodology utilized in Paper C.

Paper E. Efficient Resource Configuration for Grant-Free Ultra-Reliable Low Latency Communications

This work is built on top of the findings from the previous papers. A radio resource management solution comprising multiple grant-free configurations is presented. The solution encompasses different allocated sub-bands associated with MCSs and power control settings. The MCS selection scheme enforces users in favorable channel condition, or high coupling gain, to apply smaller sub-bands, using higher MCS and higher power density. This reduces the overlapping with transmissions from users in poor channel condition. The study includes also the impact of minimum mean square error (MMSE)-interference rejection combining (IRC) with 2 and 4 receive antennas on the achievable load. System level simulations are used for the evaluations taking into account recent considerations for UE processing time.

Paper F. Multi-cell Reception for Uplink Grant-Free Ultra-Reliable Low-Latency Communications

In this work the potential of multi-cell reception for grant-free URLLC is studied. The purpose is to exploit the additional diversity and combining gain provided by the joint reception mechanisms, for improving the URLLC outage capacity. Different combining methods are studied, namely selection-combining, chase-combining and a hybrid-combining scheme. The impact of the multi-cell reception parameters, such as assisting cell selection threshold and number of assisting cells, are evaluated for the different combining methods. Furthermore, multi-cell reception aware RRM solutions are evaluated to further improve the resource utilization. The solutions are analyzed through detailed system level simulations for a NR urban macro evaluation scenario.

4 Main Findings and Recommendations

Impact of power control optimized for URLLC

Paper D shows that, in order to improve the URLLC performance in the system, it is important to take into account the URLLC KPIs instead of just employing typical cell throughput oriented settings. Open loop power control using fractional path loss compensation ($\alpha < 1$), for example, does not show any benefit for URLLC. And with optimized P_0 setting, targeting to reduce the outage probability at 1 ms, the achievable URLLC load in the cell can be more than doubled.

The usage of power boosting for HARQ retransmissions is beneficial for reducing the outage probability. However the gain is limited for scenarios in which most of the UEs operate close the power limit. In the considered urban macro scenario, the gain in the URLLC capacity when using power boosting is approximately 20%. With the error probability for the initial transmission being very low ($\approx 10^{-3}$), retransmissions with power boosting occur very rarely, therefore not causing a harmful interference level.

Multiple grant-free configurations

The RRM solution presented in Paper E, provides multiple grant-free configurations for the URLLC UEs. With that, UEs in good average channel condition can be set to use higher MCS and lower sub-bands. This reduces the collision probability and the overlapping with UEs transmitting in wider band. However, the signal to interference-and-noise ratio (SINR) target for reliable decoding increases when using higher MCS orders. These UEs should then use their power headroom for compensating with increased power spectral density (PSD). So, there is a trade-off on reducing the collision probability

versus increasing the required SINR.

The determination on when a UE should switch between the configurations can be based on a coupling gain threshold. This threshold can be taken from the point where the UEs tend to experience a degradation on the average SINR. For the evaluated scenario with two active configurations, using MCS QPSK1/8 and QPSK1/2 with associated sub-bands and power control settings, the observed gain in terms of achievable URLLC load is around 90% compared to using a single configuration with MCS QPSK1/8.

Multi-antenna receiver diversity and interference rejection

The use of MMSE receiver with IRC capability is an attractive solution due to the receiver simplicity and maturity. It does not employ an iterative processing, as in a successive interference cancellation (SIC) receiver for example, which can be time consuming. Paper E results using MMSE-IRC demonstrate that, by changing from a 2 antenna receiver to 4 antenna receiver, the aggregated URLLC load in the network can be increased by a factor of 7. This is due to the higher degrees of freedom, which gives diversity gain for improving reliability or allows to suppress interference from overlapping transmissions. That 4-antenna configuration also provides the required diversity level for meeting the URLLC target with grant-based single-shot transmissions, though the grant-free procedure with HARQ reaches better performance. Using higher number of antennas is naturally beneficial, however it increases the cost and necessity for accurate channel estimation. Moreover, the performance depends on having time alignment with users within the cell and from other cells.

Multi-cell reception combining gain

Paper F shows that multi-cell reception can greatly improve the reliability of grant-free transmissions, thus the outage capacity for URLLC in uplink. With the multi-cell connectivity, the signal from UEs in neighbor cells is considered as useful information, instead of being just treated as interference. This is beneficial specially for cell edge users, which are the ones that cause higher inter-cell interference, while at the same suffer with lower SINR when operating in power limit. It is observed that, for a fixed load, the outage probability clearly reduces when using up to 2 assisting cells. However no major improvement is obtained by further increasing the number of assisting cells. The reference signal received power (RSRP) window, for selecting the assisting cells, is also determinant for the performance. Higher window thresholds allow more UEs to benefit from multi-cell reception, with the cost of a higher backhaul load.

Even with the simplest selection combining scheme, more than 20% gain

in the achievable URLLC load can be obtained. Major performance improvement is unleashed by performing chase-combining, using collected soft-bit information of the desired signal from the assisting cells. While the soft-combining method provides the higher gains, the backhaul load is however approximately 50 times higher compared to selection combining. Multi-cell reception aware RRM enhancements, including power control and MCS selection, further improves the URLLC outage capacity. Nevertheless, it is important to mention that multi-cell reception mechanisms have more relevance when other diversity mechanisms are not in place. For instance, with only 2-antenna receivers and without HARQ, multi-cell reception can improve the URLLC outage capacity by up to 440%. However, with 4-antenna receivers and HARQ, the improvement goes down to 22% in the studied scenario.

Main recommendations

The following recommendations are made according the presented findings:

- Open loop power control should be applied with full path loss compensation and optimized P_0, focusing on reducing outage probability for URLLC transmissions.
- For reducing the failure probability of the retransmission, power boosting can be employed while the error probability of the initial transmission should be kept low, to avoid an interference increase.
- Multiple grant-free configurations should be used, with UEs in favorable average channel condition switching to smaller sub-bands, for reducing the overlapping and improve URLLC achievable load.
- For supporting higher loads in grant-free shared resources, linear MMSE-IRC with at least 4 antennas can be utilized, giving more degrees of freedom for diversity combining and interference rejection.
- Multi-cell reception with up to 2 assisting cells should be used to obtain macro-diversity gain, boosting URLLC outage capacity specially in cases of few receive antennas and when HARQ cannot be used.

The Figure III.1 shows a summary of the achieved URLLC load and the resource utilization from some of the studied schemes. As discussed in Part II, grant-free access using K-repetitions is the simplest solution and allows very low latency, however, it supports very limited load. The usage of HARQ permits improved resource utilization since retransmissions are issued only when needed. As discussed in this part, power control optimized for URLLC results in higher achievable load. Power boosting retransmissions give limited gain in the wide area case. Using multiple grant-free configurations

chosen according average channel condition further improves the resource utilization. MMSE-IRC with higher number of antennas allows much higher loads with the cost of more complex receiver. Similarly, multi-cell reception provides great performance gains, though the solution complexity is high given the dependence on soft information exchanged through the backhaul.

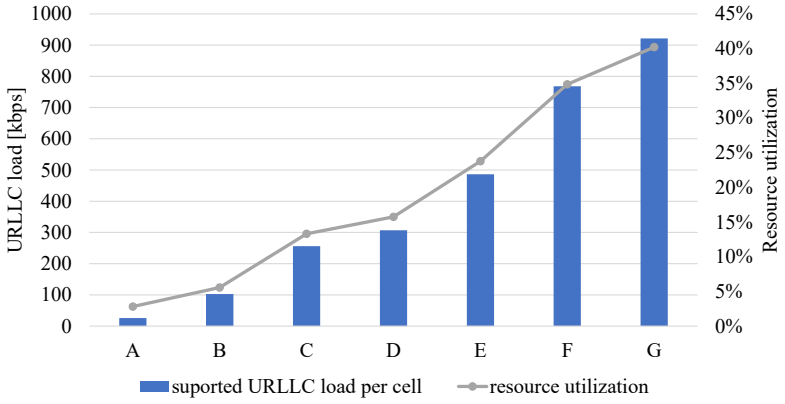


Fig. III.1: Summary performance of the studied grant-free schemes in the considered urban macro scenario: (A) K-repetitions, K=2; (B) HARQ and fractional power control; (C) HARQ and power control for URLLC; (D) HARQ with power boost; (E) Multiple GF configurations; (F) 4 antennas single transmission; (G) Multi-cell, HARQ, 2 antennas.

References

[1] A. Simonsson and A. Furuskar, "Uplink Power Control in LTE - Overview and Performance, Subtitle: Principles and Benefits of Utilizing rather than Compensating for SINR Variations," in *2008 IEEE 68th Vehicular Technology Conference*, Sep. 2008.

[2] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1468–1489, Jul. 1999.

[3] P. Popovski, J. J. Nielsen, C. Stefanovic, E. d. Carvalho, E. Strom, K. F. Trillingsgaard, A. S. Bana, D. M. Kim, R. Kotaba, J. Park, and R. B. Sørensen, "Wireless Access for Ultra-Reliable Low-Latency Communication: Principles and Building Blocks," *IEEE Network*, vol. 32, no. 2, pp. 16–23, Mar. 2018.

[4] B. Soret, P. Mogensen, K. I. Pedersen, and M. C. Aguayo-Torres, "Fundamental Tradeoffs among Reliability, Latency and Throughput in Cellular Networks," in *2014 IEEE Globecom Workshops*, Dec. 2014.

[5] G. Durisi, T. Koch, and P. Popovski, "Toward Massive, Ultrareliable, and Low-Latency Wireless Communication With Short Packets," *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.

[6] C. Rosa, D. L. Villa, C. U. Castellanos, F. D. Calabrese, P. H. Michaelsen, K. I. Pedersen, and P. Skov, "Performance of Fast AMC in E-UTRAN Uplink," in *IEEE ICC*, May 2008, pp. 4973–4977.

[7] G. Pocovi, B. Soret, K. I. Pedersen, and P. Mogensen, "MAC Layer Enhancements for Ultra-Reliable Low-Latency Communications in Cellular Networks," in *2017 IEEE International Conference on Communications Workshops*, May 2017.

[8] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

Paper D

Power Control Optimization for Uplink Grant-Free URLLC

Renato Abreu, Thomas Jacobsen, Gilberto Berardinelli, Klaus Pedersen, István Z. Kovács, Preben Mogensen

The paper has been published in the
2018 IEEE Wireless Communications and Networking Conference (WCNC)

Abstract

Ultra-reliable and low latency communication (URLLC) presents the most challenging use cases for fifth generation (5G) mobile networks. Traditionally the focus for mobile broadband has been to optimize the system throughput for high speed data traffic. However the optimization criteria for URLLC should focus on achieving small packets transmissions under strict targets such as 99.999% reliability within 1 ms. Power control is one candidate technology component for improving reliability and latency. In this work we investigate the power control for grant-free URLLC transmissions through extensive system level simulations in a urban outdoor scenario. We initially compare different settings for open loop power control (OLPC) with full and with fractional path loss compensation. Then we evaluate whether power boosting the retransmission can reduce the probability of packets delays under the 1 ms constraint. We also discuss the practical implication of applying power boosting. With full path loss compensation and boosting retransmissions, we show that a URLLC load such as 1200 small packets per second per cell can be achieved in the considered scenario.

1 Introduction

The fifth generation (5G) radio access technology should support ultra-reliable and low-latency communication (URLLC) use cases, which include applications such as traffic safety, remote tactile control, distribution automation in smart grid, etc. [1]. The third Generation Partnership Project (3GPP) has set strict requirements for URLLC in New Radio (NR), such as 32 bytes packet transmissions to be delivered in 1 ms with 99.999 % reliability [2]. It is well established that URLLC will demand enhancements of several technology components to perform well beyond the capabilities of Long-Term-Evolution (LTE) technologies, including link-adaptation, transmission-schemes and power control.

Grant-free (GF) schemes have been considered as a solution for reducing the latency of uplink (UL) initiated transmissions, by skipping the steps of scheduling request and granting [3]. In case of unpredictable traffic, configured resources can be shared by a number of users to reduce waste [4]. GF studies have focused mainly on the massive machine-type communications (mMTC) use cases [5]. In that context, non-orthogonal multiple access (NOMA) is applied to improve the system capacity by serving a massive number of devices.The cost is on the receiver complexity with algorithms that have not been optimized for low latency and ultra reliability. Different candidate schemes for NR are listed in [6, 7]. For URLLC use cases, a system level analysis of GF transmissions considering three different hybrid automatic repeat request (HARQ) schemes is presented in [8].

Power control is an important component for UL transmissions which has not yet been thoroughly studied with the focus on satisfying the strict URLLC requirements. In CDMA systems power control is used to equalize the received power and combat the near-far problem [9]. Standard power control for LTE is defined by 3GPP in [10], known as Fractional Power Control (FPC). FPC combines Open Loop Power Control (OLPC) and closed loop power corrections with fractional path-loss compensation. It allows to reduce the transmit power of cell edge users diminishing their interference on neighbouring cells, at the cost of a lower experienced performance of this users. In general, the goal of FPC is to optimize cell throughput for mobile broadband (MBB) traffic, and its performance is well investigated in e.g. [11, 12].

Traditional FPC optimization criteria focusing on throughput might not be adequate for URLLC given the different targets (latency and reliability) [13]. In this work we first investigate the suitability of LTE alike OLPC for GF URLLC. We aim at optimizing power control settings based on URLLC performance indicators. Further, we evaluate whether a power boosting mechanism for retransmissions is attractive for quickly compensating unexpected Signal-to-Interference-plus-Noise Ratio (SINR) degradations at initial transmissions. Performance is evaluated by means of detailed system level simulations. As in [8], here we use the assumptions for the NR evaluation using cyclic prefix orthogonal frequency division multiplexing (CP-OFDM) and baseline with a minimum mean square error interference rejection combining (MMSE-IRC) receiver to focus particularly on the impact of power control for GF URLLC transmissions.

The rest of the paper is organized as follows: Section 2 sets the scene of the study. Section 3 presents an overview of power control strategies and power boosting for URLLC retransmissions. The simulation assumptions are described in section 4. Section 5 presents the numeric results followed by a discussion in section 6. Finally, section 7 brings the main conclusions and some ideas about future work.

2 Setting the Scene

2.1 System description

The considered system is a single layer cellular network with synchronized base stations (BSs). The deployed BSs provides coverage to the URLLC user equipments (UEs) which are uniformly distributed in the scenario. The UEs are connected and synchronized to the serving cell. For the GF transmissions, the UEs are configured by radio resource control (RRC) signaling (as Type 1 UL [14]). The semi-static configuration includes time and frequency resource allocation, modulation and coding scheme (MCS), power control settings and

HARQ related parameters.

The traffic generated by each UE consists of small packets arriving according to a Poisson process. The transmissions occur in a frame based system like LTE and occurs in transmission time intervals (TTI) of mini-slots with 2 OFDM symbols. These assumptions follows the 3GPP NR URLLC evaluation agreements [6]. Using the 15 kHz subcarrier spacing, the length of the TTI is 0.143 ms. When a data packet arrives to the UE layer 3 buffer queue, if the queue is empty, it gets immediately passed to the layer 2 HARQ buffer which handles the transmission on GF resources. Prior to a transmission the UE might have to wait for until the start of the next TTI. This waiting time is denoted as frame alignment. If the packet is successfully decoded the BS sends an ACK feedback, otherwise it sends a NACK. After having received and decoded the feedback, the UE can decide to perform a retransmission.

Layer 1 signaling for (re)configuration and other aspects of link adaptation rather than the power control are not considered here, therefore the UE uses the entire pre-configured bandwidth for its UL data transmissions.

2.2 Problem formulation and Objectives

The objective with power control for the network of URLLC users is to increase the capacity of the system while achieving the URLLC performance requirements. The URLLC performance indicator is the user plane latency and the corresponding reliability of transmitting the packets within a latency target. We adopt the 3GPP baseline reliability target of $1 - 10^{-5}$ with latency of 1 ms [2].

In the considered system, the GF resource allocation can be shared by multiple UEs which makes the GF transmissions susceptible not only to inter-cell interference, but also to intra-cell interference. Power control is an essential mechanism to manage both intra- and inter-cell interference levels [9].

Given the described network, this means that the use of retransmissions should be minimized in order to keep the latency down. Our hypothesis is that power control settings can be tunned to improve the system performance for GF URLLC transmissions. Also, that power boosting retransmissions can reduce the retransmission probability and hence improve the system capacity for URLLC traffic.

3 Power Control with Power Boosting

In LTE, fractional power control is used to regulate the power level of the received signal at the BS, as well as to limit the inter-cell interference. The transmit power P at the UE is determined by the following expression:

$$P[\text{dBm}] = min\{P_{max}, P_0 + 10log_{10}(M) + \alpha PL + \Delta_{mcs} + f(\Delta_i)\}, \qquad \text{(D.1)}$$

where P_{max} is the maximum transmit power, M is the number of assigned Resource Blocks (RBs), P_0 is the target receive power per RB, PL is the downlink path-loss estimate calculated at the UE based on the reference signal power, Δ_{mcs} is a MCS based power offset signaled in the uplink grant, Δ_i is a closed loop correction factor, α is a fractional path-loss compensation factor and $f()$ indicates if closed loop power control are cumulative or absolute commands. The P_0 and α parameters can be cell broadcasted.

The open loop part of the power control is used to compensate for systematic offsets and large scale fading. The effect of the α factor is larger on UEs with higher path-loss which are present at cell-edge, since these UEs are also the ones which contribute the most to the inter-cell interference. The closed loop part of the power control can be used to compensate errors for the UE transmit power and possibly optimize the system performance. The way it is implemented depends on the manufacturer. Closed loop power corrections $f(\Delta_i)$ and Δ_{mcs} will not be further considered in this study.

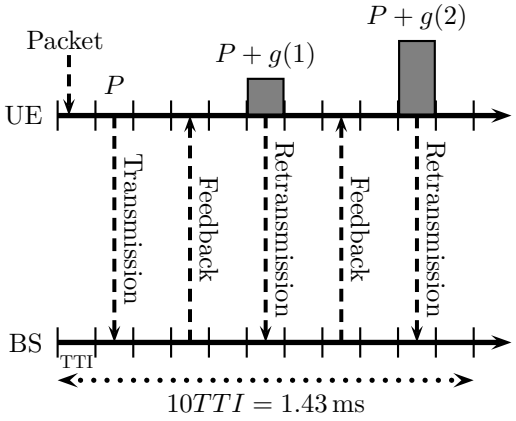


Fig. D.1: URLLC Uplink Grant-Free Transmission with Reactive HARQ and Power Boosting for the retransmissions. P is the transmit power without power boosting and $g()$ indicates the requested power boost.

The considered transmission scheme with power boosting is illustrated in Fig. D.1. In order to reach the 1 ms latency budget, there is only time for two transmission attempts. This means that if the packet is not successfully received in the first attempt, it needs to succeed in the retransmission with a very high probability. Besides using soft combining, the success probability of a retransmission can increase by enhancing the signal level and managing the interference. Like in LTE, power control can be used to manage the inter-cell

interference. And as in CDMA systems, in case the time-frequency resources are shared by multiple UEs, it can also manage intra-cell interference.

To enhance the signal level, power boosting is applied through a mapping function $g(\Delta_{PB})$, where Δ_{PB} is a power boosting index and $g()$ maps the index to a power boosting value PB_{step} in dB and is defined in (D.3). The considered uplink power control algorithm considered in this study then simplifies from (D.1) to the following:

$$P[\text{dBm}] = min\{P_{max}, P_0 + 10log_{10}(M) + \alpha PL + g(\Delta_{PB})\}, \tag{D.2}$$

where $g()$ is defined as:

$$g(\Delta_{PB}) = PB_{step} \cdot \Delta_{PB}. \tag{D.3}$$

This definition of $g()$ works as power ramping of retransmissions as $\Delta_{PB} = 0$ for the initial transmission and hence increment by 1 for each retransmission. This is also illustrated in Fig. D.1, where the value of $g()$ increases at each retransmission attempt. This can be seen as a form of link-adaptation based on the single-bit HARQ feedback. The impact of $g(\Delta_{PB})$ on the transmit power is limited by P_{max}, from (D.2).

4 Simulation Methodology

In this work the effect of power control and power boosting for GF URLLC are evaluated using system level simulations. The simulations permit to study effects that would be difficult or even unfeasible to evaluate all together with analytical models. This includes, inter- and intra-cell interference, queuing and the effects of a time-frequency variant channel. The simulation assumptions are summarized in Table D.1. The used assumptions follow the main guidelines regarding simulation for URLLC defined in [6].

The system layout is an urban macro-cellular network composed by 7 three-sector sites with 500 meters inter-site distance (ISD) including wrap-around [15]. The BS uses a Minimum Mean Square Error Interference Rejection Combining (MMSE-IRC) receiver with 2 antennas. The IRC receiver is capable of suppress inter- or intra-cell interference from a simultaneous transmission. It is assumed that the receiver can ideally estimate the channel of all superimposed transmissions. However, whether it can successfully decode the transmissions depends on the post-detection SINR after interference rejection. The decoding probability for the applied MCS is given by the link-to-system interface which is based on mutual-information effective SNR mapping (MI-ESM). As in the previous work [8], in this study the UEs are deployed only outdoor.

Table D.1: Simulation assumptions

Parameters	**Assumption**
Layout	Hexagonal grid, 7 sites, 3 sectors/site, wrap-around [6]
Propagation scenario	3D Urban Macro (UMa), 500 m ISD
UE distribution	Uniformly distributed outdoor, $3\,\mathrm{km\,h^{-1}}$ UE speed, no handover
Carrier and Bandwidth	4 GHz, 10 MHz (48 RBs) in uplink
PHY numerology	15 kHz sub-carrier spacing, 2 OFDM symbols per TTI, 12 subcarriers/RB
Timing	1 TTI (0.143 ms) to transmit and 1 TTI to process by UE and BS
HARQ configuration	4 TTIs HARQ RTT, 4 SAW channels, maximum 8 HARQ retransmissions
Uplink receiver	MMSE-IRC with 1x2 antenna configuration
Thermal noise density	$-174\,\mathrm{dBm\,Hz^{-1}}$
Receiver noise figure	5 dB
Max UE TX power	23 dBm
Traffic model	FTP Model 3 with 32 B packet and Poisson arrival of 10 PPS per UE
Link adaptation	MCS fixed to QPSK 1/8 and open loop power control
Performance target	1 ms with 10^{-5} outage probability

The system is evaluated at different loads by varying the number of UEs deployed in the network. Each UE generate a small packet of 32 Bytes following a Poisson arrival process with an average of 10 packets per second (PPS). Multiple drops of Monte Carlo simulations are conducted. At each drop the UEs are uniformly deployed in the network and stay connected until the end of the simulation. Initial random access procedures, control signaling errors and reference signal overhead are not considered.

The physical layer numerology and frame structure is inline with 3GPP NR evaluation agreements and uses CP-OFDM with mini-slots of 2 OFDM symbols [6] for transmissions in short TTI (0.143 ms). Grant-free transmissions use all available 48 resource blocks (RB) in a bandwidth of 10 MHz, to transmit the small packet with MCS fixed to QPSK 1/8. The transmissions duration and the processing time are assumed to take 1 TTI, giving a round-trip time (RTT) of 4 TTIs as the time between one transmission can be followed by a retransmission. As in [16], the simulation time is config-

ured to collect at least 5×10^6 samples from several drops to ensure sufficient confidence level on the 10^{-5} quantile.

5 Results

The evaluation is done in two steps: First by focusing on the OLPC parameters P_0 and α, where P_0 is chosen to optimize URLLC performance indicators and secondly, evaluating the gains of using power boosting, which includes selecting suitable PB_{step} values.

5.1 Power control settings

We start by analyzing the OLPC settings for α and P_0 which can satisfy URLLC performance requirements. Fig. D.2 shows the outage probability, namely the probability that the transmissions in the system does not succeed within 1 ms latency target, as a function of P_0. Fig. D.2a is with full path-loss compensation ($\alpha = 1$) and Fig. D.2b is with fractional path-loss compensation ($\alpha = 0.8$). Four different loads are being considered and are defined as the average packet generation rate per second per cell.

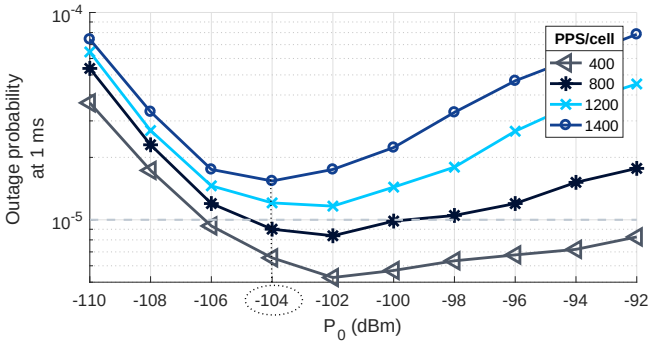
The comparison of fractional and full path-loss compensation is done in two different ranges of P_0 found by an initial sampling of a large P_0 range. It was found that $\alpha = 0.8$ provided the best performance for $-90\,\text{dBm} \leq P_0 \leq -72\,\text{dBm}$, while for $\alpha = 1$ the best range of P_0 is $-110\,\text{dBm} \leq P_0 \leq -92\,\text{dBm}$, i.e. 20 dB offset.

The best choice of P_0 is the one that provides the lowest outage probability. This is load dependent and varies less than 4 dB for the considered loads. It is also clear that the outage probability slope is steeper for P_0 values smaller than the optimum rather than higher. The penalty of being offset from the optimum P_0 becomes more significant when the load increases, meaning that particular for higher loads, it is critical to use a P_0 as close to the optimum as possible.

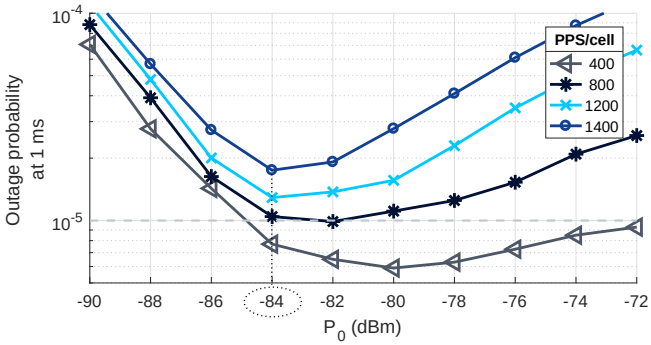
Comparing Fig. D.2a and Fig. D.2b it can be noted that the outage is slightly more sensitive to the P_0 setting for fractional path-loss compensation than for full path-loss compensation. This is due to the higher penalty to cell edge devices caused by fractional path-loss compensation, so operating with optimum P_0 setting becomes more critical in this case.

The choice of P_0 used throughout the rest of the paper is the one that provides the lowest outage probability for the highest considered load (1400 PPS). This is selected to be $P_0 = -104\,\text{dBm}$ for $\alpha = 1$ and $P_0 = -84\,\text{dBm}$ for $\alpha = 0.8$.

Previous work done on LTE, such as the one presented in [17], shows that the optimum setting of P_0 for the system performance in terms of cov-



(a) Full path-loss compensation ($\alpha = 1.0$)



(b) Fractional path-loss compensation ($\alpha = 0.8$)

Fig. D.2: Outage probability at 1 ms as a function P_0 for different traffic loads.

erage and throughput is load dependent. Taking the differences in scenarios and assumptions into account, this tendency is also present in our results, but not as significant as presented in [17]. This is expected to be due to the lack of link-adaptation with adaptive transmission bandwidth, given that the resources allocation and MCS are fixed for the pre-configured GF transmissions.

In the previous work on GF URLLC transmissions schemes [8], similar assumptions were used, but did not consider power control optimizations. The settings used was fractional power control and $P_0 = -85\,\text{dBm}$ with a resulting outage capacity of 400 PPS/cell. In this paper achieves, with the optimized power control parameters, an outage probability at at least 800 PPS/cell corresponding to a 100 % gain. This is even without using power boosted retransmissions. This underlines that deviating from the optimal P_0, particularly when using fractional path-loss compensation, can considerably impact the URLLC network performance.

Table D.2: Power headroom for boosting retransmissions

	Headroom for retransmissions		
	>0 dB	**>3 dB**	**>10 dB**
$\alpha = 0.8, P_0 = -84\,\text{dBm}$	61%	41%	8%
$\alpha = 1.0, P_0 = -104\,\text{dBm}$	35%	31%	16%

5.2 Power boosting evaluation

Fig. D.3 shows the Cumulative Distribution Function (CDF) of used transmit power for packets that were decoded using only one transmission (solid lines) and using more than one transmission (dashed lines), for both fractional and full path-loss compensation with the found optimal P_0 values. The load is 800 PPS per cell which is performing close to the acceptable baseline outage for URLLC (as seen in Fig. D.2).

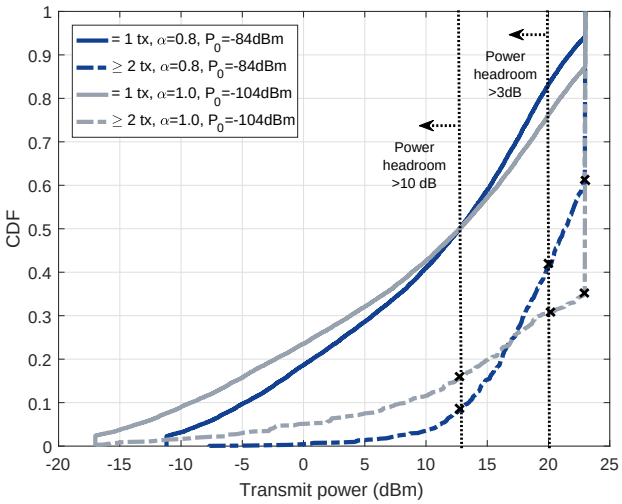


Fig. D.3: CDF of the transmit power according number of required transmissions and power control setting (load of 800 PPS/cell).

First of all it is noted that, for packets succeeding in one transmission, the probability of using full transmit power is relatively small for both $\alpha = 0.8$ ($\leq$ 6 %) and $\alpha = 1$ ($\leq$ 13 %). However, for packets requiring 2 or more transmissions ($\geq 2tx$), the probability of using full transmit power increases to 39 % and 65 % of the cases for $\alpha = 0.8$ and $\alpha = 1$, respectively. This observation matches the intuition that fractional power control allows for a larger power headroom, especially for devices with higher path-loss, i.e. close to the cell edge.

The intention with power boosting is to use some, or all, of the power headroom available after initial transmission, to increase the SINR on the retransmissions. Table D.2 shows the fractions of retransmissions occurrences which have different ranges of power headroom. For instance, taking the case with full path-loss compensation, an aggressive boosting step of 10 dB can be fully applied on approximately 16 % of the retransmission occurrences. While in a moderate configuration, with $PB_{step} = 3$ dB, approximately 31 % of the retransmissions occurrences are boosted with limited step. This can prevent UEs very close to the BS to transmit with very high power. The referred boosting steps of 3 dB and 10 dB are evaluated as values of PB_{step} along with 0 for reference and P_{max} which will cause maximum transmit power for the retransmissions.

It is worth mentioning that, in practice, a very high transmission power from a UE that is closer to the BS can increase the adjacent channel interference. A very strong signal can also overshoot the receiver and suppress the detection of other simultaneous GF transmissions in the same channel. However, such effects are not considered in this study. For this reason, the maximum PB_{step} value is included for completeness of the two extremes of power boosting (0 and P_{max}).

5.3 Performance summary

Having determined a optimal P_0 for fractional and full path-loss compensation and a set of values for PB_{step} it is time to evaluate the resultant performance for the different power control configurations. Fig. D.4 shows the Complementary Cumulative Distribution Function (CCDF) of the one-way latency as a function of PB_{step} for a load of 1200 PPS/cell. The offset between 0 and ~ 0.3 ms is caused by the transmission and processing time. The slope which follows the initial step at 0.4 ms is caused by frame alignment which is a uniform random variable of maximum length of 1 TTI. The steps are caused by the HARQ RTT between the transmissions.

It can be noted that there is just sufficient time for one retransmission in the 1 ms latency budget to reach 10^{-5} outage probability. We can also see, after the slope of the initial transmission, that the retransmission slope starts below the 10^{-3} quantile. This indicates that retransmissions occur very rarely and that power boosting has a very low impact on the interference level.

It is observed that the power boosting reduces the tails of the latency distribution in the very low quantile, i.e. in the region where the performance of the retransmission is observed. The boost of 3 dB has the lowest impact on the tail, while boosting to maximum power does not present a visible difference compared to $PB_{step} = 10$ dB.

Fig. D.5 shows the achieved outage probabilities at 1 ms as a function of the load for the different α, P_0 and PB_{step}. This figure shows clearly

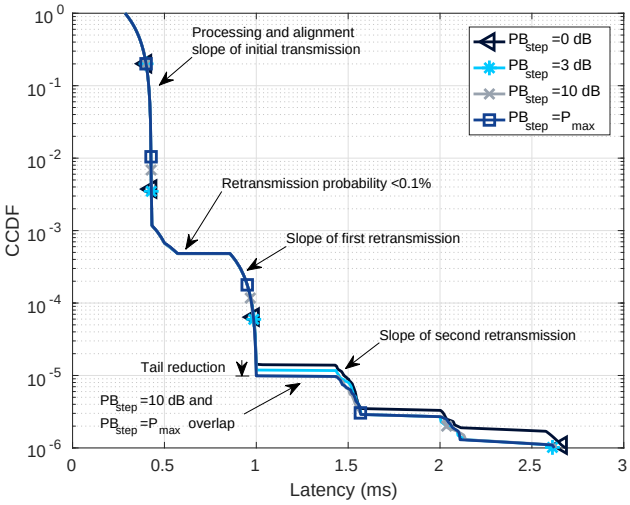


Fig. D.4: Latency CCDFs with 1200 PPS/cell.

that without power boosting the outage capacity is close to 800 PPS/cell for fractional path-loss compensation in accordance to the observations from Fig. D.2. While with optimal power control setting $\alpha = 1$, $P_0 = -104\,\text{dBm}$ and power boosting with $PB_{step} = 10\,\text{dB}$, a load of 1200 PPS/cell is achievable. The $PB_{step} = 3\,\text{dB}$ approaches an achievable load of 1100 PPS/cell. It can be seen that full path-loss compensation is generally providing the lowest outage probabilities.

Also for higher loads such as 1400 PPS/cell, the use of fractional path-loss compensation seems not beneficial, which is likely due to the higher failure probability of packets transmitted from the cell edge. It can be also seen that $PB_{step} = 10\,\text{dB}$ and $PB_{step} = P_{max}$ provides similar performance in all the cases, making the smaller step preferable in practice to lower co-channel and adjacent channel interference.

6 Discussion

In this work we considered GF parameters with fixed MCS configured by higher layers (e.g. RRC). We observed that optimum power control setting is slightly sensitive to the traffic load. A possible inclusion of link adaptation with fast reconfiguration by layer 1 signaling (e.g. Type 2 option in [14]) can modify the allocation bandwidth according to the channel conditions. Then load adaptive power control algorithms like in [17] can be beneficial for network performance.

