Radio Resource Management for Uplink Grant-Free Ultra-Reliable Low-Latency Communications

Jacobsen, Thomas Haaning

Publication date:
2019

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

RADIO RESOURCE MANAGEMENT FOR UPLINK GRANT-FREE ULTRA-RELIABLE LOW-LATENCY COMMUNICATIONS

BY
THOMAS HAANING JACOBSEN

DISSERTATION SUBMITTED 2019

AALBORG UNIVERSITY
DENMARK

Radio Resource Management for Uplink Grant-Free Ultra-Reliable Low-Latency Communications

Ph.D. Dissertation

Thomas Haaning Jacobsen

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

Dissertation submitted: May 2019

PhD supervisor: Prof. Preben Mogensen
Aalborg University

Assistant PhD supervisor: Senior Research Engineer István Z. Kovács
Nokia Bell Labs

PhD committee: Associate Professor Cedomir Stefanovics (chairman)
Aalborg University

Professor Olav Emerik Tirkkonen
Aalto University

Associate Professor Zhibo Pang
ABB Corporate Research

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

Department: Department of Electronic Systems

ISSN (online): 2446-1628
ISBN (online): 978-87-7210-448-5

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

Thomas Haaning Jacobsen

Thomas Haaning Jacobsen received his M.Sc. in engineering (network and distributed systems) from Aalborg University, Denmark, in 2015. He has been pursuing the PhD degree since 2015 in Wireless Communications in the Wireless Communications Networks (WCN) section at Aalborg University in collaboration with Nokia Bell Labs. He is currently with Nokia Bell Labs as a device standardization research expert. His research focus on industrial internet of things related topics and radio resource management for uplink grant-free ultra-reliable low-latency communications.

Abstract

The support for Ultra-Reliable Low-Latency Communications (URLLC) in fifth generation (5G) New Radio (NR) will fuel the next industrial revolution, the tactile internet and autonomous vehicle communication, by reliably connecting sensors, actuators and controllers with, for example, a maximum of 1 ms latency with at least 99.999% probability of success. The challenging latency and reliability requirement calls for significant improvements of multiple components of the radio access network (RAN).

Grant-free (GF) is one of the main enablers for uplink URLLC as it reduces the latency by omitting the conventional dynamic scheduling procedure. However, GF transmissions present a critical trade-off; while it reduces the latency budget it also removes scheduling flexibility and introduces intra-cell interference when GF radio resources are shared among URLLC devices. With a combination of revisited and novel radio resource management (RRM) mechanisms, this thesis proposes a concept with recommendations on uplink GF URLLC use in 5G NR.

The first part proposes both GF and grant-based (GB) transmission schemes for uplink URLLC and studies their achievable performance for sporadic traffic URLLC traffic. Insights are provided on the latency and URLLC capacity trade-off with a GF repetition-based, a GF retransmission-based and a GB retransmission-based transmission scheme. It is demonstrated that a repetition-based scheme is a desirable solution when the more resource efficient retransmission-based scheme cannot meet the latency requirement. It is observed that decreasing the latency requirement from 1 ms to 0.7 ms or 0.5 ms comes at a spectral efficiency degradation by a factor of 10 or 20 respectively. Results indicate that GB can reach the highest URLLC capacity when the latency requirement is relaxed from 1 ms to 1.4 ms, depending on frame-numerology, processing times and receiver capability assumptions.

GF transmission over shared radio resources pose an unprecedented challenge with the presence of sporadic intra-cell interference. In the second part of the thesis we propose RRM enhancements for uplink GF with the purpose of increasing the URLLC capacity while fulfilling the URLLC service requirements. Uplink power control is found to be an essential RRM

mechanism for URLLC with parameters optimized for GF transmissions. A novel RRM mechanism is proposed which combines resource allocation with a modulation and coding scheme (MCS) selection algorithm. This mechanism is shown to dramatically improve the URLLC reliability and further enhance the URLLC capacity by reducing the probability of fully overlapping GF transmissions.

The third part focus on diversity enhancement techniques for uplink GF URLLC. Transmission, antenna and receiver diversity is studied, where the latter is achieved by the technique of multi-cell reception. Multi-cell reception shows strong performance improvements, even with a simple multi-cell combining scheme. Novel multi-cell aware RRM techniques are presented and demonstrated to be capable of unleashing the full potential of uplink GF URLLC with multi-cell reception.

The fourth part focus on how to efficiently support enhanced Mobile Broadband (eMBB) and URLLC on the same carrier, while satisfying the strict URLLC requirements. The eMBB and URLLC service capacity trade-off is studied using both spatial and frequency domain multiplexing techniques. Service differentiated uplink power control is proposed and demonstrated to be an essential technique to enable uplink GF URLLC to be multiplexed with eMBB services.

Based on the main findings, the thesis is concluded with a summary of the relation between the achieved latency and spectral efficiency. With this, a set of concrete recommendations on how to achieve efficient support of uplink URLLC is provided along with proposals for further studies.

Resumé

Den femte generation (5G) af global mobilkommunikationsteknologi kaldet New Radio (NR), vil supportere en ny serviceklasse kaldet Ultra-Reliable Low-Latency Communications (URLLC). URLLC forventes at være en af grundpillerne til realiseringen af en række af de næste store teknologitrends såsom; den næste industrielle revolution, det taktile internet og samarbejdende selvkørende biler. URLLC forbinder sensorer, aktuatorer og kontrollere med en forsinkelse på eksempelvis 1 ms med mindst 99.999% sandsynlighed. At opnå denne lave forsinkelse med så høj sandsynlighedsgaranti kræver signifikante forbedringer for flere komponenter i eksisterende radio access networks (RAN).

Grant-free (GF) er en af de vigtigste komponenter til realiseringen af de strænge URLLC krav, da det reducere forsinkelsen på en data transmission ved at undlade den konventionelle dynamiske skeduleringsprocedure. Tilgengæld præsentere GF også et kritisk trade-off; imens GF frigør tid til flere transmissioner for at øge transmissionspålideligheden, så fjerner det også skeduleringsfrihed og introducere interferens mellem transmissioner i samme celle når URLLC enheder transmittere over de samme radio resourcer. I denne afhandling præsenteres et koncept med tilhørende anbefalinger om brugen af uplink GF URLLC i 5G NR, som er udarbejdet ved både at genevaluere kendte radio resource management (RRM) teknikker og ved udarbejdelse af nye of forbedrede teknikker.

I den første del af afhandlingen præsenteres både GF og grant-based (GB) transmissionsprotokoller til håndtering af sporadisk uplink URLLC traffic. Protokollernes trade-off mellem forsinkelse og maximum URLLC traffikbelastning hvor URLLC kravene kan overholdes (URLLC kapacitet) studeres i detaljer. Det bliver demonstreret at en repetitions-baseret GF protokol er at foretrække den mere spektral effektive retransmission-baseret GF protokol ikke kan overholde forsinkelsekravet. Det er observeret at reducere forsinkelseskravet fra 1 ms til 0.7 ms eller 0.5 ms kommer på bekostning af en faktor 10 eller 20 i spektral effektivitet. Resultater indikerer at en retransmission-baseret GB protokol kan opnå de højeste URLLC kapaciteter, men først når forsinkelseskravet resuceres til omkring 1.4 ms, afhængigt

af den anvendte frame-numerology, processeringstider og radioreceiverens egenskaber.

GF over delte radio resourcer udgør en hidtil uset udfordring grundet chancen for sporadisk interference mellem URLLC enheder i samme celle. Den anden del af denne afhandling foreslås RRM teknikker til at forbedre URLLC kapaciteten i denne konfiguration. Det er indentificeret at uplink power control er en essentiel RRM teknik men som kræver parametertuning for at opnå den maksimale URLLC kapacitet. En ny RRM teknik bliver præsenteret som kombinere en radio resource allokering strategi med en modulations og kodningsrate (MCS) udvælgelsesstrategi. Evalueringer viser at den nye teknik dramatisk kan forbedre URLLC transmissionspålideligheden og URLLC kapaciteten ved at reducere sandsynligheden for at GF transmissioner fuldt ud overlapper.

Den tredje del af denne afhandling fokusere på diversitet forbedrende teknikker til uplink GF URLLC. Transmission, antenna og receiver diversitet er studeret, hvor den sidstnævnte opnås ved hjælp af multi-cell reception. Med multi-cell reception er der observeret signifikante forbedringer selv med simple pakkekombineringsteknikker. Nye RRM teknikker præsenteres som er designet til at maximere URLLC kapaciteten med multi-cell reception.

Den sidste del af denne afhandling fokusere hvordan man effektivt kan servicere two enhanced Mobile Broadband (eMBB) og URLLC på den same radiokanal. Anvendelsen af service specifik uplink power control parametre er derfor foreslået og vises at være essentiel for at supportere uplink GF URLLC multiplexing med eMBB.

Afhandlingen konkluderes en illustration af forholdet mellem spektral effektivitet og forsinkelseskrav baseret på afhandlingens hovedresultater. Ud fra dem, gives en række konkrete anbefalinger til hvordan fremtidens cellulære nekværk effektivt kan understøtte uplink URLLC services og tilsidst gives en række forslag til relevante fremtidige studier.

Contents

Curriculum Vitae iii

Abstract v

Resumé vii

List of Abbreviations xiii

Thesis Details xvii

Preface xix

I Introductory Chapters 1

Introduction 3
1 5G New Radio Overview 4
2 Ultra-Reliable Low-Latency Communications 5
3 Anatomy of a wireless communication system 6
4 Scope and Objectives of the Thesis 8
4.1 Research Methodology 13
4.2 Research Questions and Hypothesis 14
5 Contributions 15
6 Thesis Outline 20
References 22

When is grant-free transmission an efficient option? 25
1 Transmission scheme latency budgets 25
2 Latency budget comparison 28
3 Latency and reliability influencing factors 29
4 Take-aways and outlook 31
References 32

II Transmission schemes for Uplink Ultra-Reliable Low-Latency Communications 33
1 Problems and solution space 35
2 Objectives 37
3 Included Articles 38
4 Main Findings 38
5 Recommendations and follow-up studies 39
References 40

A System Level Analysis of Uplink Grant-Free Transmission for URLLC 41

B System Level Analysis of K-Repetition for Uplink Grant-Free URLLC in 5G NR 57

III Grant-free radio resource management enhancements 71
1 Problems and solution space 73
2 Objectives 74
3 Included Articles 75
4 Main Findings 76
5 Recommendations and follow-up studies 79
References 79

C Power Control Optimization for Uplink Grant-Free URLLC 81

D Joint Resource Configuration and MCS Selection Scheme for Uplink Grant-Free URLLC 97

E Efficient Resource Configuration for Grant-Free Ultra-Reliable Low Latency Communications 113

IV Diversity and multi-cell reception 117
1 Problems and solution space 119
2 Objectives 121
3 Included Articles 122
4 Main Findings 122
5 Recommendations and follow-up studies 125
References 126

F Multi-cell Reception for Uplink Grant-Free Ultra-Reliable Low-Latency Communications 127

V Multiplexing of URLLC and eMBB services 131
1 Problems and solution space 133
2 Objectives 135
3 Included Articles 135
4 Main Findings 136
5 Recommendations and follow-up studies 137
References 138

G System Level Analysis of eMBB and Grant-Free URLLC Multiplexing in Uplink 139

H On the Multiplexing of Broadband Traffic and Grant-Free Ultra-Reliable Communication in Uplink 153

VI Conclusions 171
1 Summary of the main findings 173
2 Future work 177
References 179

VII Appendix 181

I On the Performance of One Stage Massive Random-Access Protocols in 5G systems 185

J Generic Energy Evaluation Methodology for Machine Type Communication 201

List of Abbreviations

1G first generation

2G second generation

3G third generation

3GPP 3rd Generation Partnership Project

4G fourth generation

5G fifth generation

ACK positive acknowledgement

AMC adaptive modulation and coding

BLER block error rate

BS base station

CC Chase combining

CCDF complementary cumulative distribution function

CCH control channel

CDF cumulative distribution function

CG configured-grant

CL closed loop

CN core network

CoMP coordinated multipoint

CSI channel state information

DMRS demodulation reference sequence

E2E end-to-end

eMBB enhanced Mobile Broadband

eMTC enhanced machine type communication

F5G Fantastic 5G

FDD frequency division duplex

GB grant-based

GF grant-free

gNB fifth generation NodeB

GSM Global system for Mobile Communication

HARQ hybrid automatic repeat request

IMT-2020 International Mobile Telecommunications for 2020 and beyond

IRC interference rejection combining

ITU international Telecommunication Union

JT joint transmission

KPI key performance indicator

L2S link-to-system

LA link adaptation

LTE Long Term Evolution

LTE-A LTE-Advanced

LTE-A Pro LTE-Advanced Pro

MAC medium-access-control layer

MBB mobile broadband

MCS modulation and coding scheme

METIS Mobile and wireless communications Enablers for Twenty-twenty (2020) Information Society

MIMO multiple-input multiple-output

MMIB mean mutual information per coded bit

mMIMO massive MIMO

MMSE minimum mean square error

mMTC massive Machine Type Communication

MRC maximal-ratio combining

MTC machine-type communication

MU multi user

NACK negative acknowledgement

NB-IoT Narrowband Internet of Things

NR New Radio

OFDM orthogonal frequency-division multiplexing

OFDMA orthogonal frequency-division multiple access

OL open loop

PDCCH physical downlink control channel

PDSCH physical downlink shared channel

PDU packet data units

PF proportional fair

PHY physical layer

PRB physical resource block

QoS quality of service

RAN radio access network

RE resource element

RLC automatic repeat request

RLC radio link control

RRH remote radio head

RRM radio resource management

RSRP reference signal received power

RTT round-trip time

RU resource utilization

SCS sub-carrier spacing

SIC successive interference cancellation

SINR signal to interference-and-noise ratio

SNR signal to noise ratio

SPS semi-persistent scheduling

SU single user

TB transport block

TDD time division duplex

TTI transmission time interval

UE user equipment

UMTS Universal Mobile Telecommunications System

URLLC Ultra-Reliable Low-Latency Communications

WCDMA Wideband Code Division Multiple Access

Thesis Details

Thesis Title:	Radio Resource Management for Uplink Grant-Free Ultra-Reliable Low-Latency Communications.
PhD Student:	Thomas Haaning Jacobsen.
Supervisors:	Prof. Preben Mogensen. Aalborg University. István Z. Kovács. Nokia Bell Labs.

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) in collaboration with Nokia Bell Labs. The work was carried out in parallel with mandatory courses required to obtain the PhD degree.

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

Paper A: 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*, December 2017.

Paper B: 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". In *European Wireless*, May 2019. Accepted / in press.

Paper C: 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)*, April 2018.

Paper D: 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*, December 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". In *IEEE Transactions of Vehicular Technology*, 2019. Submitted for publication.

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

Paper G: R. Abreu, T. Jacobsen, G. Berardinelli, N. H. Mahmood, K. Pedersen, I. Z. Kovács and P. Mogensen. "System Level Analysis of eMBB and Grant-Free URLLC Multiplexing in Uplink". In *IEEE Vehicular Technology Conference (VTC) Spring*, April 2019. Accepted / in press.

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". In *IEEE Vehicular Technology Conference (VTC) Spring*, April 2019. Accepted / in press.

With two articles included in an appendix:

Paper I: N. H. Mahmood, N. Pratas, T. Jacobsen, and P. Mogensen. "On the Performance of One Stage Massive Random-Access Protocols in 5G systems". In *2016 9th International Symposium on Turbo Codes and Iterative Information Processing (ISTC)*, September 2016.

Paper J: T. Jacobsen, I. Z. Kovács, M. Lauridsen, L. Hongchao, P. Mogensen, and T. Madsen. "Generic Energy Evaluation Methodology for Machine Type Communication". In *Multiple Access Conference (MACOM)*, November 2016.

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 reproduced directly or indirectly in the extended summary of the thesis. According to the Ministerial Order no. 1039 of August 27, 2013, regarding the PhD Degree § 12, article 4, co-author statements have been provided to the PhD school prior to the submission of this thesis. The co-author statements have also been made available for the assessment committee.

Preface

With a great desire learn for the purpose of making an impact and to make a positive difference, I have always been seeking engagement in projects with potential to impact a stake-holder or for big personal and professional development. Despite aiming for the private industry after having obtained a master diploma in computer networking, I got inspired by my supervisor Preben Mogensen to start a PhD in the area of wireless communications.

This dissertation is the result of two intense years work on uplink URLLC after 1.5 year with initial focus on massive Machine Type Communication (mMTC). The change of research topic came after it was decided in 3rd Generation Partnership Project (3GPP) to down prioritize mMTC for 5G NR. I decided to refocus on the URLLC topic due to the much greater potential to make a positive impact through my PhD study. A close collaboration with fellow PhD student, Renato Abreu, was established to accelerate the research on uplink URLLC. With our joint knowledge and experience of the topic and the Nokia Bell Labs system level simulator, we managed to establish a performance baseline for uplink URLLC with 5G NR compliant scenario assumptions, and started to develop the uplink GF URLLC framework. Most of the papers included in this thesis is a result of this great scientific cooperation.

During the PhD, particular before the topic change, the main scientific knowledge dissemination was conducted through student project supervision and involvement in a smart city demonstration testbed at Aalborg University. A significant effort was put in to particular two demo development projects of which I was responsible for. The latest went to the Nokia Bell Labs booth in the largest annual event for the mobile industry, Mobile World Congress (MWC), in 2018. The projects has been a personal and professional challenge and required strong leadership and communicational skills to be successful.

The PhD has been a life changing personal and professional journey. This is particular the result of the great support from my supervisors and their guidance, which allowed mistakes to occur and has helped turning them into a learning experience. A big and full-hearted thanks for your effort and patience. I would also like to thank my wife and my family, who might not

always have known what I was doing, or why I persistently continued, but they were always ready to listen when it was needed. A special thanks to my wife for her endless love and seemingly endless patience, and for her understanding, support and caring nudges, which kept me going through the difficult periods.

Thomas Haaning Jacobsen
Aalborg University, May 2019

Part I

Introductory Chapters

Introduction

Cellular networks have since its initial proposal in the late 1940s by D. H. Ring from Bell Laboratories, had the purpose to provide wide-area wireless services to mobile devices. Both the first generation (1G) and second generation (2G) cellular network focused on voice services and were deployed throughout the 1980-2000. Global system for Mobile Communication (GSM), became the first world-wide cellular network. Increasing the average and peak data-rates has been the main goal with the third generation (3G) and fourth generation (4G) global cellular networks. The 3rd Generation Partnership Project (3GPP) 4G compliant technology is known as Long Term Evolution (LTE) which had its first version (LTE Release 8) completed in 2009. Since then, new versions of LTE has been released, up to Release 15 which was frozen in 2018. Mentionable milestones through the releases are coordinated multipoint (CoMP) in Release 8, LTE-Advanced (LTE-A) from Release 10 with focus on multiple-input multiple-output (MIMO) techniques and carrier aggregation to satisfy the ever-increasing data-rate demands for mobile broadband (MBB) and LTE-Advanced Pro (LTE-A Pro) from Release 14 which included significant coverage enhancements for machine-type communication (MTC).

The future vision for mobile networks is specified by the International Mobile Telecommunications for 2020 and beyond (IMT-2020) [1], which sets the targets for fifth generation (5G) wireless networks. The vision includes support for heterogeneous services; enhanced Mobile Broadband (eMBB) with high average data rate and very high peak data rate, massive Machine Type Communication (mMTC) which require a dramatic increase in connection density, and Ultra-Reliable Low-Latency Communications (URLLC), which require a significant reduction in one-way latency and reliability enhancements of several orders of magnitude compared to LTE [2]. The initial research of 5G started already in late 2012, with publicly funded research projects such as METIS [3] and Fantastic 5G (F5G) [4]. The standardization activities started in 2016 and the first 5G New Radio (NR) version known as NR Release 15, was frozen in the end of 2018. 3GPP has recently conducted a promising self-study on NR Release 15 capability to fulfill the IMT-2020 [5]

requirements [6].

This thesis focuses on URLLC in the uplink, which targets to provide end-to-end (E2E) delivery of a small data packets from a mobile user within a very tight latency budget (e.g. 1 ms) and with a very high probability of success (e.g. five nines (99.999%)). This dissertation closely examines the challenges, carefully evaluates state-of-the-art solutions and their trade-offs, establishes a baseline for uplink URLLC performance, and designs novel mechanisms for efficient uplink URLLC support in 5G NR networks.

1 5G New Radio Overview

Fig. I.1 illustrates the three main scenarios in service classes and highlights some of their 2020 requirements defined by IMT-2020 [5, 7].

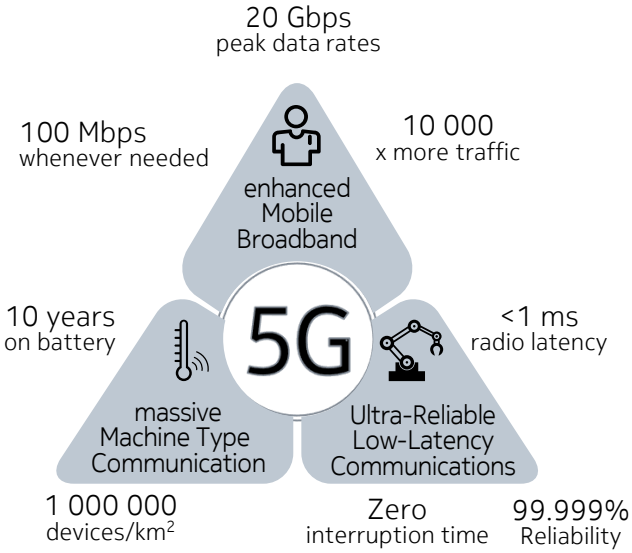


Fig. I.1: Service requirements for 5G [7].

eMBB can be seen as an enhancement from MBB services in LTE. An eMBB service is set to support up to 20 Gbps peak data rate and 100 Mbps even over a wide-area network. eMBB traffic can be considered to consist of large payloads from e.g. video streams or internet browsing. Some of the technical enablers is expected to be large bandwidth, aggregation of transmission carriers, single user (SU) and multi user (MU)-MIMO, intelligent scheduling and fast link adaptation (LA).

mMTC development started with the 3GPP LTE-A Pro Release 14, Narrowband Internet of Things (NB-IoT) and enhanced machine type communi-

cation (eMTC). mMTC services is intended to deliver infrequent small data packets with sufficient energy consumption to be equipped with batteries to last 10 years or even more, with severe coverage conditions and being able to handle at least 1.000.000 devices per km^2. The focus for mMTC is on power saving mechanisms, coverage enhancements techniques e.g. transmission repetitions, narrow bandwidth transmissions and efficient random access.

URLLC is a new service class, which targets to support unprecedented low latencies below 1 ms with ultra-high probability of success of at least 99.999%. The traffic is typically small payloads and example use cases are; control and automation systems in Industry 4.0, the tactile internet and vehicle-to-vehicle communication. The URLLC requirements sets high demands to all the related components in the radio access network (RAN) as the margins for error and delays are minimal [8].

2 Ultra-Reliable Low-Latency Communications

URLLC is anticipated to enable new use cases to fuel new business opportunities and revenue. The tactile internet enables for example humans to remotely manipulate physical or digital objects possibly in cooperation. A tactile internet enabled use case is the remote surgery [9]. Human presence in virtual reality is similar to the tactile internet, with the main difference that the cooperation is executed on virtual objects [10]. Autonomous vehicle communication require URLLC for vehicle-to-vehicle or vehicle-to-everything communication in order to enable enhanced safety operations such as alerting vehicles behind of an emergency breaking event. Conventional factory automation relies on wired communication networks for sufficient low latency and reliability. With the envisioned fourth industrial revolution (Industry 4.0), the factory hall gets more agile which will be aided by wireless URLLC [11].

While these use cases set strict E2E performance requirements, the requirements to the RAN air interface is even more challenging. A summary of the E2E and air-interface service requirements for the described use cases are listed in Table I.1. Best guesses have been added where none or vaguely formulated values were listed in the references.

Throughout this dissertation, the 3GPP requirement for 5G NR Release 15 URLLC is used, which are defined in [15] as:

- **User plane latency**. From ingress of the RAN Layer 2 at the device (user equipment (UE)) to egress of the RAN Layer 2 at the fifth generation NodeB (gNB), the target of 0.5 ms, on average if the reliability requirement is taken into account as well.

Table I.1: URLLC use cases and service requirements.

Description	E2E	Air interface	Traffic
Tele-presence and cooperation (e.g. assisted training and intervention) [12, 13]	<5 ms, 99.9999%	1-3 ms	Aperiodic
Virtual-Reality cooperation [10]	<1 ms	<0.5 ms	Periodic or aperiodic
Autonomous vehicles [12, 14]	<5 ms	3 ms, 99.999%	Periodic
Factory Automation [14]	2 ms	1 ms, 99.9999%	Periodic

- **Reliability**. The success probability of delivering a small packet of data size X, from the RAN layer 2 at the UE to the RAN layer 2 at the gNB, within Y ms, under some coverage condition. The general choice of X is 32 B and Y is 1 ms, and the success probability is $1 - 10^{-5}$.
- **Mobility**. The mobility interruption time must be 0 ms.

LTE is considered state-of-the-art deployed network with real life performance measured with estimated one-way E2E latencies below 25 ms with 50% probability and below 75 ms with 99% [7]. It is clear that fulfilling these requirements requires rethinking of several parts of the RAN, and the mechanisms in the present in layer 2 and layer 1, many of these are typically covered under the term *Radio Resource Management* (RRM). Following is a derivation of the major components which impacts the latency and reliability for URLLC in a 5G NR system, simplified in a generalized communication system.

3 Anatomy of a wireless communication system

Fig. I.2 illustrates a generalized communication system focused on the uplink direction. The system consists of U URLLC devices which are connected to at least one of C base stations (BSs), which can be distributed over a large area. Each device consists of one transmitter and each BS consists of one receiver. For each URLLC device, data arrives from higher layers to the RAN layer 2 and is stored in a queue until the lower layer is ready i.e. has a radio resources allocation. The BS may schedule these devices on demand or allow the devices to transmit on pre-allocated radio resources. The uncertainties in the wireless system are caused by noise and time and frequency-variant fading, and the interference from simultaneous transmissions. To cope with these uncertainties, diversity in both time, frequency and the spatial domain can be exploited.

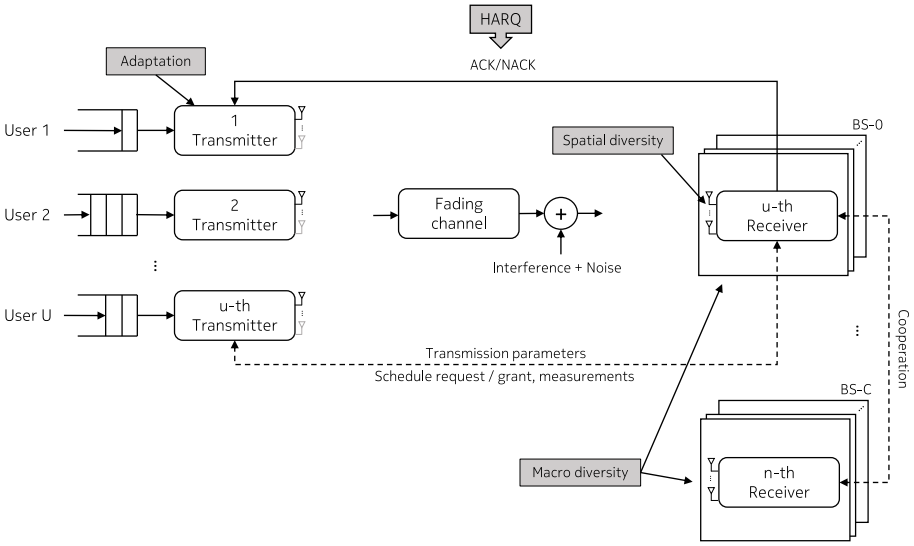


Fig. I.2: A generalized uplink wireless communication system (inspired by [16]).

The BS may use multiple receiver processes for each detected transmission. Each transmitter may be equipped with one or more transmit antenna and the receiver may be equipped with one or more receive antenna. This is typically known as MIMO antenna techniques and when multiple-devices are jointly considered, MU-MIMO. Receiver post-processing combining of the received signals can be applied to exploit the spatial dimensions from the multiple receive antenna to acquire spatial diversity translating into signal to noise ratio (SNR) or signal to interference-and-noise ratio (SINR) gains when combining takes into account the presence of interfering devices. Further, multiple BS may cooperate and jointly receive the transmissions to exploit macro diversity.