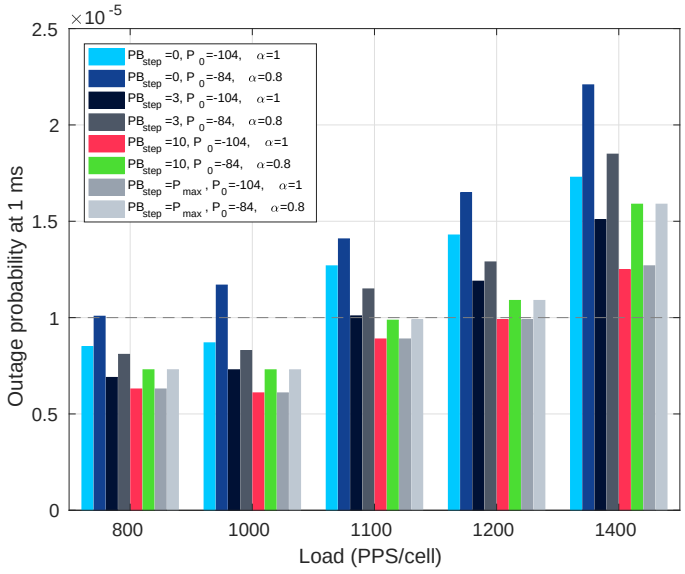


Fig. D.5: Outage at 1 ms for different power control configurations.

In GF transmission the control signaling issues for initial transmission are avoided, nevertheless the reliability of the feedback can still impact on the reactive retransmission. With power boosted retransmission, ACK/NACK false alarms can be more harmful due to possible extra interference from the provoked and boosted retransmissions. Enhancements for the feedback reliability as proposed in [18] can be employed to mitigate such issues.

As in [8], this paper assumes that the BS is capable of doing blind detection of the UEs. Orthogonal reference signals could be used for the channel estimation and UE identification. In a practical implementations the reference signal overhead and its reliability should be taken into account. More complex reception mechanisms could be applied to achieve higher GF URLLC loads. This can include NOMA schemes, and advanced receivers with higher number of antennas for improved interference suppression capabilities.

7 Conclusion

Motivated by the new requirements given for URLLC in 5G, in this paper we studied uplink power control configurations particularly for grant-free transmissions. In order to meet the strict latency and reliability constraints power control should be optimized for URLLC. Further we studied power boosting of retransmissions and evaluated this through extensive system level

simulations. Based on the observations, the take-away messages from this study are;

1. Full path-loss compensation shows better performance and less sensitivity to the choice of P_0 than fractional path-loss compensation.
2. The network performance significantly improves by using optimized power control settings. The system capacity doubles, compared with previous work.
3. The use of power boosting of retransmissions is capable of providing a further outage capacity gain of 20 %.

We emphasize that the success rate of the initial transmission should be high, such that retransmissions occur with a low probability, hence minimizing the excessive interference caused by boosting. Future studies will consider the impact of the feedback errors and the performance of the system with more advanced receivers including higher number of receiver antennas to further improve the URLLC network performance.

Acknowledgment

This research is partially supported by the EU H2020-ICT-2016-2 project ONE5G. The views expressed in this paper are those of the authors and do not necessarily represent the project views.

References

[1] International Telecommunication Union (ITU), "IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond," ITU Radiocommunication Sector, Tech. Rep., Sep. 2015.

[2] 3GPP TR 38.913 v14.1.0, "Study on Scenarios and Requirements for Next Generation Access Technologies," Mar. 2017.

[3] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, I. Riedel, A. Puschmann, A. Mitschele-Thiel, M. Muller, T. Elste, and M. Windisch, "Latency Critical IoT Applications in 5G: Perspective on the Design of Radio Interface and Network Architecture," *IEEE Communications Magazine*, vol. 55, no. 2, pp. 70–78, Feb. 2017.

[4] 3GPP TR 36.881 v14.0.0, "Study on latency reduction techniques for LTE," Jul. 2016.

[5] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy, "Massive machine-type communications in 5g: physical and MAC-layer solutions," *IEEE Communications Magazine*, vol. 54, no. 9, pp. 59–65, Sep. 2016.

[6] 3GPP TR 38.802 v14.0.0, "Study on New Radio Access Technology," Mar. 2017.

[7] H. Kim, Y.-G. Lim, C.-B. Chae, and D. Hong, "Multiple Access for 5G New Radio: Categorization, Evaluation, and Challenges," *ArXiv e-prints, arXiv:1703.09042 [cs.IT]*, Mar. 2017.

[8] T. Jacobsen, R. Abreu, G. Berardinelli, K. Pedersen, P. Mogensen, I. Z. Kovács, and T. K. Madsen, "System Level Analysis of Uplink Grant-Free Transmission for URLLC," in *2017 IEEE Globecom Workshops*, Dec. 2017.

[9] H. Holma and A. Toskala, *WCDMA for UMTS - HSPA Evolution and LTE*, 5th ed. Wiley, 2010.

[10] 3GPP TS 36.213 V14.2.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," Mar. 2017.

[11] C. U. Castellanos, D. L. Villa, C. Rosa, K. I. Pedersen, F. D. Calabrese, P. H. Michaelsen, and J. Michel, "Performance of Uplink Fractional Power Control in UTRAN LTE," in *VTC Spring 2008 - IEEE Vehicular Technology Conference*, May 2008, pp. 2517–2521.

[12] C. Rosa and K. I. Pedersen, "Performance aspects of LTE uplink with variable load and bursty data traffic," in *21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, Sep. 2010, pp. 1871–1875.

[13] B. Soret, P. Mogensen, K. I. Pedersen, and M. C. Aguayo-Torres, "Fundamental Tradeoffs among Reliability, Latency and Throughput in Cellular Networks," in *2014 IEEE Globecom Workshops*, Dec. 2014.

[14] 3GPP TSG RAN WG1 NR Ad-Hoc#2, "RAN1 Chairman's Notes," Jun. 2017.

[15] T. Hytönen, "Optimal Wrap-Around Network Simulation," Helsinki University of Technology, Tech. Rep. A432, Oct. 2001.

[16] G. Pocovi, B. Soret, K. I. Pedersen, and P. Mogensen, "MAC Layer Enhancements for Ultra-Reliable Low-Latency Communications in Cellular Networks," in *2017 IEEE International Conference on Communications Workshops*, May 2017.

[17] M. Boussif, C. Rosa, J. Wigard, and R. Müllner, "Load adaptive power control in LTE Uplink," in *2010 European Wireless Conference (EW)*, Apr. 2010, pp. 288–293.

[18] H. Shariatmadari, Z. Li, S. Iraji, M. A. Uusitalo, and R. Jäntti, "Control Channel Enhancements for Ultra-Reliable Low-Latency Communications," in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2017.

Paper E

Efficient Resource Configuration for Grant-Free Ultra-Reliable Low Latency Communications

Renato Abreu, Thomas Jacobsen, Gilberto Berardinelli, Klaus Pedersen, István Z. Kovács, Preben Mogensen

The paper has been submitted to the
IEEE Transactions of Vehicular Technology, 2019

Abstract

Achieving efficient ultra-reliable low-latency communications (URLLC) in the uplink with strict requirements such as 99.999% reliability in 1 ms is extremely challenging. Grant-free transmission is a promising access method as the error prone scheduling phase and its associated delays are avoided. For aperiodic traffic, sharing of grant-free resources is used to improve resource utilization. However, inter- and intra-cell interference caused by simultaneous transmissions from multiple users can drastically reduce the supported URLLC load in the network. In this paper, a resource configuration scheme for uplink grant-free URLLC is proposed. Our solution includes different modulation and coding scheme (MCS) options associated with sub-bands allocations and power control settings. A simple MCS selection scheme is used to assign higher MCSs and reduced sub-bands to user equipments with high coupling gains, thus reducing the impact from overlapping transmissions. Our grant-free system design includes hybrid automatic repeat request retransmissions along with a mini-slot structure and linear receiver with interference rejection capability for multi-packet reception. The performance of the proposed solution is evaluated in a multi-cell urban scenario. System level results reveal an up to 90% gain on the achievable URLLC load with respect to a single-MCS reference configuration, when a 2 antennas receiver is used. The usage of a 4 antenna receiver further boosts the achievable load by a factor of 7.

At the time of the writing, this paper is still under peer-review and pending decision by the respective publication editor. Regarding the rules for parallel publication, the paper has not been included in the public version of the thesis. The reader is encouraged to contact the author or the referred publication channel for a copy of the paper.

Paper F

Multi-cell Reception for Uplink Grant-Free Ultra-Reliable Low-Latency Communications

Thomas Jacobsen, Renato Abreu, Gilberto Berardinelli, Klaus Pedersen, István Z. Kovács, Preben Mogensen

The journal paper has been published in the
IEEE Access, 2019

Abstract

The fifth generation (5G) radio networks will support ultra-reliable low-latency communications (URLLC). In the uplink, the latency can be reduced by removing the time-consuming and error-prone scheduling procedure and instead use grant-free (GF) transmissions. Reaching the strict URLLC reliability requirements with GF transmissions is, however, particular challenging due to the wireless channel uncertainties and interference from other URLLC devices. As a consequence, the supported URLLC capacity and hence the spectral efficiency is typically low. Multi-cell reception, i.e. joint reception and combining by multiple base-stations (BS) is a technique known from Long Term Evolution (LTE), with the potential to greatly enhance the reliability. This paper propose the use of multi-cell reception to increase the URLLC spectral efficiency while satisfying the strict requirements using GF transmissions in a 5G new radio (NR) scenario. We evaluate the achievable URLLC capacity for an elaborate multi-cell reception parameter space and multi-cell combining techniques. Additionally, we demonstrate that rethinking of the radio resource management (RRM) in the presence of multi-cell reception is needed to unleash the full potential of multi-cell reception in the context of UL GF URLLC. It is observed that multi-cell reception compared to a single-cell reception, can provide URLLC capacity gains from 205% to 440% when the BSs are equipped with two receive antennas and 53% to 22% when BSs are equipped with four receive antennas, depending on whether retransmissions are enabled.

1 Introduction

The first release of the fifth generation (5G) new radio (NR) has been specified by the third generation partnership project (3GPP) in Release 15 [1]. One of the 5G use cases is Ultra-Reliable Low-Latency Communication (URLLC), which pose highly challenging service requirements [2]. URLLC is set to enable numerous services such as the tactile internet [3] and the factory of the future envisioned for the fourth industrial revolution. Here, URLLC enable controllers to wirelessly control actuators/sensors using fast control loops and mobile robots to safely and efficiently perform cooperative tasks [4]. Air-interface latency requirements for URLLC range from 0.5 ms up to 7 ms with reliability requirements from 99.9% to 99.9999%, depending on the considered use-case [5].

Several technology components have been investigated prior to the specification of 3GPP Release 15, for example a new frame-numerology with mini-slots to facilitate short transmission times [6]. Grant-free (GF) access, aka configured grant in 3GPP terminology, is an attractive solution to reduce the uplink latency by removing the time consuming steps of grant-based (GB) scheduling [1, 7]. In particular, sharing of pre-configured GF resources

among multiple User Equipments (UEs) is considered to improve the efficiency for sporadic traffic [8]. However, GF transmissions on shared radio resources are prone to inter- and intra-cell interference, which degrades the URLLC transmission reliability and limits the supported uplink URLLC capacity and hence spectral efficiency [9].

Combining multiple sources of diversity is essential to reach the high URLLC reliability requirements and improve the achieved URLLC capacity [7, 10, 11]. Diversity can be achieved in both time, frequency and the spatial domain. Frequency diversity exploits fading differences in the frequency domain and can be harvested through wide-band transmissions or sub-band channel hopping [10]. The coherence time of the radio channel is typical larger than the latency requirement, but the interference conditions can change per mini-slot. For that reason, transmission diversity through hybrid automatic retransmission request (HARQ) can be used to exploit variations in interference [12]. Further, by soft combining the retransmissions, the coverage can be improved. Spatial diversity exploits fading and interference differences by receiving copies from spatially separated antennas or receivers. It can therefore be obtained through signal combining from multiple receive antennas per base station (BS) and from spatially separated receivers with multi-cell reception [13, 14].

Multi-cell reception is a well-known technique from Long Term Evolution (LTE) Release 11, where it was known as coordinated multi-point (CoMP) reception [15, 16]. CoMP encompasses not only multi-cell combining but also interference aware and avoidance schemes. The latter is, however, not well suited for unpredictable GF transmissions. Multi-cell combining is, on the other hand, well suited for GF traffic. Combining across cells can be based on the exchange of complex in-phase and quadrature (IQ) samples, coded bits and also decoded bits [16]. Multi-cell combining based on IQ samples can be considered as a distributed antenna array system, whose complexity scales with the number of receive antennas and the number of cooperating cells [16]. As an evolution of CoMP, the concept of cloud radio access network (RAN) has been considered. Here, a centralized BS are connected to remote radio heads through a high capacity backhaul [17]. The applicability of cloud RAN and IQ based multi-cell reception is therefore best suited for smaller network deployments with few antennas per cell (i.e. below 6 GHz indoor factory or dense urban networks [18]). The complexity of combining based on coded bit exchange scales with the used modulation rate and the number of user equipments (UEs) participating in multi-cell reception [16]. Combining based on decoded bits is the simplest combining option with the lowest backhaul demands. Its usage is therefore well suited for urban macro deployment with moderate UE densities and networks with a capacity limited backhaul [19]. Multi-cell reception is also utilized in other cellular technologies such as Sigfox and LoRaWAN, targeted for non-latency sensitive

applications and extreme coverage [20, 21].

Multi-cell reception based on IQ sample exchange has been studied for LTE in [22] and when based on coded- and decoded bits in [23], with the purpose of enhancing the network throughput for GB transmissions. A more recent study is found in [14]. However, neither of these studies include the joint contribution of intra-cell and inter-cell interference. The improved signal quality from multi-cell reception leaves room for efficiency enhancements by multi-cell reception aware radio resource management (RRM) techniques. Such technique is presented in [24], which is based on the uplink power control and in [25] where modulation and coding scheme (MCS) selection is used to achieve spectral efficiency improvements. Both techniques are studied for an LTE system with the objective to maximize the average network throughput.

While the basic uplink multi-cell combining techniques are known, it remains to be understood how this can enhance the URLLC capacity, defined as the maximum tolerable aggregated offered traffic load where the challenging URLLC service requirements are still fulfilled in a 5G NR setting with sporadic traffic bursts of latency critical payloads. Additionally, the cost in terms of backhaul throughput at the achieved URLLC capacity with the different multi-cell combining techniques remains to be fully understood.

In this study, we show that rethinking of the RRM operations, can unleash the full performance gains. Specifically, the MCS configuration for the GF transmissions and the power control settings must be optimized to efficiently leverage the performance benefits of uplink multi-cell reception, both from an intra- and inter-cell interference perspective. That is, exploiting uplink multi-cell reception both for improving the robustness towards intra-cell GF collisions and for reducing the generated other-cell interference to help improve the overall URLLC performance. By doing this, we show that the use of multi-cell reception techniques offers significant gains, even when using such techniques for a sub-set of the deployed UEs to strike an attractive balance between URLLC performance benefits and network complexity.

Our conclusions are confirmed by results from advanced system-level simulations where major performance-determining effects of a multi-user multi-cell 5G NR network, with dynamic URLLC traffic, is carefully modeled according to latest industry standard agreements. That is, simulations are based on fully calibrated and recognized underlying mathematical models, allowing us to present statistically reliable results with a high degree of realism and thereby high practical relevance. Especially the physical layer transmitter and receiver chains, the medium access control (MAC) protocol and the associated parameter configurations via radio resource control (RRC) are modeled.

The reminder of this paper is structured as follows. Section 2 sets the scene for the work and presents the scenario, GF configuration and applied

MAC and RRM mechanisms. Section 3 presents the multi-cell reception combining techniques. Section 4 presents two multi-cell reception aware RRM techniques, one based on power control and the other based on MCS selection. Section 5 describes the evaluation methodology and simulation assumptions, followed by Section 6 which presents the performance evaluation. The main findings and take-aways are summarized in Section 7 which also concludes this study.

2 Setting the scene

We consider a multi-cell multi-user 5G NR urban macro network scenario as described in [26]. The network consists of multiple sites with an equal inter-site distance. Each site consists of three BSs forming sectorized cells. The BSs are assumed to be time and frequency synchronized. An average number of U URLLC UEs are deployed uniformly within each cell. The BSs transmit a cell specific reference sequence, which is used by the UEs to estimate the received signal reference power (RSRP). The UEs are assumed to connect to the cell with the highest estimated RSRP. This cell will be denoted the serving or primary cell (p-cell) in this work. Each URLLC UE is assumed to generate small packets of size P with an uncorrelated Poisson arrival process at an average rate λ. The UEs are assumed to be configured by the p-cell through radio resource control (RRC) signaling, with at least one set of periodic reoccurring radio resources for GF transmission. In order to minimize the latency, the periodicity of GF resources is set to be equal to the transmission time interval (TTI), such that all UEs may transmit in any TTI. This means that the average aggregated offered URLLC load per cell becomes $L = \lambda UP$. Each UE is configured with a unique reference sequence which is transmitted with the GF transmission. This aids identification and channel estimation at the receiving BSs.

2.1 GF resource allocation

The BS configures the UEs with at least one GF configurations through RRC signaling. A GF configuration includes the time and frequency radio resources, MCS and the periodicity where these resources are available. These GF configurations may have radio resources which overlay in time and frequency. Only one configuration can be active at a time per UE, but different UEs may have different active configurations. We use a structured resource allocation scheme as proposed in [9, 27] for fixed packet sized GF resource configuration with multiple-MCS options. In this scheme, a GF configuration with MCS_1 occupies a bandwidth of BW resource blocks (RBs), while configuration with a higher order MCS_k, use an overlaying set of radio re-

sources over a sub-band of BW/k RBs. The use of multiple sub-bands in the network for GF transmissions reduces the probability of fully overlapping transmissions, and therefore provides interference diversity [9], but requires that more energy per bit is collected.

2.2 Receiver

All cells are assumed to be equipped with a the 5G NR baseline receiver which is a linear minimum mean square error and interference rejection combining (MMSE-IRC) receiver with M receive antennas [18, 28]. With all UEs and BSs synchronized, the receiver may account for the intra-cell and inter-cell interference when computing the interference covariance matrix. With this, the desired signal can be projected into an $M-1$ dimensional subspace with minimum mean square error. The multiple receiving antennas therefore improves the receiver capabilities to handle interfering signals which is particular beneficial for GF transmission over shared resources [13, 29].

2.3 Retransmissions

A retransmission is triggered upon the reception of feedback from the p-cell. Retransmissions are a proven technique to enhance the reliability [12], but requires that the round trip time (RTT) of to a retransmission fits into the URLLC latency requirement. Retransmissions are supported in 5G NR by HARQ [1, p. 23]. We assume that retransmissions also occur on GF resources, as a GB retransmission requires a separate GB band.

2.4 Latency components

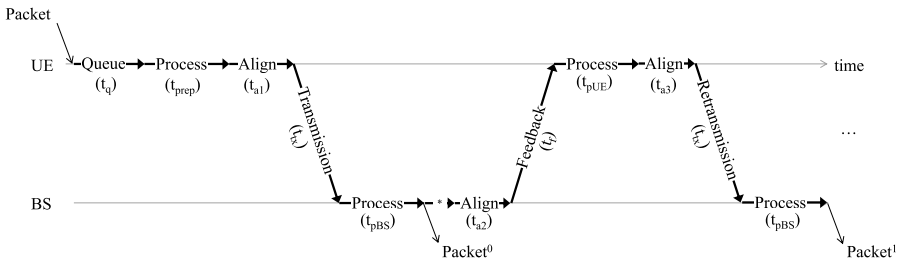


Fig. F.1: Latency components for GF transmission with a following HARQ retransmission.

The latency components involved with a retransmission are illustrated in Fig. F.1. When a URLLC packet arrives at the UE, if the HARQ queue is not empty the packet will be queued during t_q. Then, the packet is prepared for immediate GF transmission (coded and modulated). This step is assumed

to contribute with t_{prep} latency. It is assumed that a GF transmission can only commence at the start of a TTI. This waiting time is denoted as alignment time t_{a1}. The GF transmission delay over the radio interface is defined as $t_{tx} = t_{TTI}$. The BS processes the GF transmissions in t_{pBS}. Depending on whether the packet transmission was successfully decoded by the receiving BS, a positive or negative feedback message is transmitted to the UE. This feedback transmission takes t_f and is carried out after a control channel alignment time t_{a2}. The UE processes the received feedback in t_{pUE} and after another alignment t_{a3} the UE can initiate the HARQ retransmission which also takes t_{tx}.

2.5 Uplink power control

Uplink power control for 5G NR use the following expression to calculate the total UE transmit power [30, p. 14]

$$P_u[dBm] = min\{P_{max}, P_0 + 10log_{10}(2^{\mu} \cdot BW/k) + \alpha \cdot PL + f(.)\}, \qquad \text{(F.1)}$$

where P_{max} is the UE maximum transmit power, P_0 is the target receive power per RB, BW/k is the number of used RBs for the GF transmission with MCS_k and μ is a sub-carrier spacing index [31, p. 9]. α is the path loss compensation factor and PL is the slow faded UE path loss estimate to its p-cell. The term $f(.)$ covers all closed loop terms which are used to apply UE-specific transmit power adjustments to maintain transmission reliability. In [9] this term was used to define an MCS specific offset and in this study it will be used for UE-specific multi-cell reception adjustments. The open loop power control parameters α and P_0 have been shown in [32] to have a substantial influence on the achievable URLLC capacity. In particular, the use of full path loss compensation ($\alpha = 1$) and empirically optimized P_0 as a function of the load and the scenario is demonstrated to be essential. The use of full path loss compensation is well aligned with the power control recommendations with multi-cell reception [16].

2.6 Performance metrics

The main key-performance-indicator (KPI) used in this study is the achievable URLLC capacity, which is defined as the maximum average aggregated offered URLLC load L, where the URLLC service requirements can be fulfilled. The baseline URLLC service requirements set by ITU-2020 [2] is considered, which defines that a URLLC transmission must be delivered from the UE to the BS within 1 ms latency with a minimum reliability of 99.999%. A packet transmission is said to be in outage if it is not received within the latency deadline. The outage probability is defined as the complement of the reliability at a given deadline. The average backhaul load as an indicator of

the backhaul requirements to sustain achieved URLLC loads. The backhaul load is measured as the average data rate over the backhaul used for multi-cell reception. Only backhaul exchanges between different sites are included.

3 Multi-cell reception

In order to enable multi-cell reception, a set of assisting cells needs to be configured for joint reception. In this Section this procedure is described along with the considered multi-cell combining techniques.

3.1 Assisting cell selection

The p-cell may request the UE, through RRC signaling, to report a set of N strongest cells based on RSRP measurements. A set of maximum C_{MAX} assisting cells are selected as a subset of the reported N strongest. Only cells with an RSRP not less than T dB weaker than the p-cell RSRP are selected as assisting cells. Both cells from the same site and cells from different sites can be assisting cells. The p-cell is responsible for configuring the assisting cells and collecting data from them over the backhaul for multi-cell combining.

3.2 Combining schemes

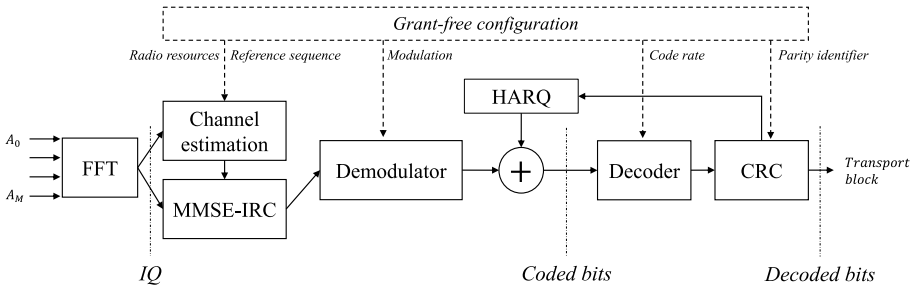


Fig. F.2: Simplified 5G NR uplink OFDM receiver chain for GF transmissions.

Fig. F.2 illustrates a simplified 5G NR uplink receiver chain for GF transmission reception. The signals received from M antennas are sampled and converted from the time domain to the frequency domain. One possibility for multi-cell combining is combining of the frequency IQ samples from the M receive antennas [16]. For each detected GF transmission, the interference covariance matrix is calculated, and MMSE-IRC processing is conducted. Note that joint detection using MMSE-IRC in large cell-clusters has been widely studied in the literature [16, 22]. It is, however, not considered in this work.

After MMSE-IRC processing, the IQ representations of the OFDM symbols are demodulated to estimate the coded bits. Each coded bit is represented by a soft value. Combining based on soft values of coded bits is another option for multi-cell combining. This is sometimes referred to as maximum ratio combining [14] or soft value information combining [23]. This technique is also used for HARQ retransmissions using chase combining [33]. In this work we denote this combining option chase-combining.

After demodulation and potential HARQ combining, the coded bits are decoded. Parity bits are used to check the decoded bits integrity by a cyclic redundancy check (CRC). If the CRC check fails, the coded bits soft representations are saved for combining with the retransmission if HARQ is enabled.

Hard combining, or selection of a successfully decoded packet from one of the assisting cells is a third option of multi-cell combining. We refer to this option as selection-combining. A hybrid of chase-combining and selection-combining is also considered. By using chase-combining for intra-site (with the p-cell) assisting cells and selection-combining for inter-site assisting cells, this hybrid-combining scheme is expected to achieve a URLLC capacity that is between that of chase-combining and selection-combining, while maintaining the backhaul throughput from selection-combining.

4 Multi-cell aware RRM enhancements

Joint processing and combining from multiple cells, improves the signal to interference and noise ratio (SINR) compared to single-cell reception. This benefit from multi-cell reception lead to an improvement of reception reliability. When the experienced reliability is higher than the target, it leaves room for transmission parameter adjustments to relax the reliability and improve the URLLC capacity. Two proposals for multi-cell reception aware RRM enhancements are:

- Closed loop power control (CLPC), to reduce the receive power density target and hence reduce the transmit power P_u.
- MCS selection, to transmit the GF data with a higher order MCS based on the experienced improvement.

The aim with the proposed CLPC strategy is to reduce the generated interference by relaxing the UE receive power density target (P_0), when the SINR exceeds a predefined target. This target needs to account for a margin for transmission collisions and fading. The MCS selection strategy attempts to make use of the improved signal quality, to increase the MCS order and hence increase the spectral efficiency.

For sporadic GF transmissions, the p-cell cannot a-priori acquire uplink channel states and set the optimal transmission parameters. As an alterna-

tive, we use the post-combining experienced SINR as an indicator for adjusting the transmission parameters.

We propose to define the closed loop function $f(.)$ from (F.1) as

$$f(\Gamma_s)\ [dB] = \begin{cases} \Gamma_d - \Gamma_s, & \text{if } \Gamma_s > \Gamma_d \\ 0, & \text{otherwise} \end{cases}, \tag{F.2}$$

where Γ_s is the experienced post-combining SINR and Γ_d is a predefined desired post-combining SINR. As the instantaneous post-combining SINR is subject to fading and sporadic interference, Γ_s is obtained by low-pass filtering the instantaneous SINR samples. Intentionally, this strategy does not allow an increase in transmission power, as it targets to effect UEs which systematically obtain SINR enhancements from multi-cell reception.

The MCS selection strategy is inspired from the study in [9] and the framework for resource allocations is already described in Section 2. The idea is that UEs operating with higher order MCS options occupy a smaller bandwidth BW/k given their higher spectral efficiency. Further, such UEs have the option of selecting different sub-bands, hence reducing the probability of having fully overlaying GF transmissions. However, the usage of higher order MCS requires a higher SINR to maintain the transmission reliability.

The choice of using a default MCS_1 or a higher order MCS_k, is done by comparing Γ_s with a threshold Γ_t such that the chosen MCS is given by

$$\text{MCS} = \begin{cases} \text{MCS}_k, & \text{if } \Gamma_s > \Gamma_t \\ \text{MCS}_1, & \text{otherwise} \end{cases}. \tag{F.3}$$

A too low Γ_t may jeopardize the future GF transmission reliability if the SINR cannot reach the transmission reliability with the higher order MCS. On the other hand, a too high Γ_t is a missed opportunity to increase the spectral efficiency and eventually a chance of increasing the URLLC capacity.

5 Evaluation methodology

The performance evaluation is designed to be of high practical relevance and is therefore based on extensive simulations using an advanced system level simulator designed for providing a high degree of realism. The simulations are based on widely recognized mathematical models, calibrated to industry standards. This simulator includes detailed models of the physical layer procedures, the transmitter and receiver chains, the MAC protocol, MCS selection, power control, UE measurements and cell selection. The simulation assumptions are aligned with 5G NR evaluation methodology for URLLC [18, p. 112] and are summarized in Table F.1.

Table F.1: Simulation assumptions

Parameters	Assumption
Layout	Hexagonal grid with $C = 21$ cells distributed at 7 sites, world wrap-around
Inter-site distance	500 m
Carrier-bandwidth	4 GHz
Channel model	3D Urban Macro (UMa)
UE distribution	Uniformly distributed outdoor, 3 km/h quasi-static fading model
UE transmitter	23 dBm, 1 omni-directional transmit antenna
BS receiver	MMSE-IRC, single panel with 1 or 2 columns and two polarizations to acquire $M = 2$ or $M = 4$ receive antennas
Noise figure	5 dB
Thermal noise	-174 dBm/Hz
Bandwidth	10 MHz, FDD
PHY numerology	30 kHz sub-carrier spacing, 4 symbols/TTI, 12 sub-carriers/RB,
GF configuration	4 symbol periodicity, 24 RB
Traffic model	FTP Model 3 with 32 B packet and Poisson arrival rate of $\lambda = 10$ PPS per UE
Power control	Open loop power control (α=1), scheme and load optimized P_0 and $f(.)$ from (F.2)
MCS selection	MCS_1 = QPSK1/8 as baseline. Optional selection of MCS_4 = QPSK1/2 using (F.3)
Timing	In symbols; $t_{tti} = 4$ (0.143 ms), $t_{tx} = t_{tti}$, $t_{a1} = [0, t_{tti}]$, $t_{a2} = 1$, $t_{a3} = 0$, $t_{prep} = t_{pBS} = t_{pUE} = 3$
HARQ	Maximum of 4 retransmissions when enabled
Soft value	5 bit per coded bit

The simulated network consists of $C = 21$ synchronized BSs distributed over 7 sites with an inter-site distance of 500 m. The BS directional antenna consists of a single antenna panel with one or two cross-polarized antennas to acquire a total of two or four receive antenna elements. The antenna gains is modeled according to [34]. The UEs are uniformly distributed outdoors within the network. Wrap-around is used to avoid world-edge effects. The channel model follows the 3D urban macro model [26, p. 12]. The UEs are semi-stationary with 3 km/h speed for fast fading calculations. The carrier frequency is 4 GHz with an uplink bandwidth of 10 MHz, using frequency division duplex (FDD). The sub-carrier spacing (SCS) is assumed to be 30 kHz

($\mu = 1$) which results in 24 RB with 12 sub-carriers per RB following 5G NR numerology [31, p. 9]. A mini-slot is assumed to consist of 4 OFDM symbols (0.143 ms). GF transmission opportunities are available in every mini-slot and denotes a TTI. Propagation delays are assumed to be negligible. However, it is noted that a longer cyclic prefix than the default 5G NR cyclic prefix of 2.38 ţs for 30 kHz SCS might be needed for assisting cells located further away than the second tier from the p-cell. The UEs are assumed to be time-aligned with the p-cell.

The BSs are configured to transmit a cell-specific reference sequence with a periodicity of 100 ms. The UEs estimates the wide-band RSRP measurement based on the reference sequences. The RSRP measurements are low-pass filtered using a moving average (MA) filter, averaging over the 20 most recent samples. These filtered RSRP value is signaled to the p-cell which configures the assisting cells.

Each URLLC UE generate a $P = 32$ B uplink packet according to a homogeneous Poisson Point Process with an average generation rate of $\lambda = 10$ packets per second (PPS) per UE. The load in the network is configured by adjusting the number of UEs U. Deployed UEs are ideally synchronized with the network and are configured with shared GF configurations. The UEs are configured with GF configurations which use either MCS_1 or MCS_4 when the MCS-selection scheme is used. We choose $\text{MCS}_1 = \text{QPSK}1/8$ (QPSK with code rate $1/8$) and $\text{MCS}_4 = \text{QPSK}1/2$ (QPSK with code rate $1/2$) as used in recent work [9]. The UEs are configured to use MCS_4 when the threshold $\Gamma_s > \Gamma_t$ according to F.3. When this scheme is not enabled, MCS_1 is used.

The BSs are equipped with an MMSE-IRC receiver following the model presented in [28, 35]. BS c (a p-cell or an assisting cell) equipped with M receive antennas calculate the receiver filter g_c for a desired signal from UE u as

$$g_c = H_{u,c}^H R^{-1}, \tag{F.4}$$

where $(.)^H$ is the Hermitian operator, $H_{u,c} \in C^{M \times 1}$ is the channel matrix of the desired signal from UE u to BS c and R is the is the IRC interference covariance matrix given by

$$R = P_u H_{u,c} H_{u,c}^H + \sum_{i \in I} P_i H_{i,c} H_{i,c}^H + \sigma_{n,c}^2, \tag{F.5}$$

where $i \in I$ denotes interfering signals from the set of simultaneously transmitting UEs, $H_{i,c}$ is the channel matrix from UE i to BS c, P_u and P_i are the transmit powers from UE u and i respectively and $\sigma_{n,c}^2$ is the total background noise power received by BS c. All UEs are assumed to be uniquely identified based on its reference sequence, as a demodulation reference sequence (DMRS). The DMRS is also assumed to be used to acquire an ideal channel estimate from all simultaneously transmitting UEs when calculating the interference covariance matrix R.

The post-receiver instantaneous SINR from UE u to a receiving BS c can then be expressed as

$$\Psi_{u,c} = \frac{\Omega_{u,c} \left\| g_c H_{u,c} \right\|^2 P_u}{\sum_{i \in I} \Omega_{i,c} \left\| g_c H_{i,c} \right\|^2 P_i + \sigma_{n,c}^2}, \tag{F.6}$$

where $\Omega_{u,c}$ and $\Omega_{i,c}$ denotes the large scale fading from UE u and i respectively. The SINR value $\Psi_{u,c}$ is calculated for each sub-carrier and each symbol and then combined to an effective SINR in the mutual information domain, following the models presented in [36, 37]. The combining of soft bits which is used for the chase-combining multi-cell combining technique and for retransmissions follows the chase-combining principle [33]. Chase combining is modeled by a linear summation of the obtained effective SINRs and can be expressed as

$$\Phi_{u,E} = \sum_{s \in E} \Phi_{u,s} \eta^s, \tag{F.7}$$

where $\Phi_{u,s}$ denotes the effective SINR of transmission s for device u and $\eta \in \mathbb{R}_{0<x\leq 1}$ denotes the combining efficiency which is set to 1 in this work. The selection-combining is modeled as a selection of the maximum effective SINR which can be expressed as

$$\Phi_{u,E} = max_{s \in E} \left(\Phi_{u,s} \right). \tag{F.8}$$

A link-to-system model obtained through extensive link level simulations is then used to determine the error probability for the transmission depending on the $\Phi_{u,E}$ and the applied MCS.

Uplink power control with full path loss compensation ($\alpha = 1$) is used, based on the findings in [32]. The uplink power control parameter P_0 which leads to the maximum observed URLLC capacity is selected for each combination of M, multi-cell combining scheme, HARQ and the aggregated offered URLLC load L. Identifying the optimum uplink power control parameter values when interference is present, is a well-known NP-hard problem [38], and hence simulation based sensitivity studies are used to find the optimum values. A maximum interval of 2 dB is used for sensitivity studies of P_0 and a maximum interval of 5% is used when conducting a sensitivity study on the maximum L where the URLLC requirements can be satisfied.

The value of the SINR thresholds Γ_d and Γ_t from (F.2) and (F.3), depends on the used MCS, the reliability target, L and the experienced intra- and inter-cell interference. Sensitivity studies are conducted to determine the values of Γ_d or Γ_t that maximize the URLLC capacity. The filter chosen for Γ_s is a MA filter over the past 20 post-combining effective SINR samples ($\Phi_{u,E}$), collected with a periodicity of 100 ms.

Statistically reliable results are ensured by multiple Monte Carlo simulations drops. A total of 5 million samples (1 per generated GF packet) are collected to reliably determine the latency and outage probability relation. This

gives a theoretical statistical confidence interval (assuming Gaussian residuals) of 27% around the 10^{-5} quantile with 95% confidence [39].

Assuming the UE and BS processing times from [40] and that the URLLC packet arrives at the worst time instance, the latency estimate for the initial GF transmission is 3.25 TTIs which corresponds to 0.6 ms. When adding a HARQ retransmission, the latency increases to 6.25 TTIs, corresponding to 0.9 ms. The additional delay from queuing is captured with the system level simulations, but the additional delays from backhaul and the multi-cell combining processing delays are omitted for simplicity.

Backhaul load calculations only account the exchanged data between sites. For chase-combining each receiving BS transfers 2 coded bits per OFDM symbol due to QPSK modulation. Each coded bit is represented by a soft value which is assumed to be 5 bits [16]. With MCS_1 this gives 11520 b per GF transmission. With selection and hybrid-combining, only the URLLC packet is exchanged, which is 32 B or 256 b.

6 Performance evaluation

The performance evaluation is structured into three parts. In the first part, we examine the trade-off between the achievable URLLC capacity and backhaul load using chase-, hybrid- and selection-combining for different choices of the RSRP window T and the maximum number of assisting cells C_{MAX}. Based on the findings from the first part, a T and C_{MAX} is selected and used for the remaining parts. In the second part the maximum achievable URLLC capacity with multi-cell reception is quantized when each BS is equipped with $M = \{2, 4\}$ receive antennas per BS and with the use of retransmissions. In the third part, the performance of the two proposed multi-cell aware RRM enhancement schemes based on either CLPC or MCS-selection is examined.

6.1 Trade-off between URLLC capacity and backhaul load

The probability distribution of a URLLC UE having either 0 (single-cell), 1, 2 or 3 configured assisting cells as a function of the RSRP threshold T is shown in Fig. F.3. It is observed that when increasing T, cells with larger RSRP differences can be selected as assisting cells and the average number of configured assisting cells increases. It is also noticed that by increasing T, the probability of having 0 assisting cells decreases. These devices are typically located at the cell-center which are served only by their p-cell.

Fig. F.4 shows the achievable URLLC capacity for the considered multi-cell reception combining schemes with parameters $C_{MAX} = \{1, 2, 3\}$ and $T = \{4, 6, 8, 10, 12\}$ dB. $M = 4$ receive antennas is used and retransmissions are not enabled. The corresponding empirically optimized power control pa-

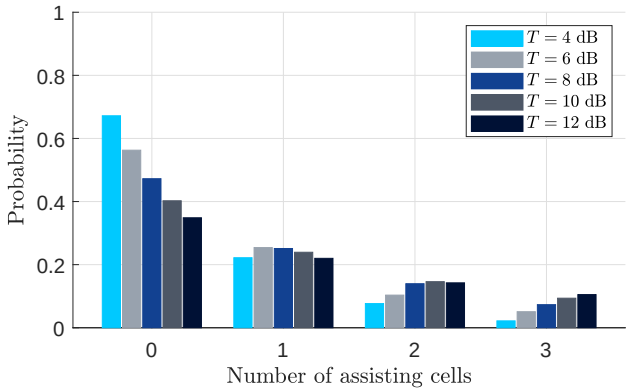


Fig. F.3: Probability distribution of the number of configured assisting cells per URLLC UE.

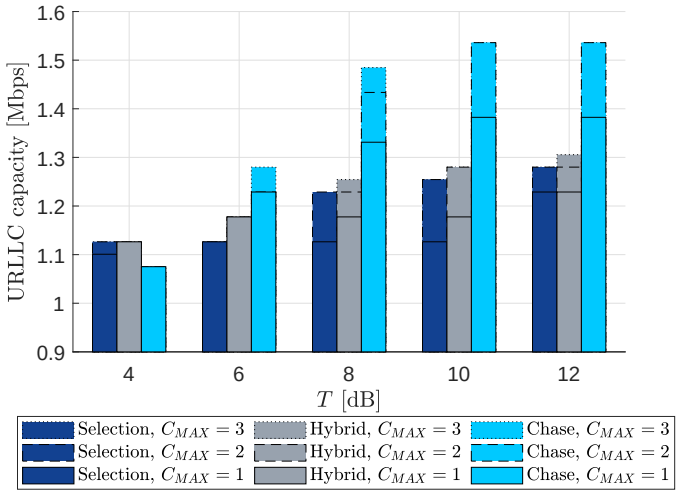


Fig. F.4: The maximum achievable URLLC capacity as a function multi-cell parameters T and C_{MAX}. Each BS is equipped with $M = 4$ receive antennas and retransmissions are not enabled.

rameters are selected using $T = 8$ dB and $C_{MAX} = 2$, which are found to be $P_0 = \{-101, -101, -98\}$ dB for selection-, hybrid-, and chase-combining respectively. The achievable URLLC capacity is observed to generally increase with T and C_{MAX} for all three combining schemes. The largest URLLC capacity enhancement by increasing C_{MAX} and T is observed when using chase-combining. The impact of increasing the maximum number of assisting cells is more evident at $T > 8$ dB, where changing C_{MAX} from 1 to 2 results in up to 10% increase in URLLC capacity. The impact of increasing C_{MAX} from 2 to 3 maximum assisting cells is almost indistinguishable.

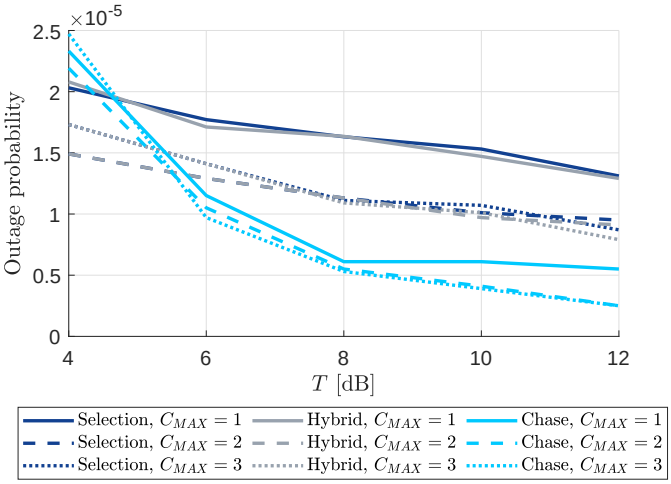


Fig. F.5: Outage probability with $L = 1.28$ Mbps aggregated URLLC load.

Fig. F.5 shows the outage probability for an aggregated URLLC load of $L = 1.28$ Mbps for the same parameter space used in Fig. F.4. This further strengthens the observations drawn from Fig. F.4. A clear reduction in the outage probability is observed when increasing T and C_{MAX}, and the largest reduction is observed for $C_{MAX} > 1$ and for chase-combining $T > 8$ dB. It is also worth to notice the similarities when using selection- and hybrid-combining, and their generally worse URLLC capacity compared to chase-combining. For $T = 4$ dB the outage probability for chase-combining is higher than for selection- and hybrid-combining. The reason for this is likely due to the different power control parameters.

Fig. F.6 shows the corresponding backhaul load for the same parameter space used in Fig. F.4 and Fig. F.5. Firstly, it is observed that the backhaul load difference between selection- or hybrid-combining and chase-combining is almost two orders of magnitude, where the corresponding URLLC capacity difference is in the order of 50%. Secondly, the backhaul load increases from $T = 4$ dB to $T = 12$ dB almost by a factor of 6 for all combining schemes, with a corresponding URLLC capacity increase of 16% for selection- or hybrid-combining and up to 43% for chase-combining. It is observed that the majority of the observed URLLC capacity gains (32% of maximum 43% for chase and 12% of 16% for selection- and hybrid-combining) is achieved with $T = 8$ dB and with $C_{MAX} = 2$ assisting cells. These parameters are used for the remaining two parts of the performance evaluation.

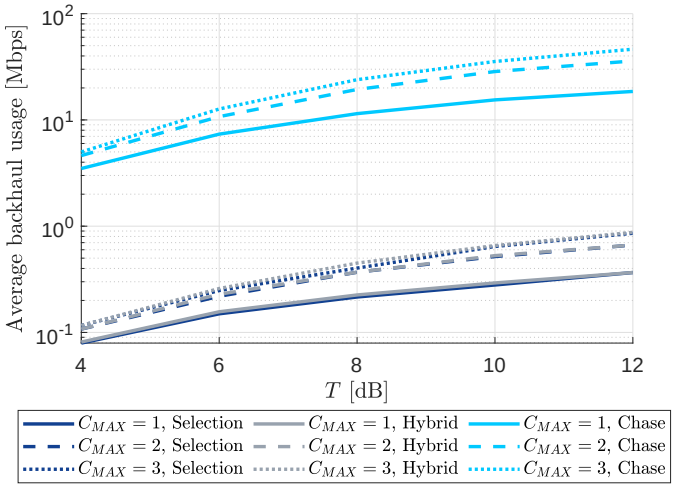


Fig. F.6: Backhaul load for the parameter space and URLLC capacity from Fig. F.4.

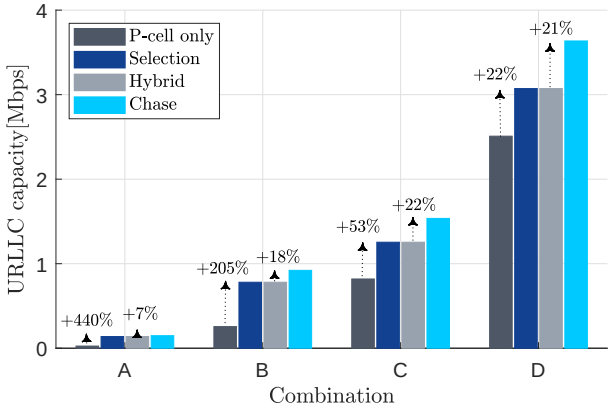


Fig. F.7: Achievable URLLC capacity for multi-cell reception combining options and parameter combinations given in Table F.2.

6.2 URLLC capacity summary

Fig. F.7 shows the achievable URLLC capacity with parameters $C_{MAX} = 2$ and $T = 8$ dB, for the three combining schemes and with four combinations of retransmissions and receive antennas per BS, labeled according to Table F.2. For reference, a configuration where URLLC UEs are only served by a single-cell (the p-cell) is also included. Notice that combination B corresponds to the one used in the first part of the performance evaluation.

It is observed that multi-cell reception improve the URLLC with $M = 2$

Table F.2: Combination labels

Combination	Receive antennas	Retransmissions
A	$M = 2$	Disabled
B	$M = 2$	Enabled
C	$M = 4$	Disabled
D	$M = 4$	Enabled

receive antennas is used by 205% when retransmissions are enabled (combination B) and 440% when retransmissions are disabled (combination A). When $M = 4$ receive antennas are used, multi-cell reception is observed to improve the URLLC capacity by 22% when retransmissions are enabled (combination D) and 53% when retransmissions are disabled (combination C). The obtained gains indicate that, despite the use of retransmission and increased spatial diversity from multiple receive antennas, there is still room for significant enhancements by jointly receiving GF transmissions at multiple BS. Thirdly, it is observed that the additional improvement in terms of URLLC capacity by using chase-combining compared to selection- or hybrid-combining for combination B, C and D is in the range from 7% to 22%. The smallest improvement is observed for combination A, which also achieves the lowest URLLC capacity overall.

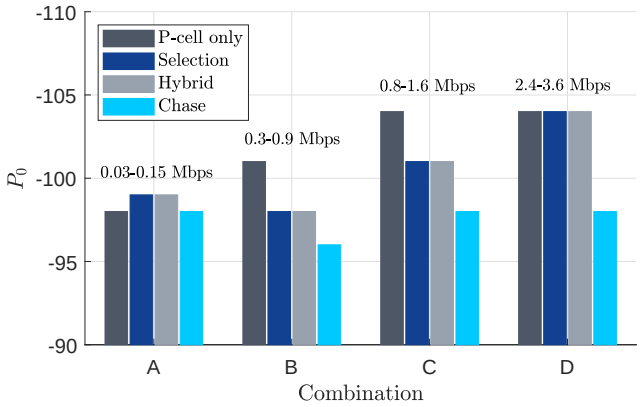


Fig. F.8: Corresponding empirically optimized P_0 value used to generate Fig. F.7.

Fig. F.8 includes the corresponding P_0 values as used in Fig. F.7. When the UEs are served only by a single-cell (p-cell), the identified optimal P_0 decreases with the URLLC capacity, which is in line with related observations on the optimum choice of P_0 [32]. This tendency is also clear with multi-cell reception with hybrid- or selection-combining, but when chase-combining

is used this tendency is not present. One explanation is the usefulness of the reception of transmissions in assisting cells. For selection- and hybrid-combining, these transmissions can be considered useful only if the transmission can be decoded by the assisting cell. With chase-combining, the received transmission at an assisting-cell is useful even if it cannot be decoded at the assisting cells alone. As a consequence, the highest URLLC capacity is observed with chase-combining and selection- and hybrid-combining is observed to perform indistinguishably, contrary the expectation.

6.3 Multi-cell aware RRM enhancements

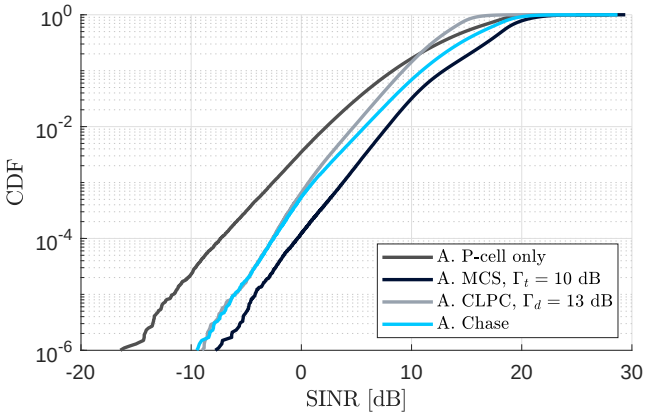


Fig. F.9: SINR CDF for the two proposed RRM enhancement schemes evaluated for combination A ($L = 0.03$ Mbps) defined in Table F.2.

Finally, the performance of the two proposed multi-cell reception aware RRM enhancement schemes are evaluated. Fig. F.9 and Fig. F.10 show the cumulative distribution function (CDF) of the SINR for combinations A and C, respectively, with single-cell reception, multi-cell reception (chase-combining) and with the CLPC and MCS schemes on top. The loads for the combinations are chosen from Fig. F.7, being $L = 0.03$ Mbps for combination A and $L = 0.8$ Mbps for combination C. The choices of Γ_d used in (F.2) and Γ_t used in (F.3) are those which is identified to provide the maximum URLLC capacity. A significant SINR enhancement is observed by enabling multi-cell reception with chase-combining, as also noticed in the initial part of the performance evaluation. Additionally, it is observed that for combination A, the MCS-selection scheme is capable of providing 2-3 dB SINR improvement. The CLPC based scheme, though reducing the probability of experiencing a high SINR (> 3 dB), it slightly improves the SINR at low quantiles ($\leq 10^{-5}$). Based on these observations the two RRM enhancements schemes are expected to

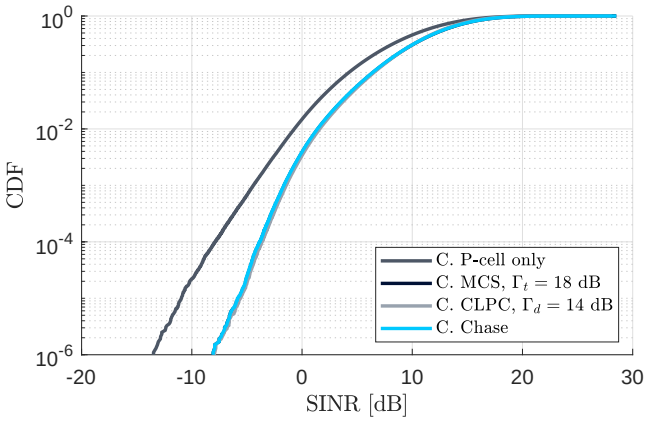


Fig. F.10: SINR CDF for the two proposed RRM enhancement schemes evaluated for combination C ($L = 0.8$ Mbps) defined in Table F.2.

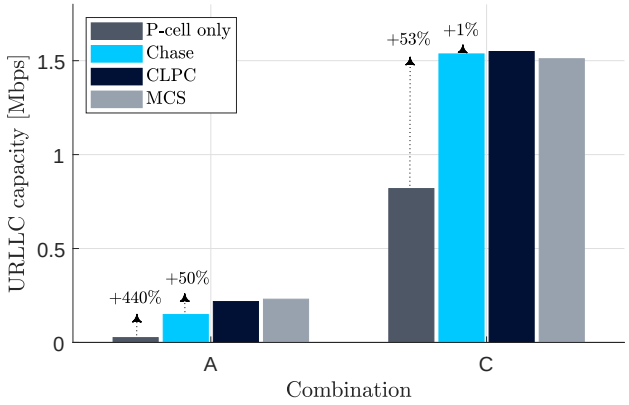


Fig. F.11: Achieved URLLC capacity with the two proposed RRM enhancement schemes evaluated for combination A and C defined in Table F.2.

provide a URLLC capacity gain for combination A. For combination C, no obvious improvement in the SINR tail is observed for the two RRM enhancement schemes.

Fig. F.11 shows the achieved URLLC capacity for combination A and C with single-cell reception, multi-cell reception (chase-combining) and with the CLPC and MCS-selection scheme on top. We observe gains of CLPC and MCS in combination A, with up to 50% URLLC capacity increase. For combination C, using $M = 4$ receive antennas, the gains of the schemes are negligible. With $M = 4$, the RRM enhancement schemes intended for the

$\approx$ 50% deployed URLLC devices who are benefiting from multi-cell reception (from Fig. F.3), are not giving a positive effect on the overall experienced URLLC reliability in the network. This indicates that the reliability bottleneck is to be found within the remaining 50% single-cell served UEs.

7 Conclusion

In this study we have proposed and studied the potential of applying multi-cell reception as technique to improve the URLLC capacity for UL GF URLLC transmissions in a 5G NR scenario. Detailed insights into the sensitivity of multi-cell reception parameters and combining techniques are provided on the URLLC capacity and the backhaul throughput. On top, two multi-cell reception aware RRM schemes have been proposed. Performance evaluations are conducted with advanced system level simulations to provide a high level of realism in a multi-user multi-cell NR network. The main findings are summarized as:

- Multi-cell reception can provide substantial gains in URLLC capacity when compared to single-cell reception. Even the simplest multi-cell combining technique, referred to as selection-combining, is observed to provide capacity gains from 205% to 440% when BSs are equipped with two receive antennas and 53% to 22% when the BSs are equipped with four receive antennas and depending on whether HARQ retransmissions are used. Soft multi-cell combining gives additional capacity gains of +7% to +21%.

- A large improvement is observed by allowing 2 instead of 1 assisting cells per UE, but no benefit is found when increasing to 3 allowed assisting cells. The largest URLLC capacity gains are observed for RSRP thresholds of 8-10 dB.

- While soft combining may provide the largest URLLC capacity gains with multi-cell reception, it also requires almost two orders of magnitude larger backhaul load and requires networks with high backhaul capacity.

- The full URLLC capacity gains can be achieved on top of soft combining with the proposed multi-cell aware RRM enhancements based on closed-loop power control or MCS-selection. The highest gains are observed in low diversity order configurations.

Future work will study the potential of multi-cell reception for GF in an indoor factory scenario where higher backhaul capacity can be assumed along with the use of even more receive antennas per cell. In this scenario, the

IQ based multi-cell reception technique becomes interesting and should be studied. Additionally, the potential of increasing the SCS and the bandwidth to reduce the mini-slots duration and further reduce the latency should be studied. With shorter mini-slot durations, the use of dynamically scheduled transmissions can be studied as a technique to achieve even higher URLLC capacities for comparable URLLC latency and reliability requirements.

Acknowledgments

This research is partially supported by the EU H2020-ICT-2016-2 project ONE5G. The views expressed in this paper are those of the authors and do not necessarily represent the project views.

References

[1] 3GPP TS 38.300 V15.2.0, "NR; NR and NG-RAN Overall Description," Jun. 2018.

[2] ITU-R, "Report ITU-R M.2410-0 - Minimum requirements related to technical performance for IMT-2020 radio interface(s)," International Telecommunication Union (ITU), Tech. Rep., Nov. 2017.

[3] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5G-Enabled Tactile Internet," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 460–473, Mar. 2016.

[4] 3GPP TR 22.804 v2.0.0, "Study on Communication for Automation in Vertical Domains," May 2018.

[5] 3GPP TR 38.824 v1.0.0, "Study on physical layer enhancements for NR ultra-reliable and low latency case (URLLC)," Nov. 2018.

[6] K. I. Pedersen, G. Berardinelli, F. Frederiksen, P. Mogensen, and A. Szufarska, "A Flexible 5G Frame Structure Design for Frequency-Division Duplex Cases," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 53–59, Mar. 2016.

[7] P. Popovski, J. J. Nielsen, C. Stefanovic, E. d. Carvalho, E. Strom, K. F. Trillingsgaard, A. S. Bana, D. M. Kim, R. Kotaba, J. Park, and R. B. Sørensen, "Wireless Access for Ultra-Reliable Low-Latency Communication: Principles and Building Blocks," *IEEE Network*, vol. 32, no. 2, pp. 16–23, Mar. 2018.

[8] 3GPP TR 36.881 v14.0.0, "Study on latency reduction techniques for LTE," Jul. 2016.

[9] T. Jacobsen, R. B. Abreu, G. Berardinelli, K. I. Pedersen, I. Kovács, and P. E. Mogensen, "Joint Resource Configuration and MCS Selection Scheme for Uplink Grant-Free URLLC," in *2018 IEEE Globecom Workshops*, 2018, (Accepted/in press).

[10] C. She, C. Yang, and T. Q. S. Quek, "Radio Resource Management for Ultra-Reliable and Low-Latency Communications," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 72–78, Jun. 2017.

[11] N. A. Johansson, Y. P. E. Wang, E. Eriksson, and M. Hessler, "Radio Access for Ultra-Reliable and Low-Latency 5G Communications," in *IEEE ICC Workshop (ICCW)*, Jun. 2015.

[12] T. Jacobsen, R. Abreu, G. Berardinelli, K. Pedersen, P. Mogensen, I. Z. Kovács, and T. K. Madsen, "System Level Analysis of Uplink Grant-Free Transmission for URLLC," in *2017 IEEE Globecom Workshops*, Dec. 2017.

[13] G. Berardinelli, R. Abreu, T. Jacobsen, N. H. Mahmood, K. Pedersen, I. Z. Kovács, and P. Mogensen, "On the Achievable Rates over Collision-Prone Radio Resources with Linear Receivers," in *2018 IEEE 29th PIMRC*, Sep. 2018.

[14] A. Wolf, P. Schulz, D. Oehmann, M. Doerpinghaus, and G. Fettweis, "On the Gain of Joint Decoding for Multi-Connectivity," in *2017 IEEE Global Communications Conference*, Dec. 2017, pp. 1–6.

[15] 3GPP TR 36.819 v11.2.0, "Coordinated multi-point operation for LTE physical layer aspects," Sep. 2013.

[16] P. Marsch and G. P. Fettweis, *Coordinated Multi-Point in Mobile Communications: From Theory to Practice*, 1st ed. Cambridge University Press, 2011.

[17] A. Karimi, K. I. Pedersen, N. H. Mahmood, J. Steiner, and P. Mogensen, "5G Centralized Multi-Cell Scheduling for URLLC: Algorithms and System-Level Performance," *IEEE Access*, vol. 6, pp. 72 253–72 262, Dec. 2018.

[18] 3GPP TR 38.802 v14.0.0, "Study on New Radio Access Technology," Mar. 2017.

[19] 3GPP TS 22.261 v16.5.0, "Service requirements for the 5G system," Sep. 2018.

[20] Sigfox, "Sigfox main website," Apr. 2019. [Online]. Available: https://www.sigfox.com

[21] J. de Carvalho Silva, J. J. P. C. Rodrigues, A. M. Alberti, P. Solic, and A. L. L. Aquino, "LoRaWAN ? A low power WAN protocol for Internet of Things: A review and opportunities," Jul. 2017.

[22] C. Hoymann, L. Falconetti, and R. Gupta, "Distributed Uplink Signal Processing of Cooperating Base Stations Based on IQ Sample Exchange," in *2009 IEEE International Conference on Communications*, Jun. 2009, pp. 1–5.

[23] L. Falconetti, C. Hoymann, and R. Gupta, "Distributed Uplink Macro Diversity for Cooperating Base Stations," in *2009 IEEE International Conference on Communications Workshops*, Jun. 2009, pp. 1–5.

[24] Y. Ding, D. Xiao, and D. Yang, "Performance analysis of an improved uplink power control method in LTE-A CoMP network," in *2010 3rd IEEE International Conference on Broadband Network and Multimedia Technology (IC-BNMT)*, Oct. 2010, pp. 624–628.

[25] A. Muller, P. Frank, and J. Speidel, "Performance of the LTE Uplink with Intra-Site Joint Detection and Joint Link Adaptation," in *2010 IEEE 71st Vehicular Technology Conference*, May 2010, pp. 1–5.

[26] 3GPP TR 38.913 v14.1.0, "Study on Scenarios and Requirements for Next Generation Access Technologies," Mar. 2017.

[27] T. Jacobsen, R. Abreu, G. Berardinelli, K. Pedersen, I. Kovács, and P. E. Mogensen, "System Level Analysis of K-Repetition for Uplink Grant-Free URLLC in 5G NR," 2019.

[28] K. Pietikainen, F. D. Carpio, H. L. Maattanen, M. Lampinen, T. Koivisto, and M. Enescu, "System-Level Performance of Interference Suppression Receivers in LTE System," in *2012 IEEE 75th Vehicular Technology Conference (VTC Spring)*, May 2012.

[29] Y. Léost, M. Abdi, R. Richter, and M. Jeschke, "Interference rejection combining in LTE networks," *Bell Labs Technical Journal*, vol. 17, no. 1, pp. 25–49, Jun. 2012.

[30] 3GPP TS 38.213 v15.3.0, "Physical layer procedures for control (Release 15)," Sep. 2018.

[31] 3GPP TS 38.211 v15.4.0, "Physical channels and modulation," Dec. 2018.

[32] R. Abreu, T. Jacobsen, G. Berardinelli, K. Pedersen, I. Z. Kovács, and P. Mogensen, "Power control optimization for uplink grant-free URLLC," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2018.

[33] D. Chase, "Code Combining - A Maximum-Likelihood Decoding Approach for Combining an Arbitrary Number of Noisy Packets," *IEEE Transactions on Communications*, vol. 33, no. 5, pp. 385–393, May 1985.

[34] 3GPP TR 36.873 v12.7.0, "Study on 3D channel model for LTE," Dec. 2017.

[35] G. Pocovi, K. I. Pedersen, and P. Mogensen, "Joint Link Adaptation and Scheduling for 5G Ultra-Reliable Low-Latency Communications," *IEEE Access*, vol. 6, pp. 28 912–28 922, May 2018.

[36] 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 *2005 IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 4, Sep. 2005, pp. 2306–2311 Vol. 4.

[37] R. Srinivasan, J. Zhuang, L. Jalloul, R. Novak, and J. Park, "IEEE 802.16m Evaluation Methodology Document (EMD)," IEEE 802.16 Broadband Wireless Access Working Group, Tech. Rep. IEEE 802.16m-08/004r2, Jul. 2008.

[38] Z.-Q. Luo and S. Zhang, "Dynamic Spectrum Management: Complexity and Duality," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, pp. 57 – 73, 03 2008.

[39] L. D. Brown, T. T. Cai, and A. DasGupta, "Interval Estimation for a Binomial Proportion," *Statistical Science*, vol. 16, no. 2, pp. 101–133, May 2001.

[40] R1-1807825, "Summary of Maintenance for DL/UL Scheduling," May 2018.

Part IV

Multiplexing of eMBB and Grant-free URLLC

Multiplexing of eMBB and Grant-free URLLC

1 Problem Description

The previous part of the thesis focuses mainly on the support of Ultra-Reliable Low-Latency Communications (URLLC) assuming a portion of the bandwidth dedicated to this service type. Nevertheless, fifth generation (5G) networks are expected to support multiple services with an efficient usage of the radio interface resources [1]. For that, it is desirable to have a common pool of radio resources which can be used to multiplex the traffic of the heterogeneous services.

URLLC and enhanced Mobile Broadband (eMBB) services are very distinct in terms of traffic characteristics and quality of service (QoS) requirements. URLLC is characterized by small packets arriving sporadically, which should be delivered with a reliability of $1 - 10^{-5}$ in 1 ms. While eMBB is characterized by large data volumes which should be transmitted with very high data rates. The dynamic multiplexing of these traffic in the downlink has been well specified for New Radio (NR) in 3GPP Release-15, which allows puncturing scheduling mechanisms [2]. In the uplink however, the multiplexing of URLLC with eMBB is only applicable through grant-based procedures. In this case, the URLLC user equipment (UE) should first send an indication of a packet available to the base station, which should send a cancellation to the eMBB UE transmitting over the resources before the URLLC transmission can occur [3, 4].

For grant-free procedures, aiming at meeting tight latency requirements without the use of scheduling signals, the problem presents other aspects. In the case of sporadic URLLC traffic, eMBB can be scheduled over the grant-free frequency bands for reducing resource wasting when no URLLC data is transmitted. However, an ongoing eMBB transmission cannot be interrupted for favoring URLLC, since there is no indication about when a grant-free transmission will occur. Hence, URLLC and eMBB transmissions will in-

evitably overlap, potentially jeopardizing the performance of the services.

Power control is an option to manage the interference levels between the transmissions, e.g. by increasing the URLLC and/or reducing the eMBB transmit power. The impact of the power control configuration on the performance of each service should therefore be understood. The reception and decoding architecture is also determinant for the system performance. Advanced receivers with multiuser detection should have sufficient degrees of freedom for receiving an overlaying eMBB stream and, at the same time, allowing reliable reception of collision prone URLLC transmissions. Successive interference cancellation (SIC) can be employed for the benefit of eMBB by removing the interference from the initially decoded URLLC signal. However, URLLC can not take advantage of SIC, because it would require eMBB to be firstly decoded with the same level of reliability, prior to the decoding of the URLLC signal. Besides, due to the latency restriction, the URLLC data should be decoded before processing the large eMBB block length [5]. Due to the mentioned issues and trade-offs involved, the methods for multiplexing eMBB and URLLC in the uplink should be carefully investigated.

2 Objectives

A summary of the main objectives of this part of the work follows below:

- Quantify the impact of multiplexing eMBB over grant-free URLLC shared resources, using different power control configurations.
- Obtain insights on required power control enhancements required for the support of efficient multiplexing of heterogeneous traffic.
- Determine whether using separate resources or overlaying resources is preferable for multiplexing eMBB and grant-free URLLC for different operation regimes.

3 Included Articles

The main body of this part of the thesis includes the following articles:

Paper G. System Level Analysis of eMBB and Grant-Free URLLC Multiplexing in Uplink

This paper considers the problem of multiplexing eMBB traffic overlaying with the grant-free URLLC shared resources. The eMBB traffic is modeled as full-buffer and URLLC sporadic traffic follows a Poisson arrival process. Different open loop power control configurations are assumed for eMBB

and URLLC. The analyses are conducted through system level analysis for a multi-user multi-cell network following the same simulation guidelines as in the previous section. Baseline MMSE-IRC receiver is utilized, i.e. SIC is not considered. The impact over URLLC is shown in terms of outage probability for low and high URLLC loads. Besides, the impact on the signal to interference-and-noise ratio (SINR) for eMBB is also shown for 1 and 2 users streams.

Paper H. On the Multiplexing of Broadband Traffic and Grant-Free Ultra-Reliable Communication in Uplink

In this work, two allocation strategies for multiplexing eMBB and grant-free URLLC are compared. One utilizes separate frequency resources for each traffic, avoiding their mutual interference. While in the other, the resources are shared, using overlaying allocation for both traffic types. An analytical framework for deriving the outage probability depending on the receiver type and the operation regime is presented, extending the work from Paper L in appendix. The method is based on the outage probability for minimum mean square error (MMSE) receiver. In addition, MMSE with SIC is also considered, in which the interference from firstly decoded URLLC transmissions are removed from the eMBB signal. The evaluation for different operation regimes takes into account the reliability requirements, data size, traffic volume, bandwidth, average SNR, TTI size, number of antennas, among other settings. Based on that, the achievable loads is calculated for each traffic considering 5G NR assumptions.

4 Main Findings and Recommendations

Impact on eMBB and URLLC

The system level evaluation in Paper G demonstrates that the outage probability for the URLLC transmissions becomes 10 to 100 higher when eMBB is overlaying, depending on the power control settings and number of streams. With 1 overlaying eMBB stream, a low URLLC load can be supported as long as URLLC uses a P_0 value $\sim$5 dB higher then eMBB. A load 30 times higher for URLLC can be achieved without eMBB overlaying, assuming MMSE-IRC with 4 receive antennas. However, in this case the resource utilization is only $\sim$35%.

eMBB typically benefits of fractional path loss compensation for improving cell throughput. However, the usage of $\alpha < 1$ and higher P_0 can increase the intra-cell interference on overlaying URLLC transmissions, leading to higher outage probability. Therefore, eMBB should be preferably configured with full path loss compensation as well. The obvious cost of using

$\alpha = 1$ and lower P_0 value for eMBB while overlaying with URLLC is the reduced capacity. Nevertheless, the resource utilization can approximate the 100% in case of high data volumes.

In case of low URLLC load, the degradation on eMBB SINR is naturally low. However, this is not the case if the URLLC load increases. A SIC receiver could then be exploited for removing the interference from URLLC. Otherwise, in case the latency requirement can be relaxed and a reliable control channel is available, preemption mechanisms should be considered for avoiding the mutual interference between the transmissions.

Overlaying versus separate allocations

Paper H shows that the achievable rate for URLLC over shared resources saturates at high SNR, as also discussed in appendix Paper K. In addition, it is shown that if overlaying allocation with one eMBB stream is considered, the achievable rate is lower and saturates earlier even with more advanced receivers, as MMSE with 4 receive antennas. This is because one degree of freedom from the receiver is devoted for suppressing the eMBB interference. Hence, reducing diversity and capability for resolving collisions among transmissions from URLLC UEs. For allowing higher achievable rates for URLLC, the power of interfering eMBB should be reduced compared with URLLC, specially when simpler receivers are being used. For instance, with MMSE with 2 antennas, if URLLC has 10 dB more power than eMBB, the achievable rate can be almost doubled.