The success of a transmission can be signaled to the user as a positive acknowledgement (ACK) or a negative acknowledgement (NACK) and in the latter case, error recovery mechanisms can be triggered such as hybrid automatic repeat request (HARQ), which in that case can trigger a retransmission to benefit from transmission diversity. Alternatively, the user may be configured to transmit the same packet multiple times in an attempt to increase the transmission reliability. The BS may signal adjustments of the users transmission parameters, e.g. to boost the power of a retransmission, reduce in its transmission power if the signal quality is larger than necessary to maintain a predefined transmission reliability target. The BS can also signal adjustments of the modulation and coding scheme (MCS) to increase the spectral efficiency, or fit more devices in the available frequency spectrum in the de-

fined time resources used for transmission, also denoted as a transmission time interval (TTI).

Further, if the devices are moving, they may have to change serving BS which triggers a radio handover event, which may result in URLLC packets being buffered and hence delayed while the device is not configured.

4 Scope and Objectives of the Thesis

Fig. I.3 illustrates the mapping from URLLC service requirements to the potential solutions. Next, these potential solutions will be described in detail one by one.

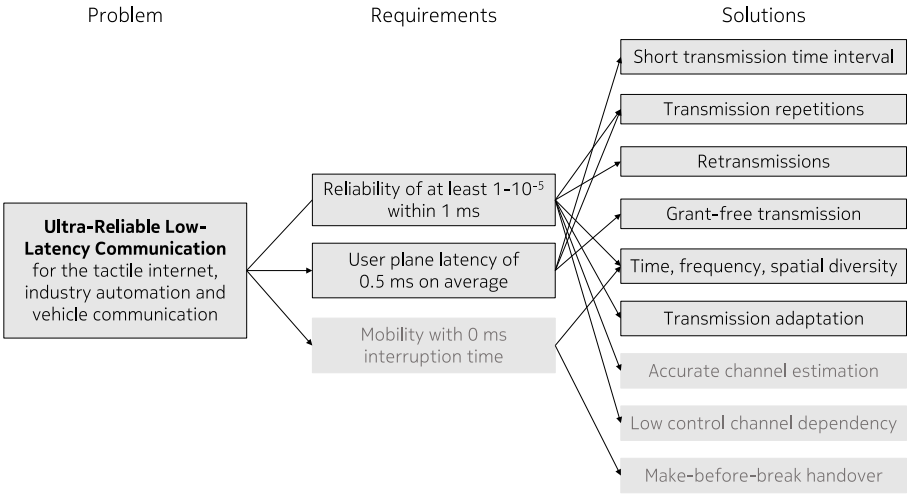


Fig. I.3: Mapping from problem to requirements to potential solutions. Requirements and solutions not covered in this dissertation are grayed out.

Dynamic scheduling is the traditionally procedure to allocate radio resources in a cellular networks. This is carried out through a scheduling procedure which involves the device to transmit a scheduling request and the BS to transmit a scheduling grant. A data transmission is then carried out on scheduled resources, which can be referred to as a grant-based (GB) transmission. With 5G NR assumptions, the scheduling procedure takes at least 0.5 ms [17], which has fueled the concept of grant-free (GF) transmissions which omits the scheduling procedure and carries out the URLLC payload transmission on pre-configured radio resources. GF is a enabler to reduce the latency and allow to fit more retransmissions or repetitions within the latency requirement.

Grant-free (GF) transmissions has been possible in LTE through semi-persistent scheduling (SPS) framework and in 5G NR through a more flex-

ible configured-grant (CG) framework. Both SPS and CG builds on the same principle, namely by pre-allocating periodic radio resources along with pre-configured transmission parameters. Periodic allocations, when aligned with periodic characteristics of the URLLC traffic, the utilization of the pre-allocated resources can be high. However, when the traffic is less predictable (i.e. sporadic) there can be GF transmission opportunities which are not used or the URLLC has more packets than what fits into the pre-allocated resources, and hence some of them have to wait (increasing the latency). To overcome this limitation with sporadic traffic, sharing of transmission opportunities, can increase the utilization of opportunities as well as reduce the latency compared to dedicated resource allocation [18]. The drawback of sharing GF transmission opportunities is the risk that multiple devices transmit on the same pre-allocated radio resources simultaneously causing mutual interference which harms the transmission reliability.

Achieving high degrees of diversity is considered to be essential to reach the ultra-high reliability as required for URLLC [8]. Retransmissions can be considered a mean to obtain transmission diversity in the time domain by exploiting the changing interference and fading conditions while combining the received energy. Diversity can also be achieved in the frequency domain and the spatial domain. In the frequency domain, multiple frequency channels (with assumed independent fading) can be obtained as a function of the coherence bandwidth. In the spatial domain multiple receive antennas can be used to acquire more energy and to exploit fading differences between the antennas by combining the received signal from each antenna. When receiving simultaneous transmissions from different devices, combining mechanisms can utilize the estimated channel response to increase the separation between received signals [19–21]. Spatial diversity can hence improve the reliability while improving the receiver capability to receive multiple transmissions on shared radio resources. Another diversity technique is combining of received transmissions from different BS, which is known as receiver diversity, macro-diversity reception or multi-cell reception. Multi-cell reception can aid the handover procedure from one BS to the another as possible target BS might already by configured to receive the desired transmissions.

Shortening the TTI means that more transmission opportunities can be fitted within the same latency budget. This is essential in order to fit multiple transmission opportunities in latency requirement [22]. Short TTIs can also reduce the latency spent on waiting for the start of the next TTI to start a transmission. However, shortening of the TTI, without changing the MCS, requires a larger bandwidth to fit the entire coded packet. The control plane overhead might also increase. Shortening of the receiver processing times at both the BS and device is another way to reduce the latency. The processing times affects for example the latency used by the BS to receive the uplink URLLC transmission or a dynamic scheduling request and the latency used

by the device to receive transmission feedback or a dynamic scheduling grant.

Transmission repetitions is a simple technique to enhance the transmission reliability, for example by combining the repetitions to acquire more energy per received bit at the receiver. In 5G NR this technique is introduced as K-repetitions, where the transmission is repeated K times in consecutive TTI. Contrary retransmissions, K-repetitions does not rely on control channel (CCH) signaling to provide reliability enhancements, once it is configured. However, once configured, all K-repetitions are transmitted if no feedback is provided, also if it is not needed for a successful reception. The use of K-repetitions can therefore result in unnecessary generated interference.

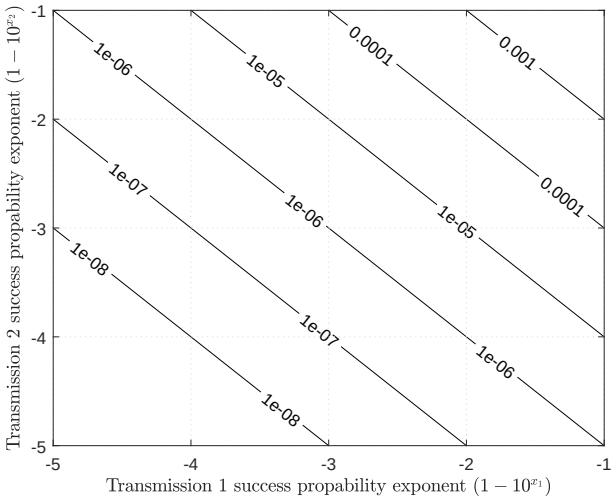


Fig. I.4: Ideal combined failure rate when combining two independent transmissions of failure exponent x_1 and x_2.

Retransmission of the initial transmission is an alternative approach to repetitions to increase the reliability e.g. by acquiring more energy per bit by combining transmissions. However, a retransmission is only commenced when the device has received a NACK of the initial transmission. Retransmissions are therefore only carried out when needed, contrary K-repetitions. Additionally, retransmissions as well as repetitions can be coded to to be self-decode-able or as an extension to a previous transmissions. The latter has slightly better performance when all transmissions are correctly received, but is less robust if some of the transmissions are lost [23]. This is also known as redundancy versions. Fig. I.4 illustrates the ideal failure rate when two independent transmissions with a success probability of $1-10^{x_1}$ and $1-10^{x_2}$ are combined in the reliability domain. In this case, a packet with success probability $1-10^{-2}$ is received, but did not succeed decoding, and the re-

transmission has a success probability of $1 - 10^{-3}$, but combined they reach a success probability of $1 - 10^{-5}$ on average. In a wireless network, the initial transmission and its retransmission are not necessarily independent in terms of fading or interference. Further, as retransmissions are triggered by the BS, the feedback can be used to adjust transmission parameters for the retransmission or even reserve radio resources for the retransmission.

Adapting the transmission parameters based on channel state information (CSI), is also known as link adaptation (LA). The idea is to adapt transmission parameters based on knowledge of the CSI and possible also the interference conditions. For sporadic traffic, acquiring up-to-date CSI can lead to a high resource overhead. When the GF transmissions are sporadic and carried out over shared radio resources, the adaptation of transmission parameters needs to account for the event of multiple interfering GF transmissions. For that reason, adaptation based only on the latest instantaneous CSI can have a negative impact on the following GF transmission because of unpredicted interference conditions. Instead it is better to base the transmission adaptation on longer-term CSI to capture the varying interference conditions [24]. Transmission adaptation requires feedback from the receiver which means that the URLLC transmission becomes dependent on the downlink CCH reception.

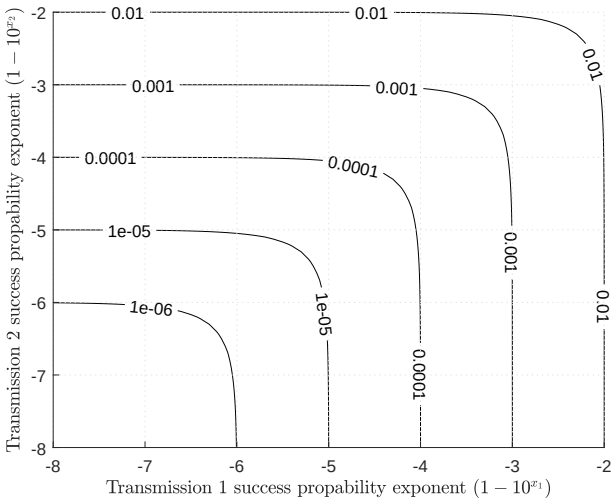


Fig. I.5: Total error probability of a two transmissions with mutual dependence.

Control channel (CCH) can enhance the URLLC capacity, by allowing adaptation of transmission parameters, but also introduces a dependence between the URLLC transmission reliability and the control message. For traffic in the uplink, the reliability of the downlink CCH is used for transmission parameter adaptation and transmission feedback. With dynamic scheduling

the uplink CCH is used as well to carry scheduling request messages. Basic reliability theory can be used to illustrate the consequence when the success of a packet directly depends on the success of another, this could be a URLLC transmission and a control packet used to either trigger a retransmission (NACK) or a scheduling grant. This dependence is illustrated in Fig. I.5, where the average success probability of the first and second message is denoted $1 - 10^{x_1}$ and $1 - 10^{x_2}$ respectively. This effect is exaggerated the more control messages the URLLC transmission depends on. However, when the control message is for transmission parameter adjustments, a wrong or lost control message might cause suboptimal parameters and the depicted example can be considered worst case. The negative impact from the transmission feedback can be reduced by considering explicit ACK or asymmetric robustness of ACK and NACK transmission [25]

Accurate **channel estimation** for demodulation at the BS is essential for transmission reliability, as a poor channel estimate, might cause an erroneous channel equalization and hence demodulation and decoding errors. Accurate channel estimates is also important for the combining efficiency at the receiver for spatial diversity and hence for the efficiency of the receiver to separate overlapping transmissions. On top of accurate channel estimation, detecting when a device is transmitting on GF resources as well as identification of the transmitting device is also important for the URLLC reliability, as it enables bookkeeping of which transmissions should be combined using HARQ.

Mobility. When the URLLC device is moving within the network, it will experience at some point that it will be more beneficial to change its serving BS. However, in order to achieving 0 ms interruption time during a handover between serving BS, the target BS needs to be configured and ready to serve the device before it reaches a point where the current serving BS cannot ensure a satisfying URLLC performance. This technique is referred to as soft and softer handover in Wideband Code Division Multiple Access (WCDMA) [26] and is also known as the "make-before-break" handover procedure. The make-before-break technique can be achieved through soft or softer handover with multi-cell reception in the uplink. Throughour this dissertation, static URLLC devices are considers and therefore the aspect of mobility will not be further considered.

Multiplexing of service classes with URLLC is important to consider for URLLC to meet its requirement in a scenario where radio resources are shared between service classes. One example can be the remote presence 5G use case which includes the combination of large data streams which can be served efficiently as an eMBB service and a real-time control communication system of between sensors, actuators and a controller which can be served with a URLLC service [13]. Identifying efficient multiplexing solutions, radio resource management (RRM) mechanisms and evaluating the performance

compromises to be made between the multiplexed services, while still satisfying their respective service requirements, is important for efficient co-existence of heterogeneous services.

Summary and thesis focus

GF transmissions is identified as a key enabler for uplink URLLC as it provides latency headroom for reliability enhancing techniques such as transmission repetitions and retransmissions. However, intra-cell interference is also introduced when GF radio resources are shared by the URLLC devices. No performance baseline has previously been established for either GF or GB transmissions when optimizing towards the URLLC requirements in 5G NR compliant networks. Therefore one of the focus points in this dissertation is to quantize the achievable performance of GF transmission through the development and evaluation of RRM mechanisms and transmission strategies. This includes identifying the benefits of selected diversity techniques and comparing the performance with GB based transmission strategies.

This dissertation targets to study how to fulfill the latency and reliability requirement for uplink URLLC under realistic settings. That is; the evaluation should include as many of the depicted reliability impacting factors in I.2 as possible in order to provide realistic results at the 99.999% reliability level, such as multi-path frequency selective and time-varying fading, antenna signal combining capturing the channel characteristics subject to the desired transmission and interfering transmissions including those from other-cells, line-of-sight probabilities, as well as major RRM mechanisms and RAN protocol stack latency.

4.1 Research Methodology

The research methodology adopted in this thesis generally follows five steps:

1. **Problem identification:** Problems are identified based the defined use cases and requirements when compared to an established performance baseline. These use cases, requirements and baselines, are identified by consulting the open literature. Performance baselines are established if they do not already exists. Based on the identified use cases and requirements and potential gap from the performance baseline, the open literature is revisited in search for existing solutions and related work.

2. **Hypothesis and potential solutions:** Based on related work from the open literature, a set of potential solutions are identified. Their benefits and drawbacks are studied in detail. New ideas might be generated at this stage. An hypothesis is formulated on the expected outcome when applying the potential solution on the problem.

3. **Validation:** Validation of the hypothesis, requires selection of the right evaluation methodology. For many of the problems tackled in this study, the right methodology has been system level Monte Carlo simulations using a proprietary Nokia Bell Labs simulator. The simulator includes many of the major performance impacting algorithms present in the radio access network and physical factors such as propagation, fading and interference. Particular acquiring analytical expressions and closed-form expressions for interference in a multi-cell multi-user networks can be proven to be a NP-hard problem [27] and even simpler scenarios implies significant simplifications, such as single-cell or single-user networks [28]. Further this simulator is calibrated against industry standards and is capable of modeling a 5G NR radio access network. Novel features and mechanisms have been implemented on demand during the study, in order to validate the hypothesis based on realistic performance evaluations.

4. **Iteration:** Based on the findings from the validation step, the hypothesis is either validated or rejected. In many cases the findings gave rise for a reformulation of the hypothesis following a revisit of the the validation step, along with the choice of evaluation methodology. This iteration process continue until the problem is fully understood and the hypothesis is firmly rejected or accepted.

5. **Dissemination:** The findings are then disseminated through research publications and presentations in relevant forums such as international conferences, journal publications, international research collaborations between universities and in international collaboration projects within the partner company Nokia Bell Labs. With the partner company, the main findings and concepts have been contributed to the standardization body 3GPP.

4.2 Research Questions and Hypothesis

The main research questions and hypothesis addressed in this dissertation are:

Q1 Given a certain reliability and latency requirement, which uplink transmission strategy for sporadic uplink URLLC traffic can achieve the highest URLLC capacity?

H1 A transmission scheme can be based on repetitions or retransmissions and can be either GF or GB. GF transmission schemes based on repetitions are simple and can be utilized for even very strict latency requirements. A retransmission-based transmission scheme is expected to reach a higher URLLC capacity as it only transmits “on-demand”.

GB schemes is expected to be able to reach a higher URLLC capacity, due to its ability to dynamically schedule transmissions, but at the cost of latency.

Q2 How can RRM be used to enhance the URLLC capacity of sporadic uplink GF URLLC traffic?

H2 RRM mechanisms optimized for sporadic GF transmissions which account for the intra-cell interference will enhance the URLLC capacity. This includes revisiting known mechanisms such as uplink power control but also new schemes such as long-term MCS selection and a strategic radio resource allocation schemes to minimize the probability of fully overlaying transmissions.

Q3 What is the expected benefit of multi-cell reception and how much can it improve the URLLC capacity when combined with transmission and spatial diversity at the receiver?

H3 Multi-cell reception will significantly enhance the URLLC capacity for sporadic GF transmissions on shared radio resources as it can acquire more energy per bit with multiple receive antennas, but also experience different fading and interference conditions at each BS. Multi-cell reception aware RRM which accounts for the experienced signal quality gains will further enhance the achieved URLLC capacity. Transmission diversity and spatial diversity through multiple receive antennas per BS is expected to benefit all devices throughout the network.

Q4 How can a URLLC service most efficiently be multiplexed with an eMBB service on the same carrier in the uplink?

H4 URLLC and eMBB are two service classes with very distinct service requirements. An ongoing eMBB transmission of a large packet, can delay the transmission of a URLLC short packet. By avoiding a trunking efficiency loss, allowing simultaneous overlaying of eMBB and URLLC transmissions will provide a favorable capacity trade-off between eMBB and URLLC.

5 Contributions

The main contributions of this study are as follows:

1. **System concept for sporadic uplink GF URLLC**. The concept includes; GF transmission scheme for a given latency requirement, a combination of revisited and new RRM mechanisms such as optimized uplink power control and a joint resource allocation and MCS selection scheme and

interference aware linear receivers equipped with multiple receive antennas. Further it is shown how multiple receive antennas, retransmission and multi-cell reception can provide additional URLLC capacity gains for sporadic GF traffic.

2. **Established a transmission scheme baseline performance for uplink URLLC**. Uplink transmission scheme performance baselines for URLLC have been established using a state-of-the-art system level simulator. The system level simulator captures the reliability and latency influencing factors to ensure adequate details are covered in order to provide realistic performance results for a 5G NR network. This includes; inter- and intra-cell radio interference, HARQ process modeling, queuing, channel models, power control and link-to-system mapping. Further a simulation methodology has been designed to ensure that sufficient number of samples are acquired to make mature conclusions.

3. **Uplink power control recommendations and validation**. The optimum power control parameters for a 5G NR network have been studied using detailed system level simulations. Insights to the importance of power control parameter tuning for URLLC is provided along with a feasibility study of power boosting retransmissions.

4. **Novel joint resource allocation and MCS selection scheme for uplink GF**. Strategic overlaying of transmission is proposed and validated with significant gains in terms of URLLC capacity to serve uplink GF URLLC. The MCS selection and resource allocation scheme is implemented, tested and used for validation of the proposed GF concept.

5. **Studying the potential of diversity and multi-cell reception for GF uplink URLLC**. The potential of transmission, antenna and receiver diversity is studied. The latter is achieved with multi-cell reception. The technique are studied for uplink GF URLLC in a 5G NR compliant network setting, as techniques to improve transmission reliability and hence the URLLC capacity. On top, two novel multi-cell reception aware RRM adaptation schemes are proposed and demonstrated to be capable of unleashing the full URLLC capacity potential when using multi-cell reception with GF transmission schemes.

6. **Recommendations on uplink URLLC and eMBB service multiplexing with URLLC**. Detailed analysis of deployment options for combined eMBB and URLLC services are provided. The use of service differentiated uplink power control and multiple receive antennas with linear receivers is studied as enablers for allowing spatial domain multiplexing. Then insights into the capacity trade-off between spatial domain

multiplexing and frequency domain multiplexing of a shared radio carrier are provided.

7. Outlying the cost in terms of spectral efficiency of supporting URLLC services at latencies from 1 ms down to 0.5 ms in a 5G NR compliant network.

Most of the findings of the PhD study have been contributed to the 3GPP standardization body through the partner company and have assisted shaping the specification of the 5G NR Release 15. Some of the presented concepts are currently being discussed for Release 16 [29].
Five co-authored patent applications have been successfully filed during the study. Two of these are publicly disclosed:

Patent Application 1: Radio Configuration for machine type communications (No. WO2018054475).

Patent Application 2: Method and apparatus for Wireless Device Connectivity Management (No. WO2018083368).

The main findings of this study are presented through a collection of the following articles:

Paper A: 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*, December 2017.

Paper B: 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". In *European Wireless*, May 2019. Accepted / in press.

Paper C: 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)*, April 2018.

Paper D: 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*, December 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". In *IEEE Transactions of Vehicular Technology*, 2019. Submitted for publication.

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

Paper G: R. Abreu, T. Jacobsen, G. Berardinelli, N. H. Mahmood, K. Pedersen, I. Z. Kovács and P. Mogensen. "System Level Analysis of eMBB and Grant-Free URLLC Multiplexing in Uplink". In *IEEE Vehicular Technology Conference (VTC) Spring*, April 2019. Accepted / in press.

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". In *IEEE Vehicular Technology Conference (VTC) Spring*, April 2019. Accepted / in press.

Paper I: N. H. Mahmood, N. Pratas, T. Jacobsen, and P. Mogensen. "On the Performance of One Stage Massive Random-Access Protocols in 5G systems". In *2016 9th International Symposium on Turbo Codes and Iterative Information Processing (ISTC)*, September 2016.

Paper J: T. Jacobsen, I. Z. Kovács, M. Lauridsen, L. Hongchao, P. Mogensen, and T. Madsen. "Generic Energy Evaluation Methodology for Machine Type Communication". In *Multiple Access Conference (MACOM)*, November 2016.

The article A-H constitutes the main part of the thesis, whereas article I-J are included in an appendix. During the study, the following scientific publications have been co-authored, but are not included in the thesis. The reader is therefore kindly referred to their respective publication channels:

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

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

- I. Z. Kovács, P. E. Mogensen, M. Lauridsen, T. Jacobsen, K. Bakowski, P. Larsen, N. Mangalvedhe and R. Ratasuk. *LTE IoT Link Budget and*

Coverage Performance in Practical Deployments. In *IEEE 28th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, October 2017.

A large part of the Ph.D. study has been dedicated to system-level simulation development by designing a proper model for new features, implementing them, conducting proper testing and applying it for evaluation. The simulator used is a Nokia Bell Labs proprietary platform. The platform is implemented in object oriented C++ and has been developed for evaluating both LTE and 5G NR. It has has been calibrated against several 3GPP industry members. The simulator includes detailed modeling of RRM mechanisms such as packet scheduling, LA and HARQ. It use industry standard models for propagation, fading and conducts explicit online interference calculations. Parts of the development of the simulator has been done in collaboration with other PhD students which is reflected in the co-author statements. The main contributions to the development of the platform are:

- **GF transmission schemes**: Design, implementation and validation of uplink GF transmission schemes (K-repetitions and Reactive) and early termination of repetitions upon reception of a positive (ACK) feedback (Proactive). At least one of the mentioned schemes are used in Paper A-G.
- **GF radio resource scheduler**: A uplink GF radio resource scheduler which allows overlaying uplink transmissions has been developed and validated for proper capturing of intra- and inter-cell interference and for different receive combining techniques. This scheduler was used in Paper A and C.
- **Enhanced GF radio resource scheduler**: An enhanced uplink GF resource scheduler with support of multiple-MCS options and allocations according to a predefined resource grid of overlapping sub-bands. The scheduler included interfaces for MCS selection algorithms and MCS dependent uplink power control adjustments. This scheduler was used in Paper B, D-G.
- **Frequency hopping**: Sub-band hopping when multiple sub-band options existed. This was used for K-repetitions sub-band frequency hopping as used in Paper B, or for initial and retransmission sub-band hopping, as used in Paper D-F.
- **Uplink power control semi-static offset**: An uplink power control offset used to enforce a higher receive power density target for power boosting retransmissions as used in Paper C, or as a function of the used MCS, as used in Paper D-F.

- **Semi-static transmission adaptation**: Transmission adaptation based on coupling gain and average SINR for uplink GF transmissions was implemented and evaluated in Paper D and E.
- **Multi-cell reception**: Three multi-cell combining techniques was implemented and evaluated; chase-, selection- and hybrid-combining techniques. The three techniques was developed to account for co-located cells. The performance of multi-cell reception and the three schemes was evaluated in combination with different number of receive antennas and HARQ for uplink GF URLLC in Paper F.
- **Multi-cell reception RRM enhancements**: A adaptation of uplink power control (closed loop (CL)) was proposed, implemented and evaluated, along with an MCS adaptation strategy, based on the estimated multi-cell reception signal condition enhancements. Moving average filtered experienced SINR was used as adaptation input. The enhancement schemes are presented in Paper F.
- **Detailed uplink statistics**: Statistics has been implemented for both RAN layer 2 and 3 to properly capture and calculate whether the URLLC requirements are satisfied. Statistics on power overhead and overlapping transmissions has also been added.
- **Uplink URLLC evaluation methodology**: In order to conduct reliable conclusions for the very strict latency and reliability requirements set for URLLC, the evaluation methodology has to be carefully designed, such that a sufficient number of samples can be collected, from a representative set of distributed devices throughout the network, while providing decent simulation execution times. A higher number of deployed devices, would provide a good coverage sample distribution, but is also computational heavy when calculating the mutual generated interference. Further, the packet generation rate per device can be set high to collect many packets per second, but with the drawback of a higher intra-device queuing probability.

6 Thesis Outline

The thesis is structured in six parts and an appendix. The main articles are presented in Part II-V. A visual representation of the included articles and their respective Part is provided in Fig. I.6. Each part contains a brief description of the problem, the main contributions and findings to aid the readers understanding on how the articles relate. Part VI summarizes the main findings and concludes the thesis.

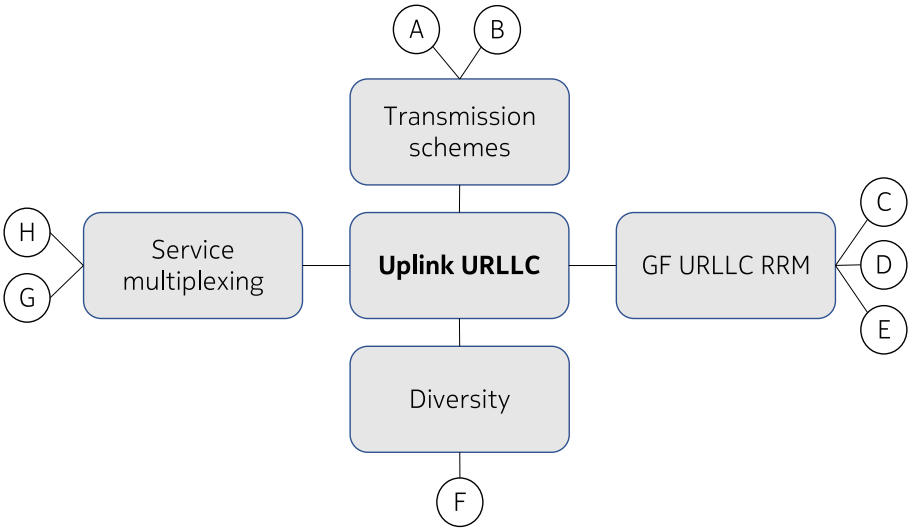


Fig. I.6: Thesis outline and mapping to the included articles.

- **Part I: Introduction -** This part consists of two chaptors. In the present chapter the work is motivated, the main problems is outlined and a description of the contributions is presented. The following chapter, provides an in-depth description and brief analysis of GF and GB transmission schemes for uplink URLLC and their trade-off between latency and achievable URLLC capacity.

- **Part II: Transmission schemes for uplink URLLC -** In this part, transmissions schemes for uplink URLLC are proposed. The schemes are; GF K-repetitions, retransmission-based, a hybrid referred to as proactive and a GB transmission schemes. The schemes are carefully evaluated with the purpose of quantizing their trade-off between latency and URLLC capacity defined as the maximum aggregated offered URLLC load where the URLLC requirements are fulfilled (hence the achievable URLLC spectral efficiency). Detailed analysis of the K-repetition scheme parameter space is included and includes multiple sub-bands and frequency hopping.

- **Part III: RRM for GF URLLC -** In this part, RRM enhancements for uplink GF URLLC are presented and evaluated with the purpose to enhance the URLLC capacity. This includes a detailed study of the uplink power control parameter optimization for reliability enhancements. The parameters for URLLC are quite different from those typically used in LTE for MBB. Then a novel joint resource allocation and MCS selection scheme is presented, which adapts the MCS based on multiple previous

transmission quality measurements. The resource allocation scheme allows transmission with different MCS options to partially overlap GF transmissions and reduce the probability of fully overlapping transmissions in order to increase the URLLC capacity.