When eMBB and URLLC use separate frequency bands, it is clear that the best performance in terms of achieved load is obtained when a different power control configuration is used for each service, allowing the power spectral density to be increased as much as the allocated bandwidth decreases. This should be preferred when targeting higher URLLC loads in detriment of eMBB, in low SNR conditions. However, if SIC is used and the targeted URLLC load is low, a better performance is achieved with eMBB and grant-free URLLC overlaying. And in high SNR conditions, overlaying allocation is also better even for high URLLC load. In stringent operation regimes, defined for instance by larger URLLC payloads, higher reliability requirements, and low number of receive antennas, using separate bands is more favorable.

It is noted the difference in the performance when using overlaying versus separate bands is much smaller in low SNR compared with higher SNR conditions. This means that in high SNR scenarios, it is worthier to employ costly solutions using advanced receivers with SIC, in order to take advantage of overlaying allocations. Another important observation is that the usage of separate bands requires more control signaling for reconfiguring the frequency resources and power control for all UEs when the target supported load for each service varies.

The poor performance for URLLC multiplexed with eMBB in high load motivates the adoption of preemption mechanisms. However, those are only feasible when grant-based allocation can be utilized, i.e. for relaxed latency and available reliable control channels. Figure IV.1 shows extended results considering grant-based mechanisms. The results are based on the same framework and definitions from Paper H. Preemption mechanisms are assumed with URLLC being scheduled on fraction of resources prior to eMBB. It is also considered the option where URLLC is scheduled over eMBB, puncturing an ongoing transmission, assuming the case where eMBB users cannot monitor a preemption indication in short TTIs. For grant-based schemes, the main advantage is that URLLC does not suffer with collisions. Therefore, much higher URLLC loads can be achieved compared with the results from Paper H, specially at medium to high SNR.

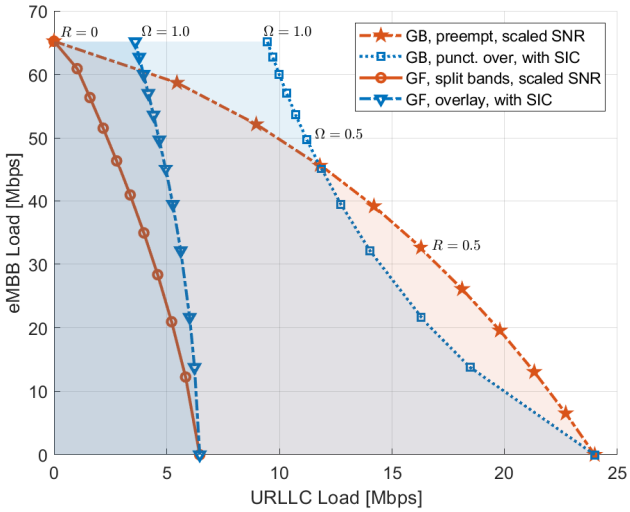


Fig. IV.1: Achievable loads for URLLC and eMBB, including grant-based (GB) options (preemption/puncturing), and different receive strategies on medium/high average SNR $\bar{\gamma}_u = 10\,\text{dB}$. $W = 10\,\text{MHz}$, $D = 256\,\text{bits}$, $N_u = 50$, $N_e = 2$ and $M = 4$. Definitions and grant-free (GF) reference results follow Paper H.

Main recommendations

The following list presents the recommendations based on the study:

- With linear receivers, the multiplexing of eMBB traffic overlaying grant-free allocations should be only employed in cases of very low URLLC load and with lower transmit power for the eMBB UEs (∼5 dB less) compared to the URLLC UEs.

- Not only URLLC but also eMBB UEs should use power control with full path loss compensation, in order to reduce the outage probability for URLLC transmissions.
- When MMSE receiver with SIC is in place, the usage of overlaying allocations for eMBB and sporadic URLLC traffic should be preferred, as long as the URLLC load is low or the signal to noise ratio (SNR) condition is medium/high (e.g. 10 dB).
- The use of separate frequency bands for eMBB and URLLC is preferable at least in case of low degrees of freedom in the receiver, large URLLC packets or stricter reliability requirement for initial transmission.

References

[1] K. I. Pedersen, G. Berardinelli, F. Frederiksen, P. Mogensen, and A. Szufarska, "A Flexible 5G Frame Structure Design for Frequency-Division Duplex Cases," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 53–59, Mar. 2016.

[2] K. I. Pedersen, G. Pocovi, and J. Steiner, "Preemptive scheduling of latency critical traffic and its impact on mobile broadband performance," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, June 2018.

[3] A. Anand, G. de Veciana, and S. Shakkottai, "Joint Scheduling of URLLC and eMBB Traffic in 5G Wireless Networks," *ArXiv e-prints*, Dec. 2017. [Online]. Available: http://arxiv.org/abs/1712.05344

[4] R1-1805629, "Summary of handling UL multiplexing of transmission with different reliability requirements," Apr. 2018.

[5] P. Popovski, K. F. Trillingsgaard, and G. D. Osvaldo Simeone, "5G Wireless Network Slicing for eMBB, URLLC, and mMTC: A Communication-Theoretic View," *CoRR*, vol. abs/1804.05057, 2018. [Online]. Available: http://arxiv.org/abs/1804.05057

Paper G

System Level Analysis of eMBB and Grant-Free URLLC Multiplexing in Uplink

Renato Abreu, Thomas Jacobsen, Klaus Pedersen, Gilberto Berardinelli, Preben Mogensen

The paper has been published in the
2019 IEEE 89th Vehicular Technology Conference (VTC Spring)

Abstract

5th generation radio networks should efficiently support services with diverse requirements. For achieving better resource utilization, the sharing of the radio channel between the different services is an attractive solution. While the downlink multiplexing can be well accomplished with dynamic scheduling, efficient multiplexing of enhanced mobile broadband (eMBB) and ultra-reliable low-latency communications (URLLC) in uplink is still an open problem. In particular, we consider the case of URLLC using grant-free allocation for sporadic transmissions, multiplexed on shared resources with eMBB with high data volume. Since the moment in which a grant-free transmission occurs is not known, URLLC and eMBB transmissions overlay. Power control settings are then assessed as a way to manage the performance trade-off between the services. Due to the complexity of 5G NR, the evaluation is based on advanced system level simulations. Insights regarding the configuration of fractional power control settings upon the coexistence of the different services are presented.

1 Introduction

The recent 5th generation (5G) new radio (NR) specifications include features for conveying traffic with different characteristics and requirements. One example is enhanced mobile broadband (eMBB) which focuses on high volume of data transmissions, demanding high spectral efficiency. Ultra-reliable low-latency communications (URLLC) target instead, to deliver intermittent small payloads with high success probability in a short time interval. A baseline target for URLLC is to enable transmissions over the air interface of 32 bytes payloads within 1 ms and a $1 - 10^{-5}$ reliability [1]. The initial support of each of these services is readily provided by the 3GPP Release-15 specification [2]. However, the multiplexing of uplink traffic with different reliability requirements has gained attention, given the need of supporting heterogeneous services while ensuring efficient use of the radio resources [3]. The efficient multiplexing of eMBB and URLLC in downlink can be achieved by dynamic scheduling, with the high priority URLLC transmissions puncturing the eMBB allocation [4]. In uplink, similar concept can be employed with preemption schemes, both for intra-UE (for the same UE) and for inter-UE (between different UEs) traffic multiplexing. With this, eMBB transmission is paused while URLLC is granted to transmit. While this solution is valid for dynamic scheduled transmissions, the same is not applicable when grant-free schemes are utilized. Grant-free transmissions, specified as configured grants in NR [5], is one of the main enablers of uplink URLLC with very stringent requirements. In that, the resource allocation settings, as well as other physical layer parameters, are pre-configured by radio resource control (RRC) signaling. Thus, the regular handshake process, of sending a schedul-

ing request and waiting for a grant allocation for every transmission, can be avoided. This reduces not only the delay, but also the dependence of error-prone control signaling for every transmission. For reducing the resource wastage caused by sporadic URLLC transmissions, the base station (BS) can configure the same resources to multiple user equipments (UE). However, this leads to augmented intra-cell interference when transmissions overlap. The problem becomes more evident if the grant-free resources are overlaid for multiplexing abundant eMBB traffic. Since it is not known a priori if a grant-free URLLC transmission will occur, it is not possible to timely interrupt an ongoing transmission for avoiding a collision, potentially degrading the reliability.

Different studies have considered the problem of multiplexing heterogeneous traffic in uplink. In [6], a joint eMBB and URLLC scheduler is proposed, with superposition of ongoing transmissions. The overlaying multiplexing between resource greedy broadband traffic and sporadic small data is considered in [7] and evaluated with basic information theoretical tools for a single cell scenario. An heterogeneous non-orthogonal multiple access approach is studied in [8] using a theoretic model, however, multiple URLLC transmissions over the shared resource are not considered. In [9], a theoretical analysis of overlaying versus separate allocation is presented. Minimum-mean square error (MMSE) is considered for the reception of multiple URLLC and eMBB transmissions. Detailed analysis considering the aspects of a multi-cell 5G NR system are not considered in previous works.

In this work we present system level performance evaluation for the inter-UE multiplexing of eMBB and URLLC uplink transmissions. We consider the case of sporadic grant-free URLLC, with shared resource allocations, overlaying with full-buffer eMBB streams, in a multi-cell system. We discuss the aspects of open loop power control and identify the criteria for setting the relevant parameters in order to manage the trade-off between URLLC reliability and eMBB capacity. Results from detailed simulation campaigns following 5G NR assumptions are presented in terms of URLLC outage probability and eMBB SINR.

The reminder of the work is organized as follows. The considered system is presented in Section 2 and the power control aspects in Section 3. Section 4 describes the methodology and assumptions. Results are presented in Section 5 and discussed in Section 6. Section 7 concludes the paper.

2 System model

We consider a multi-cell radio network composed of C cells with synchronized base stations (BS). A fixed number of URLLC UEs N_u are deployed in each cell. Besides, N_e eMBB UEs can be active in the same cell. The UEs

are considered to be connected and synchronized with the serving BS for their uplink data transmission. Fig. G.1 illustrates the considered multiplexing scheme. The eMBB UEs are assumed to have a large amount of data to transmit. Their traffic follows a full buffer model, ensuring a permanent flow of eMBB data to be scheduled over the time slots. The N_e eMBB UEs are scheduled over the full carrier bandwidth W. The BS exploits then multi-user reception solutions by employing an M_r antennas receiver, for retrieving overlaying signals.

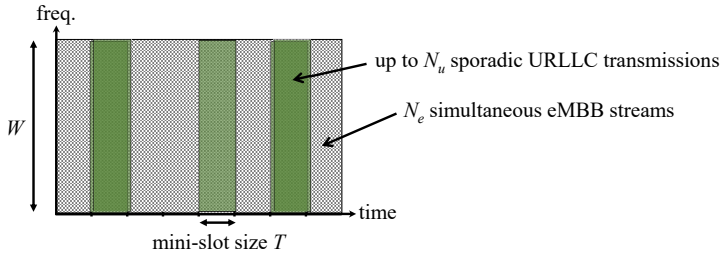


Fig. G.1: Overlaying eMBB and grant-free URLLC allocations in a cell.

The URLLC UEs have sporadic traffic consisted of small payloads of size B. Such traffic is modeled as a Poisson arrival process with packet arrival rate λ. In order to serve the URLLC traffic with minimum latency, a short-TTI of duration T is employed. The serving BS configures also the URLLC UEs to transmit with grant-free resources over the bandwidth W. We assume that the N_u UEs share the same resource configuration, therefore their transmissions are susceptible to mutual collisions, in addition to the interference from eMBB traffic being multiplexed over the same resources. A wide-band allocation allows harvesting frequency diversity. It also permits the use of a robust modulation and coding scheme (MCS) to cope with fading and potential interference from simultaneous transmissions.

A linear minimum-mean square error with interference rejection combining (MMSE-IRC) receiver is assumed in the BS. Since the UEs and the BSs are fully synchronized, it permits the receiver to take into account intra- and inter-cell interference signals for computing the interference covariance matrix. Then, the MMSE-IRC receiver operates on the degrees of freedom offered by the multiple receive antennas to retrieve multiple overlaid transmissions. Still, in case the interference level is too severe the reception can be compromised. This motivates the use of careful power control settings for reducing the penalty in the URLLC reliability or eMBB capacity.

3 Power control setting for overlaying transmissions

The 3GPP Release-15 specification defines the power control for the uplink channels in [10]. The transmit power (in dBm) over the physical uplink shared channel (PUSCH) is described, in simplified notation, as

$$P = min \begin{cases} P_{max} \\ P_0 + 10log_{10}(2^{\mu}M) + \alpha PL + \Delta_{mcs} + f(i) \end{cases}, \tag{G.1}$$

where P_{max} is the maximum transmit power of the UE, P_0 is a UE specific parameter related to the power per resource block (RB), the exponent μ is set according the sub-carrier spacing (0 for 15 kHz, 1 for 30 kHz, and so on), M is the number of RBs allocated, α is a path-loss compensation factor, PL is the estimated path-loss between the UE and the BS. Δ_{MCS} is a quality requirement parameter depending on the MCS that can be configured by upper layers and $f(i)$ is a parameter for closed loop power control adjustments; these were not considered in this study.

The use of fractional power control is known for improving the capacity for broadband communication [11]. For such, $\alpha < 1$ is applied, as well as a correspondent increase in P_0, improving the SINR, and hence, the throughput of cell center UEs. However, as discussed in [12], the usage of full path-loss compensation is more attractive for URLLC to avoid an outage penalty in cell edge. In the case of overlaying allocations, the performance of eMBB and URLLC presents a trade-off, i.e. power control settings that benefits eMBB penalizes URLLC and vice-versa. Thus, in our proposal the settings are applied on a service basis. With that, eMBB UEs are configured with P_0^e and α^e, while URLLC UEs are configured with P_0^u and α^u. Here we assume that, for each service, all UEs in the cell use the same parameters. These parameters should be carefully selected for meeting the service requirements. As a simple example, for $\alpha^u = \alpha^e$ setting $P_0^e >> P_0^u$ potentially increases the interference of eMBB over URLLC compromising the reliability. While $P_0^e << P_0^u$ can deteriorate the eMBB capacity.

4 Evaluation Methodology

The impact on the performance of overlaying grant-free URLLC and eMBB is evaluated through extensive system level simulations for different power control settings. The evaluation methodology is based on NR assumptions as defined in [13]. The simulator uses commonly accepted models and is calibrated according to 3GPP NR guidelines [14]. The main parameters for the network configuration and the main simulation assumptions are summarized in Table G.1.

Table G.1: Simulation assumptions

Parameters	Assumption
Layout	Hexagonal grid with 21 cells (7 sites and 3 sectors/site), world wrap-around
Inter-site distance	500 meters
Carrier frequency	4 GHz
Channel model	3D Urban Macro (UMa)
UE distribution	Uniformly distributed outdoor, 3 km/h UE speed fading model
UE transmitter	$P_{max} = 23$ dBm, $M_t = 1$ antenna
BS receiver	MMSE-IRC, $M_r = 4$ antennas
Receiver noise figure	5 dB
Thermal noise	-174 dBm/Hz
Bandwidth	$W = 10$ MHz in uplink, FDD
PHY configuration	15 kHz sub-carrier spacing, 2 symbols mini-slot ($T = 0.143$ ms), 12 sub-carriers/RB
Grant-free configuration	MCS QPSK1/8, periodicity of 2 symbols, 48 RBs for data transmission, HARQ disabled
eMBB UEs per cell	0 (no eMBB interference baseline), 1 (single stream) and 2 (MU-MIMO streams)
eMBB traffic model	full-buffer
URLLC UEs per cell	10 for low load, and 300 for high load
URLLC traffic model	FTP Model 3, $B = 32$ bytes, Poisson arrival rate of $\lambda = 10$ packets per second per UE

A 3D urban macro scenario is assumed, consisting of $C = 21$ synchronized cells (7 sites with 3 sectors each). The inter-site distance is 500 meters. World wrap around is used for avoiding edge effects. We consider different load conditions for URLLC. For low load, 10 URLLC UEs per cell are uniformly distributed in the scenario. And for high load, 300 URLLC UEs per cell are distributed. Each URLLC UE transmits payloads of $B = 32$ bytes following a Poisson arrival process with average arrival interval of 100 ms, i.e. $\lambda = 10$ packets per second. This leads to a load $L = 25.6$ kbps per cell for low URLLC load, and $L = 768$ kbps for high URLLC load. One and two eMBB UEs are also deployed in each cell, equivalent to a single stream and two multi-user MIMO streams. The eMBB UEs use full-buffer traffic model, being continuously scheduled over the full bandwidth. The UEs are deployed at the beginning of the simulation drop. Each UE connects to the cell with highest reference signal received power (RSRP) and remains in connected state until the simulation finishes.

The URLLC UEs are configured for transmission in mini-slots of 2 OFDM

symbols, with sub-carrier spacing of 15 kHz which leads to a $T = 0.143$ ms TTI. The allocation for grant-free transmissions uses a bandwidth $W = 10$ MHz, giving 48 RBs for data, with 2 symbols periodicity. This allows a transmission opportunity in full-band at every TTI in order to minimize latency. The grant-free transmissions use a conservative MCS QPSK 1/8, fitting the 32 bytes payload in one-shot transmission without segmentation. Considering latest processing time assumptions (capability 2 in [10]), a transmission can be received and processed within 1 ms. HARQ retransmissions are not considered.

The BSs are equipped with MMSE-IRC with $M_r = 4$ receive antennas. Channel estimation is assumed ideal for the desired and interference signals. The successful reception of a packet depends on the obtained post-processing SINR at the receiver and the used MCS. For every detected transmission, the post-processing SINR after the MMSE-IRC receiver combining is calculated for each sub-carrier. That is used to compute the symbol-level mutual information metric according to the applied modulation as described in [15]. Then, given the used code rate, a look-up table obtained from extensive link level simulations is used to map the metric value to a block error probability.

Multiple simulation drops are executed for collecting 5 million URLLC transmission samples, in order to obtain statistically significant results in the low quantiles [16]. The main key performance indicator analyzed for URLLC is the outage probability, i.e. the complement of the reliability (targeting 10^{-5}). The latency of each transmission is used for determining an empirical complementary cumulative distribution functions (CCDF). The outage probability is then read at the 1 ms from the latency CCDF. For the eMBB performance, we collect the 5th percentile and the 50th percentile SINR values. These reference metrics indicate the cell edge and the near to average performance, respectively.

5 Performance evaluation

The power control settings P_0 and α for eMBB and URLLC UEs were varied for the different simulation campaigns, in which were collected the one-way latency of the URLLC packets and the SINR of the eMBB transmissions. The power control settings for URLLC were chosen as the ones that allow the highest URLLC load while fulfilling the requirements [12]. Full path-loss compensation is used for URLLC, i.e. $\alpha^u = 1$. For eMBB, full and fractional path-loss compensation are used, i.e. $\alpha^e = 1$ and $\alpha^e = 0.7$ respectively. The P_0 values are set equal or lower than the URLLC ones, except when fractional path-loss compensation is used. For reference, the empirical cumulative distribution function (CDF) of the coupling gain for the evaluated outdoor scenario is shown in Fig. G.2. The CDFs of the URLLC and the eMBB transmit

power are also shown for each utilized setting. For both, URLLC and eMBB using $\alpha^u = \alpha^e = 1$ and $P_0^u = P_0^e = -108\,\text{dBm}$, 3% of the UEs transmit with maximum power P_{max}. For URLLC configured with conservative power control settings, $\alpha^u = 1$ and $P_0^u = -103\,\text{dBm}$, 15% of the URLLC UEs transmit with P_{max}. For eMBB with $\alpha^e = 0.7$ and $P_0^e = -78\,\text{dBm}$, as well as with $\alpha^e = 1$ and $P_0^e = -113\,\text{dBm}$, virtually no eMBB UE reaches P_{max}.

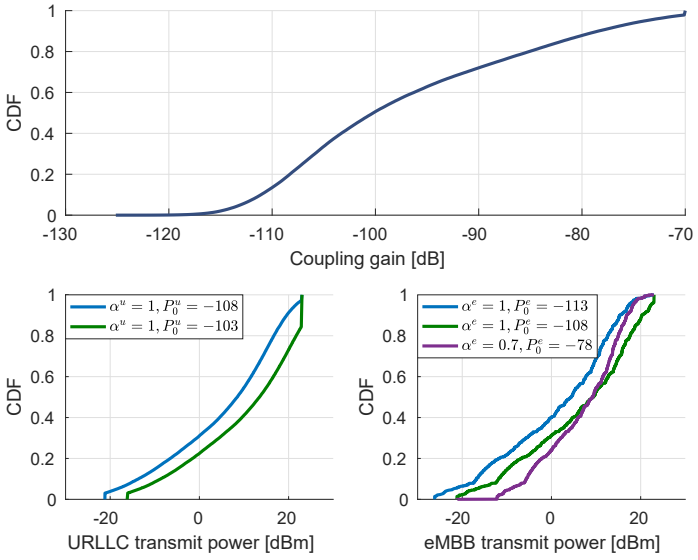
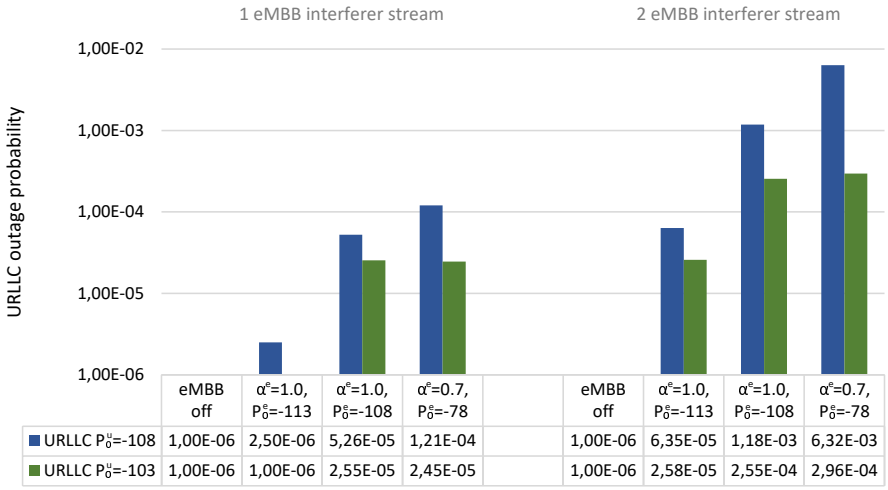


Fig. G.2: Coupling gain distribution in evaluated urban macro scenario outdoor (top). Transmit power distribution for URLLC UEs (bottom left), and eMBB UEs (bottom right).

Fig.G.3 shows the outage probability for the case of 10 URLLC UEs per cell, with their transmissions being multiplexed with 1 and with 2 eMBB interferer streams. Baseline cases without eMBB interference are also shown as "eMBB off". It is observed that the URLLC target is satisfied if no eMBB UEs are present, leading to an outage probability $< 10^{-6}$. Reducing the power of eMBB with $P_0^e = -113\,\text{dBm}$ (i.e. 5 dB lower than for the URLLC UE) also allows URLLC to reach the target, when only 1 eMBB stream is present. For the cases where eMBB uses the same power control settings as URLLC, the outage probability rises to the order of 10^{-4}. With 2 simultaneous eMBB streams, the penalty for URLLC is obviously higher due to the increased interference. The use of fractional path-loss compensation for eMBB does not help, since the cell center eMBB UEs generates higher intra-cell interference. The outage probability for high URLLC load, with 300 URLLC UEs per cell, is shown in Fig.G.4. In this case the URLLC requirement is nearly met only when eMBB UEs are not transmitting, i.e. without eMBB interference a URLLC



	eMBB off	α^e=1.0, P_0^e=-113	α^e=1.0, P_0^e=-108	α^e=0.7, P_0^e=-78	eMBB off	α^e=1.0, P_0^e=-113	α^e=1.0, P_0^e=-108	α^e=0.7, P_0^e=-78
URLLC P_0^u=-108	1,00E-06	2,50E-06	5,26E-05	1,21E-04	1,00E-06	6,35E-05	1,18E-03	6,32E-03
URLLC P_0^u=-103	1,00E-06	1,00E-06	2,55E-05	2,45E-05	1,00E-06	2,58E-05	2,55E-04	2,96E-04

Fig. G.3: Outage probability of grant-free URLLC for $L = 25.6$ kbps.

load of ≈ 0.77 Mbps per cell is supported. However, the outage probability of URLLC increases by a factor of 10 to 100 when eMBB is present. For both load situations, the use of a high P_0^u makes URLLC more robust to the presence of eMBB interference. However, when eMBB is not present, the lower P_0^u results in a lower outage due to reduced interference among URLLC UEs. Using lower P_0^e values reduces the impact on URLLC, however it comes with the cost of lower SINR for eMBB, which converts to a capacity loss.

Fig.G.5 and Fig.G.6 shows the impact on the eMBB SINR for the different power control settings. For the lower URLLC load there is little difference on eMBB performance for the different URLLC P_0^u settings. As expected, the eMBB SINR is low in the case of a low P_0^e. And from full to fractional path-loss compensation, there is an improvement in the 50th percentile SINR and a degradation in the 5th percentile SINR. The same observation can be drawn for one and for two eMBB streams. With the higher URLLC load there is a clear impact in the eMBB SINR (up to 3.1 dB for $P_0^u = -108$ dBm). Besides, the 5 dB increase in P_0^u, causes up to 1.67 dB of degradation in eMBB SINR. The low 5th percentile SINR values, getting down to -5 dB, indicates the very limited eMBB capacity in the cell edge even with high P_0^e.

It is worth to mention that the resource utilization without eMBB, for low URLLC load is 1.4%, and for high URLLC load is 35%. This means that a big share of the resources is wasted in detriment of URLLC. This demonstrates the importance of multiplexing eMBB together with the URLLC traffic for the feasibility of the 5G system.

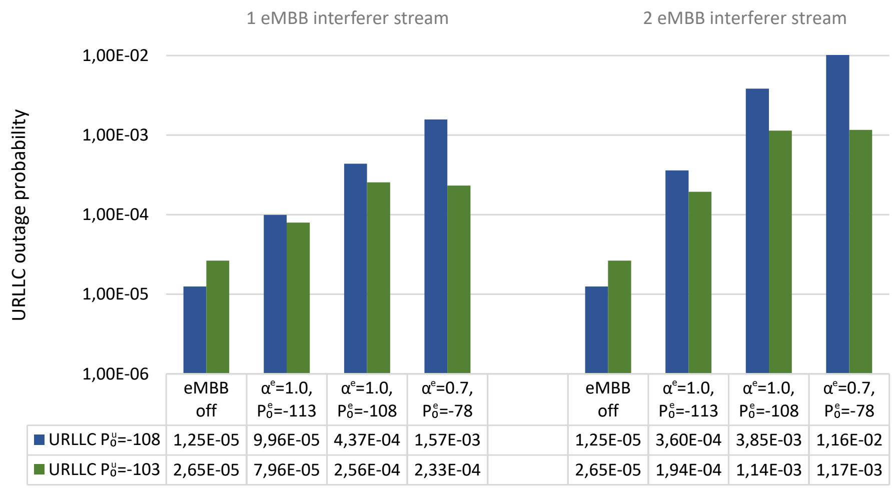


Fig. G.4: Outage probability of grant-free URLLC for $L = 768$ kbps.

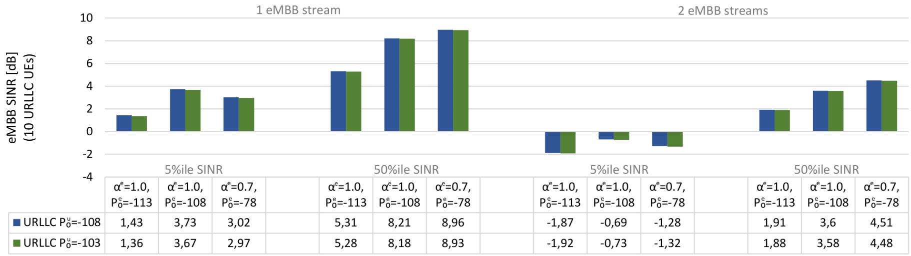


Fig. G.5: eMBB SINR with grant-free URLLC load of $L = 25.6$ kbps.

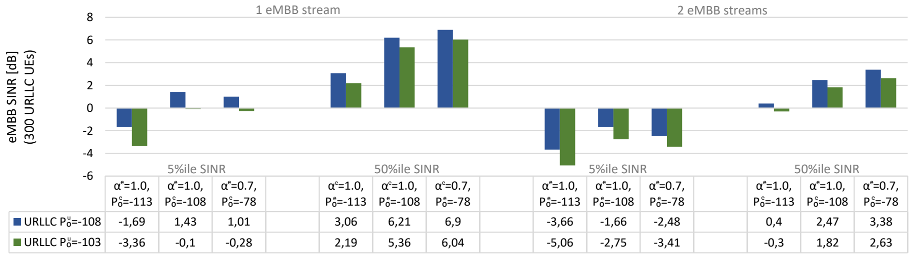


Fig. G.6: eMBB SINR with grant-free URLLC load of $L = 768$ kbps.

6 Discussion

It is worth noting that, despite the potential of fractional path-loss compensation for improving eMBB average throughput, cell center eMBB UEs with ele-

vated transmit power further penalizes the URLLC transmissions. Therefore, full path-loss compensation and lower P_0 values should be also preferred for eMBB when multiplexing with URLLC.

The presence of a high URLLC load in the cell imposes a reduced capacity for eMBB. The use of the receiver capability for MU-MIMO is compromised due to the limitation on degrees of freedom for suppressing all the mutual interference. The system performance can be enhanced e.g., by utilizing MMSE-IRC with higher number of antennas, which improves the diversity order and interference rejection capability. Besides, successive interference cancellation (SIC) can be employed for subtracting the signal from decoded URLLC transmissions from the received signal. This can mainly reduce the interference over the eMBB transmissions [8, 9].

For applications in which the latency requirement can be relaxed, preemption schemes enabled by dynamic downlink control signal should be preferred [17]. Those are able to interrupt on-going eMBB transmissions for scheduling URLLC data. eMBB can be potentially resumed after the URLLC transmission. With that, both URLLC and eMBB should be benefited from the reduced interference. Besides, dynamic scheduling permits accurate resource allocation and adaptation per-user transmission basis. This results in guaranteed quality of service with efficient usage of resources.

7 Conclusions

In this paper, we studied the performance of grant-free URLLC and eMBB multiplexing in uplink. We considered the overlaying of eMBB transmissions with the grant-free URLLC transmissions over the same resources. Different uplink transmit power control settings are proposed for managing the trade-off between the URLLC outage probability and the eMBB capacity. Detailed evaluation of the settings was conducted through extensive system level simulations following 5G NR assumptions. We observe that overlaying URLLC and eMBB transmissions is only feasible for low URLLC loads (e.g. 0.26 Mbps). Even though, it requires restrictions which impose severe performance loss for eMBB, such as, reduced capability for co-scheduling users and 5 dB lower P_0 value. Higher URLLC load of e.g. $\approx$ 0.77 Mbps is supported when no eMBB UE is multiplexed over the same resources. However it results in a poor resource utilization (35%). The insights obtained for the power control configuration can be utilized as reference for the setup of 5G deployments with heterogeneous services. The results demonstrate the severe penalty caused by eMBB transmissions over URLLC. This motivates the application of preemption mechanisms for avoiding collisions when URLLC traffic can be dynamic scheduled.

Future work should consider dynamic scheduling solutions of the uplink

URLLC transmissions suspending on-going eMBB transmissions, as well as the impacts of the control channel overhead and imperfections.

8 Acknowledgments

This research is partially supported by the EU H2020-ICT-2016-2 project ONE5G. The expressed views are those of the authors and do not necessarily represent the project views.

References

[1] ITU-R, "Report ITU-R M.2410-0 - Minimum requirements related to technical performance for IMT-2020 radio interface(s)," International Telecommunication Union (ITU), Tech. Rep., Nov. 2017.

[2] 3GPP TS 38.300 V15.2.0, "NR; NR and NG-RAN Overall Description," Jun. 2018.

[3] RP-181477, "SID on Physical Layer Enhancements for NR URLLC," Jun. 2018.

[4] K. I. Pedersen, G. Pocovi, and J. Steiner, "Preemptive scheduling of latency critical traffic and its impact on mobile broadband performance," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, June 2018.

[5] 3GPP TS 38.331 V15.2.1, "NR; Radio Resource Control (RRC) protocol specification," Jun. 2018.

[6] A. Anand, G. de Veciana, and S. Shakkottai, "Joint Scheduling of URLLC and eMBB Traffic in 5G Wireless Networks," *ArXiv e-prints*, Dec. 2017. [Online]. Available: http://arxiv.org/abs/1712.05344

[7] G. Berardinelli and H. Viswanathan, "Overlay transmission of sporadic random access and broadband traffic for 5G networks," in *2017 International Symposium on Wireless Communication Systems (ISWCS)*, Aug. 2017, pp. 19–24.

[8] P. Popovski, K. F. Trillingsgaard, and G. D. Osvaldo Simeone, "5G Wireless Network Slicing for eMBB, URLLC, and mMTC: A Communication-Theoretic View," *CoRR*, vol. abs/1804.05057, 2018. [Online]. Available: http://arxiv.org/abs/1804.05057

[9] R. B. Abreu, T. Jacobsen, G. Berardinelli, K. I. Pedersen, and P. E. Mogensen, "On the Multiplexing of Broadband Traffic and Grant-Free Ultra-Reliable Communication in Uplink," in *2019 IEEE 89th Vehicular Technology Conference (VTC Spring)*, 2019, (Accepted/in press).

[10] 3GPP TS 38.214 v15.1.0, "NR; Physical layer procedures for data," Mar. 2018.

[11] C. U. Castellanos, D. L. Villa, C. Rosa, K. I. Pedersen, F. D. Calabrese, P. H. Michaelsen, and J. Michel, "Performance of Uplink Fractional Power Control in UTRAN LTE," in *VTC Spring 2008 - IEEE Vehicular Technology Conference*, May 2008, pp. 2517–2521.

[12] R. Abreu, T. Jacobsen, G. Berardinelli, K. Pedersen, I. Z. Kovács, and P. Mogensen, "Power control optimization for uplink grant-free URLLC," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2018.

[13] 3GPP TR 38.802 v14.0.0, "Study on New Radio Access Technology," Mar. 2017.

[14] RP-180524, "Summary of calibration results for IMT-2020 self evaluation," Mar. 2018.

[15] R. Srinivasan, J. Zhuang, L. Jalloul, R. Novak, and J. Park, "IEEE 802.16m Evaluation Methodology Document (EMD)," IEEE 802.16 Broadband Wireless Access Working Group, Tech. Rep. IEEE 802.16m-08/004r2, Jul. 2008.

[16] G. Pocovi, K. I. Pedersen, and P. Mogensen, "Joint Link Adaptation and Scheduling for 5G Ultra-Reliable Low-Latency Communications," *IEEE Access*, vol. 6, pp. 28 912–28 922, May 2018.

[17] R1-1900931, "Solution for UL inter-UE multiplexing between eMBB and URLLC ," Jan. 2019.

Paper H

On the Multiplexing of Broadband Traffic and Grant-Free Ultra-Reliable Communication in Uplink

Renato Abreu, Thomas Jacobsen, Gilberto Berardinelli, Nurul H. Mahmood, Klaus Pedersen, István Z. Kovács, Preben Mogensen

The paper has been published in the
2019 IEEE 89th Vehicular Technology Conference (VTC Spring)

Abstract

5G networks should support heterogeneous services with an efficient usage of the radio resources, while meeting the distinct requirements of each service class. We consider the problem of multiplexing enhanced mobile broadband (eMBB) traffic, and grant-free ultra-reliable low-latency communications (URLLC) in uplink. Two multiplexing options are considered; either eMBB and grant-free URLLC are transmitted in separate frequency bands to avoid their mutual interference, or both traffic share the available bandwidth leading to overlaying transmissions. This work presents an approach to evaluate the supported loads for URLLC and eMBB in different operation regimes. Minimum mean square error receivers with and without successive interference cancellation (SIC) are considered in Rayleigh fading channels. The outage probability is derived and the achievable transmission rates are obtained based on that. The analysis with 5G new radio assumptions shows that overlaying is mostly beneficial when SIC is employed in medium to high SNR scenarios or, in some cases, with low URLLC load. Otherwise, the use of separate bands supports higher loads for both services simultaneously. Practical insights based on the approach are discussed.

1 Introduction

The support for services with heterogeneous requirements is one of the goals of fifth generation (5G) new radio (NR). In particular, the enhanced mobile broadband (eMBB) and ultra-reliable low-latency communications (URLLC) service classes have distinct characteristics in terms of traffic type and key performance indicators. While eMBB tolerates a moderate reliability and focus on high data rates, URLLC targets highly reliable small packets transmissions with short latency deadlines, such as 1 ms with 99.999% reliability [1].

In uplink, the eMBB traffic can be dynamically scheduled using large block lengths. However, the scheduling request and grant procedure required for a packet transmission are source of delays and errors, which can jeopardize the latency and reliability [2]. Therefore grant-free access, which allows immediate access to the channel without the scheduling procedure, is considered for URLLC [3]. Multiple users can share the same grant-free allocation to improve the radio resource utilization [4]. In a 5G network, the same carrier may need to support both grant-free URLLC and scheduled eMBB traffic. One option is to split the available bandwidth between each service class. However, this may lead to poor spectral efficiency in case of sporadic URLLC transmissions. Sharing the same radio resources for grant-free URLLC and eMBB traffic, with overlaying allocations, might improve the spectral efficiency. The consequence is the mutual interference between the two service classes, which may compromise the reliability of URLLC or degrade the eMBB data rate. Power control schemes and multi-antenna re-

ceivers, including successive interference cancellation (SIC), are potential solutions to mitigate the interference [5]. Our interest is then to study whether separate bands or overlaying allocations is preferred for ensuring efficient multiplexing of both services, depending on the scenario, traffic load and receiver characteristics.

Previous works have formed the bases for studying the coexistence of multiple traffic. The capacity of multi-antenna systems with spatial multiplexing is provided in [6], with and without SIC. The work in [7] derives the reliability of the minimum mean square error (MMSE) receiver in Rayleigh channel including multiple interferers. In [8], the overlaying of broadband traffic and sporadic transmissions is studied using basic information theoretic tools. The dynamic multiplexing of URLLC and eMBB traffic is evaluated considering preemption [9] and superposition schemes [10], which can be applied for scheduled transmissions. The recent work in [11] investigates the potential of non-orthogonal multiple access (NOMA) for heterogeneous services, though collisions between URLLC transmissions are not considered. The achievable rates in collision prone resources is discussed in [12] for sporadic URLLC transmissions and linear receivers. Collisions between multiple URLLC transmissions and eMBB transmissions is not considered in the related works.

In this paper we study the multiplexing of eMBB and grant-free URLLC traffic using an analytical framework. The presented methodology is based on the findings in [7] and [8], where achievable rates in different interference scenarios and with different receiver types have been derived. The performance of both service classes is compared using overlaying allocations and separate bands. We describe the outage probability in each case, i.e. the complement of the reliability, considering linear MMSE receiver, and also MMSE with SIC for the case of overlaying transmissions. Numerical analysis is conducted considering NR requirements and numerology. The required rate for URLLC transmissions is obtained and the impact on the supported loads for eMBB and URLLC is evaluated with different settings. Further the paper discusses the implications when either of the multiplexing options are used and comes with concrete recommendations for 5G NR operation with heterogeneous services.

The rest of the paper is organized as follows. Section 2 describes the system model. Section 3 presents the outage and achievable load calculation. Numerical results are shown in Section 4 and discussed in Section 5. Finally, conclusions are drawn in Section 6.

2 System Model

We consider a scenario where users are connected and synchronized to one serving cell for uplink data transmission. N_e active users have eMBB service, while N_u users have URLLC service. The total available bandwidth W can either be split to each service class or be shared for overlaying transmissions, as illustrated in Fig. H.1. The users transmit over a flat i.i.d Rayleigh fading channel with additive Gaussian noise. Users with a specific traffic type operate over the same resources.

For separate bands, we define a bandwidth split ratio R. With that, a bandwidth $W_u = WR$ is used for URLLC and a bandwidth $W_e = W(1-R)$ is used for eMBB, with $0 \leq R \leq 1$. For overlaying transmissions, it is assumed that both services use the full band W, so $W_u = W_e = W$. In this case, eMBB signals have an average interferer power relative to URLLC expressed as Ω, i.e. for URLLC users with average receive power $\bar{p_u}$ and eMBB with average receive power $\bar{p_e}$ over the same band, $\Omega = \bar{p_e}/\bar{p_u}$. It is assumed that the users from each service class are power controlled so that they are received with the same average signal-to-noise ratio (SNR). To meet strict latency requirements, the URLLC transmissions occur in a short transmission time interval (TTI) of duration T. Whereas eMBB transmissions use long TTIs which allows to benefit from larger coding gains [13].

The eMBB traffic is resource greedy, inducing an uninterrupted interference to other users that are transmitting simultaneously over the same band. $N_e > 1$ can be seen as the case of multi-user MIMO, in which multiple users are scheduled to transmit over the same time-frequency resources, exploiting the spatial dimension of a multi-antenna receiver [6]. The traffic from each URLLC user is assumed to follow a Poisson distribution with packet arrival rate λ per TTI and fixed payload size of D bits. The outage probability targeted for URLLC transmissions is ϵ_u, while for eMBB transmissions it is ϵ_e. For 5G NR use cases the value of ϵ_u should reach 10^{-5} in one or more transmission attempts, to satisfy the strict reliability requirement. Whereas, in cellular networks such as LTE the value of ϵ_e is in the order of 10^{-1}, for the sake of high throughput [14]. The effect of HARQ retransmissions is not considered in this work.

An MMSE receiver with M antennas is assumed. In the case that the URLLC transmissions overlay eMBB streams, we consider two different approaches: conventional MMSE receiver, and MMSE with SIC. For the latter, we assume that the URLLC transmissions should be identified, e.g. using a reference signal, and decoded first, considering the low latency requirement. Then SIC is employed, assuming that the interference of URLLC transmissions over the eMBB streams is completely canceled out.

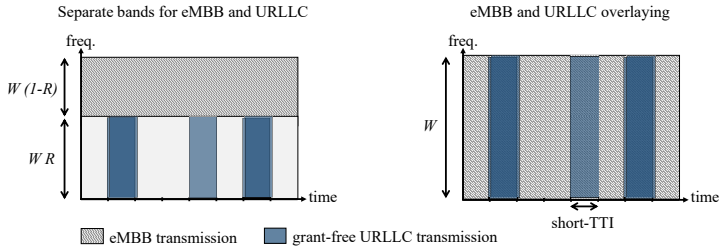


Fig. H.1: Separate bands vs. overlaying transmissions for eMBB and URLLC.

3 Analysis of overlaying and separate bands

In this section we present an analytical approach to evaluate the multiplexing of eMBB and sporadic URLLC traffic. The approach builds on top of closed-form solutions that models the reliability for an ideal MMSE receiver with additive interference channels. The model presented in [7] allows to consider each signal source with a different average interferer power relative to a desired source. The outage probability with randomly active sources with the same power characteristics are described and numerically validated in [12]. In this work, we distinguish two classes which can possibly have different average receive SNR, from a total of $v + w$ interferers. v of them have an average interferer power relative to the desired source given by Γ_v. And w interferers have an average interferer power relative to the desired source denoted by Γ_w. We later relate the v interferers as the URLLC ones, and the w interferers as the eMBB ones. The desired source can be either an eMBB or an URLLC signal, that can suffer with interference coming from users of the same or different class. The outage probability for the transmissions subject to interference is calculated as follows [7]:

$$P_f(\bar{\gamma}, v, w, \Gamma_v, \Gamma_w) = 1 - e^{\psi/\bar{\gamma}} \sum_{n=1}^{M} \frac{A_n}{(n-1)!} \left(\frac{\psi}{\bar{\gamma}}\right)^{n-1}, \tag{H.1}$$

where $\bar{\gamma}$ is the average SNR of the desired source signal at the receiver input, and ψ is the post-combining SINR required for receiving with an outage probability P_f. With the two classes of interferers, we have that

$$A_n = \begin{cases} 1 & \text{if } v + w \leq M - n \\ \dfrac{1 + \sum_{i=1}^{M-n} C_i \psi^i}{(1 + \psi\Gamma_v)^v (1 + \psi\Gamma_w)^w} & \text{if } v + w > M - n \end{cases}, \tag{H.2}$$

where C_i is the coefficient of ψ^i in the expansion of $(1 + \psi\Gamma_v)^v (1 + \psi\Gamma_w)^w$.

In a collision prone scenario the resultant outage probability, can be calculated by combining the collision probability and the outage probability for

the given number of interferers [12]. This outage probability can be interpreted as a long term error rate. The probability of having x simultaneous transmissions generated by other y users that are randomly active is

$$P_c(x,y) = \binom{y}{x} P_a^x (1-P_a)^{y-x}, \tag{H.3}$$

where P_a is the probability of each user to transmit. In the case of Poisson arrival traffic with arrival rate λ, as we assume for the URLLC users, $P_a = 1 - e^{-\lambda}$.

From that, we describe the outage probability for eMBB and URLLC transmissions for the case of separate bands and for overlaying transmissions.

3.1 MMSE receiver and separate bands

In the case that a separate band is reserved for each service class, URLLC and eMBB transmissions do not interfere with each other, and their outage probabilities can be derived independently. However, sporadic URLLC transmissions can still collide with each other within the URLLC band. With power control, all the URLLC interferers are assumed to have the same average power at the receiver input as the desired URLLC source. Given that, we assign $v = N_u - 1$ and $\Gamma_v = 1$, while $w = 0$ and $\Gamma_w = 0$ since there is no other type of interferer in the same band. The outage probability for the URLLC transmissions is then given by

$$P_{f,u} = \sum_{z=0}^{N_u-1} P_c(z, N_u - 1) P_f(\bar{\gamma}_u, z, 0, 1, 0), \tag{H.4}$$

where $\bar{\gamma}_u$ is the average SNR of the URLLC users. Note that (H.4) is equivalent to the result obtained in [12].

For eMBB, transmission streams from different users can mutually interfere when they are scheduled in the same time-frequency resources, as in the case of multi-user MIMO. Assuming that the eMBB users have the same power control configuration, which leads to the same average power at the receiver as the desired eMBB source, we set $\Gamma_w = 1$. Assuming that all the available resources are simultaneously used by the N_e active users, we have that $w = N_e - 1$. The outage probability of eMBB without URLLC interference can be expressed as

$$P_{f,e} = P_f(\bar{\gamma}_e, 0, N_e - 1, 0, 1), \tag{H.5}$$

where $\bar{\gamma}_e$ is the average SNR of the eMBB users.

3.2 MMSE receiver and overlaying transmissions

When URLLC and eMBB have overlaying allocations, the reliability of the URLLC transmissions is not only affected by collisions with sporadic URLLC interferers, but also by the continuous eMBB interferers. Hence, we set $w = N_e$ and $\Gamma_w = \Omega$, besides $\Gamma_v = 1$. With that, the outage probability for the URLLC transmissions is calculated as

$$P_{f,u} = \sum_{z=0}^{N_u-1} P_c(z, N_u - 1) P_f(\bar{\gamma}_u, z, N_e, 1, \Omega). \tag{H.6}$$

Likewise, eMBB is also affected by the transmissions from the N_u URLLC users in the same band. Given that $\Omega = \bar{p_e}/\bar{p_u}$ as described in Section 2, the average URLLC interferer power relative to the desired eMBB source is the inverse of Ω. Hence, we set $\Gamma_v = 1/\Omega$ and $\bar{\gamma} = \bar{\gamma}_e = \bar{\gamma}_u\Omega$. At the same time, with other eMBB streams present with the same average interferer power, we have that $w = N_e - 1$ and $\Gamma_w = 1$. Then, the outage probability of the eMBB transmissions is given by

$$P_{f,e} = \sum_{z=0}^{N_u} P_c(z, N_u) P_f(\bar{\gamma}_u\Omega, z, N_e - 1, 1/\Omega, 1). \tag{H.7}$$

3.3 MMSE with SIC receiver and overlaying transmissions

With SIC we assume that URLLC traffic has to be decoded first, due to its strict latency. Then its interference contribution is removed from the receive signal. This means that only eMBB actually benefits from SIC. Given that, the outage probability of URLLC transmissions in this case can be also expressed by (H.6).

Assuming that $\epsilon_u << \epsilon_e$, the interference from failing URLLC transmissions, which cannot be canceled by SIC, is negligible. With eMBB not suffering from URLLC interference, the outage probability of the eMBB transmissions can be calculated with (H.5).

3.4 Achievable rate and load calculation

Using the described outage probability for each case, we can calculate numerically the minimum value for the SINR ψ to meet a given requirement. Here, we find ψ that satisfy $P_{f,u} = \epsilon_u$ for the URLLC cases, and $P_{f,e} = \epsilon_e$ for the eMBB cases. For a certain rate r in bps/Hz, the outage probability is expressed as $Prob[\log_2(1+\psi) < r]$. From this relation we can obtain the maximum rate corresponding to the outage probability requirement as

$$r[\text{bps/Hz}] = \log_2(1+\psi). \tag{H.8}$$

The achievable eMBB load, which corresponds to the maximum throughput with a given ϵ_e, is calculated as

$$L_e[\text{bps}] = rW_eN_e(1-\epsilon_e). \tag{H.9}$$

For URLLC transmission of a packet of size D in a bandwidth W_u and in a TTI of duration T, the transmission rate is given by

$$r_u[\text{bps/Hz}] = D/T/W_u. \tag{H.10}$$

With the correspondent SINR for this rate, i.e. $2^{r_u}-1$, we calculate numerically the maximum arrival rate $\hat{\lambda}$ that is allowed for a given number of URLLC users meeting the outage probability requirement. Then, the achievable URLLC load can be calculated as

$$L_u[\text{bps}] = D\hat{\lambda}N_u/T. \tag{H.11}$$

Given that ϵ_u is very low, the impact of transmission failures in the resultant load is considered negligible.

4 Numerical analysis

In this section we first present the achievable rate for URLLC transmissions overlaying a eMBB stream. We then find the achievable load for both kind of services, considering NR assumptions. Finally, a comparison between the allocation approaches is provided for different operation regimes.

4.1 Achievable rates for URLLC

For eMBB we consider $\epsilon_e = 10^{-1}$, whereas $\epsilon_u = 10^{-3}$ for URLLC. These values are usual block error rate targets for the initial transmission of these services, considering that a higher reliability is more efficiently achieved after retransmission [2]. We consider the case of MMSE with $M = 2$ and $M = 4$ receive antennas. A URLLC load is imposed with $N_u = 50$ users and packet arrival rate $\lambda = 10^{-2}$ per TTI for each user. Different relative receive power of eMBB with respect to the URLLC signals are assumed with $\Omega = \{1, 0.5, 0.1, 0\}$. Setting $\Omega = 0$ is equivalent to no eMBB, i.e. $N_e = 0$.

The achievable rate for URLLC depending on the SNR $\bar{\gamma}_u$ is shown in Fig. H.2. The interference-free curve denotes a benchmark assuming dedicated resources for each user. It is observed that the rate practically saturates after $\bar{\gamma}_u = 10\,\text{dB}$ for $M = 2$, i.e. a higher SNR does not yield on higher URLLC capacity. This is due to the eMBB interference and collisions with the imposed URLLC load. The achievable rate obviously increases with lower

values of Ω, since the SINR of URLLC increases. This means that, for guaranteeing high URLLC capacity, the power of URLLC signals should be higher than the ones of eMBB in the overlaying band. It is evident that $M = 4$ allows the highest rates due to the better interference rejection capability of the receiver. At $\bar{\gamma}_u = 10\,\text{dB}$ and $\Omega = 1$, it allows a rate just 3.3 times lower than the interference-free benchmark, compared to the 10 times lower with $M = 2$. The higher number of receive antennas allows higher URLLC rates and gives possible room for multiple eMBB streams.

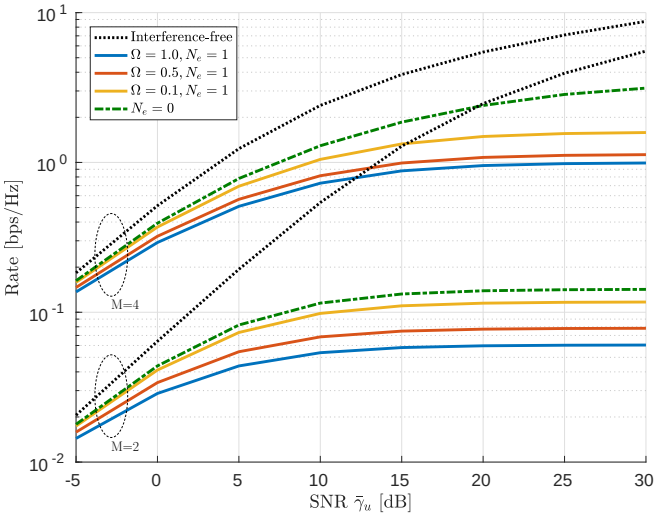


Fig. H.2: Achievable rates for URLLC overlaying one eMBB stream with different Ω, considering $N_u = 50$, $\lambda = 10^{-2}$, and MMSE with 2 and 4 antennas. For the interference-free curve it is assumed dedicated resources.

4.2 Achievable loads

Now we compare the resource allocation options for multiplexing URLLC and eMBB traffic, considering particular NR assumptions [4]. For that, we calculate the achievable load for each service according to the receiver type, average SNR, average interferer power relative to source, and allocated band. We consider a bandwidth $W = 10\,\text{MHz}$. For separate bands, we assume $R = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$, corresponding to full band for eMBB until full band for URLLC. For overlaying transmissions, we assume $\Omega = \{0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$, which corresponds to no eMBB until eMBB with same average receive power as URLLC. Given the

higher priority of URLLC, we do not consider the option of eMBB with higher average receive power than URLLC.

URLLC users transmit payloads of $D = 256$ bits using a short-TTI of 0.143 ms. This may represent the case of a NR mini-slot numerology with 4 symbols per TTI and 30 kHz sub-carrier spacing. The eMBB users transmit large volume of data exploiting capacity-achieving codes. In the following examples we assume $M = 4$ and $N_e = 2$, i.e. two eMBB streams are simultaneously active in the same band, as in MU-MIMO.

Four operation modes are considered:

- Separate bands and equal SNR: the average SNR is $\bar{\gamma}_u = \bar{\gamma}_e = \bar{\bar{\gamma}}$ for URLLC and eMBB, where $\bar{\bar{\gamma}}$ is the average SNR over the bandwidth W. It refers to a system in which users keep the same power spectral density (PSD) regardless of the operational bandwidth.

- Separate bands and scaled SNR: $\bar{\gamma}_u = \bar{\bar{\gamma}}/R$ for URLLC and $\bar{\gamma}_e = \bar{\bar{\gamma}}/(1-R)$ for eMBB, i.e. the average SNR is increased as much as the associated bandwidth decreases. It refers to a system where users maintain the same output power regardless of the operational bandwidth.

- Overlay with SIC: overlaying transmissions considering MMSE with ideal SIC and different values of Ω.

- Overlay without SIC: overlaying transmissions with MMSE receiver and different values of Ω.

Fig. H.3 and Fig. H.4 show the achievable loads for eMBB and URLLC in a low SNR scenario ($\bar{\gamma}_u = 0$ dB in full band) and medium SNR scenario ($\bar{\gamma}_u = 10$ dB in full band), respectively. Each line delimits the maximum load that can be achieved depending on R or Ω, while meeting the requirements given by ϵ_e and ϵ_u. The region to the left of the line represents lower load combinations that can be supported. The maximum supported URLLC load is denoted by $\hat{L}_u$. At 20% of $\hat{L}_u$ is indicated a low URLLC load regime, and at 80% of $\hat{L}_u$ is indicated a high URLLC load regime. The maximum gain of overlaying allocation relative to using separate bands in terms of eMBB throughout is denoted by $G_{o,e}$.

In the low SNR scenario as it is shown in Fig. H.3, we observe that the separate bands and equal SNR operation (dashed red line) shows the lowest achievable loads. For example with $R = 0.5$, only up to 1 Mbps can be reliably supported for URLLC, and up to 11 Mbps for eMBB. This performance can happen when same power control settings are used for both services. On the other hand, for separate bands and service SNR scaling with R (solid red line), the performance is generally better. For overlay without SIC (dashed blue line), a lower achievable load is experienced for both services compared to the use of separate bands as in the previous case. For

example with $\Omega = 0.8$ and 2 Mbps URLLC load, up to 14 Mbps can be reliably supported for eMBB, while 17 Mbps can be reached if traffic is conveyed in separate bands. While for overlay with SIC (solid blue line), there is an advantage of overlaying when the URLLC load is lower than 2.4 Mbps, due to the reduced interference in this condition. Anyway, it can be noted that overlaying is generally not a good option in low SNR cases.

For the medium SNR scenario in Fig. H.4, there is a clear advantage of overlaying when MMSE with SIC is used. Without noise limiting and canceled URLLC interference, the antenna combining can strength the eMBB signal boosting its throughput. However, without SIC the achievable load for both services is higher if separate bands are allocated. This avoids that the mutual interference between the traffic penalizes the performance of each other. Given that the URLLC rate saturates, the result for a high SNR scenario is omitted here, though the same observations as for medium SNR are valid.

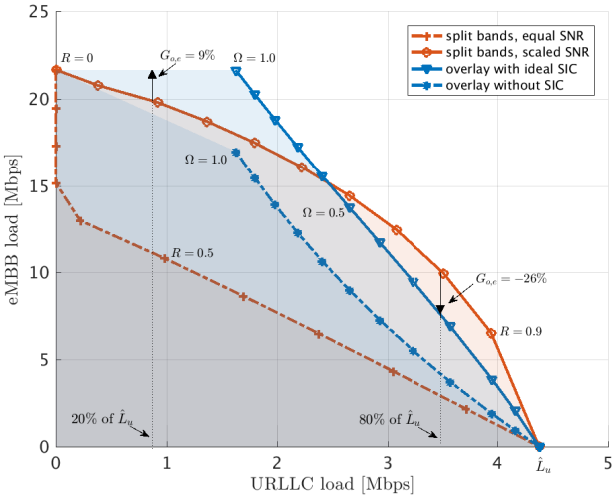


Fig. H.3: Achievable loads for URLLC and eMBB considering different receive strategies and low average SNR $\bar{\gamma}_u = 0$ dB. $W = 10$ MHz, $D = 256$ bits, $N_u = 50$, $N_e = 2$ and $M = 4$.

4.3 Comparison for different regimes

Fig. H.5 shows the gain $G_{o,e}$ of overlaying relative to separate bands allocation in terms of eMBB throughput, for low and high URLLC load regimes. Two packet sizes, $D = 256$ bits and $D = 1600$ bits, are assumed for URLLC. Besides, we also assume two values for the outage probability targeted for URLLC. $\epsilon_u = 10^{-3}$ refers to a system in which a higher reliability can be achieved after a retransmission, and $\epsilon_u = 10^{-5}$ refers to a system where the

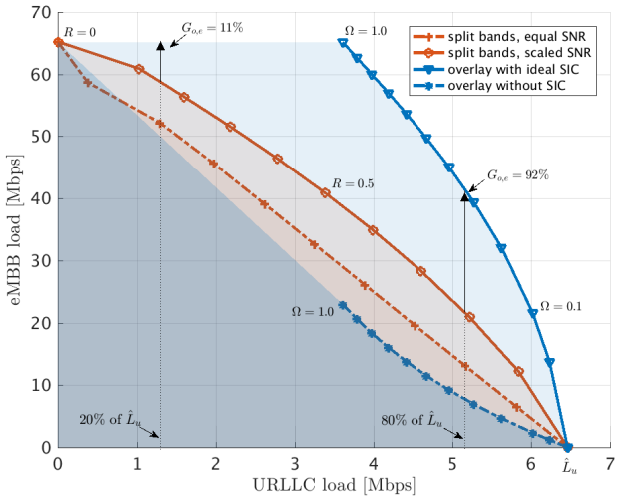


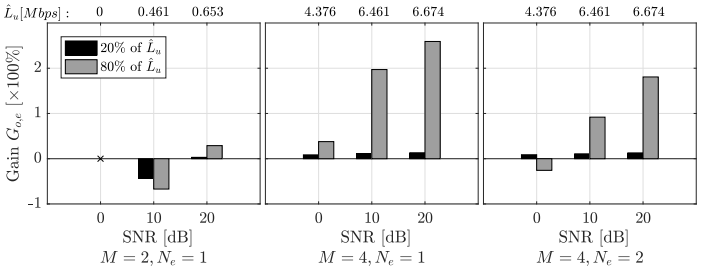
Fig. H.4: Achievable loads for URLLC and eMBB considering different receive strategies and moderate average SNR $\bar{\gamma}_u = 10\,\text{dB}$. $W = 10\,\text{MHz}$, $D = 256\,\text{bits}$, $N_u = 50$, $N_e = 2$ and $M = 4$.

reliability target should be achieved with a single shot transmission. The absolute values of the maximum supported URLLC load $\hat{L}_u$ for each case are shown on the top of the plots.

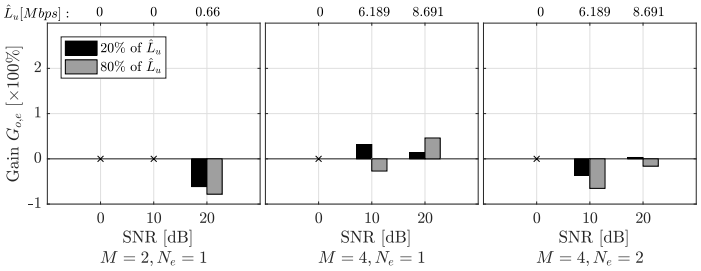
In many cases marked with "x", we note that no URLLC load can be supported. This is observed in most cases for $M = 2$ in low SNR scenarios, independent of the allocation scheme. As can be seen in Fig. H.5a and Fig. H.5c, for small packet size there is a significant gain of overlaying at high SNR, specially for 4 receive antennas and high URLLC load regime (up to +260%). In case of large packets as shown in Fig. H.5b and Fig. H.5d, overlaying allocation may lead to losses, while minor gains appears only in case of $M = 4$ antennas and $N_e = 1$ eMBB stream, at high SNR. For stricter reliability such as 10^{-5}, the gain of overlaying is reduced, and losses get more evident with the 1600 bits packets.

5 Discussion

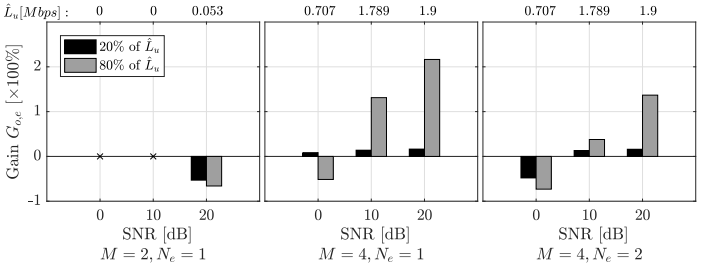
In many cases the allocation of separate bands for each service class shows to be more efficient, specially when SIC is not employed. In practice, it implies that the bandwidth needs to be reconfigured for all grant-free users whenever the target supported load changes. This results in additional control signaling overhead. To avoid this issue, for instance in a scenario where the URLLC load varies very often, it would be recommended to proactively allocate a



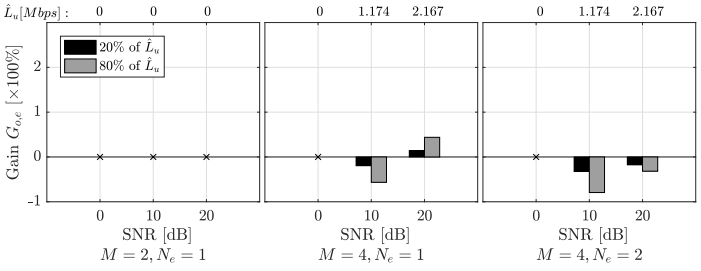
(a) $D = 256\,\text{bits}, \epsilon_u = 10^{-3}, T = 0.143\,\text{ms}$.



(b) $D = 1600\,\text{bits}, \epsilon_u = 10^{-3}, T = 0.143\,\text{ms}$.



(c) $D = 256\,\text{bits}, \epsilon_u = 10^{-5}, T = 0.250\,\text{ms}$.



(d) $D = 1600\,\text{bits}, \epsilon_u = 10^{-5}, T = 0.250\,\text{ms}$.

Fig. H.5: Gain of overlaying relative to separate bands allocation in terms of eMBB throughput for different settings.

larger share of the bandwidth for URLLC to cope with the load variation, to the detriment of the eMBB capacity.

For scenarios with low average SNR, e.g. macro deployments, the gains of overlaying transmission using SIC are insignificant compared to operating with a simple MMSE receiver. Besides, even when SIC is available, the crossing regions indicate that it is beneficial to switch between separate bands and overlaying mode depending on the load aimed for each service. On the other hand, in a dense deployment with medium/high SNR, the application of a more complex receiver with SIC is more relevant, given the higher achievable loads.

It is important to note also that, for a network with users that have multiple traffic types, as for eMBB and URLLC services, it is beneficial to use different transmission parameters for each kind of service. This means, for example, that one user should be configured with a power control setting for eMBB and another for URLLC.

The proposed approach presented in this paper can be also relevant for feasibility analysis and decision making. For example, by assigning costs to each traffic, one can find the optimal load balance policy that results in the highest profit, and select the corresponding bandwidth shares or the power control settings for that.

6 Conclusion

In this work we studied how to efficiently multiplex grant-free URLLC and eMBB services in the uplink. Two possible options of multiplexing are considered, namely, separate bands and overlaying transmissions. We describe the outage probability for each service and for each multiplexing option considering MMSE receiver and MMSE with SIC. With this approach we can compare the achievable load that can be supported for each traffic. The resource allocation considers different shares of the bandwidth for each traffic in separate bands, or different relative receive power when the transmissions are overlaying. Numerical analyses considering NR assumptions are carried out. The results show that overlaying provides better performance generally using MMSE with SIC either in high SNR or for low URLLC loads. Separate bands for each service class is better when a SIC processing is not employed, the URLLC packet size is large and higher reliability levels are required for URLLC. Future work should consider traffic bursts and the effect of power limitation for overlaying transmissions.

7 Acknowledgments

This research is partially supported by the EU H2020-ICT-2016-2 project ONE5G. The views expressed in this paper are those of the authors and do not necessarily represent the project views.

References

[1] ITU-R, "Report ITU-R M.2410-0 - Minimum requirements related to technical performance for IMT-2020 radio interface(s)," International Telecommunication Union (ITU), Tech. Rep., Nov. 2017.

[2] G. Pocovi, H. Shariatmadari, G. Berardinelli, K. Pedersen, J. Steiner, and Z. Li, "Achieving Ultra-Reliable Low-Latency Communications: Challenges and Envisioned System Enhancements," *IEEE Network*, vol. 32, no. 2, pp. 8–15, Mar. 2018.

[3] P. Popovski, J. J. Nielsen, C. Stefanovic, E. d. Carvalho, E. Strom, K. F. Trillingsgaard, A. S. Bana, D. M. Kim, R. Kotaba, J. Park, and R. B. Sørensen, "Wireless Access for Ultra-Reliable Low-Latency Communication: Principles and Building Blocks," *IEEE Network*, vol. 32, no. 2, pp. 16–23, Mar. 2018.

[4] 3GPP TR 38.802 v14.0.0, "Study on New Radio Access Technology," Mar. 2017.

[5] R1-1803659, "UL multiplexing between URLLC and eMBB," Apr. 2018.

[6] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[7] H. Gao, P. J. Smith, and M. V. Clark, "Theoretical reliability of MMSE linear diversity combining in Rayleigh-fading additive interference channels," *IEEE Transactions on Communications*, vol. 46, no. 5, pp. 666–672, May 1998.

[8] G. Berardinelli and H. Viswanathan, "Overlay transmission of sporadic random access and broadband traffic for 5G networks," in *2017 International Symposium on Wireless Communication Systems (ISWCS)*, Aug. 2017, pp. 19–24.

[9] C.-P. Li, J. Jiang, W. Chen, T. Ji, and J. Smee, "5G ultra-reliable and low-latency systems design," in *2017 European Conference on Networks and Communications (EuCNC)*, Jun. 2017, pp. 1–5.

[10] A. Anand, G. de Veciana, and S. Shakkottai, "Joint Scheduling of URLLC and eMBB Traffic in 5G Wireless Networks," *ArXiv e-prints*, Dec. 2017. [Online]. Available: http://arxiv.org/abs/1712.05344

[11] P. Popovski, K. F. Trillingsgaard, and G. D. Osvaldo Simeone, "5G Wireless Network Slicing for eMBB, URLLC, and mMTC: A Communication-Theoretic View," *CoRR*, vol. abs/1804.05057, 2018. [Online]. Available: http://arxiv.org/abs/1804.05057

[12] G. Berardinelli, R. Abreu, T. Jacobsen, N. H. Mahmood, K. Pedersen, I. Z. Kovács, and P. Mogensen, "On the Achievable Rates over Collision-Prone Radio Resources with Linear Receivers," in *2018 IEEE 29th PIMRC*, Sep. 2018.

[13] K. I. Pedersen, G. Berardinelli, F. Frederiksen, P. Mogensen, and A. Szufarska, "A Flexible 5G Frame Structure Design for Frequency-Division Duplex Cases," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 53–59, Mar. 2016.

[14] P. Wu and N. Jindal, "Coding versus ARQ in Fading Channels: How Reliable Should the PHY Be?" *IEEE Transactions on Communications*, vol. 59, no. 12, pp. 3363–3374, Dec 2011.

Part V

Conclusions

Conclusions

1 Summary of the Main Findings

The stringent requirements of Ultra-Reliable Low-Latency Communications (URLLC) in 5G systems demand new radio interface solutions, different from the ones usually applied for high throughput broadband services. This dissertation focused on the design of radio access and resource management solutions for efficiently enabling URLLC in the uplink. Grant-free access is placed as an enabler for fast uplink transmissions. However, this is accompanied by challenges in terms of efficient usage of the radio resources or potential interference between users. In this research, solutions were presented for achieving the URLLC requirements with improved aggregated load in the system. The evaluations were carried by means of analytical tools and detailed system level simulations.

Resource preallocation mechanisms should be utilized for reducing the dependency on control information exchanges. With semi-persistent scheduling, dedicated radio resource can be assigned per user for reliable transmissions. Moreover, the proposed schemes shown in Part II, based on preallocation of retransmission resources for group of users, allow reducing the signaling also for retransmissions. Resource efficiency gains in the order of 20% are achieved comparing with conservative single transmission solutions. And using blind retransmissions and SIC, latency reductions of at least 40% are obtained with the suppression of the feedback and the effect of HARQ round trip time.