- **Part IV: Diversity and multi-cell reception -** This part study the potential of transmission, antenna and spatial diversity for uplink GF URLLC. The latter is achieved by multi-cell reception, which is proposed as a technique to enhance the URLLC capacity. Two multi-cell aware RRM mechanisms are proposed to unleash the full potential of multi-cell reception.
- **Part V: Multiplexing of eMBB and URLLC -** This part address the challenge of achieving efficient multiplexing of URLLC and eMBB on a single carrier. Two sharing options are considered; either both services can simultaneously use the same frequency band with the risk of having transmissions utilizing the same radio resources, or the frequency is split into two dedicated parts. The feasibility of the first option is studied with service differentiated uplink power control and multiple receive antennas at the BS with linear receivers.
- **Part VI: Conclusion -** This part summarizes the main findings from Part II-V and based on these, provides a) an estimation of the relation between latency and achievable spectral efficiency and b) a set of recommendations on how 5G can efficiently support URLLC in the uplink. The conclusion is finalized with an outlook on future work for RRM techniques for uplink GF URLLC.

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] P. Popovski, "Ultra-reliable communication in 5G wireless systems," in *1st International Conference on 5G for Ubiquitous Connectivity*, Nov. 2014, pp. 146–151.

[3] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka *et al.*, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 26–35, 2014.

[4] Deliverable D6.2, "Final Report ? Outcomes, Exploitation and Disseminatio," Jun. 2017.

[5] 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.

[6] 3GPP TR 37.910 v1.0.0, "Study on Self Evaluation towards IMT-2020 Submission," Sep. 2018.

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

[8] 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.

[9] 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.

[10] R1-1812069, "Summary of 7.2.6.1 Remaining details on evaluation methodology," Oct. 2018.

[11] 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.

[12] ITU-T, "Technology Watch Report," International Telecommunication Union (ITU), Tech. Rep., Aug. 2014.

[13] R1-1812110, "LS on TSN requirements evaluation," Nov. 2018.

[14] 3GPP TS 22.186 v15.4.0, "Enhancement of 3GPP support for V2X scenarios," Sep. 2018.

[15] 3GPP TR 38.913 v14.1.0, "Study on Scenarios and Requirements for Next Generation Access Technologies," Mar. 2017.

[16] 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.

[17] R1-1806016, "Evaluation of UP latency in NR," May 2018.

[18] 3GPP TR 36.881 v14.0.0, "Study on latency reduction techniques for LTE," Jul. 2016.

[19] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[20] 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.

[21] 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.

[22] 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.

[23] P. Frenger, S. Parkvall, and E. Dahlman, "Performance comparison of HARQ with Chase combining and incremental redundancy for HSDPA," in *IEEE 54th Vehicular Technology Conference. VTC Fall 2001.*, Oct. 2001.

[24] 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.

[25] 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.

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

[27] 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.

[28] 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.

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

When is grant-free transmission an efficient option?

Throughout this dissertation, four distinct transmission schemes are considered as candidates for uplink URLLC. GF repetition-based (K-repetitions), GF retransmissions-based through HARQ, a K-repetition scheme with feedback for early termination and a GB retransmission-based scheme.

Intuitively, these schemes, should perform differently under comparable assumptions, such that one will achieve a higher URLLC capacity than the others and achieve different latencies. This is also reflected with research question Q1. Further, the configurations of each scheme can be chosen depending on the latency and reliability requirement, which for the sake of generality, may be different from the 1 ms with $1 - 10^{-5}$ reliability. This present chapter provides an introduction into comparing the three main schemes (GF K-repetitions, GF with retransmissions and GB with retransmissions), are provided. Then, an overview of the main latency and reliability influencing factors are provided together with a brief description of how these affect each scheme differently.

1 Transmission scheme latency budgets

The latency components involved in a URLLC uplink data transmission using the GF K-repetition scheme is illustrated in Fig. I.7. The focus here is the medium-access-control layer (MAC) and physical layer (PHY) of the RAN, which are the layers involved in the 3GPP definition of URLLC requirements [1], when assuming that no radio link control (RLC) retransmissions is used. When a packet enters the MAC layer, the first latency block "packet arrival" starts. The packet gets assigned a transmission channel (like stop-and-wait HARQ channel). Depending on the availability of the transmission

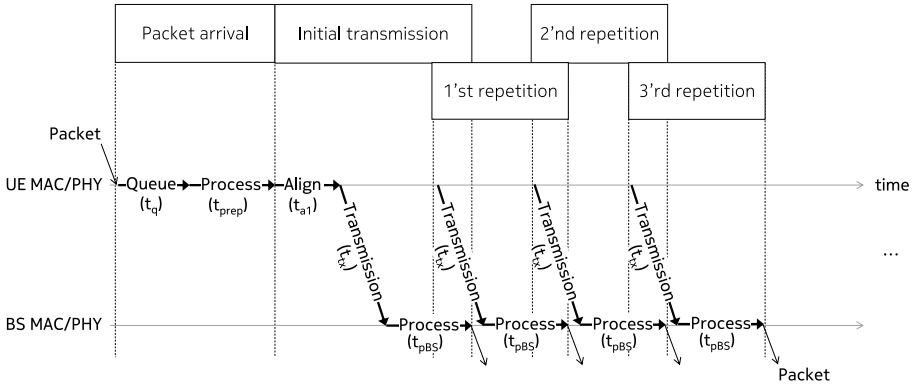


Fig. I.7: Grant-free (GF) K-repetitions latency budget with $K = 4$ repetitions

channel, it will have to wait in a queue (t_{queue}) until it is available, and until a corresponding physical resource allocation and MCS is assigned. Then the packet is prepared for transmission (encoded then modulated and parity bits are added (t_{prep})). For GF transmissions, the physical resource allocation and MCS is pre-configured. It is assumed throughout this thesis, that pre-configured allocations are available in every TTI. The next latency block is the initial transmission and can commence in the start of a TTI. The transmission occupies the entire allocation of an TTI (t_{tx}). At the BS, the transmission is received and processed (t_{pBS}). For K-repetitions, the transmitter does not wait for feedback from the receiver before commencing the next of a total of K repetitions which are transmitted in consecutive TTIs. Each transmission is processed by the BS and combined. When the last transmission is decoded and combined with the previous transmissions, it is determined whether the packet has been successfully received (by a parity bit check). If it is, the packet is forwarded up to the higher layers of the RAN on the BS-side.

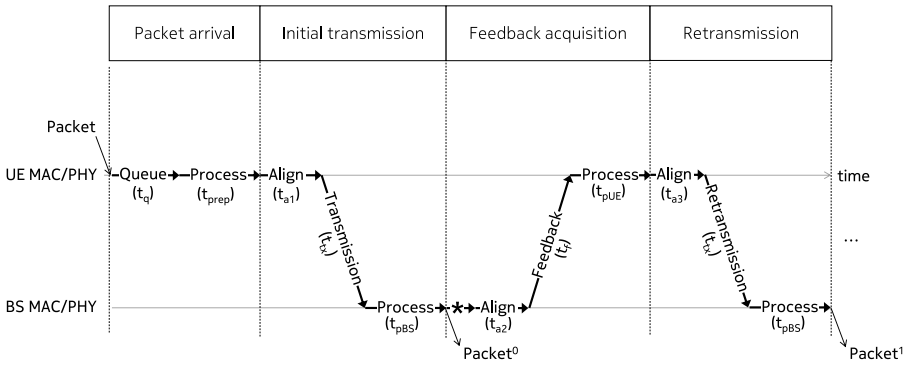


Fig. I.8: Grant-free (GF) retransmission-based latency budget with one retransmission

The latency budget for a URLLC transmission using the GF retransmission-based scheme is illustrated in Fig. I.8. The two first latency blocks "Packet arrival" and "Initial transmission" are the same as for the K-repetition scheme. The latency block "Feedback acquisition" is triggered if the initial transmission is not successfully received. In that case, the BS can attempt to schedule a retransmission if a GB retransmission is used. This latency component is marked with a $*$ and can be omitted if GF retransmissions is used. The reception feedback is transmitted back to the device (UE) in the next TTI (t_f after t_{a2}). The feedback is processed by the device (t_{pUE}). A "retransmission" latency block can then commence in the next TTI after another alignment (t_{tx} after t_{a3}). The retransmission is processed and combined with the first transmission (t_{pBS}) and forwarded to the higher layers if successfully decoded.

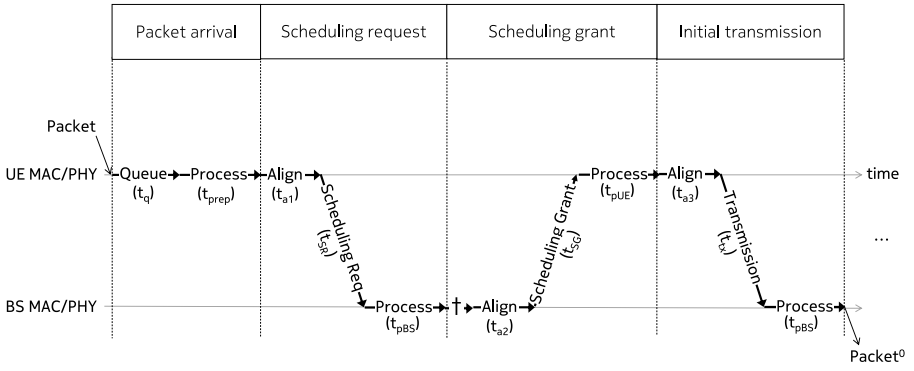


Fig. I.9: Grant-based (GB) retransmission-based latency budget

The latency budget for a GB retransmission-based scheme is illustrated in Fig. I.9. Prior to the initial transmission block (which is assumed to consist of similar components and each of similar size as the equivalent from the GF schemes, two new latency blocks needs to be completed first, which are the "scheduling request" and "scheduling grant". Upon the packet arrival and initial preparation and alignment, a scheduling request is transmitted by the device (t_{SR}). The BS processes the request (t_{pBS}), schedules the initial transmission for the device (with a latency contribution marked with), and transmits a scheduling grant to the device (t_{SG}). The device processes and prepares its initial transmission which can commence at the allocated resources after t_{a3}. Notice that this alignment can be longer than a TTI, if the allocation is not in the consecutive TTI. This can be the case due to the BS packet scheduler, which might not be able to fit all requested transmissions in earliest possible TTI.

2 Latency budget comparison

In this part, the latency assumptions from [2] are used for the sake of comparison with the assumptions used in Paper B and Paper E. As a reference, the latest assumptions for 5G NR are provided in [3].

Table I.2: Latency component assumptions from [2].

Latency component	Value
t_{symbol}	0.036 ms
t_{TTI}	$4 \cdot t_{symbol} = 0.143$ ms
t_{tx}	t_{TTI}
t_f	t_{symbol}
t_{pBS}	$3 \cdot t_{symbol}$
t_{pUE}	t_{pBS}
t_{SR}	t_{TTI}
t_{SG}	t_{symbol}
GF transmission opportunity	Every TTI
Scheduling periodicity	Every TTI
Scheduling decision	$3 \cdot t_{symbol}$

The used latency assumptions are summarized in Table I.2, and the latency for the transmission schemes are illustrated in Fig. I.10. To get a realistic latency budget estimate at the 10^{-5} quantile, maximum initial transmission alignment time t_{a1} of the packet to the MAC layer is assumed. This occurs when t_{prep} is finished right after the start of a TTI.

It is observed that for latency requirement below 1 ms, GF K-repetitions with $K < 4$, and GF with maximum 1 retransmission are feasible options. If the retransmission is assumed to be scheduled, the additional latency component renders GF with 1 retransmission unable to stay within 1 ms. Increasing the latency requirement to 1.4 ms means that GB with maximum 1 retransmission becomes a feasible option, as well as GF with a maximum of 2 retransmissions and K-repetitions with $K \leq 7$. With GF and scheduled retransmissions only 1 retransmission fits into this example 1.4 ms latency requirement. This calculation does not account for the additional latency from queuing, and only partly accounts for the reliability dimension of the feasible schemes. A latency comparison like this is therefore not sufficient to select the most efficient scheme based on the latency requirement. In Fig. I.10, the best initial guess of latency thresholds where it will be more efficient to change from GF repetitions-based to GF retransmission-based ($L_{feedback}$) then to GB retransmission-based (L_{GB}) are marked. The initial guess of these thresholds are to set them where the retransmission-based schemes can fit a retransmis-

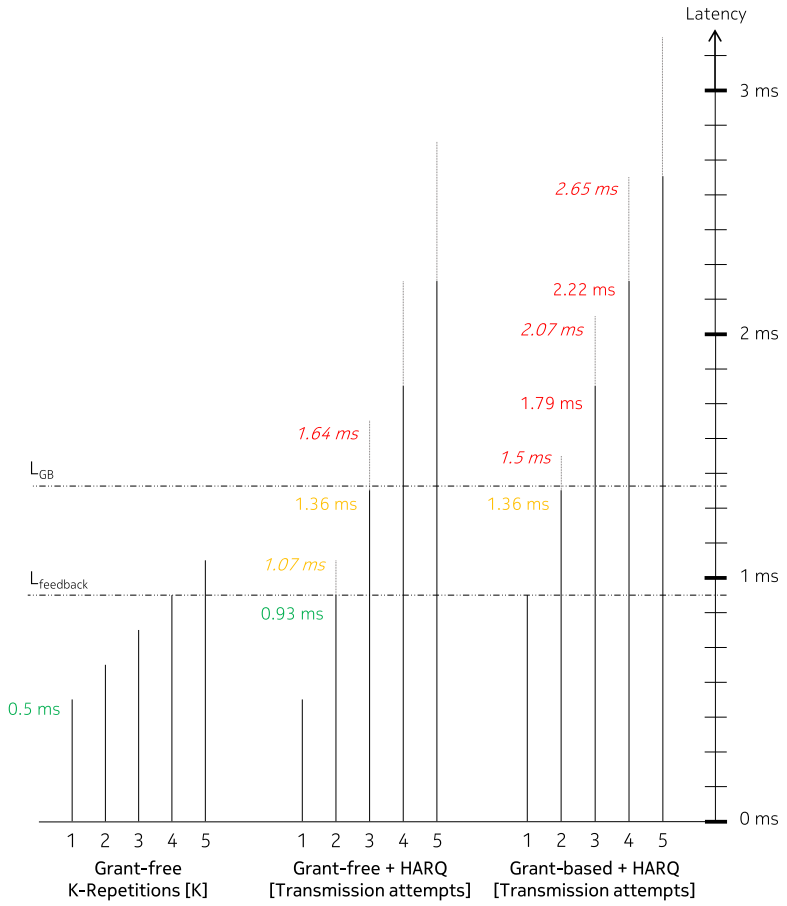


Fig. I.10: Transmission scheme latency comparison with possible scheme selection latency thresholds $L_{feedback}$ and L_{GB}. Latencies which includes scheduling decision times are written in *italic*

sion in the latency requirement. However, this pen-and-paper exercise does not capture all factors which influence the achieved reliability and latency which can affects these thresholds. The next section will discuss these major influencing factors.

3 Latency and reliability influencing factors

A brief overview of the latency and reliability influencing factors for each of the considered transmission schemes is summarized in Fig. I.11.

Interference is a known limitation for the capacity of wireless cellular networks. Both GF schemes are subject to intra-cell interference as no coordi-

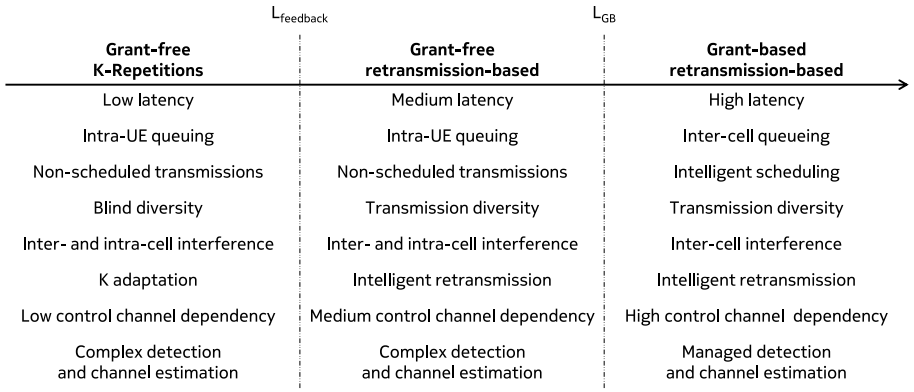


Fig. I.11: Summary and comparison of latency and reliability influencing factors

nation is carried out between devices in the network. GF K-repetitions relies on transmitting K times, and hence the generated interference scales with the choice of K. Both retransmission based schemes (GF and GB) rely on feedback to determine whether a retransmission should be carried out and therefore only generates additional interference when needed. The GF schemes are, however, subject to intra-cell interference, but may rely on granted retransmissions to avoid this on the retransmissions, at the cost of additional scheduling delay and the need for a granted transmission band which either requires more bandwidth or excluding radio resources which could otherwise have been used for GF transmission. The consequences of the latter is either an increased latency budget (e.g. from alignment or queuing) or reduced transmission bandwidth.

Queuing is an important factor in URLLC latency evaluations. At the 10^{-5}- quantile, even an otherwise small probability of queuing can have a considerate latency contribution. Queuing occurs when the RAN is unable to serve arriving URLLC packet immediately. This can be when the BS cannot schedule all requested URLLC packets in the next possible TTI or when the device has more packets than configured GF transmission opportunities. The queuing probability and latency contribution scales with rate of generated URLLC data packets. For the GF schemes, the queuing probability scales with the generated traffic per device (intra-device), whereas for the GB scheme the queuing probability scales with the accumulated traffic for all served devices as they need to request resources prior to transmitting the URLLC payload.

Adaptation-capability of the transmission parameters to match the experienced channel conditions is another difference between the three transmissions schemes. The experienced channel conditions change due to the time- and frequency-varying fading channel and due to interference variations. On

top of fading, the GF transmission schemes using shared radio resources, are subject to fast varying intra-cell interference causing a so-called flash-light effect. The retransmission-based schemes has the benefit of being able to trigger a retransmission when the initial transmission fails and can, even if scheduled retransmissions are used, configure the transmission parameters for the retransmission dynamically. This is not an option for the K-repetitions scheme, which must rely on semi-static adaptation of the number of repetitions K. Adaptation of transmission parameters for the GF schemes using shared radio resources are limited to a semi-static, whereas the GB can adapt dynamically. There are two primary reasons for this; firstly, the contribution of the strong intra-cell interference should be taken into account to avoid over- or under-compensation with the cost of reliability losses or spectral efficiency degradations and secondly, wide-band channel quality estimation requires coordinated uplink sounding transmission.

Diversity is essential to reach high reliability. One of the most essential diversity technique for URLLC is frequency diversity [4, 5]. Frequency diversity utilize the independent fading channels (defined by the coherence bandwidth) across the available frequency bands and can be exploited with wide-band transmissions. On the other hand, reducing the transmission band, allows more transmissions in the same bandwidth. GF schemes can utilize a reduced transmission band to reduce the probability of collisions, while GB can utilize it to minimize the queuing effect. K-repetitions can even form frequency hopping patterns, to combine frequency diversity and interference-diversity into what we have called blind-diversity in Fig. I.11. Retransmissions allows for transmission diversity, if the initial transmission fails, e.g. due to fading or interference. Another important diversity technique is spatial diversity, which can be achieved by applying multiple receive antennas. Multiple receive antennas allow the receiver to capture more of the energy and exploit the fading differences between the antennas [6]. Increasing the number of receive antennas can therefore improve the received SNR and improve the network coverage [6]. Further combining across multiple receive antennas with interference awareness, can be used to improve the SINR of the desired transmission [7]. The latter is particular important for GF schemes when sharing of the transmission opportunities across multiple devices is utilized and the former property is an important property for GB scheduling.

4 Take-aways and outlook

The take-aways from this chapter are summarized as:

- The transmission scheme which can achieve the highest URLLC capacity and spectral efficiency depends on the latency requirement.

- Even small differences, such as scheduling decision delays for scheduled retransmissions may cause a transmissions scheme to violate the latency requirements.
- Evaluation of the transmission schemes should include as many realistic performance determining factors as possible that affects the latency and reliability, such as intra- and inter-cell interference, device and BS queuing, receiver capture effects and RRM mechanisms to ensure that solid conclusions can be drawn.
- Awareness of the choice of assumptions and how they affect each scheme differently is necessarily to make generalized conclusions.

Based on these take-away messages, Part II evaluates the proposed transmission schemes for 5G NR with the purpose of proposing the most efficient strategies for uplink URLLC traffic. In Part III, novel RRM enhancements designed for GF transmissions, to mature the judgment of the feasibility of GF transmission schemes for uplink URLLC are provided.

References

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

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

[3] R1-1808449, "IMT-2020 self-evaluation: UP latency analysis for FDD and dynamic TDD with UE processing capability 2 (URLLC)," Aug. 2018.

[4] P. Popovski, "Ultra-reliable communication in 5G wireless systems," in *1st International Conference on 5G for Ubiquitous Connectivity*, Nov. 2014, pp. 146–151.

[5] 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.

[6] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency," *Foundations and Trends® in Signal Processing*, vol. Nov., no. 3-4, pp. 154–655, 2017. [Online]. Available: http://dx.doi.org/10.1561/2000000093

[7] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

Part II

Transmission schemes for Uplink Ultra-Reliable Low-Latency Communications

Transmission schemes for Uplink Ultra-Reliable Low-Latency Communications

In this part of the dissertation, uplink transmission schemes for Ultra-Reliable Low-Latency Communications (URLLC) are proposed.

1 Problems and solution space

Compared to previous generations, fifth generation (5G) New Radio (NR) has enabled new transmissions schemes to reduce the latency such as grant-free transmissions on pre-allocated radio resources, and enhance the reliability by transmission repetitions and retransmissions. The choice of transmission scheme is important as this dictates the achievable latency, reliability and maximum spectral efficiency. The following transmission schemes are considered for uplink URLLC:

- Repetition-based grant-free (GF) (also denoted GF K-repetitions), where up to K repetitions are carried out in consecutive slots. The repetitions can be exact replicas for soft-combining support.
- Repetitions-based GF with early termination is similar to the repetition-based GF, but listens for transmission feedback to possible stop the series of repetitions before reaching K. This scheme is also denoted GF proactive.
- Retransmission-based GF. After an initial GF transmission, the device will wait for feedback from the serving base station (BS) before carrying out a GF retransmission. This scheme is also denoted GF reactive.
- Retransmission-based grant-based (GB). A scheduling procedure (schedul-

ing request in uplink and a scheduling grant in downlink) is carried out prior to the initial URLLC transmission. Retransmissions are triggered with a feedback message and carried out on granted resources.

With the GF transmissions, the following parameters needs to be pre-configured a-priori; physical radio resources allocation, modulation and coding scheme (MCS) and allocation periodicity. As a typical assumption throughout this dissertation, all devices use the same MCS, which is chosen to maximize the transmission bandwidth in the uplink band for maximum frequency diversity [1]. The uplink URLLC traffic is sporadic which means that the URLLC data arrival is unpredictable. To minimize the probability of queuing and to improve the resource utilization, all devices are configured with the same pre-configured radio resources reoccurring every transmission time interval (TTI).

With the K-repetitions, there can be a benefit of reducing the transmission bandwidth, to form multiple sub-bands. The choice is a trade-off between frequency diversity by using wide-band transmissions and interference diversity by using multiple sub-bands in the frequency domain to fit multiple transmissions in the frequency domain. Increasing the number of repetitions K or increasing the transmission power is needed to compensate for the increased energy per bit requirements with the higher-order MCS when using smaller sub-bands. Increasing K leads to a increased probability of queuing and increasing the transmission power increases the generated interference in the network.

Prior-art is found in [2, 3], where uplink GF transmission schemes are evaluated. They, however, use a simplified set of assumptions such as a single-cell environment, idealized uplink power control and a simplified receiver. Another example is found with Paper I in this dissertations appendix, where both a probabilistic and signal to interference-and-noise ratio (SINR) based model is used to study the performance of contention-based transmission schemes in the scope of massive Machine Type Communication (mMTC). As outlined in Part I, an evaluation of these transmission schemes should capture the main latency and reliability influencing factors to draw solid conclusions. These factors are inter- and intra-cell interference, a multi-user multi-cell network, hybrid automatic repeat request (HARQ) with realistic latency components and a linear minimum mean square error (MMSE) receiver. The combination of these factors, is to the best of our knowledge, not used to quantify the performance of each of the considered transmission schemes before.

2 Objectives

Based on the problem description and identified solution space, the following research objectives are targeted with this part of the dissertation:

- Quantify the trade-off between latency and maximum aggregated URLLC traffic (URLLC capacity) where the URLLC service requirements are fulfilled for the proposed transmission schemes.
- Quantify the trade-off between wide-band transmissions and transmission using smaller sub-bands for GF K-repetitions.
- Evaluate under which assumptions and latency requirements GF transmission schemes can achieve a higher URLLC capacity than GB transmission schemes.

As mentioned in Part I, a realistic performance evaluation of transmission schemes for URLLC should capture as many of the latency and reliability influencing factors as possible. In order to capture the effects of propagation, fading and interference in a multi-user multi-cell environment and the radio access network (RAN) medium-access-control layer (MAC) and physical layer (PHY) layer procedures, which includes radio resource scheduling, transmission repetitions and retransmissions, the performance evaluation is performed using advanced system level simulations. The simulator includes detailed modeling of the radio resource management (RRM) techniques such as uplink power control, scheduling, transmission adaptation and receiver combining algorithms. The simulator uses industry standard models for propagation, fading and receiver capabilities to provide a realistic performance evaluation which includes the aspects of intra-cell and other-cell interference in a multi-user multi-cell 5G NR network.
Addressing the research objectives has lead to the following sub-objectives:

- Design and implement the considered GF and GB transmission schemes in an advanced state-of-the-art system level simulator which properly captures the effect of propagation, interference and fading, while modeling the major latency and reliability impacting factors in the RAN MAC and PHY layer procedures.
- Design a suitable system level evaluation methodology for uplink URLLC, which ensures that statistically reliable conclusions can be made.
- Establish a performance baseline for GF and GB uplink URLLC transmission schemes in a 5G NR compliant network, using advanced system level simulations.

3 Included Articles

The main findings of this part are disseminated in the following articles:

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

This article propose uplink URLLC transmission schemes and studies their performance in terms of the latency, URLLC capacity, and channel usage. The following GF transmission schemes are considered: GF K-repetitions, GF retransmission-based and a combination of the two denoted GF proactive. The GF schemes are compared with a simple GB transmission scheme. The performance evaluation is carried out using an advanced system level simulator and state-of-the-art 3rd Generation Partnership Project (3GPP) NR Release 15 latency assumptions are considered.

Paper B. System Level Analysis of K-Repetition for Uplink Grant-Free URLLC in 5G NR

This paper studies the design of a GF K-repetition transmission scheme and the achievable URLLC capacity in a 5G NR network. That is; the used MCS, the number of transmission repetitions K and the receiver power density target set by uplink power control. Selection of these parameters represents a multiple trade-offs; while higher order MCS allows more sub-bands and reduces the probability of overlapping transmissions, the increased MCS order also require a higher receive power density to maintain the average transmission reliability. Another option is to increase the number of repetitions and increases the blind-diversity, but it also increases the generated interference and the probability of a packet being queued.

4 Main Findings

The main findings from the articles are:

GF retransmissions-based schemes is more efficient than GF repetition-based schemes, when the latency budget allows.

Paper A shows how the GF retransmission-based scheme is capable of reaching higher URLLC capacity than the GF repetition-based scheme. That is, when the latency with one retransmission is within the latency requirement. This tendency is supported in Paper B, where the GF K-repetition scheme is explored. Despite a significant improvement to comparable performance

with the retransmission-based scheme, by utilizing power control optimizations, as will be further introduced in Part III. However, it is also observe that the K-repetitions transmission scheme is subject to queuing which limits its achieved URLLC capacity. It is observed that GF retransmission-based can achieve a URLLC capacity of 0.26 Mbps while fulfilling the 1 ms latency requirement. K-Repetition with $K = 2$ can achieve 0.05 Mbps but also with a latency of 0.7 ms latency due to queuing.

The lowest outage-probability (the compliment of the reliability) achieved with K-repetitions, was found when introducing a higher-order MCS and multiple sub-bands were formed. Hopping between the sub-bands achieves blind-diversity (both interference and frequency diversity by blindly utilizing the shared channel). Fixing the latency budget by K, a too high-order MCS and hence more sub-bands, was harming the outage probability and a too low-order MCS was similarly harming the outage probability. Increasing the number of sub-bands, means more interference diversity can be achieved, but it also needs a larger K is needed for frequency diversity and to acquire sufficient energy per bit for a successful decoding.