For sporadic URLLC traffic, grant-free access using shared resource allocations shows to be a viable solution for fast uplink transmissions. Also in this case, the chosen transmission and retransmission strategy have a critical role in respect to the latency and reliability performance. Detailed system level simulations were carried to include realistic effects of a multi-cell 5G urban macro scenario. Repetition schemes have the potential to guarantee the reliability within the shortest latencies. However, the increased interference and queuing effect caused by multiple replicas limit the supported URLLC load. With mini-slot structures and fast user equipment (UE) processing ca-

pability assumed for 5G systems, short round-trip time can be reached. Considering that, an HARQ retransmission scheme should be preferable, since it allows to achieve the target reliability with reduced channel usage. As an example, by permitting one HARQ retransmission the URLLC load can be at least four times higher compared with the usage of two repetitions.

The presented studies in Part III show that the achievable reliability in the URLLC system is very sensitive to the transmit power control settings. It is clear that, in order to achieve the best URLLC system performance, the settings should be optimized considering reliability and latency rather than throughput. As a starting point, full path loss compensation should be utilized to avoid a reliability penalty at the cell-edge. Moreover, the target receive power per resource block should be optimized so that the transmission compensate for the fading variations, while not causing excessive interference. The analysis shows that, by changing from usual throughput oriented setting to optimized settings for reliability, the supported URLLC load in the system could be virtually doubled. The use of power boosting mechanisms for retransmissions reduces the outage probability. The improvement is limited in macro scenarios due to the UEs operating close to maximum power.

Based on the learnings from the system level study, a solution encompassing multiple active grant-free configurations, i.e. with different sub-band sizes, associated modulation and coding scheme (MCS) and power control settings, is proposed. The assignment of high MCS and smaller sub-bands for UEs in better average channel condition reduces the transmission overlapping, improving the overall URLLC capacity by approximately 90% compared to conservative approaches that use a fixed robust MCS. The importance of exploiting spatial diversity and interference rejection capability given by multi-antenna receivers is also highlighted. By employing linear receiver structures such as MMSE-IRC along with low MCS, multiple colliding transmissions can be resolved. The results show that the URLLC load can be increased by up to 7 times just by changing from 2 to 4 receive antennas.

Multi-cell reception of grant-free transmissions was thoroughly studied for the sake of improving the URLLC system performance. It permits harvesting spatial and interference diversity, and improve the robustness by combining soft information received by assisting cells. It was observed that soft combining has great potential for improving the URLLC supported load with the cost of higher backhaul traffic. For the cases where the base stations are equipped with 2 receive antennas and HARQ retransmissions are not employed, the usage of multi-cell reception allows more than 4 times higher URLLC load. And even in cases of 4 receive antennas and HARQ retransmissions are enabled, multi-cell reception can still improve the URLLC supported load by 40%.

Finally, the issues related to the multiplexing of eMBB and sporadic URLLC transmissions using grant-free resources were studied in Part IV. Overlaying

allocation is considered for improving the resource utilization. In this case, transmit power control configuration plays again an important role. The system level evaluation shows that only low URLLC loads are supported when overlaying with eMBB, while a capacity penalty is still imposed to the latter. An analytical method is derived for calculating the outage probability and achievable loads for eMBB and URLLC, and used for comparing overlaying versus separate bands allocation for the services. The numerical analysis reveals that overlaying allocation leads to better performance in certain operation regimes like, at high SNR, low URLLC load, and with the use of MMSE receivers with successive interference cancellation capability. Otherwise, the use of separate resources for each service is recommended, including the cases of strict reliability required in a single transmission, for low SNR, and for higher payload size. This motivates the development of preemption schemes to be employed when the URLLC service requirements give room for grant-based scheduling.

2 Recommendations

The following main recommendations are provided addressing the research questions stated in Part I:

Q1 What transmission/retransmission schemes should be utilized for achieving the best URLLC performance in uplink?

R1 Grant-free transmissions should be employed for achieving the best latency performance, and retransmissions should be used for enhancing the reliability. Preallocation of retransmission resources for group of users can be employed to avoid rescheduling signaling issues. Blind repetitions are recommended when latency target is very strict, i.e. below 1 ms, with the cost of lower supported load. While HARQ is recommended when short TTI and fast processing time allows for at least one retransmission before the deadline.

Q2 How to improve the resource efficiency of URLLC for supporting higher achievable loads in the system?

R2 Power control with full path loss compensation should be used with P_0 optimized for URLLC performance. Power boosting retransmission can be employed for further reducing the outage. Employing multiple grant-free configurations is recommended, associating users to subbands and MCS according their average channel condition, for reducing the probability of fully overlapping transmissions. MMSE-IRC receivers with, e.g., 4 antennas, should be used providing interference rejection

capability and diversity. Multi-cell reception can be used with 3 cooperating cells, giving diversity combining.

Q3 How to multiplex enhanced Mobile Broadband (eMBB) and grant-free URLLC to support both services with improved resource utilization?

R3 eMBB and grant-free URLLC should be multiplexed in overlaying resource allocation in the following conditions: in high SNR, with successive interference cancellation receiver, with low URLLC load and lowered eMBB power. While eMBB and URLLC should be allocated in separate resources when: in low SNR, strict URLLC reliability to be achieved in a single transmission, without interference cancellation receiver and for high URLLC load target.

3 Future Work

Many aspects regarding the support of uplink URLLC in next generation cellular networks are still to be investigated, despite of the findings presented herein. This section describes some of these aspects, which could not be addressed in this work due to the limited time and restricted scope of the research project.

So far it has being considered in NR that the misdetection probability of uplink transmissions using configured grants is sufficiently low. However, non-idealities of the demodulation reference signal (DMRS) can impact on the detection performance, as discussed in [1]. The study in [2] takes detection issues into account in a simplified analytical model. However, detailed link level simulations may be required to understand the detection performance depending on the DMRS configuration, false alarm target, sequence lengths and number of simultaneous transmissions.

As was shown, the multiplexing of traffic with different characteristic using overlaying allocation can lead to a high performance degradation for URLLC. Other multiplexing options based on preemption mechanisms, should be further studied. The support of such solutions, requires improvements on the control channels reliability and monitoring capability by the eMBB UEs being preempted. Therefore, further studies should consider the reliability of the signaling used for indicating the URLLC transmissions, as well as the impact of pausing the eMBB transmission flow.

The aimed performance for URLLC has shown to be feasible and sufficient for satisfying the baseline IMT-2020 requirements. Yet, some future use cases impose specific requirements which demands further improvements of the radio access network. For example, time sensitive networking (TSN) utilized for factory automation presents latency requirements down to 0.5 ms [3]. Moreover, such systems are characterized by isochronous and deterministic

communication, in which an absolute time cycle for the transmissions should be respected. Other factors like mean time between failures, jitter, availability and survival time are also critical. Various technical components should be enhanced to support these systems, including accurate timing reference throughout the network, strict QoS assurance and compatible interfacing with standard industrial Ethernet systems [4].

Advanced multi-antenna techniques, such as massive-MIMO, should be exploited for improving the reliability through channel hardening and beamforming gain. This solution should be employed in sub-6 GHz band for preventing channel blockage issues [5]. It should be observed that these techniques can leverage their full performance if accurate channel estimation can be obtained. Therefore, it turns to be efficient mainly in static channel conditions or when exploiting channel reciprocity in time division duplex (TDD) mode, which brings latency issues.

References

[1] R1-1901330, "Summary of 7.2.6.3 Enhanced configured grant PUSCH transmissions," Jan. 2019.

[2] G. Berardinelli, N. H. Mahmood, R. Abreu, T. Jacobsen, K. Pedersen, I. Z. Kovács, and P. Mogensen, "Reliability Analysis of Uplink Grant-Free Transmission Over Shared Resources," *IEEE Access*, vol. 6, pp. 23 602–23 611, Apr. 2018.

[3] 3GPP TS 22.104 v16.1.0, "Service requirements for cyber-physical control applications in vertical domains," Mar. 2019.

[4] A. Neumann, L. Wisniewski, R. S. Ganesan, P. Rost, and J. Jasperneite, "Towards integration of Industrial Ethernet with 5G mobile networks," in *2018 14th IEEE International Workshop on Factory Communication Systems (WFCS)*, Jun. 2018.

[5] E. Bjornson, L. Van der Perre, S. Buzzi, and E. G. Larsson, "Massive MIMO in Sub-6 GHz and mmWave: Physical, Practical, and Use-Case Differences," *IEEE Wireless Communications*, Jan. 2019.

Part VI

Appendix

Paper I

System Level Analysis of K-Repetition for Uplink Grant-Free URLLC in 5G NR

Thomas Jacobsen, Renato Abreu, Gilberto Berardinelli, Klaus Pedersen, István Z. Kovács, Preben Mogensen

The paper has been published in the
European Wireless, 2019

Abstract

Ultra-reliable low-latency communications (URLLC) sets high service requirements for the fifth generation (5G) new radio (NR) standard. Grant-free (GF) transmissions is considered a promising technique for reducing the latency in the uplink. To achieve efficient radio resources utilization, sharing of resources is required for sporadic uplink traffic. Repetitions based transmission schemes aims to enhance the reliability of GF transmissions. However, repetitions may also generate excessive interference and cause additional queuing, harming the reliability and latency. In this work, we explore radio resource management (RRM) configurations for repetition based transmission schemes. That includes the number of repetitions, the allocation size per transmission (sub-band), sub-band hopping and uplink power control. Evaluations are conducted in a 5G NR compliant multi-user multi-cell simulation scenario with sporadic uplink GF URLLC transmissions. Our findings suggest that repetitions based schemes can, with a careful selection of the sub-band size and uplink power control parameters, achieve comparable URLLC performance with retransmission based schemes when the effect of queuing is disregarded.

1 Introduction

The fifth generation (5G) new radio (NR) standard target to support the challenging Ultra-Reliable Low-Latency Communication (URLLC) service requirements [1]. The third generation partnership project (3GPP) has adopted the baseline URLLC requirement which is 1 ms one-way latency deadline for transmitting a packet with a reliability of 99.999% [2]. Grant-free (GF) is a recognized approach to reduce the latency in uplink transmissions, by skipping the scheduling request procedure. With unpredictable URLLC traffic, GF transmissions over orthogonal preallocated resources becomes resource inefficient as resources can be left unused. Sharing of preallocated resources between URLLC sources, can enhance the resource efficiency [3]. The price to pay, is that GF transmissions become subject to intra-cell interference. Retransmission schemes such as hybrid automatic repeat request (HARQ) are known for improving the transmission reliability. However, it comes at the expense of an increased latency as the terminal needs to wait for the feedback before performing a retransmission, being affected by the feedback round-trip-time (RTT) [4].

Different transmission schemes have been considered for enabling GF URLLC. The use of repetitions is one simple way of enhancing the reliability, by transmitting consecutive replicas of the packet without waiting for feedback prior to transmitting the next one. The 3GPP NR Release-15 standard has established the configuration of GF transmissions, known as configured grant, through radio resource control (RRC) with possible activa-

tion via downlink control channel [5]. The framework allows the configuration of the physical layer parameters including the settings of K-repetitions, i.e. K consecutive transmissions of the same packet. Our recent work [6] evaluated three schemes for sporadic GF URLLC transmissions in uplink; K-repetitions, Reactive HARQ and Proactive (repetitions with early termination), along with a grant-based reference. Results strongly indicated that the K-repetitions scheme was subject to high interference from the excessive channel use. Full-band transmission repetitions was used, hence not considering the use of higher order modulation and coding scheme (MCS) and hopping between sub-bands. Contention-based transmission schemes using repetitions are studied in [7], where the optimum number of consecutive transmissions is found. A simplified scenario and reception model are considered. In [8] deterministic access patterns based on combinatorial code design are utilized and shows promising gains compared to transmission in random chosen access slots, when ideal interference cancellation of decoded replicas is assumed. Recent work [9] evaluates a repetition based scheme along with two feedback based schemes using analytical tools in a single-cell scenario. The contribution does not consider the effect of inter-cell interference, NR system settings for evaluation and the possibility of transmission repetitions to finish earlier than the feedback based schemes.

This work conducts a thorough evaluation of the transmission repetition parameters; number of repetitions, the chosen MCS and resource allocation in multiple sub-bands, hopping through the allocated sub-bands, along with optimized uplink power control settings. A feedback stop-and-wait retransmission scheme referred to as Reactive HARQ is included as baseline. The evaluation is done using detailed system level simulations capturing the major performance influencing factors in both, the multiple-access protocol layer and physical layer in the radio access network stack, with commonly agreed models in 3GPP. The simulator is also used e.g. in [10, 11].

The reminder of the paper is structured as follows. Section 2 describes the network and traffic model. Section 3 presents the K-repetition transmission scheme with intra-slot frequency hopping. The simulation assumptions and methodology are described in Section 4. Section 5 presents the performance evaluation, followed by Section 6, which concludes the work.

2 Setting the scene

We consider a multi-user multi-cell synchronous network consisting of C cells and N URLLC user equipments (UE) uniformly distributed per cell. We assume that the UE connect to the strongest cell, and acquires full synchronization with the network in both time and frequency. Each URLLC UE generates a small packet of size B according to a Poisson arrival process with average

rate λ. The aggregated URLLC offered load per cell is therefore given by $L = \lambda \cdot N \cdot B$.

The URLLC UEs are configured for GF transmission over a set of preallocated radio resources. These resources can span multiple sub-bands and are available in every transmission time interval (TTI). We consider an OFDM uplink channel with a bandwidth composed of BW resource blocks (RB) available in the frequency domain. The BW RBs are divided into n sub-bands. Short TTI of duration T are used for GF transmissions. The base station configures the UEs to transmit K consecutive replicas of the packet, hopping to a randomly selected sub-band at each transmission attempt. Note that the same sub-band can be selected with a certain probability, limiting the gain in terms of frequency diversity. However, the potential of interference diversity is kept in this case. It is also important to observe that, this approach is different from the hopping mechanism specified in 3GPP Release-15 [12], which only allows alternate hopping between two sub-bands. Besides, the support of intra-slot repetition within the 14 symbols slot is still under discussion in 3GPP for Release-16 [13].

With the fixed packet size B and bandwidth BW, increasing n also mean that the size of each sub-band is reduced, which implies that the transmission MCS needs to be increased, as illustrated in Fig. I.1 for different options of n and for $BW = 48$ RBs. Open loop power control is utilized to regulate the target receive power density at each cell as defined in [14].

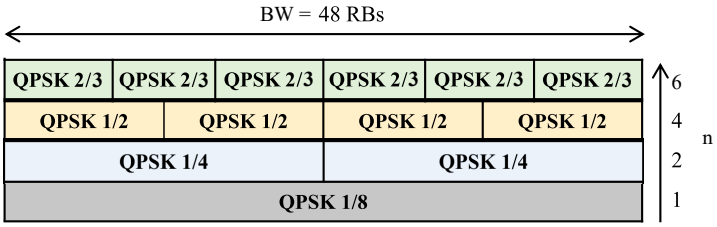


Fig. I.1: Examples of radio resource allocations of n sub-bands and corresponding MCS over BW RBs [15].

3 K-repetitions scheme

Upon arrival of a URLLC packet for immediate transmission at the UE, the packet is prepared for transmission and when ready, the data transmission is performed in the next TTI. For $K > 1$ the repetitions are assumed to be carried out in consecutive TTIs. Upon the end of each transmission, the receiving cell needs to process the received packet and for $K > 1$, combine the received repetitions. A maximum of one transmission can be carried out per TTI per UE. Therefore, ongoing transmissions may force a new packet arrival to wait

until its completion, hence being subject to queuing. The latency of a packet that is decoded after $1 < k < K$ replicas is therefore given by

$$t_k = t_{queue} + t_{prep} + t_{align} + k \cdot t_{TTI} + t_{proc}. \tag{I.1}$$

While the latency contributions t_{prep}, t_{proc}, total transmission time $k \cdot t_{TTI}$ and t_{align} are either known or its upper bound are given, t_{queue} upper bound is not straight forward to determine as it depends on the UE load subject to λ and the number of repetitions K. It should be noted from (I.1) that, the latency is counted from the moment that the packet is generated, until the moment that any replica is successfully received. The latency of packets that are not received after K-repetitions is accounted as infinite.

Different realizations of GF transmissions are shown in Fig. I.2 where GF transmissions are carried out with $K = 2$ and for different number of sub-bands n using sub-band hopping. Increasing the number of sub-bands means that, for unchanged L and the number of transmission repetitions K, the probability of overlaying transmissions is reduced. Further, with $K > 1$ and multiple sub-bands ($n > 1$), frequency hopping can be applied to randomize and reduce systematic transmission overlaying. The total collision probability, i.e. that all K repetitions from a UE have an overlaying transmission, as a function of K and n is shown in Fig. I.3 using (9) from [7]. The load in this case is generated by $N = 100$ UEs and $\lambda = 10$ packets per second (PPS). From Fig. I.3 we observe that the collision probability is reduced when $K > 1$ and $n > 1$.

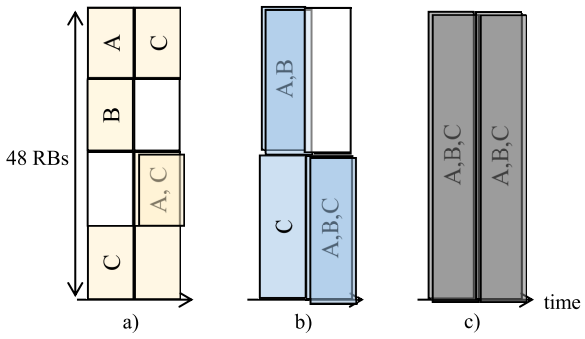


Fig. I.2: Realizations of GF transmissions with n sub-bands over K repetitions using sub-band hopping for UEs A, B and C.

Though the total collision probability tends to decrease with K and n, this does not necessarily lead to a reliability improvement. Increasing n and the corresponding MCS, also implies that a higher energy per bit is needed to sustain a transmission reliability target. This can be obtained either by increasing K or increasing the receive power density target through uplink

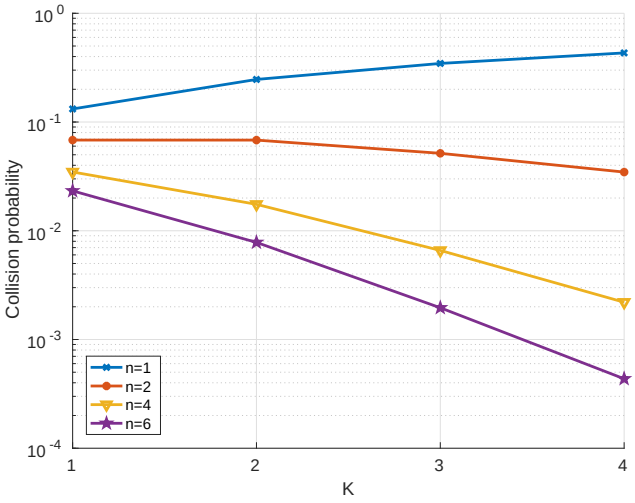


Fig. I.3: Collision probability as a function of the number of sub-bands n and repetitions K using (9) from [7]. The load is given by $N = 100$ UEs with $\lambda = 10$ PPS.

power control, which both implies an increase in channel usage or interference power. Further, the choice of K is bounded by the URLLC latency requirement. And the received power density target is bounded by the UE maximum transmission power. It is therefore not a trivial optimization problem to maximize the URLLC performance, while accounting the diversity gains of using repetitions on sub-bands, the additional interference generated by the repetitions and the uplink power control.

4 Evaluation Methodology

For the performance evaluation we use system level simulations. The evaluation assumptions are in line with URLLC evaluations for 5G NR defined in [16], and are summarized in Table I.1. A network consisting of $C = 21$ cells is used. The cells are distributed at 7 sites with 3 sectors each, resulting in a regular hexagonal urban macro layout with an inter-site distance of 500 m. URLLC UEs are uniformly distributed outdoors. The uplink bandwidth is 10 MHz, spanning $BW = 48$ RBs. Each RB has 12 sub-carriers with a spacing of 15 kHz. A mini-slot of 2 OFDM symbols is used giving a TTI length of $T = 0.143$ ms. The 3D Urban Macro (UMa) channel model is used.

Traffic is generated with a Poisson arrival rate $\lambda = 10$ PPS per UE and $B = 32$ bytes. The packet generation rate was chosen as a trade-off between queuing, number of deployed UEs and simulation time. The offered load

Table I.1: Simulation assumptions

Parameters	**Assumption**
Layout	Hexagonal grid composed of 7 sites with 3 sectors/site (21 cells), 500 meters of inter-site distance, wrap-around enabled
Channel model	3D Urban Macro (UMa)
Carrier frequency	4 GHz
UE distribution	100% uniformly distributed outdoor, 3 km/h for modeling fading channel
Base station receiver	MMSE-IRC with 2 antennas
Receiver noise figure	5 dB
Thermal noise	-174 dBm/Hz
UE transmitter	1 antenna, max. transmit power of 23 dBm
Bandwidth	10 MHz
Frame numerology	15 kHz sub-carrier spacing, $t_{TTI} = 0.143$ ms short-TTI (2 symbols mini-slot), 12 sub-carriers/RB
Latency contributions	$t_{prep} = t_{TTI}$, $t_{proc} = t_{TTI}$ and $t_{align} = [0, t_{TTI}]$,
Configured grant	2-symbols periodicity (every TTI), $n = 1$ use 48 RBs (QPSK1/8), $n = 2$ use 24 RBs (QPSK1/4), $n = 4$ use 12 RBs (QPSK1/2), $n = 6$ use 8 RBs (QPSK3/4). Random sub-band hopping is allowed.
URLLC traffic model	FTP Model 3 with Poisson arrival rate of $\lambda = 10$ packets/sec per UE and $B = 32$ bytes payload

is varied by changing the number of UEs per cell. It is assumed that each generated replicas is transmitted using the same redundancy version, and that the receiver combines them using chase combining.

A minimum-mean square error with interference rejection combining (MMSE-IRC) receiver with 2 antennas is assumed. The successful reception of a transmission sample depends on the SINR after the receiver combining. The post-processing SINR values for all sub-carrier including inter- and intra-cell interference are calculated and converted, according to the modulation, to a symbol-level mutual information metric as described in [17]. This metric is mapped through a link-to-system table, depending on the coding rate, to a block error probability value. This value is used for determining if the packet was successful or not. The latency of the packet is then registered, counting from the moment the packet arrived in transmitter buffer until the moment it was successfully received.

The key performance indicator is the achieved outage probability, i.e. the complement of the reliability, which the target for URLLC is 10^{-5} before 1 ms. The evaluation methodology is conducted in two steps. Firstly, a sensitivity

study on the achieved outage probability according the number of sub-bands n relative to the receive power density target P_0, is conducted. This is made for both, $K = 2$ and $K = 4$, as they fit with 1 ms latency requirement given the adopted numerology. Secondly the maximum load L, of which the reliability requirement can be met is found for $K = 2$, $K = 4$ when the best choices of n and P_0 found in the first step are applied. The sensitivity study is conducted using a similar methodology as the one presented in [11], where it is applied on the reactive HARQ baseline scheme.

5 Performance evaluation

Firstly, we search empirically for the optimal power control setting that leads to the lowest outage probability for each scheme. Four different numbers of sub-bands are considered with $n = \{1, 2, 4, 6\}$. This means sub-bands size of 48, 24, 12 and 8 RBs using MCSs QPSK1/8, QPSK1/4, QPSK1/2 and QPSK3/4 respectively. The offered load is $L = 0.256$ Mbps per cell, equivalent to $N = 100$ UEs per cell transmitting $B = 32$ bytes packets with $\lambda = 10$ PPS each. This load was observed to be the highest URLLC load achievable with the baseline reactive HARQ scheme in this scenario [11].

Fig. I.4 shows the obtained outage probability after K-repetitions for $K = 2$. It possible to note that the lowest outage probability obtained are comparable for QPSK1/8 with $P_0 = -107$ dBm, QPSK1/4 with $P_0 = -104$ dBm and QPSK1/2 with $P_0 = -98$ dBm. The optimal P_0 value naturally increases with the MCS given the higher SINR requirement for reliable decoding. The outage probability value in the order of 10^{-4} indicates that the URLLC reliability target can not be met with any of the settings for the applied load. This means that the gain from applying more sub-bands does not sufficiently compensate for the extra interference caused with the repeated transmission.

The same analysis is carried for K-repetitions with $K = 4$ in Fig. I.5. In this case we can note an considerable improvement in the outage probability, when comparing the best performance obtained with QPSK1/8 and the performance with a higher order MCS such as QPSK1/2. The achieved outage probability using QPSK1/2 with $P_0 = -98$ dBm gets down to the order of 10^{-5} after the 4 repetitions. The better performance is due to the higher diversity and combining gain obtained with the repetitions in detriment of the higher interference caused by the replicas. With $K = 4$ more energy per bit can be accumulated in time improving the robustness.

The cumulative distribution function (CDF) of the SINR for each scheme, using the configuration that allows the lowest outage probability, is shown in Fig. I.6. The increase on 50th percentile SINR between HARQ, $K = 2$ (2-rep) and $K = 4$ (4-rep) corresponds respectively to the increase in optimum P_0 value. 2-Rep has similar SINR tail as HARQ, however due to higher MCS

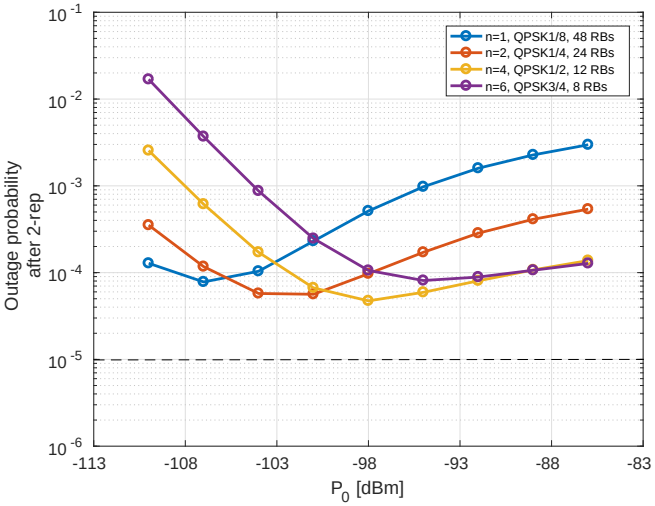


Fig. I.4: Sensitivity of outage probability in relation to P_0 and n for $K = 2$.

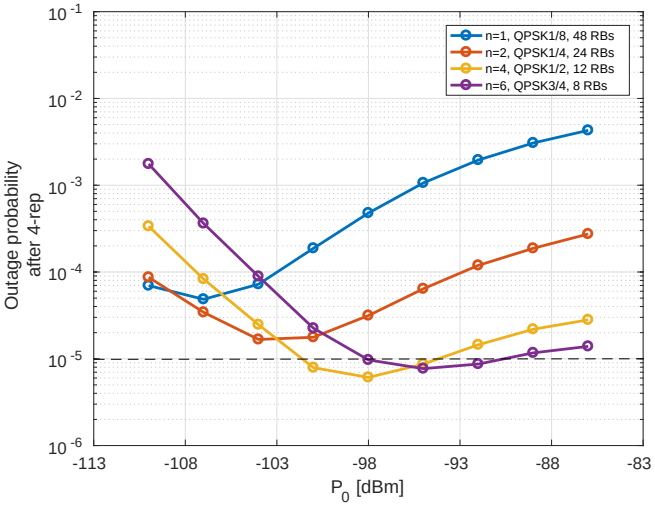


Fig. I.5: Sensitivity of outage probability in relation to P_0 and n for $K = 4$.

the achieved reliability tends to degrade. It important to note that both, HARQ and 2-rep permit two transmission attempts. 4-rep shows an SINR degradation of $\approx 1\,\mathrm{dB}$ on the low quantiles $< 10^{-4}$, but the combination of the 4 repetitions increases the resultant reliability.

Fig. I.7 shows the complementary cumulative distribution function (CCDF)

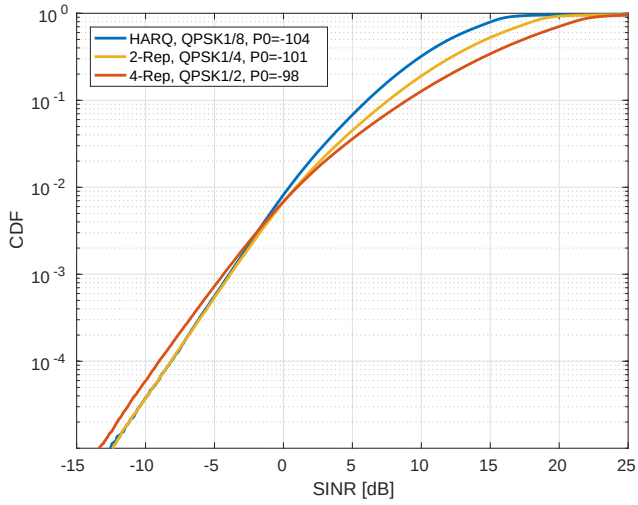


Fig. I.6: CDF of the SINR for the different schemes.

of the latency for the baseline Reactive HARQ and for the K-repetition schemes. For the considered load and packet arrival rate, it can be noted that target latency of 1 ms and reliability of $1 - 10^{-5}$ can only be reached with the HARQ scheme. Though with 4 repetitions a low outage can be achieved, queuing delays caused by the replicas in the transmission buffer prolong the tail of the latency distribution. As for the illustrated example, considering an average of $\lambda = 10$ PPS generated by the higher layers, it rises to $\lambda = K \cdot 10$ PPS with K repetitions. This can causes an increased queuing such that the latency deadline is exceeded if an early replica is not promptly received. For HARQ, it is important to mention that a retransmission has priority over the initial transmission. So it is very unlikely that a packet retransmission is queued.

The bar plot in Fig. I.8 summarizes the maximum URLLC load which can be achieved with each transmission scheme while meeting the $1 - 10^{-5}$ reliability target, disregarding queuing delays. K-repetitions with $K = 2$ supports the lowest load of 0.051 Mbps, while with $K = 4$ a load of 0.307 Mbps, 20% higher than with reactive HARQ, can be supported. It is important to highlight that, satisfying the latency constraint such as 1 ms will depend on the traffic. Transmissions from UEs with higher packet arrival rates are more susceptible to queuing delays for higher values of K.

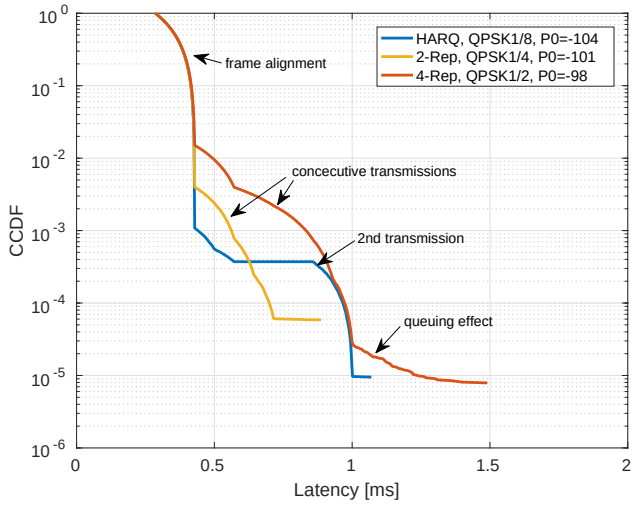


Fig. I.7: Complementary cumulative distribution function of the latency for K-repetitions with $K = 2$, $K = 4$ and the HARQ baseline ($L = 0.256$ Mbps).

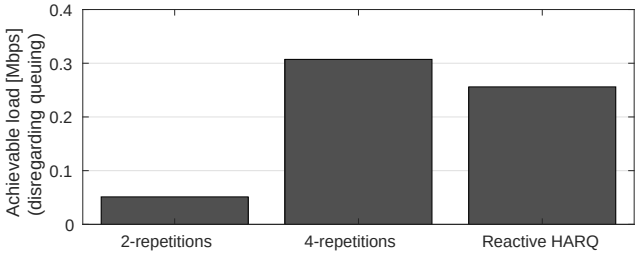


Fig. I.8: Maximum loads supported with $K = 2$, $K = 4$ and reactive HARQ, neglecting queuing delays.

6 Conclusion

In this work we have studied the performance of K-repetitions with intra-slot frequency hopping schemes for URLLC. An extensive exploration of the parameter space involved in GF transmissions with K-repetitions was conducted. That involves the number of transmission repetitions, the sub-band allocation size per transmission, the usage of sub-band hopping and uplink power control RRM mechanism.

By increasing the number of sub-bands, and the number of transmission repetitions, gains can be harvested from a reduced interference probability and with frequency diversity through sub-band hopping. However, when

a larger number of sub-bands is used, a higher receive power density or number of repetitions is also needed, which also increase the generated interference.

Our evaluations are conducted in a multi-user multi-cell network to include the effects of intra-cell and inter-cell interference within a 5G NR compliant scenario with sporadic uplink GF URLLC transmissions. Our findings show that K-repetitions can, with a similar latency budget, reach lower outage probabilities than a GF HARQ baseline, with optimized power control settings, number of repetitions and number of sub-bands. However, the queuing effect, potentially cause K-repetitions to violate the latency requirement.

Acknowledgments

This research is partially supported by the EU H2020-ICT-2016-2 project ONE5G. The views expressed in this paper are those of the authors and do not necessarily represent the project views.

References

[1] 3GPP TS 38.300 V15.2.0, "NR; NR and NG-RAN Overall Description," Jun. 2018.

[2] 3GPP TR 38.913 v14.1.0, "Study on Scenarios and Requirements for Next Generation Access Technologies," Mar. 2017.

[3] 3GPP TR 36.881 v14.0.0, "Study on latency reduction techniques for LTE," Jul. 2016.

[4] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*, 2nd ed. Wiley, 2011, pp. 108–120.

[5] 3GPP TS 38.331 V15.2.1, "NR; Radio Resource Control (RRC) protocol specification," Jun. 2018.

[6] T. Jacobsen, R. Abreu, G. Berardinelli, K. Pedersen, P. Mogensen, I. Z. Kovács, and T. K. Madsen, "System Level Analysis of Uplink Grant-Free Transmission for URLLC," in *2017 IEEE Globecom Workshops*, Dec. 2017.

[7] B. Singh, O. Tirkkonen, Z. Li, and M. A. Uusitalo, "Contention-Based Access for Ultra-Reliable Low Latency Uplink Transmissions," *IEEE Wireless Communications Letters*, Apr. 2017.

[8] C. Boyd, R. Vehkalahti, and O. Tirkkonen, "Combinatorial code designs for ultra-reliable IoT random access," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Oct 2017, pp. 1–5.

[9] G. Berardinelli, N. H. Mahmood, R. Abreu, T. Jacobsen, K. Pedersen, I. Z. Kovács, and P. Mogensen, "Reliability Analysis of Uplink Grant-Free Transmission Over Shared Resources," *IEEE Access*, vol. 6, pp. 23 602–23 611, Apr. 2018.

[10] G. Pocovi, K. I. Pedersen, and P. Mogensen, "Joint Link Adaptation and Scheduling for 5G Ultra-Reliable Low-Latency Communications," *IEEE Access*, vol. 6, pp. 28 912–28 922, May 2018.

[11] R. Abreu, T. Jacobsen, G. Berardinelli, K. Pedersen, I. Z. Kovács, and P. Mogensen, "Power control optimization for uplink grant-free URLLC," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2018.

[12] 3GPP TS 38.214 v15.3.0, "NR; Physical layer procedures for data," Sep. 2018.

[13] R1-1813118, "On Configured Grant enhancements for NR URLLC," Nov. 2018.