The findings presented in Paper A and Paper B are supported by the recent work presented in [4] which presents a capacity analysis using a simpler receiver type (maximal-ratio combining (MRC)) on a interference channel with channel usage probabilities. More recent work presented in [5], conducts an information theoretic analysis on the latency and capacity trade-off when using fixed length transmissions and retransmission-based (HARQ). Despite both analysis are based on a simplified single-cell scenario, the results are in-line with our the findings.

GB requires a relaxed URLLC latency requirement

Paper A clearly demonstrates that the GB transmission scheme, has a significant worse latency to reliability tendency than the GF schemes. At equivalent URLLC aggregated traffic loads, of which GF can satisfy the 1 ms latency requirement with $1 - 10^{-5}$ reliability, the GB needs between 50-100% larger latency budget to reach the $1 - 10^{-5}$ reliability. In Part I, it was predicted that the latency requirement increase could be in the order of 40%. This estimate did not consider the aspect of reliability or latency from queuing.

5 Recommendations and follow-up studies

Based on the presented findings, the following recommendations are made:

- GF retransmission-based schemes achieve the highest URLLC capacity of 0.26 Mbps when the URLLC latency requirement is 1 ms and should be used as the baseline uplink URLLC transmission scheme.

- GF K-repetitions is recommended when the latency requirement is below 1 ms. With $K = 2$ a 0.7 ms latency requirement can be satisfied with a URLLC capacity of 0.05 Mbps. Tuning of the sub-band size and uplink power control parameters is recommended to maximize the gain from blind-diversity.
- GF transmission schemes is preferred over GB transmission schemes for sporadic uplink URLLC traffic when the URLLC latency requirement is in the order of 1 ms.

The findings presented in this part of the dissertation indicates that the GB transmission scheme is severely impacted by queuing. Techniques to minimize this effect could improve the GB scheduler performance with sporadic URLLC traffic. Further, our findings indicates that there is room for RRM enhancements for GF transmission schemes to make them capable of reaching even higher URLLC capacities. Such schemes should include optimized uplink power control and GF MCS selection strategies.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] J. Östman, R. Devassy, G. C. Ferrante, and G. Durisi, "Low-Latency Short-Packet Transmissions: Fixed Length or HARQ?" *CoRR*, vol. abs/1809.06560, Oct. 2018. [Online]. Available: http://arxiv.org/abs/1809.06560

Paper A

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 GlobeCom Workshops, 2017

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.

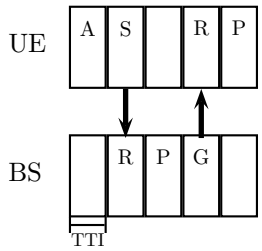


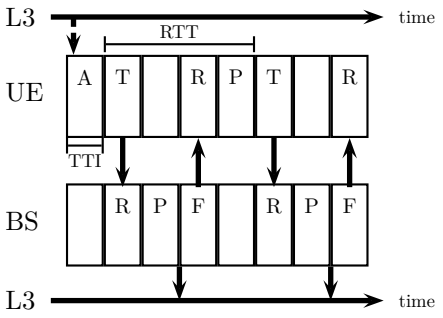
Fig. A.1: Scheduling request model used for Grant-Base access. Legend: A = Frame alignment, S = Scheduling Request, R = Reception, P = Processing, G = Scheduling Grant.

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. A.1 and with the conventional HARQ scheme (reactive Fig. A.2(a)).

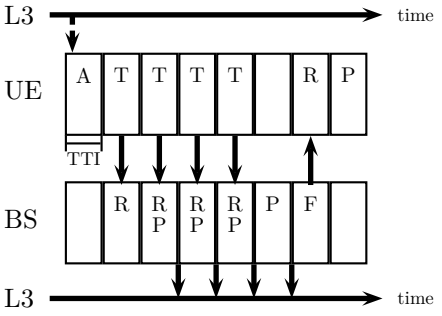
When using the GB approach, each UL transmission is coordinated by the base-station (BS). Upon a packet arrival on layer 3 (L3), a UE waits for the next subframe occurrence for transmitting a scheduling request (SR) signal (S). After processing the SR, the BS transmits a scheduling grant (G) which indicates 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.

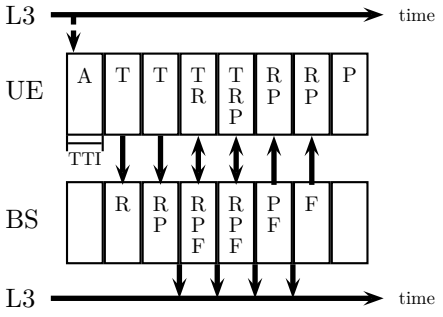
Three HARQ schemes are considered for GF transmissions, namely a Reactive, K-Rep and Proactive scheme. The Reactive scheme is illustrated in Fig. A.2(a). When the UE has finalized its initial uplink data transmissions, its signal is processed at the BS, which will transmit a positive or negative acknowledgement. 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. A.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. A.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. A.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 A.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 A.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.

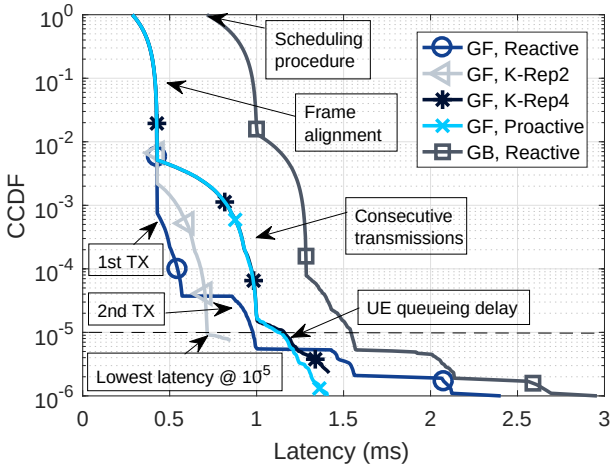
The adopted Minimum Mean Square Error Interference Rejection Combining (MMSE-IRC) receiver is assumed to be able to ideally estimate the interference covariance matrix for suppressing intra-cell and inter-cell interference. Given the 2 receive antennas, up to one interfering stream can be suppressed. 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 $\approx 10^{-6}$.

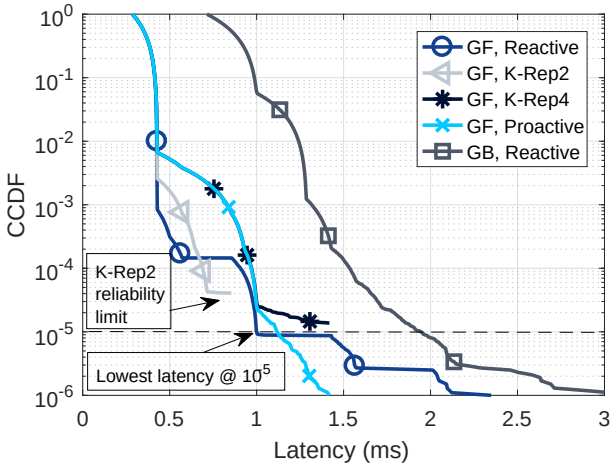
4 Results

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. A.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. A.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. A.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 A.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 A.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

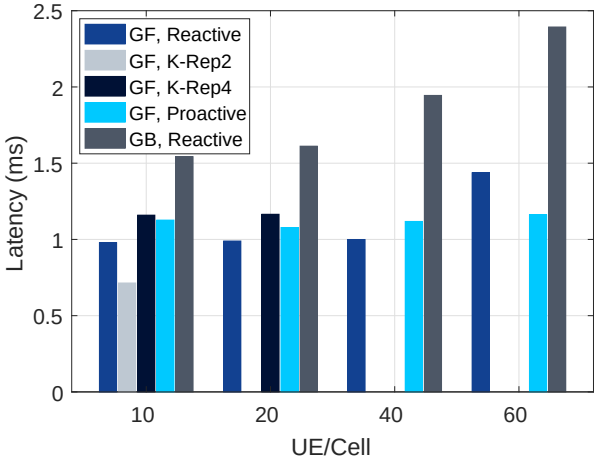


Fig. A.4: Achieved latency at $1 - 10^{-5}$ reliability as a function of load.

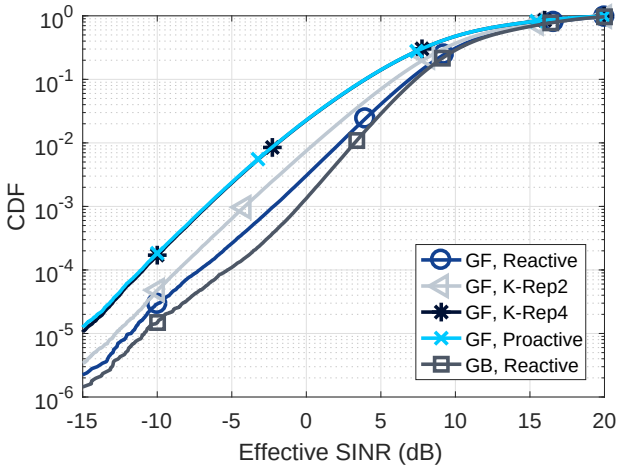


Fig. A.5: Average effective SINR per RB for GF and GB (40 UE / cell).

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 A.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-

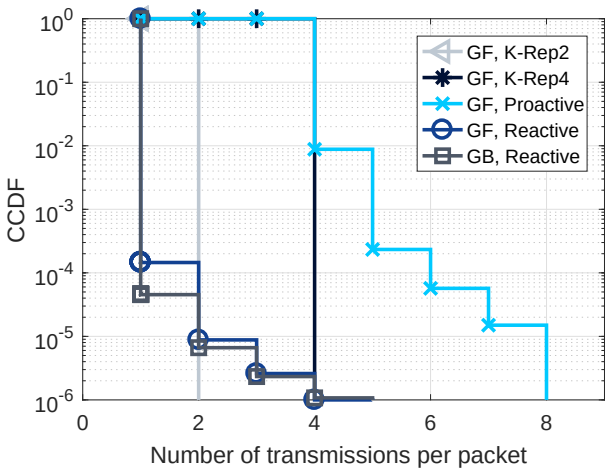


Fig. A.6: CCDF of the number of transmissions per packet (40 UE / cell).

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.

Paper B

System Level Analysis of K-Repetition for Uplink Grant-Free URLLC in 5G NR

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

The paper has been accepted for publication in
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. B.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].

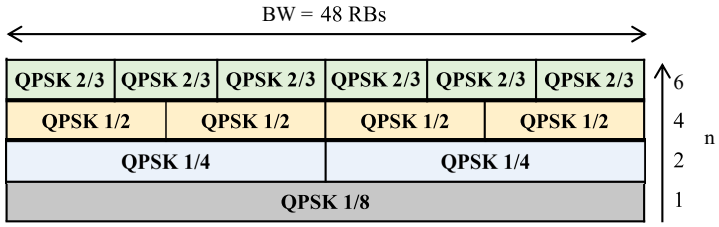


Fig. B.1: Examples of radio resource allocations of n sub-bands and corresponding MCS over BW RBs [15].

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

$$t_k = t_{queue} + t_{prep} + t_{align} + k \cdot t_{TTI} + t_{proc}. \tag{B.1}$$

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 (B.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. B.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. B.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. B.3 we observe that the collision probability is reduced when $K > 1$ and $n > 1$.

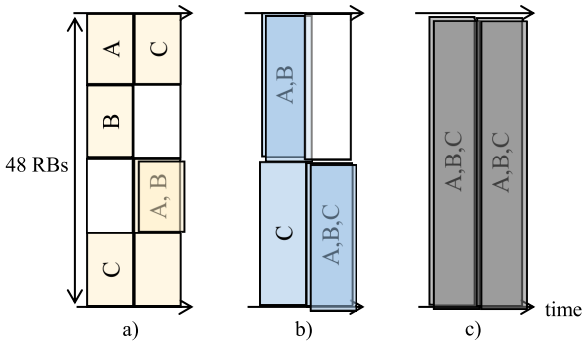


Fig. B.2: Realizations of GF transmissions with a) $n = 4$, b) $n = 2$ and c) $n = 1$ sub-bands over K repetitions using sub-band hopping for UE A, B and C.

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

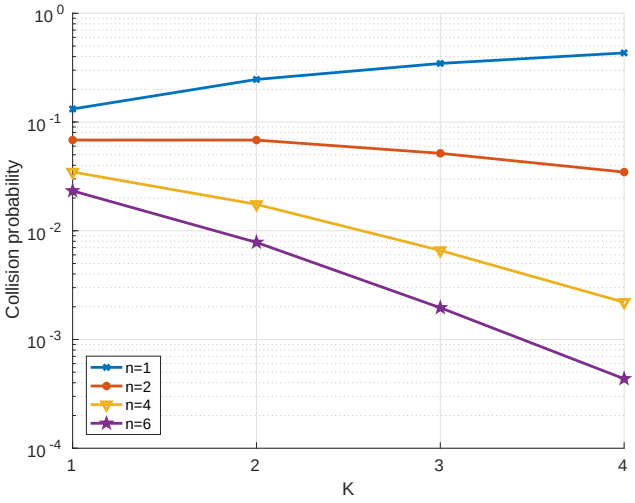


Fig. B.3: Collision probability as a function of the number of sub-bands n and repetitions K using (9) from [7]. The load is given by $N = 100$ UEs with $\lambda = 10$ PPS.

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 B.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 B.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. B.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. B.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. B.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

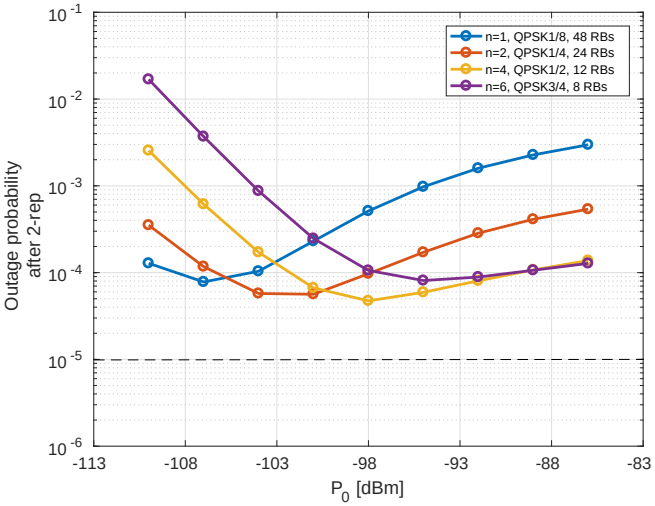


Fig. B.4: Sensitivity of outage probability in relation to P_0 and n for $K = 2$.