[14] 3GPP TS 38.213 v15.3.0, "Physical layer procedures for control (Release 15)," Sep. 2018.

[15] T. Jacobsen, R. B. Abreu, G. Berardinelli, K. I. Pedersen, I. Kovács, and P. E. Mogensen, "Joint Resource Configuration and MCS Selection Scheme for Uplink Grant-Free URLLC," in *2018 IEEE Globecom Workshops*, 2018, (Accepted/in press).

[16] 3GPP TR 38.802 v14.0.0, "Study on New Radio Access Technology," Mar. 2017.

[17] R. Srinivasan, J. Zhuang, L. Jalloul, R. Novak, and J. Park, "IEEE 802.16m Evaluation Methodology Document (EMD)," IEEE 802.16 Broadband Wireless Access Working Group, Tech. Rep. IEEE 802.16m-08/004r2, Jul. 2008.

Paper J

Joint Resource Configuration and MCS Selection Scheme for Uplink Grant-Free URLLC

Thomas Jacobsen, Renato Abreu, Gilberto Berardinelli, Klaus Pedersen, István Z. Kovács, Preben Mogensen

The paper has been published in the
IEEE 2018 GlobeCom Workshops

Abstract

Ultra-reliable and low-latency communications (URLLC) addresses the most challenging set of services for 5G New Radio. Uplink grant-free transmissions is recognized as a promising solution to meet the ambitious URLLC target (1 ms latency at a 99.999% reliability). Achieving such a high reliability comes at the expense of poor spectral efficiency, which ultimately affects the load supported by the system. This paper proposes a joint resource allocation solution including multiple modulation and coding schemes (MCSs) and power control settings for grant-free uplink transmissions on shared resources. The scheme assigns smaller bandwidths parts and higher MCS to the UEs in good average channel conditions, reducing the probability of fully overlapping transmissions. The performance analysis shows that the scheme is capable of increasing the system outage capacity by ∼90%, compared to prior art solutions using a conservative single-MCS configuration with fully overlapping transmissions.

1 Introduction

One of the major goals of 5G New Radio (NR) is the support of ultra-reliable and low-latency communication (URLLC) to enable mission-critical applications. Meeting the strict URLLC requirements with a 10^{-5} packet failure probability within 1 ms is very challenging [1]. Many technology components towards achieving this have been investigated such as short transmission time intervals (TTIs) [2], semi-persistent scheduling (SPS) [3], fast hybrid automatic repeat request (HARQ) [4], and robust error correction coding [5].

For meeting the URLLC requirements in uplink, grant-free (GF) solutions have been found to be attractive, as time-consuming steps of grant-based scheduling and its potential errors are avoided [6, 7]. For 5G NR (Release-15) it has been agreed that GF transmissions happen according to a predefined configuration which includes power control settings, modulation and coding scheme (MCS), time-frequency resource allocation, among others. At most one GF configuration per bandwidth part is active at a time [8]. This is communicated to the user equipment (UE) by radio resource control (RRC) with possible activation via downlink control channel [9]. For GF transmissions, it is further assumed; that a configuration can be shared by multiple UEs [10], the MCS and transmission bandwidth is fixed [11, 12] and open loop power control is used [13].

It is known from numerous LTE uplink studies that dynamic link adaptation is beneficial. Using a combination of open and closed loop power control, and fast adaptive modulation and coding (AMC) based on channel state information (CSI) acquired by sounding brings clear benefits for mobile broadband traffic [14, 15]. This is found to be the case for dynamically

scheduled transmissions, adjusting the MCS on a TTI basis. However, for GF URLLC cases, the situation is different. First, the URLLC traffic per UE is sporadic with small payloads appearing infrequently at the users for immediate uplink transmission. This means that there are no steady transmissions from the users that the base station nodes can utilize for CSI estimation. Secondly, as GF URLLC rely on fast uplink access without grant, there is no downlink signaling for conveying MCS and transmission bandwidth adjustments per transmission event. Finally, URLLC target transmissions where one URLLC packet is included in each transmission, as segmentation of URLLC payloads over multiple transmissions risks jeopardizing the latency targets of URLLC. Our hypothesis is therefore that a new joint MCS and transmission bandwidth selection method for GF URLLC transmission could help boosting the aggregated URLLC traffic that can be tolerated in the network.

We therefore propose a solution encompassing a hierarchical resource configuration that facilitates uplink transmissions of URLLC payloads (of fixed size) using different MCS schemes and transmission bandwidths. The idea is to allow partly overlapping transmissions with corresponding adjustments of the users MCS and power control settings. In short, we propose a solution where users are assigned to use different GF transmission settings according to a predefined resource grid, consisting of MCSs and different transmission sub-bands. The scheme allows to efficiently leverage the trade-offs between reducing the uplink collision probabilities by using lower transmission bandwidth per user versus the cost in terms of higher required signal-to-interference-plus-noise ratio (SINR) from using higher order MCS. The value of the proposed scheme is studied in a dynamic multi-user, multi-cell environment in line with the 3GPP NR assumptions.

Due to the high degree of complexity of the system model, we rely on state-of-the-art system level simulations to preserve the high degree of realism, which would otherwise be jeopardized if imposing simplifications to allow analytical performance analysis. The simulations are based on the widely accepted models agreed in 3GPP for NR studies, and were also used for the works in [16, 17]. Finally, special care is given to ensure that statistically reliable performance results are generated, such that mature conclusions can be drawn.

The rest of the paper is structured as follows: Section II outlines the system model and objectives of the study. Section III presents the proposed resource configuration. Section IV outlines the simulation assumptions, while Section V presents the performance results. Section VI concludes the study.

2 System Model and Performance Metrics

2.1 Network and transmission model

A multi-cell synchronous network is assumed, following the 3GPP guidelines as in [10, 16, 17]. A fixed number of U URLLC UEs are deployed in the cells and are assumed to be uplink synchronized and in connected state. Small packets of fixed size B bytes are generated by each UE according to independent Poisson arrival processes with an average packet arrival rate λ. Grant-free uplink transmissions occur in a framed structure based on OFDM, frequency-division duplexing (FDD) and short-TTI [2]. The GF resources are shared by the U UEs in the cell. In this sense, transmissions can occur simultaneously on the same time/frequency resources (collisions). The successful reception of the packets depends on the used MCS and the post-processing SINR achieved after the receiver combining. Multi-user detection is assumed, therefore overlapping transmissions can be received depending on the resultant SINR [18]. If the reception fails the UE issues a HARQ retransmission after processing the feedback from the base station (BS) [17]. Chase-combining is used to improve the decoding performance after each retransmission.

2.2 Power control

Power control is utilized to regulate the transmit power in order to meet a target receive power and limit the generated interference in the network. We assume open-loop power control for the transmissions as in LTE [19], such that the UE transmit power is given by

$$P[\text{dBm}] = min\{P_{max}, P_0 + 10log_{10}(M) + \alpha PL + \Delta_{MCS}\}, \qquad \text{(J.1)}$$

where P_{max} is the maximum transmit power, P_0 is the target receive power per resource block (RB), M is the number of used RBs, α is the fractional pathloss compensation factor, PL is the slow faded pathloss and Δ_{MCS} is a power offset per RB that can be applied depending on the MCS. The Δ_{MCS} setting will be further discussed in this paper. As discussed in [13], we apply full pathloss compensation ($\alpha = 1$).

2.3 Performance metric

We adopt the performance target for URLLC defined by 3GPP [1]; a success probability of $1 - 10^{-5}$ to receive a small packet (32 bytes) in the radio interface with a maximum one-way latency of 1 ms. The prior-art solutions use a conservative single-MCS, to meet the performance target [11, 13, 17]. In the baseline case, all UEs transmit using the full band in an entire TTI, using QPSK1/8 as the conservative single-MCS. Our target is to improve the achievable load per cell ($L[\text{b/s}] = \lambda \cdot U \cdot B \cdot 8$) in the network, which meets the URLLC performance target, compared to the baseline. This load is referred to as the system outage capacity.

3 Joint resource allocation and MCS selection

3.1 Resource allocation

The proposed hierarchical resource allocation scheme encompasses multiple transmission bandwidths and power control settings associated with the MCSs for grant-free transmissions. The scheme uses the resources within a bandwidth part of size BW. Each MCS is univocally associated to a specific sub-band size $\leq BW$. The supported set of MCS, $\mathbb{M}$, includes N MCS options denoted by $MCS_n(k)$, with index $n \in [1, N]$ and k is the ratio between the bandwidth BW and the sub-band size associated to the MCS. Shortened MCS notation can omit k. $\mathbb{M}$ is sorted such that $MCS_1(1)$ has the lowest modulation and coding rate, i.e. the most conservative option and use the full bandwidth BW. Higher MCS options form a set $\mathbb{M}_{1+} \subset \mathbb{M}$ for $n > 1$, which are mapped to sub-bands of size $BW \cdot k^{-1}$ with $k > 1$. Considering the strict latency requirement for URLLC traffic, the MCS options and k are chosen such that the URLLC payload can be fully transmitted in the corresponding sub-bands without segmentation. The UEs are pre-configured via RRC signaling with the resource allocation scheme, defining the sub-bands RBs, the set of corresponding MCSs and the power offsets.

Fig. J.1 shows an example configuration of the resource grid, i.e. the sub-bands and MCS options, where the set $\mathbb{M} = \{MCS_1(1), MCS_2(2), MCS_3(4)\} = \{\text{QPSK1/8}, \text{QPSK1/4}, \text{QPSK1/2}\}$ is supported. Each MCS has an associated Δ_{MCS_n}. Transmissions with MCS_1 use all the 48 RBs, while transmissions with MCS_2 or MCS_3 use sub-bands of size 24 and 12 RBs respectively. Fig. J.2 illustrates examples of GF transmissions and their overlap which can occur using the configuration illustrated in Fig. J.1. Fully overlapping transmissions can occur for transmissions using the same MCS whereas transmissions using different MCS can partially overlap.

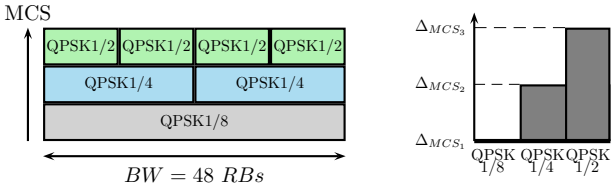


Fig. J.1: Example configuration of MCS, corresponding power spectral density offsets and frequency allocations for grant-free transmissions

The BS can estimate and decide, e.g. based on infrequent UE reports, the MCS and corresponding sub-band to be used and indicate it to the UE through downlink signaling. If multiple sub-bands are associated to the MCS,

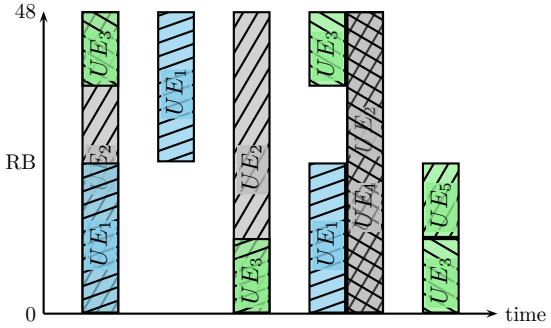


Fig. J.2: Example resource allocations for grant-free transmissions from five UEs using the example configuration from Fig. J.1

either the BS assigns one or allows the UE to randomly select. By knowing the possible combinations of transmitting UEs, $\mathbb{M}$ and the associated sub-bands, the blind decoding complexity at the receiver side is bounded. UEs in good average channel condition can be signaled to use one of the higher MCS options ($\mathbb{M}_{1+}$) instead of the conservative MCS_1. Since higher MCSs are leveraged through smaller bandwidth parts, the collision probability is reduced among the sub-bands, while UEs operating simultaneously with lower order MCSs are only partly overlapped. This can be of mutual benefit to the UEs in the network and potentially increase their achieved reliability and in the end the system outage capacity. The price to pay for UEs using $\mathbb{M}_{1+}$ is that they need a corresponding higher power spectral density in order to maintain the reliability of their transmissions, which means that the interference in the used sub-band is increased. The power spectral density offset can be configured for the power control defined in (J.1), but due to the transmit power limitation P_{max}, it can not be guaranteed that Δ_{MCS} can be fully applied. For this reason, only UEs with sufficient transmit power headroom to fully apply Δ_{MCS} should use $\mathbb{M}_{1+}$.

The choice of Δ_{MCS} should consider the higher SINR targets for $\mathbb{M}_{1+}$, the power headroom, and the generated interference. Further, the values can be predetermined from the difference in required SINR to maintain a block error rate (BLER) target, which can be found using BLER/SINR curves obtained using extensive link-level simulations. As an initial setting we propose to use

$$\Delta_{MCS_n}[\text{dB}] = 10log_{10}(k), \tag{J.2}$$

such that the target transmit power is maintained, and apply fine-tuning based on the observed outage performance.

3.2 MCS selection scheme

We propose a simple MCS and correspondent bandwidth selection scheme which is defined using a set of $N-1$ coupling gain thresholds $\mathbb{C}_T = \{C_{T_1}, ..., C_{T_{N-1}}\}$ sorted in ascending order. The MCS_n is selected according to $n = \arg\min_i \left(C_{T_i} | C \leq C_{T_i}\right)$, where C is the experienced coupling gain, which is defined as the long-term channel gain between the UE and base station antenna ports [20]. The selection is done such that the lower the coupling gain is, the more conservative is the used MCS. For $C > C_{T_{N-1}}$, MCS_N is used. Note that the idea of grouping the UEs based on coupling gain thresholds is similar to the one used in NB-IoT [21].

The choice of $\mathbb{C}_T$ depends on the scenario, $\mathbb{M}$ and the power control settings. For this reason an expression valid for all deployment scenarios is not straightforward. We propose that $\mathbb{C}_T$ is chosen based on outage statistics computed using one-way latency measurements collected at the BS, prior to applying the joint resource and MCS selection scheme, and sorted into coupling gain intervals. Good candidates for threshold values are found between intervals where the outage probability increases significantly.

3.3 Example of partly overlapping transmissions

In this section we give an example of how a resource configuration with $\mathbb{M}_{1+}$ can give SINR improvements compared to a single-MCS configuration. Consider the simple example illustrated in Fig. J.3, where two UEs transmit with fully overlapping transmissions on the left and the alternative configuration on the right. For simplicity, this example does not consider the effect of fading.

In the first case, UE_a and UE_b use MCS_1 in full band with w RBs. In the alternative configuration, UE_b is configured to use a higher MCS $MCS_2 \in \mathbb{M}_{1+}$ and hence uses a smaller bandwidth of m RBs, ensuring that when both UEs transmit simultaneously their transmissions only partly overlap. UE_b use Δ_{MCS_2} to increase its power spectral density. The post detection SINRs of the used RBs are averaged per RB for computation of the effective SINR of the data stream.

The resultant SINR of the two fully overlapping transmissions for UE_a and UE_b can be expressed by $\gamma_a = P_a/(N_0 + P_b)$ and $\gamma_b = P_b/(N_0 + P_a)$ respectively, where N_0 is the Gaussian noise spectral density, P_a and P_b are the power spectral density (PSD) from UE_a and UE_b respectively, giving $\gamma_a = \gamma_b$ for $P_a = P_b$. With the partial overlapping configuration, the transmission from UE_b uses a higher spectral density $\hat{P}_b = P_b \cdot 10^{\Delta_{MCS_2}/10}$, resulting in an SINR expressed by $\hat{\gamma_b} = \hat{P}_b/(N_0 + P_a)$. The SINR for UE_a maintaining MCS_1 and P_a can be expressed by

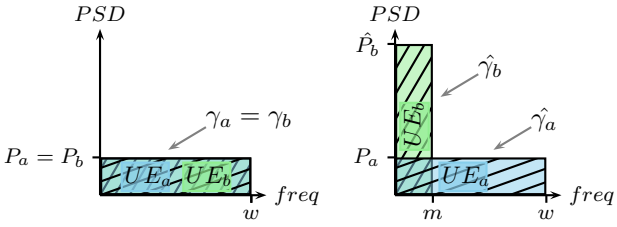


Fig. J.3: Two fully overlapping transmissions (left) versus two partial overlapping transmissions (right)

$$\hat{\gamma}_a = \frac{w-m}{w} \cdot \frac{P_a}{N_0} + \frac{m}{w} \cdot \frac{P_a}{N_0 + \hat{P}_b}. \tag{J.3}$$

An evaluation of the SINR gain $\hat{\gamma}_a/\gamma_a$ using (J.3) is shown in Fig. J.4 considering different PSDs $\hat{P}_b/P_a$ and sub-band size ratios m/w. It is assumed $w = 48$ RBs, $N_0 = -126\,\text{dBm/RB}$ and $P_a = -131\,\text{dBm/RB}$. At a given power density ratio, the respective SINR gain for UE_a decreases with the increase of the overlapping ratio. The dashed line follows the performance when Δ_{MCS_2} is selected according to (J.2). An SINR gain for UE_a is achieved in the $\hat{\gamma}_a/\gamma_a > 0\,\text{dB}$ region. The performance with the initial Δ_{MCS_2} for all m/w is found to be in this region. UE_b mutually experiences an SINR gain, i.e. $\hat{\gamma}_b/\gamma_b > 0$ for $\hat{P}_b > P_b$, nevertheless it has a capacity penalty with the reduced bandwidth. The vertical dotted line shows the example of $m/w = k^{-1} = 12/48 = 0.25$ meaning $k = 4$ gives an initial $\Delta_{MCS_2} = 10log_{10}(4) \approx 6\,\text{dB}$ marked in the point **X**. Following the dotted line for $\Delta_{MCS_2} > 6\,\text{dB}$, the SINR of UE_b increases together with the ratio $\hat{P}_b/P_a$, however the SINR gain of UE_a reduces. It should be observed that, for low overlapping m/w ratios, the increase of $\hat{P}_b$ in relation to P_a has lower impact on the SINR gain of UE_a. However, for ratios such as $m/w = 0.5$ or higher, there is not much room to adjust Δ_{MCS_n} without causing a loss in SINR for UE_a. Notice that this example does not include the effect of intra sub-band interference, as only 1 UE is considered per MCS, which would affect the observed gains. For this reason, after applying the initial Δ_{MCS_n}, fine-tuning it can be beneficial, as mentioned in Section 3.1.

4 Simulation Methodology

An advanced system-level simulator is used for assessing the performance of the proposed resource allocation scheme. The simulator models the 5G NR design, adopting the commonly agreed mathematical models in 3GPP for radio propagation, traffic models, key performance indicators, etc [10].

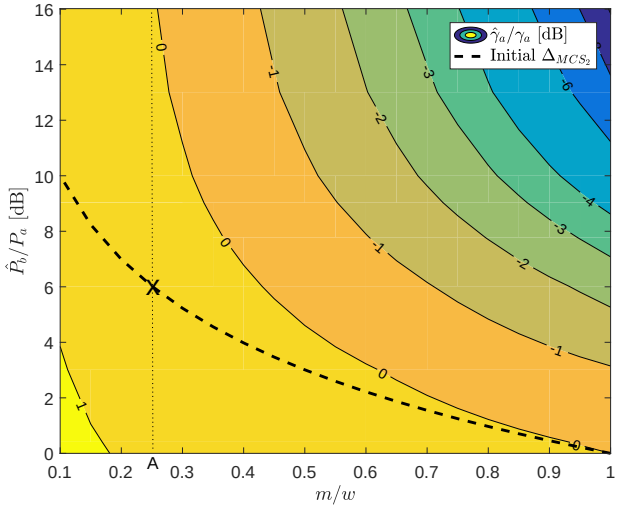


Fig. J.4: SINR gain $\hat{\gamma}_a/\gamma_a$ in dB of UE_a using the MCS_1 as a function of m/w and $\hat{P}_b/P_a$ ratios

The same simulator was also used in the earlier URLLC studies published in [4, 13, 17]. The network layout is a single layer urban macro network consisting of 7 sites, each having 3 sectors composing a regular hexagonal grid topology with 500 meters of inter-site distance (ISD), using wrap-around [22]. UEs are random distributed (all outdoor), following a spatial uniform distribution. The traffic per UE follows a Poisson arrival process in line with system model in Section 2. The offered URLLC traffic load is adjusted by varying the number of users U per macro-cell area, while keeping $\lambda = 10$ packets per second (PPS) and $B = 32$ bytes fixed. The time-granularity of the simulator is one OFDM symbol, and the frequency resolution is one subcarrier. The main simulation assumptions are described in Table J.1.

For each GF transmission from a UE to a BS, the received post detection SINR is calculated (accounting for both inter- and intra-cell interference) assuming a two-antenna receiver and Minimum Mean Square Error Interference Rejection Combining (MMSE-IRC) which is the baseline detector for NR evaluation [10, 23]. Ideal channel estimation of both the desired and the interfering signals is assumed. Based on [24, 25], the SINR values are mapped to the mutual information domain, taking the applied modulation scheme into account. Given the mean mutual information per coded bit (MMIB) and the used coding rate of the transmission, the error probability of the transmission is determined from look-up tables that are obtained from extensive link level simulations.

The simulations of the GF URLLC transmissions are in line with the pre-

Table J.1: Simulation assumptions

Parameters	Assumption
Layout	Hexagonal grid, 7 sites, 3 sectors/site, 500 m ISD
UE distribution	Uniformly distributed outdoor, 3 km/h speed, no handover
Channel model	3D Urban Macro (UMa)
Carrier and bandwidth	4 GHz, FDD, 10 MHz (48 RBs) UL
PHY numerology	15 kHz sub-carrier spacing, 2 symbols/TTI, 12 sub-carriers/RB
Timing	1 TTI (0.143 ms) to transmit and 1 TTI to process by UE and BS [17]
HARQ configuration	4 TTIs HARQ RTT, 4 SAW channels, up to 8 HARQ transmissions using chase combining
Max. UE TX power	23 dBm
BS receiver noise figure	5 dB
Thermal noise density	-174 dBm/Hz
BS receiver type	MMSE-IRC, 1 TX x 2 RX UL
Traffic model	FTP Model 3 with 32 B packet and Poisson arrival rate of 10 PPS per UE
Power control	Open loop power control (α=1, P_0=$-$104 dBm) and variable Δ_{MCS}
MCS selection	Coupling gain based with threshold $\mathbb{C}_T$

sented system model; including open loop power control, HARQ with chase combining, queuing, etc. Results from the simulator have been benchmarked against calibration results shared in 3GPP for the NR macro simulation scenario, confirming a good match. To ensure statistical reliable results, information is collected from at least $5 \cdot 10^6$ completed URLLC payload transmissions. With this amount of independent samples the outage probability can be said to be within a 27 % error margin around the 10^{-5} quantile with 95 % confidence using the interval estimation of a binomial proportion [26].

5 Results

This section evaluates a two MCS resource allocation configuration $\mathbb{M} = \{MCS_1(1), MCS_2(4)\} = \{\text{QPSK1/8}, \text{QPSK1/2}\}$. QPSK1/8 is used as the conservative MCS option (as in [13, 17]) and QPSK1/2 as the higher MCS option. We set the initial power spectral density offset $\Delta_{MCS_2} = 6$ dB by

following (J.2).

Fig. J.5 shows the outage probability at 1 ms per coupling-gain interval for the baseline and for the proposed scheme. The offered load is 486.4 kbps per cell. To get high accuracy per coupling gain interval, $50 \cdot 10^6$ transmission latency samples have been collected in the network for this result. The percentage of samples per interval is $\sim 6\,\%$. Each marker is placed on the maximum coupling gain of the interval. This means, for example, that the marker on coupling gain $-110\,\text{dB}$ represents the outage in the interval $(-113\,\text{dB}, -110\,\text{dB}]$. The MCS selection threshold set $\mathbb{C}_T$ is defined based on outage probability statistics of one-way latency measurements calculated per coupling gain interval. The threshold $\mathbb{C}_T = C_{T_1} = -110\,\text{dB}$ is chosen by observing that below this value the outage probability increases significantly for the baseline configuration, as indicated in the figure.

With the chosen C_{T_1}, fine-tuning of Δ_{MCS_2} is performed. Fig. J.5, also shows the performance of the proposed scheme with $\Delta_{MCS_2} = \{6\,\text{dB}, 10\,\text{dB}, 20\,\text{dB}\}$. Increasing Δ_{MCS_2} from the initial setting improves the reliability for the UEs using MCS_2, while also degrading the reliability for the UEs using MCS_1. For $\Delta_{MCS_2} = \{6\,\text{dB}, 10\,\text{dB}\}$ the reliability in the intervals using MCS_2 are comparable, which indicates that the UEs in these intervals are able to apply the full PSD offset through power control. For a very high PSD offset ($\Delta_{MCS_2} = 20\,\text{dB}$) the variation on reliability indicates that not all coupling gain intervals are capable of applying the full offset and reaching the reliability requirement.

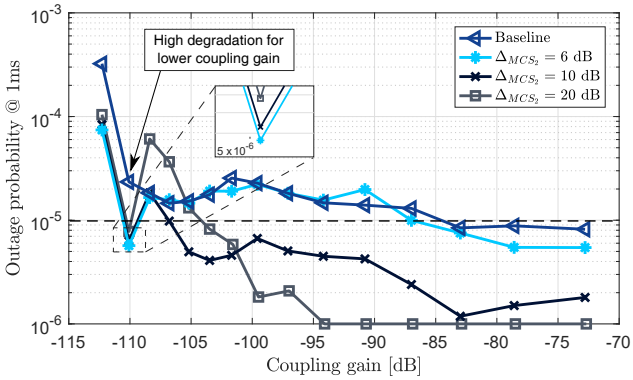


Fig. J.5: Outage probability in coupling gain intervals with $\approx 6\,\%$ of all transmission latency samples per interval. $L = 486.4$ kbps/cell.

The reliability statistics per coupling gain interval in Fig. J.5 does not show the systems overall reliability when combining all latency samples. For that, the latency CCDF for the system is shown in Fig. J.6, for both the baseline and the considered scheme with $\Delta_{MCS_2} = \{6\,\text{dB}, 10\,\text{dB}, 20\,\text{dB}\}$. The stair-

case behavior comes from HARQ retransmissions [17]. From the figure, it can be seen that the option with $\Delta_{MCS_2} = 10\,\text{dB}$ is capable of reaching the target outage probability of 10^{-5} within 1 ms. The baseline is only capable of reaching an outage probability of $3.7 \cdot 10^{-5}$ at the 1 ms latency deadline. Considering the fine-tuning of Δ_{MCS_2} it can be seen that $\Delta_{MCS_2} = 10\,\text{dB}$ is the best option, indicating that further increasing the offset does not improve the performance.

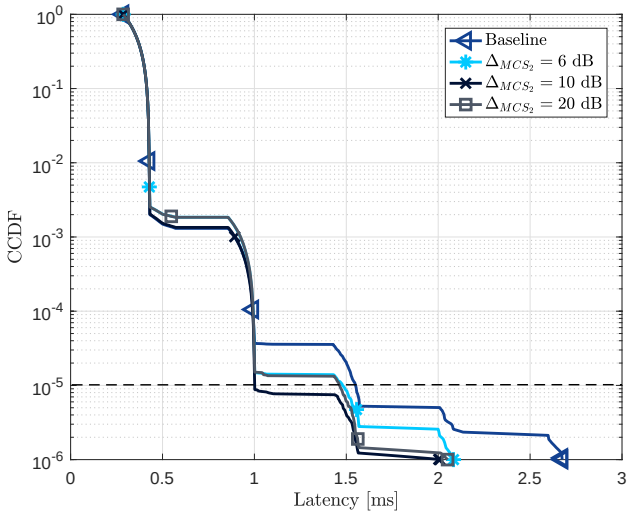


Fig. J.6: Complementary Cumulative Distribution Function (CCDF) of the latency with different MCSs configurations for $L = 486.4\,\text{kbps/cell}$

Fig. J.7 shows a sensitivity study of C_{T_1} impact on the outage probability. The threshold that gives the lowest outage for both $\Delta_{MCS_2} = \{6\,\text{dB}, 10\,\text{dB}\}$ is $C_{T_1} = -110\,\text{dB}$, confirming the earlier choice. This coupling gain threshold value corresponds to 12 % of all transmissions using the MCS_1 and 88 % using MCS_2.

Fig. J.8 summarizes the achieved overall outage probability at 1 ms comparing the baseline with the proposed joint resource allocation and MCS selection scheme with $\Delta_{MCS_2} = \{6\,\text{dB}, 10\,\text{dB}\}$. The maximum supported offered load for the baseline is 256.0 kbps/cell, which aligns with previous work done in [13]. Using the proposed scheme the supported load increases to 358.4 kbps/cell using $\Delta_{MCS_2} = 6\,\text{dB}$ and 486.4 kbps/cell using $\Delta_{MCS_2} = 10\,\text{dB}$. The proposed scheme is capable of increasing the system outage capacity up to 40 % using the initial Δ_{MCS} and a further 35 % by fine-tuning it.

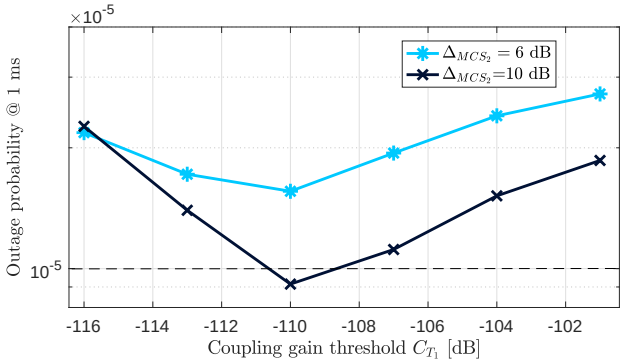


Fig. J.7: Outage probability at 1 ms versus coupling-gain threshold C_{T_1}. UEs with $C > C_{T_1}$ apply MCS_2 with a power offset Δ_{MCS_2}, otherwise MCS_1 is applied. $L = 486.4$ kbps/cell.

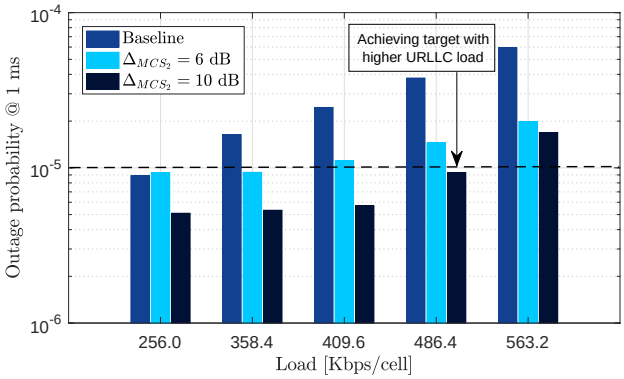


Fig. J.8: Outage probability at 1 ms as a function of offered load

6 Conclusion

In this paper we have proposed a joint resource allocation and MCS selection scheme for uplink grant-free URLLC. The scheme allows to pre-define a set of MCSs, transmission bandwidths and power offsets. The MCS selection is based on the coupling gain of the UEs. UEs in good average channel condition have reduced collision probability at the expense of eventual higher interference power in the sub-bands, while UEs in poor average channel conditional have lower degradation with partial overlapping. Compared with a conservative single-MCS configuration, the proposed scheme shows that the system outage capacity can be increased by 90 %, up to 486.4 kbps per cell, while still fulfilling the URLLC requirements.

Future work will focus on the potential of multi-site reception and re-

ceiver diversity together with the proposed joint resource allocation and MCS selection scheme to further enhance the system capacity for uplink grant-free URLLC transmissions.

Acknowledgment

This research is partially supported by the EU H2020-ICT-2016-2 project ONE5G. The views expressed in this paper are those of the authors and do not necessarily represent the project views.

References

[1] 3GPP TR 38.913 v14.1.0, "Study on Scenarios and Requirements for Next Generation Access Technologies," Mar. 2017.

[2] K. I. Pedersen, G. Berardinelli, F. Frederiksen, P. Mogensen, and A. Szufarska, "A Flexible 5G Frame Structure Design for Frequency-Division Duplex Cases," *IEEE Communications Magazine*, vol. 54, no. 3, pp. 53–59, Mar. 2016.

[3] 3GPP TR 36.881 v14.0.0, "Study on latency reduction techniques for LTE," Jul. 2016.

[4] G. Pocovi, H. Shariatmadari, G. Berardinelli, K. Pedersen, J. Steiner, and Z. Li, "Achieving Ultra-Reliable Low-Latency Communications: Challenges and Envisioned System Enhancements," *IEEE Network*, vol. 32, no. 2, pp. 8–15, Mar. 2018.

[5] N. A. Johansson, Y. P. E. Wang, E. Eriksson, and M. Hessler, "Radio Access for Ultra-Reliable and Low-Latency 5G Communications," in *IEEE ICC Workshop (ICCW)*, Jun. 2015.

[6] H. Shariatmadari, Z. Li, S. Iraji, M. A. Uusitalo, and R. Jäntti, "Control Channel Enhancements for Ultra-Reliable Low-Latency Communications," in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2017.

[7] B. Singh, O. Tirkkonen, Z. Li, and M. A. Uusitalo, "Contention-Based Access for Ultra-Reliable Low Latency Uplink Transmissions," *IEEE Wireless Communications Letters*, Apr. 2017.

[8] 3GPP TS 38.300 V15.2.0, "NR; NR and NG-RAN Overall Description," Jun. 2018.

[9] 3GPP TS 38.331 V15.2.1, "NR; Radio Resource Control (RRC) protocol specification," Jun. 2018.

[10] 3GPP TR 38.802 v14.0.0, "Study on New Radio Access Technology," Mar. 2017.

[11] C. Wang, Y. Chen, Y. Wu, and L. Zhang, "Performance Evaluation of Grant-Free Transmission for Uplink URLLC Services," in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, Jun. 2017.

[12] G. Berardinelli, N. H. Mahmood, R. Abreu, T. Jacobsen, K. Pedersen, I. Z. Kovács, and P. Mogensen, "Reliability Analysis of Uplink Grant-Free Transmission Over Shared Resources," *IEEE Access*, vol. 6, pp. 23 602–23 611, Apr. 2018.

[13] R. Abreu, T. Jacobsen, G. Berardinelli, K. Pedersen, I. Z. Kovács, and P. Mogensen, "Power control optimization for uplink grant-free URLLC," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2018.

[14] C. Rosa, D. L. Villa, C. U. Castellanos, F. D. Calabrese, P. H. Michaelsen, K. I. Pedersen, and P. Skov, "Performance of Fast AMC in E-UTRAN Uplink," in *IEEE ICC*, May 2008, pp. 4973–4977.

[15] H. Holma and A. Toskala, *WCDMA for UMTS - HSPA Evolution and LTE*, 5th ed. Wiley, 2010.

[16] G. Pocovi, B. Soret, K. I. Pedersen, and P. Mogensen, "MAC Layer Enhancements for Ultra-Reliable Low-Latency Communications in Cellular Networks," in *2017 IEEE International Conference on Communications Workshops*, May 2017.

[17] T. Jacobsen, R. Abreu, G. Berardinelli, K. Pedersen, P. Mogensen, I. Z. Kovács, and T. K. Madsen, "System Level Analysis of Uplink Grant-Free Transmission for URLLC," in *2017 IEEE Globecom Workshops*, Dec. 2017.

[18] S. Saur and M. Centenaro, "Radio Access Protocols with Multi-User Detection for URLLC in 5G," in *European Wireless 2017; 23th European Wireless Conference*, May 2017.

[19] C. U. Castellanos, D. L. Villa, C. Rosa, K. I. Pedersen, F. D. Calabrese, P. H. Michaelsen, and J. Michel, "Performance of Uplink Fractional Power Control in UTRAN LTE," in *VTC Spring 2008 - IEEE Vehicular Technology Conference*, May 2008, pp. 2517–2521.

[20] 3GPP TR 36.824 V11.0.0, "E-UTRA; LTE coverage enhancements," Jun. 2012.

[21] R. Ratasuk, N. Mangalvedhe, J. Kaikkonen, and M. Robert, "Data Channel Design and Performance for LTE Narrowband IoT," in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, Sep. 2016.

[22] T. Hytönen, "Optimal Wrap-Around Network Simulation," Helsinki University of Technology, Tech. Rep. A432, Oct. 2001.

[23] F. M. L. Tavares, G. Berardinelli, N. H. Mahmood, T. B. Sørensen, and P. Mogensen, "On the Potential of Interference Rejection Combining in B4G Networks," in *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*, Sep. 2013.

[24] 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 *2005 IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 4, Sep. 2005, pp. 2306–2311 Vol. 4.

[25] R. Srinivasan, J. Zhuang, L. Jalloul, R. Novak, and J. Park, "IEEE 802.16m Evaluation Methodology Document (EMD)," IEEE 802.16 Broadband Wireless Access Working Group, Tech. Rep. IEEE 802.16m-08/004r2, Jul. 2008.

[26] L. D. Brown, T. T. Cai, and A. DasGupta, "Confidence Intervals for a binomial proportion and asymptotic expansions," *The Annals of Statistics*, vol. 30, no. 1, pp. 160–201, Feb. 2002.

Paper K

On the Achievable Rates over Collision-Prone Radio Resources with Linear Receivers

Gilberto Berardinelli, Renato Abreu, Thomas Jacobsen, Nurul Huda Mahmood, Klaus Pedersen, István Z. Kovács, Preben Mogensen

The paper has been published in the
2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)

Abstract

In this paper, we discuss the achievable transmission rates over collision-prone radio resources shared by a number of devices, representative of novel Internet-of-Things (IoT) scenarios. We consider Maximum Ratio Combining (MRC) and Minimum Mean Square Error (MMSE) receivers at the base station, and derive the relationship between target failure probability and saturation rate, which represents the maximum achievable rate over shared resources in the interference limited regime. MRC receiver is shown to be sensitive to the presence of statistically relevant interferers operating over the same resources, rapidly leading to rate saturation. The MMSE receiver adds a tier of protection to collisions thanks to its interference suppression capabilities, suffering for a rate penalty only in case of a high number of users. A realistic system analysis in an indoor hotspot scenario validates the analytical trends and suggests insights on practical link adaptation strategies.

1 Introduction

Rate adaptation refers to the techniques for adapting the amount of information to be transmitted over a communication channel according to its quality. In current cellular networks, it is typically performed by means of adaptive modulation and coding (AMC), where the modulation and coding scheme (MCS) for encoding the data is selected according to the radio channel conditions [1]. In case of closed loop rate adaptation, the MCS selection is based on a feedback from the receive node on the measured channel quality, considering a target failure probability. In non-scheduled systems such as IEEE 802.11, reactive mechanisms for rate adaptation are used, where the MCS is adjusted on a slow basis depending on the amount of previous successful transmissions [2].

There is a recent regrown attention for collision-prone transmission over shared resources given the emergency of a plethora of novel Internet-of-Things (IoT) use cases, with a major focus on massive access [3]. Recently, transmission over shared resources has also been studied as a valid solution for uplink latency-limited services as targeted by upcoming 5th Generation (5G) radio technology, since it allows avoiding the time consuming steps of scheduling request and grant [4]. Recent studies on uplink grant-free transmission over shared resources typically assume a fixed transmission rate, or a rate selected according to a packet error probability which disregards the eventual occurrence of collisions [5]. However, the presence of unpredictable simultaneous transmissions may lead to an increase of the failure probability with respect to the target one. This may be particularly critical for delay-constrained services demanding high reliability, e.g. Ultra-Reliable Low-Latency Communication (URLLC) services. On the other hand, operat-

ing with over-conservative rate leads to poor network spectral efficiency and ultimately reduces the number of supported links. To the best of our knowledge, the selection of a transmission rate targeting a certain failure probability when operating in collision-prone shared resources is still an open problem.

In this paper, we discuss the achievable rates in collision-prone resources considering Rayleigh fading channels with linear receivers. In particular, both Maximum Ratio Combining (MRC) [6] and Minimum Mean Square Error (MMSE) receivers [7] are considered. We derive the expression of the failure probability as a function of the maximum sustainable rate in the interference limited regime, and evaluate the analytical performance with different number of users in the shared resource pool as well as with different number of receive antennas. We also complement our analytical study with a realistic system evaluation in an indoor hotspot scenario. Our aim is to obtain insights on how link/rate adaptation should be performed for transmissions over shared resources.

The paper is structured as follows. In Section II, the achievable rates over shared resources are derived analytically and analyzed for the two receiver types at different network loads. Section III presents a rate analysis in a realistic indoor office scenario and a comparison with the analytical findings. Insights on how to design link adaptation are also provided. Finally, Section IV concludes the paper and states the future work.

2 Achievable rates over shared resources

We consider N perfectly synchronized users sharing the same radio resources for their transmissions. A transmission happens in a single Transmission Time Interval (TTI), according to a packet arrival rate λ per TTI. Since the users are synchronized, their transmissions are fully overlapping in a TTI time in case of a simultaneous packet arrival, i.e. no partial collisions happen. Note that the assumption of perfect synchronization is consistent with the recently defined 5G scenarios [4]. We assume the users operating over a flat Rayleigh fading channel, and to be power controlled such that their transmissions are received at the same average power, though their instantaneous receive power may change at each transmission due to Rayleigh fluctuations. The base station is equipped with M receive antennas. The transmission rate should be selected such that the user payload can be delivered with a certain failure probability $\mathrm{P_f}$ at the first transmission. In a collision-free scenario, i.e. dedicated radio resources, the maximum rate r which still guarantees $\mathrm{P_f}$ can be derived numerically from [6]:

$$\mathrm{P_f} = 1 - e^{-\frac{2^{\mathrm{r}}-1}{\overline{\gamma}}} \sum_{k=0}^{M-1} \left(\frac{2^{\mathrm{r}}-1}{\overline{\gamma}}\right)^k \frac{1}{k!}. \tag{K.1}$$

where $\overline{\gamma}$ denotes the average SNR per antenna.

Collisions may lead to an increase of the failure probability, and a wiser selection of the transmit rate should take into account their eventual occurrence. The failure probability of a UE of interest in a collision-prone scenario can be expressed as:

$$\mathrm{P_f} = \sum_{z=0}^{N-1} \mathrm{P_{c,z}} \mathrm{P_{f,z}}, \tag{K.2}$$

where $\mathrm{P_{c,z}}$ is the probability of having z users transmitting simultaneously with the UE of interest, and $\mathrm{P_{f,z}}$ is the probability of failure in case of such z active interferers.

The collision probability $\mathrm{P_{c,z}}$ can be calculated as

$$\mathrm{P_{c,z}} = \binom{N-1}{z} \mathrm{P_a^z} \left(1 - \mathrm{P_a}\right)^{N-1-z}, \tag{K.3}$$

where $\mathrm{P_a}$ is the packet arrival probability. In case of Poisson arrivals at a rate λ, $\mathrm{P_a} = 1 - e^{-\lambda}$ [8]. Figure 1 displays the probability of having z colliding users assuming a set of $N = \{10, 50\}$ users sharing the same resources, and different arrival rates. The number of users has a major impact only at frequent arrival rates ($\lambda = 10^{-2}$). Despite of the significant number of users sharing the same resources, the probability of z simultaneous transmissions is only statistically relevant for very small values of z. Considering collision events happening at a significantly lower probability than a target failure probability $\mathrm{P_f}$, would not lead to any significant impact in the calculation in (K.2). We introduce here the empirical concept of relevant number of interferers, which is the number $\overline{z}$ such that $\mathrm{P_{f,\overline{z}}} \geq \alpha \cdot \mathrm{P_f}$, where α is set such that (K.2) can be approximated as $\mathrm{P_f} \approx \sum_{z=0}^{\overline{z}} \mathrm{P_{c,z}} \mathrm{P_{f,z}}$. Such concept will be used in the numerical analysis.

The failure probability $\mathrm{P_{f,z}}$ in case of z active interferers depends on the receiver type, and is calculated in the following.

MRC receiver

Though unable to suppress the interference, the MRC receiver strengthens the power of the user of interest and therefore adds a tier of protection with respect to eventual collisions. With the assumption of the same average SNR $\overline{\gamma}$ for all the users, the failure probability in case of z interferers can be expressed as follows [9]:

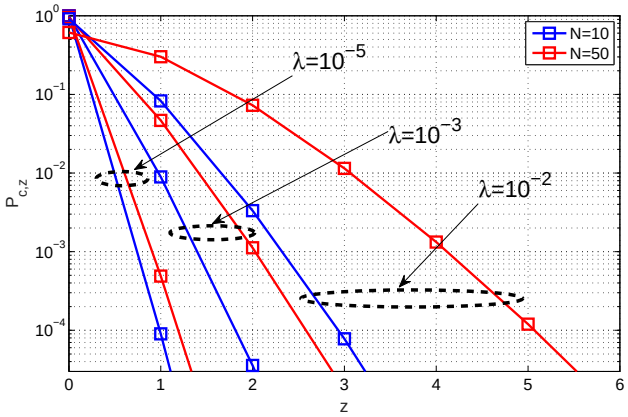


Fig. K.1: Probability of z colliding users in the same TTI.

$$\mathrm{P}_{\mathrm{f},z} = 1 - \left(\frac{1}{2^{\mathrm{r}}}\right)^{z} e^{-\frac{2^{\mathrm{r}}-1}{\overline{\gamma}}} \sum_{k=0}^{M-1} \frac{\left(2^{\mathrm{r}}-1\right)^{k}}{\overline{\gamma}^{k} k!} \cdot \sum_{i=0}^{k} \binom{k}{i} \frac{\Gamma\left(z+i\right)\overline{\gamma}^{i}}{\Gamma\left(z\right)\left(2^{\mathrm{r}}\right)^{i}}, \quad \text{(K.4)}$$

where $\Gamma\left(\cdot\right)$ denotes the Gamma function [8]. Note that $\mathrm{P}_{\mathrm{f},0}$ corresponds to the expression in (K.1). The maximum rate r that still guarantees a certain P_{f} can be then calculated numerically by applying (K.4) in (K.2).

In case no interferers are present ($N = 1$), $\mathrm{P}_{\mathrm{f}} = \mathrm{P}_{\mathrm{f},0}$ and $\mathrm{r} \to \infty$ if $\overline{\gamma} \to \infty$. In the presence of $N \geq 1$ users operating over the same resources, the achievable rates guaranteeing a certain P_{f} saturate at high SNR, i.e. $\mathrm{r} \to \mathrm{r}_{\mathrm{s}}$ if $\overline{\gamma} \to \infty$. The saturation rate r_{s} in collision-prone resources can be calculated numerically from the following expression:

$$\mathrm{P}_{\mathrm{f}} = \sum_{z=1}^{N-1} \mathrm{P}_{\mathrm{c},z} \left(1 - \left(\frac{1}{2^{\mathrm{r}_{\mathrm{s}}}}\right)^{z} \cdot \left(1 + \sum_{k=1}^{M-1} \left(\frac{2^{\mathrm{r}_{\mathrm{s}}}-1}{2^{\mathrm{r}_{\mathrm{s}}}}\right)^{k} \frac{1}{k!} \frac{\Gamma\left(z+k\right)}{\Gamma\left(z\right)}\right)\right). \quad \text{(K.5)}$$

The proof of (K.5) is given in Appendix A.

MMSE receiver

The MMSE receiver aims at suppressing a number of interferers by exploiting the knowledge of their instantaneous channel response. This subsumes a system design where the channel responses of multiple simultaneously active users can be resolved, e.g. orthogonal reference sequences are used. The authors in [10] derive a reliability function for the MMSE receiver in Rayleigh fading channels by considering an equivalent interference model for the additive Gaussian noise. According to the analytical findings in [10], $\mathrm{P}_{\mathrm{f},z}$ can be expressed as follows:

$$\mathrm{P}_{\mathrm{f},z} = 1 - e^{-\frac{\gamma}{\overline{\gamma}}} \sum_{n=1}^{M} \frac{\mathrm{A}_n(\gamma)}{(n-1)!} \left(\frac{\gamma}{\overline{\gamma}}\right)^{n-1}, \tag{K.6}$$

where γ denotes the required SNR for guaranteeing a failure probability $\mathrm{P}_{\mathrm{f},z}$, and $\mathrm{A}_n(\gamma)$ is defined as follows:

$$\mathrm{A}_n(\gamma) = \begin{cases} 1 & \text{if } z \leq M-n \\ \dfrac{1 + \sum_{i=1}^{M-n} \mathrm{C}_i \gamma^i}{(1+\gamma)^z} & \text{if } z > M-n \end{cases} \tag{K.7}$$

where C_i is the coefficient of γ^i in the expansion of $(1+\gamma)^z$. Similarly to the MRC case, the maximum rate $\mathrm{r} = \log_2(1+\gamma)$ guaranteeing P_{f} can be calculated numerically from (K.6) and (K.2).

The saturation rate r_{s} guaranteeing P_{f} at a high SNR regime can be calculated from:

$$\mathrm{P}_{\mathrm{f}} = \sum_{z=M}^{N-1} \mathrm{P}_{\mathrm{c},z} \left(1 - \frac{1 + \sum_{i=1}^{M-1} \mathrm{C}_i \gamma_{\mathrm{s}}^i}{(1+\gamma_{\mathrm{s}})^z}\right), \tag{K.8}$$

with $\gamma_{\mathrm{s}} = 2^{\mathrm{r}_{\mathrm{s}}} - 1$. The proof of (K.8) is given in Appendix B. Observe that the achievable rates saturate in case $\overline{z} \geq M$. When $\overline{z} < M$, $\mathrm{r} \to \infty$ if $\overline{\gamma} \to \infty$, i.e. no rate saturation appears. This reflects the well-known asymptotic performance of the MMSE receiver, which is able to suppress up to $M-1$ interferers at high SNR [7].

2.1 Numerical Analysis

We evaluate here the achievable rates in case of $N = \{10, 50\}$ users operating over shared resources and targeting a failure probability $\mathrm{P}_{\mathrm{f}} = 10^{-3}$ at each transmission. Such P_{f} is selected as it can lead to a final outage probability $< 10^{-5}$ upon a single retransmission, as targeted for instance by 5G NR for URLLC use cases [4]. Poisson packet arrivals are considered. The analytical expressions presented above are used, and simulation results are also included for the sake of validation. We simulate the packet arrivals and

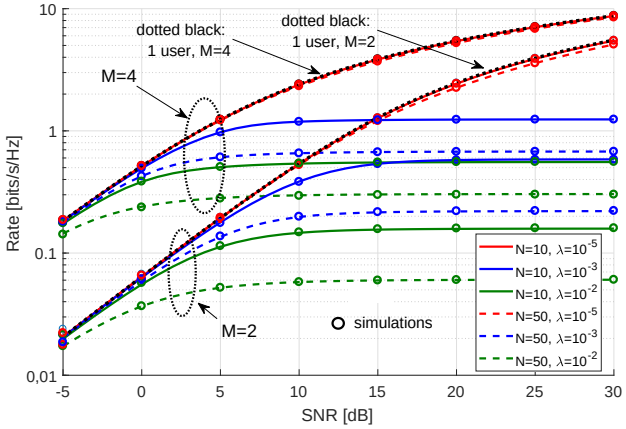


Fig. K.2: Achievable rates with an MRC receiver at the base station.

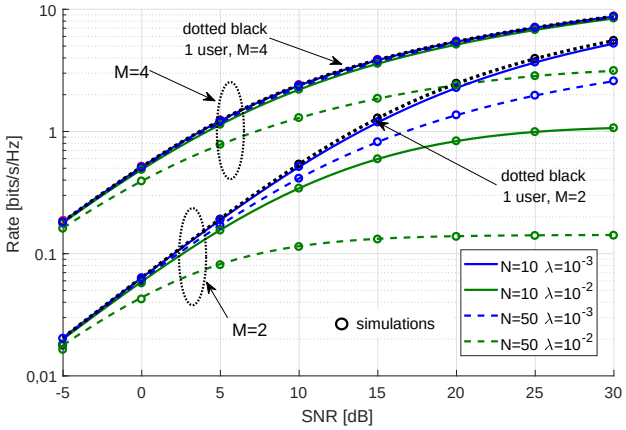


Fig. K.3: Achievable rates with an MMSE receiver at the base station.

generate random Rayleigh fading coefficients at each receive antenna. Shannon rates guaranteeing P_f are then calculated according to known signal-to-interference plus noise ratio (SINR) expressions for MRC [11] and MMSE receivers [12].

The number of relevant interferers can be calculated according to the empirical definition given in this section and from the collision probability results shown in Figure 1, considering our target $P_f = 10^{-3}$ and by assuming $\alpha = 5\%$, i.e. collisions happening at a lower rate than 5% of the target failure probability are not to be considered statistically relevant. For $\lambda = 10^{-5}$, only 1 interferer has a relevant impact on the performance. For $\lambda = 10^{-3}$, the

number of relevant interferers is 1 and 2 for $N = 10$ and $N = 50$, respectively. Finally, for $\lambda = 10^{-2}$, the number of relevant interferers is 3 and 5 for $N = 10$ and $N = 50$, respectively.

Figure 2 displays the transmission rate leading to a $P_f = 10^{-3}$ as a function of the average SNR, considering an MRC receiver at the base station. Cases of $M = 2$ and $M = 4$ receive antennas are considered, and the collision-free cases (dotted black curves) are also included as a benchmark. The transmission rate is only marginally affected by the presence of interferers at very low SNR, where performance is dominated by noise. No significant degradation is visible at low arrival rates ($\lambda = 10^{-5}$) due to the infrequent presence of interferers. The rate saturation effect clearly appears for the $\lambda = \{10^{-3}, 10^{-2}\}$ cases, and is more limiting for a high number of users. The usage of a higher order receive diversity obviously leads to higher achievable rates.

Results obtained with an MMSE receiver are shown in Figure 3. Note that the $\lambda = 10^{-5}$ cases are not displayed here, since MRC was already shown to perform closely to the single user case for such sporadic arrivals. The performance is not significantly affected by the presence of the interferers in case their relevant number is lower than M. As discussed in section 2, this is due to the interference suppression capability of the MMSE receiver at high SNR, while at low SNR the performance is limited by noise. For the $M = 4$ case, a rate degradation is indeed visible only at a frequent arrival rate ($\lambda = 10^{-2}$) and $N = 50$, where $\bar{z} = 5 > M$. The lower interference suppression capability of the $M = 2$ configuration translates to higher sensitivity to the number of users sharing the same radio resources, and the performance is still visibly affected by the saturation phenomenon at least for $\lambda = 10^{-2}$.

Simulations show a perfect match with the analytical results, thus proving their validity. The presented analysis was meant at identifying the theoretical trends for the achievable rates over shared resources with transmission over fading channels and a tight failure probability constraint. A system level view on the achievable rates will be presented in the next section.

3 System analysis

In this section, we complement the analytical findings with the results of a realistic system analysis within the scope of 5G NR. We consider a single site 120×50 m indoor hotspot NLOS scenario defined in [13], with isotropic base station antennas located at a 3 m height. The coherence bandwidth of the channel is around $\sim$4 MHz. The assumptions on traffic models and physical layer numerology are consistent with the latest 3GPP agreements for URLLC services [4], and also used in previous studies (e.g., [5]). In particular, we consider a 32 bytes payload to be transmitted in a short TTI composed by 2 OFDM symbols, with a 15 kHz subcarrier spacing. A total of 27 MCSs

ranging from QPSK1/8 to 64QAM3/4 are considered. The payload can be mapped over a TTI and a 10 MHz bandwidth in case a QPSK1/8 modulation and coding scheme is used. The bandwidth is then reduced accordingly to the rate increase for the higher order MCSs; for example, a 5 MHz bandwidth is used in case of QPSK1/4, a 2.5 MHz bandwidth in case of QPSK1/2, and so on.

We assume Poisson packet arrivals at a rate of 7 packets per second; with the selected numerology, this corresponds to an arrival rate λ per TTI equal to 10^{-3}. The base station is equipped with 2 receive antennas and MRC or MMSE receiver. Ideal channel estimation for both desired and interfering signals is considered. We assume open loop power control with full pathloss compensation. The transmit power at the user is then given by $\mathrm{P_T}\,[\mathrm{dBm}] = \mathrm{P_0}\,[\mathrm{dBm/RB}] + 10\log_{10}(N_{\mathrm{RB}}) + \mathrm{PL}\,[\mathrm{dB}]$, where $\mathrm{P_0}$ is the target receive power spectral density per resource block (RB), N_{RB} denotes the number of RBs, PL is the path-loss. By considering a target $\mathrm{P_0}$ in the range $[-115, -85]$ dBm/RB, a thermal noise power density of -174 dBm/Hz and a 5 dB noise figure at the receiver, the corresponding SNRs are in the range $[1.4, 30.4]$ dB.

We study the performance for a single user scenario (i.e., no interference in the occupied resources), as well as for the case of a number of user sharing the same resources. In the latter case, all the users are transmitting over the same bandwidth, e.g. a 10 MHz bandwidth for QPSK 1/8, a 5 MHz bandwidth for QPSK1/4, and so on. A 4 GHz carrier frequency is assumed, and a 3 kmph UE speed. Simulations are run for different power control settings and for a large number of MCS, assuming a fixed MCS per simulation. More than 10^5 packet transmissions are simulated to ensure statistical confidence. The highest order MCS which still copes with a predefined $\mathrm{P_f}$ is then identified as representative of the achievable rate with that target.

Note that analytical results presented in the previous section are based on the assumption of flat Rayleigh fading and Shannon rates, while the system level analysis captures a number of realistic effects of the InH channel model defined in [13] such as shadowing, antenna patters, delay and azimuth spread, as well as an empirical link-to-system interface based on Mutual Information Effective SINR mapping [14]. For further detail on the simulator assumptions, we refer to our previous study [5]. It is then not an objective of this study to pursue a perfect match between analytical results obtained from the model and system level performance. The comparison is rather meant to verify the expected trends and suggest insights on practical link adaptation strategies.

Table I reports the highest order MCS which guarantees a failure probability $\mathrm{P_f}$ not larger than 10^{-3}, assuming different number of users as well as different target SNRs. At low SNR, the selected MCS is the same or very similar for all the cases since the performance is mainly noise-limited. For the single user case, the rate increases as a function of the SNR, while the

Table K.1: Higher order MCS guaranteeing $P_f \leq 10^{-3}$

SNR	N=1, MRC	N=10, MRC	N=50, MRC	N=50, MMSE
1.4 dB	QPSK1/8	QPSK1/8	QPSK1/8	QPSK1/8
6.4 dB	QPSK1/4	QPSK1/4	QPSK1/5	QPSK1/4
11.4 dB	QPSK1/2	QPSK1/3	QPSK1/4	QPSK2/5
16.4 dB	QPSK2/3	QPSK2/5	QPSK1/4	QPSK1/2
21.4 dB	16QAM3/5	QPSK2/5	QPSK1/4	QPSK2/3
26.4 dB	16QAM3/4	QPSK2/5	QPSK1/4	QPSK3/4
31.4 dB	64QAM3/4	QPSK2/5	QPSK1/4	16QAM1/2

multi-user MRC cases experience the rate saturation phenomenon. In particular, the rates saturate earlier for the crowded scenario of 50 users. Rates are reduced also for the MMSE case, which however does not experience rate saturation in the SNR region of interest thanks to its interference rejection capabilities; as mentioned in Section II, the combination $\gamma = 10^{-3}$ and $N = 50$ leads to a number of relevant interferers equal to 2, and MMSE with 2 receive antennas is able to suppress one of them. This is consistent with the trends identified with the analytical studies.

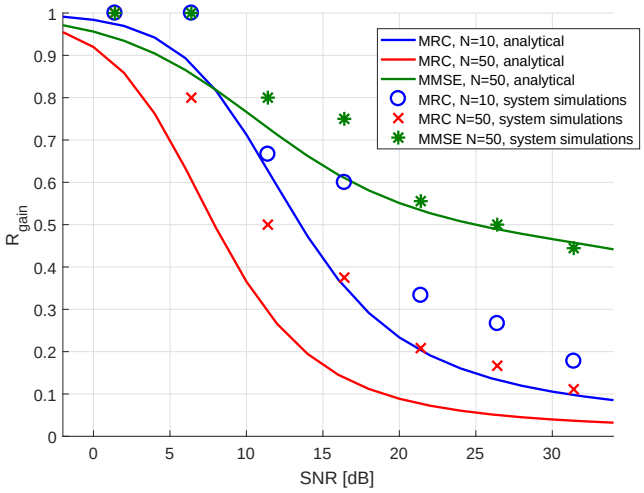


Fig. K.4: Rate gain for transmission over shared resources with respect to a single-user scenario.

Figure 4 displays the rate gain R_{gain} as the ratio between the achievable rate in shared resources over the achievable rate in single-user scenario, and compares the analytical findings and the results obtained with system simulations. The rate penalty for transmission over shared resources with respect to single user transmission is visibly lower in the realistic analysis compared

to the analytical findings, especially for the case of MRC receiver. This is due to the fact that, when operating over shared resources, low order MCSs are selected. In our system assumption, the usage of a low order MCS leads to a large occupied bandwidth, and therefore to the possibility of capturing frequency diversity gain. This translates to a rate performance improvement with respect to the flat Rayleigh fading in the analytical studies. Note that the performance gap is reduced in the case of MMSE receiver; the interference suppression capabilities translates to the usage of higher order MCSs and therefore of a smaller bandwidth, which diminishes the diversity benefits. It is worth observing that the trends of the system analysis results do not appear strictly monotonic, with a few points clearly "off-track" with respect to the trend for a given configuration, e.g. MRC $N = 10$ at $\sim$11.4 dB SNR. This is a consequence of the limited MCS granularity, which translates to the selection of an MCS leading to significantly lower P_f than 10^{-3} in case the immediate next MCS leads to a higher P_f than the target.

In a practical system, a rapid selection of the MCS to be set in a cell serving users in the same resources might not be possible since it depends on the specific propagation conditions which need to be estimated over time. The rates calculated from (K.2) only subsume knowledge of the overall load in the shared resources in terms of number of users and packet arrival rates, and appeared as a safe choice for a first MCS selection which guarantees a failure probability below P_f.

In summary, by knowing the number of users in the shared resources and their traffic profile, the base station may select the initial MCS as the closest value to the analytical rates calculated from (K.2). This avoids the usage of a "worst case" MCS which may severely affect the resource efficiency in the network. For instance, moving from lowest order MCS QPSK1/8 to QPSK1/4 already allows relinquishing half of the bandwidth to other services, or to accommodate a significant larger number of users coping with the target P_f. Upon the first choice given by (K.2), a fine tuning of the MCS to be used in the cell can be obtained, for instance, by using an outer-loop-link-adaptation (OLLA) mechanism for the sake of converging to the optimal MCS for a given scenario [15]. In current cellular systems, OLLA is mainly meant to correct link adaptation errors and operates by applying an instantaneous SNR offset for MCS selection depending on the success of the latest transmission. However, its benefits are rather poor in the presence of strong bursty interferers [16]. An OLLA mechanism for transmission over collision-prone shared resources should rather apply corrections at a slower pace and be based on long term packet failure statistics. The design of such scheme is left for future work. The possibility of using different MCSs within the group of users operating over shared resources based on, e.g. coupling gain, is also to be explored.

4 Conclusions and future work

In this paper, we have discussed the achievable transmission rates guaranteeing a target failure probability in collision-prone scenarios, assuming Rayleigh fading and linear receivers at the base station. We have introduced the concept of relevant number of interferers, and have discussed the maximum achievable rates as a function of the collision probability, number of users and number of receive antennas. In case an MRC receiver is used at the base station, a clear rate penalty with respect to a collision-free scenario appears at a medium/high SNR region. In particular, rates are saturating at medium/high packet arrival rates and remain constant regardless of the SNR. The same performance as in collision-free scenarios can instead be achieved in case a MMSE receiver is used at the base station, provided the number of statistically relevant interferers is lower than the number of receive antennas. In highly interfered scenarios or with a limited number of receive antennas, the performance of MMSE receiver also suffers from rate penalty and saturation.

We have complemented the analytical results with a realistic system analysis in an indoor hotspot scenario, which reveals similar trends. The analytical rate estimations can be used as a basis for a first MCS selection in collision-prone scenarios, to be finely tuned on a slow basis. We believe the presented findings can be used as a reference for the design of empirical link adaptation strategies for transmission over shared resources. The design of a slow Outer Loop Link Adaptation (OLLA) mechanism meant at fine tuning the MCS selection is left for future work.

Acknowledgements

This research is partially supported by the EU H2020-ICT-2016-2 project ONE5G. The views expressed in this paper are those of the authors and do not necessarily represent the project views.

1 Proof of (K.5)

We calculate here

$$\lim_{\overline{\gamma}\to\infty} \sum_{z=0}^{N-1} \mathrm{P}_{\mathrm{c},z}\mathrm{P}_{\mathrm{f},z} = \sum_{z=0}^{N-1} \mathrm{P}_{\mathrm{c},z} \lim_{\overline{\gamma}\to\infty} \mathrm{P}_{\mathrm{f},z}. \tag{9}$$

Let us first rewrite (K.4) as follows:

$$\mathrm{P}_{\mathrm{f},z} = 1 - \left(\frac{1}{2^{\mathrm{r}}}\right)^{z} e^{-\frac{2^{\mathrm{r}}-1}{\overline{\gamma}}} \cdot \cdot \sum_{k=0}^{M-1} \frac{\left(2^{\mathrm{r}}-1\right)^{k}}{k!} \sum_{i=0}^{k} \binom{k}{i} \frac{\overline{\gamma}^{i}}{\overline{\gamma}^{k}} \frac{\Gamma\left(z+i\right)}{\Gamma\left(z\right)\left(2^{\mathrm{r}}\right)^{i}} \quad (10)$$

One can observe that

$$\lim_{\overline{\gamma}\to\infty} \overline{\gamma}^{i-k} = \begin{cases} 1 & \text{if } i = k \\ 0 & \text{if } i < k \end{cases}$$

It follows that

$$\lim_{\overline{\gamma}\to\infty} \mathrm{P}_{\mathrm{f},z} = 1 - \left(\frac{1}{2^{\mathrm{r_s}}}\right)^{z} \left(\sum_{k=0}^{M-1} \left(\frac{2^{\mathrm{r_s}}-1}{2^{\mathrm{r_s}}}\right)^{k} \frac{1}{k!} \frac{\Gamma\left(z+k\right)}{\Gamma\left(z\right)}\right) = 1 - \left(\frac{1}{2^{\mathrm{r_s}}}\right)^{z} \left(1 + \sum_{k=1}^{M-1} \left(\frac{2^{\mathrm{r_s}}-1}{2^{\mathrm{r_s}}}\right)^{k} \frac{1}{k!} \frac{\Gamma\left(z+k\right)}{\Gamma\left(z\right)}\right). \quad (11)$$

Since $\lim_{\overline{\gamma}\to\infty} \mathrm{P}_{\mathrm{f},0} = 0$, by combining (11) with (9) we obtain the result in (K.5).

2 Proof of (K.8)

Let us calculate

$$\lim_{\overline{\gamma}\to\infty} \mathrm{P}_{\mathrm{f},z} = \lim_{\overline{\gamma}\to\infty} 1 - e^{-\frac{\gamma}{\overline{\gamma}}} \sum_{n=1}^{M} \frac{\mathrm{A}_n\left(\gamma\right)}{(n-1)!} \left(\frac{\gamma}{\overline{\gamma}}\right)^{n-1}. \quad (12)$$

Note that

$$\lim_{\overline{\gamma}\to\infty} e^{-\frac{\gamma}{\overline{\gamma}}} \left(\frac{\gamma}{\overline{\gamma}}\right)^{n-1} = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n > 1 \end{cases}$$

It follows that

$$\lim_{\overline{\gamma}\to\infty} \mathrm{P}_{\mathrm{f},z} = 1 - \mathrm{A}_1\left(\gamma\right), \quad (13)$$

which applied in (9) with the definition given in (K.7), leads to the expression in (K.8).

References

[1] H. Holma and A. Toskala, *LTE for UMTS: OFDMA and SC-FDMA Based Radio Access*. Wiley, 2009.

[2] P. Roshan and J. Leary, *802.11 Wireless LAN Fundamentals*. Cisco Press, 2004.

[3] L. Dai, B. Wang, Y. Yuan, I. Han, S. Chih-Lin, and Z. Wang, "Non-Orthogonal Multiple Access for 5G: Solutions, Challenges, Opportunities, and Future Research Trends," *IEEE Communication Magazine*, vol. 53, no. 9, pp. 74–81.

[4] "Study on New Radio Access Technology - Physical Layer Aspects," 3rd Generation Partnership Project, Tech. Rep. 38.802 v14.2.0, 2017.

[5] T. Jacobsen *et al.*, "System Level Analysis of Uplink Grant-Free transmission for URLLC," *IEEE Globecom*, December 2017.

[6] M. K. Simon and M. S. Alouini, *Digital Communications Over Fading Channels*. 2nd ed. Hoboken, New Jersey: Wiley-Interscience, 2005.

[7] D. N. C. Tse and O. Zeitouni, "Linear multiuser receivers in random environments," *IEEE Transactions on Information Theory*, vol. 46, no. 1, pp. 171–188, January 2000.

[8] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. McGraw Hill, 1991.

[9] V. A. Aalo and J. Zhang, "Performance Analysis of Maximal Ratio Combining in the Presence of Multiple Equal-Power Cochannel Interferers in a Nakagami Fading Channel," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 2, pp. 497–503, March 2001.

[10] G. H., P. J. Smith, and M. V. Clark, "Theoretical Reliability of MMSE Linear Diversity Combining in Rayleigh-Fading Additive Interference Channels," *IEEE Transactions on Communications*, vol. 46, no. 5, pp. 666–672, May 1998.

[11] J. Cui and A. U. H. Sheikh, "Outage Probability of Cellular Radio Systems Using Maximal Ratio Combining in the Presence of Multiple Interferers," *IEEE Transactions on Communications*, vol. 47, pp. 1121–1124, August 1999.

[12] P. Li, D. Paul, R. Narasimhan, and J. Cioffi, "On the Distribution of SINR for the MMSE MIMO Receiver and Performance Analysis," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 641–657, January 2006.

[13] "Guidelines for evaluation of radio interface technologies for IMT-Advanced," ITU-R, Tech. Rep. M-2135-1, December 2009.

[14] S. Tsai and A. Soong, "Effective-SNR Mapping for Modeling Frame Error Rates in Multiple state Channels," *3GPP2-C30-20030429-010*, April 2003.

[15] C. Rosa, D. L. Villa, C. U. Castellanos, F. D. Calabrese, P. Michaelsen, K. I. Pedersen, and P. Skov, "Performance of Fast AMC in E-UTRAN Uplink," *International Conference on Communications (ICC)*, May 2008.

[16] M. Gatnau, D. Catania, F. Frederiksen, A. F. Cattoni, G. Berardinelli, and P. Mogensen, "Dynamic Outer Loop Link Adaptation for the 5G Centimeter-Wave Concept," *21th European Wireless Conference*, May 2015.

ISSN (online): 2446-1628
ISBN (online): 978-87-7210-454-6

AALBORG UNIVERSITY PRESS