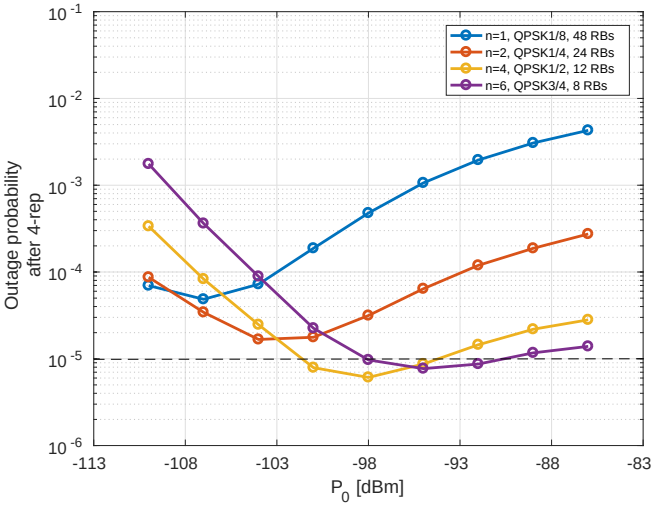


Fig. B.5: Sensitivity of outage probability in relation to P_0 and n for $K = 4$.

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. B.7 shows the complementary cumulative distribution function (CCDF)

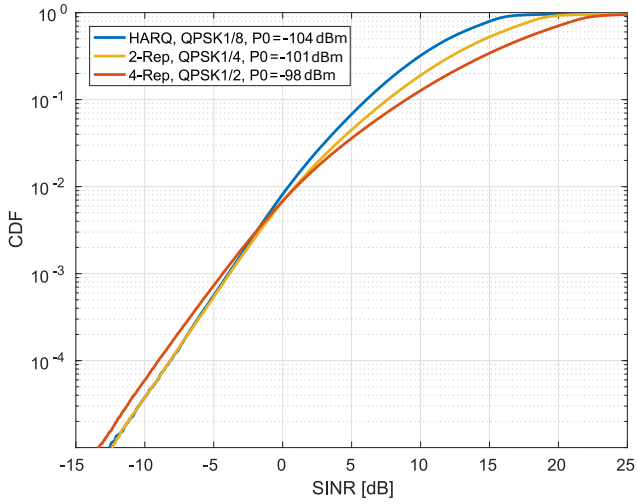


Fig. B.6: CDF of the SINR for the different schemes.

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. B.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.

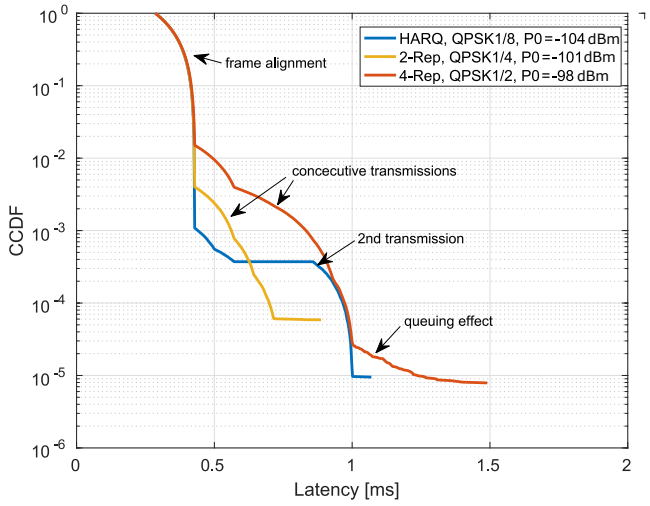


Fig. B.7: Complementary cumulative distribution function of the latency for K-repetitions with $K = 2$, $K = 4$ and the HARQ baseline ($L = 0.256$ Mbps).

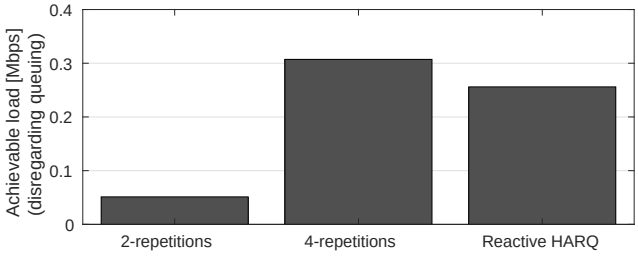


Fig. B.8: Maximum loads supported with $K = 2$, $K = 4$ and reactive HARQ, neglecting queuing delays.

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.

Part III

Grant-free radio resource management enhancements

Grant-free radio resource management enhancements

This part of the dissertation studies radio resource management (RRM) mechanisms to enhance the network performance when serving sporadic uplink Ultra-Reliable Low-Latency Communications (URLLC) traffic over shared grant-free (GF) radio resources. This includes uplink power control optimizations, a power-based retransmission boosting mechanism and a novel joint resource allocation and modulation and coding scheme (MCS) selection strategy.

1 Problems and solution space

GF on shared radio resources is identified as a key enabler for uplink URLLC, by mitigating the latency consuming scheduling procedure. However, the sharing of radio resources between devices means that transmissions becomes prone to interference from overlaying transmissions, which harms the achievable capacity and hence the spectral efficiency. The combination of the challenge of serving URLLC services with previously unprecedented requirements and the presence of intra-cell interference, means that conventional RRM mechanisms used to enhance the network performance used in Long Term Evolution (LTE) and fifth generation (5G) New Radio (NR) might require rethinking when considered for uplink GF URLLC.

One of these mechanisms is uplink power control, which has been used in LTE and is specified for 5G NR in Release 15 [1]. Uplink power control is a mechanism used to regulate the target signal strengths and to manage the generated intra- and inter-cell interference. Uplink power control achieves this by providing a set of rules used by the device to set the transmission power. Fractional path loss open loop power control has proven beneficial for mobile broadband (MBB) [2], but it remains to be understood how it may benefit the URLLC capacity with sporadic uplink traffic served using a GF transmission scheme over shared radio resources.

In order to increase the URLLC capacity for GF, techniques that does not imply larger latency but improve the reliability are desired. One such idea is to boost the transmission power of the retransmission. As a retransmission only occur about 0.1% of all URLLC transmissions (in order to reach the reliability target of $1 - 10^{-5}$ as observed in Paper A), the additional transmission power should have an small contribution to the reliability of other transmissions.

Another idea originates from adaptive modulation and coding (AMC) as used in LTE which utilize fast acquisition of channel state information (CSI) to dynamically adapt the MCS to the instantaneous channel conditions [3]. However, as described in Part I, adaptation of transmission parameters for uplink GF on shared radio resources needs to be semi-static due to the sporadic intra-cell interference, the lack of coordinated uplink sounding transmissions and the lack of a scheduling grant. Devices might, however, still experience different long-term signal conditions, which can be exploited with semi-static adaptation of the MCS.

It was demonstrated in Paper B that increasing the MCS-order can lead to interference diversity by reducing the probability of overlaying transmission (in the same frequency band). However, increasing the MCS also requires an increase in the needed energy per bit to maintain the transmission reliability. It may, however, not be all devices which have the required power headroom to meet the target energy per bit with the higher order MCS. These devices could therefore be better off by using a lower order MCS. Further, a higher order MCS is also less robust to overlaying transmissions, and hence needs an additional power or bandwidth margin to sustain the transmission reliability [4]. A practical scheme which allows GF transmissions to overlay each other in the frequency domain and overlaying with different MCS options selected based on the their long-term experienced coverage conditions, could have potential to increase the URLLC capacity with sporadic traffic transmitted over shared GF radio resources.

2 Objectives

The research objectives of this part are:

- Study the optimal uplink power control parameters for sporadic uplink URLLC traffic using shared GF radio resources.
- Quantify the potential of power boosted retransmissions for uplink GF in terms of URLLC capacity enhancements.
- Study the URLLC capacity gain potential of a MCS selection scheme for uplink GF URLLC transmissions.

The following sub-objectives have been identified to address the main research objectives:

- Design and implement adjustable power boosting levels of GF retransmissions in the system level simulator.
- Implement device power headroom statistics for both initial GF transmission and retransmissions in the system level simulator.
- Conduct a sensitivity study of the achievable URLLC capacity increase using power boosted retransmission at different levels.
- Design and implement a GF transmission resource allocation scheme with multiple MCS support and which is capable of adapting the MCS based on long-term reliability or signal to interference-and-noise ratio (SINR) statistics.

3 Included Articles

The main findings of this part are included in the following articles:

Paper C. Power control optimization for uplink grant-free URLLC

In this article open loop power control parameters for uplink GF URLLC is studied and recommendations are provided. The retransmission-based GF transmission scheme as presented in Paper A is used. With this GF scheme, a sensitivity study is conducted on the open loop power control parameters; the path loss compensation factor α and the receive power density target P_0. A retransmission power boosting mechanism is presented and the potential URLLC capacity gain is evaluated with different power boosting levels. The performance evaluation provides detailed device power headroom statistics. The evaluation is conducted using the same full-blown system level simulator as used in Part II.

Paper D. Joint Resource Configuration and MCS Selection Scheme for Uplink Grant-Free URLLC

In this article a novel joint resource configuration and MCS selection scheme for GF URLLC is presented. The frequency domain multiplexing is structured in layers of the MCS-order, with the intention to bound the blind decoding complexity at the receiver and to decrease the required control channel (CCH) signaling when an MCS change is signaled. The proposed MCS selection scheme is based on parameterized coupling gain thresholds, which is found by examination of reliability statistics. A MCS dependent power control term is introduced. Insights into the relation between the outage

probability and the coupling gain based on detailed analysis using system level simulations are provided. System level simulations are used to ensure that the effect of intra- and inter-cell interference and the capture effect at the receiver when transmissions partially overlap in the frequency domain is captured in the evaluation.

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

In this journal paper we build on top of the findings from Paper A, C and D and propose a RRM framework for sporadic uplink GF URLLC. The framework includes the GF retransmission-based transmission scheme, open loop power control, the joint resource configuration and MCS selection scheme with minimum mean square error (MMSE)-interference rejection combining (IRC) linear receiver. The proposed joint resource allocation and MCS selection scheme from Paper D is enhanced with a simplified MCS selection threshold criterion using the average SINR instead of reliability. Insights on the trade-offs of using partial overlapping transmissions are provided along with an investigation of the benefits of increasing the number of receiver antennas. Lastly a reevaluation of the achievable URLLC capacity with GF and grant-based (GB) transmission schemes are carried out, where the proposed RRM framework is applied. An enhanced GB scheduler is presented which multiplex requested transmissions on orthogonal sub-bands in the frequency domain to minimize the queuing probability.

4 Main Findings

The main findings from this part of the dissertation are summarized as:

Open loop power control with full path loss compensation

It is found that open loop power control with full path loss compensation can reach a slight improvement in the outage probability and URLLC capacity gain of 25% when compared to fractional path loss compensation. Further we find that the achievable URLLC capacity is very sensitive to the selection of P_0. The optimum choice of P_0 depends on the used transmission scheme, the aggregated URLLC load, the deployment scenario and the number of receive antennas.

Retransmission power boosting is limited by the device power headroom

It is observed that the achievable URLLC capacity can be enhanced by applying power boosting by up to 20%. Of the devices which conducted a retrans-

mission, only 69% was able to apply up to 3 dB power boost and 65% were unable to apply any power boosting at all. The power headroom is observed to be smaller for retransmissions when compared to the initial transmissions, indicating a correlation between power limitation and the probability to retransmission. When a device uses maximum transmit power, it is unable to apply the transmit power needed to compensate the path loss to the receiver, hence its received signal will be weaker, on average, than signals from non power limited devices.

Partially overlapping GF transmissions in the frequency domain can greatly improve the URLLC capacity

An RRM mechanism based on a strategic radio resource allocations for simultaneous support of multiple MCS options, is observed to reduce the probability of fully overlapping transmissions and improve the URLLC transmissions reliability significantly. This sub-band structure is illustrated in Fig. 5 in Paper E, where N MCS options is supported and each with k sub-bands and a received power spectral density offset Δ_{MCS}. The achieved URLLC capacity with joint parameter optimization is observed to be 0.5 Mbps per cell which corresponds to a gain of 90% compared to the GF baseline relying on a single MCS configuration. Further, it is observed that both transmissions using a higher and lower order MCS benefits with the proposed scheme. A large performance impact is observed by the schemes ability to handle bursts of traffic where multiple transmissions overlap.

Multi receive antennas may significantly improve URLLC performance

Spatial diversity by multiple receive antennas was expected to not only increase the coverage, but also improves the receive capability to receive overlapping transmissions. A large URLLC capacity gains by up to 700% when increasing from two to four receive antennas is observed. This will be further examined in Part IV of this dissertation.

The enhanced GB retransmission-based scheme can be superior to the GF retransmissions-based scheme when the latency requirement is relaxed.

Based on the RRM enhancements presented in this part of the thesis, we extend on the findings from Part II. The enhancements include the uplink power control optimization methodology and the joint resource configuration and MCS selection, as well as the enhanced scheduler for the GB transmission scheme. Both the GB and the GF schemes show clear improvements from Paper A to Paper E, as shown in Fig. 8 in Paper E . It is clear that the GB scheme is only able to satisfy the 1 ms latency requirement without retransmissions and at this latency achieves a significantly lower URLLC capacity.

With two receive antennas, it is found that the GB scheme is not able to meet the URLLC $1 - 10^{-5}$ reliability and 1 ms latency requirement. The reason for this is found in the enhanced GB scheduler, which increases the MCS in order to fit more transmission allocations in the same transmission time interval (TTI) in the frequency domain. While reducing the queuing probability, there can be devices in poor coverage conditions which might not be able to meet the required reliability for the higher MCS. Increasing the number of receive antennas will therefore also be beneficial for GB transmission schemes as it can improve the coverage.

An additional study was conducted with the purpose to identify at which URLLC load, the GB scheme achieves a lower outage probability, than the GF scheme. With two receive antennas, the GB scheme surpasses the GF baseline scheme from Paper C, but not the enhanced GF scheme, when the latency requirement is relaxed to 1.4 ms and the reliability requirement is relaxed to $1 - 10^{-4}$. If this study was carried out with four instead of two

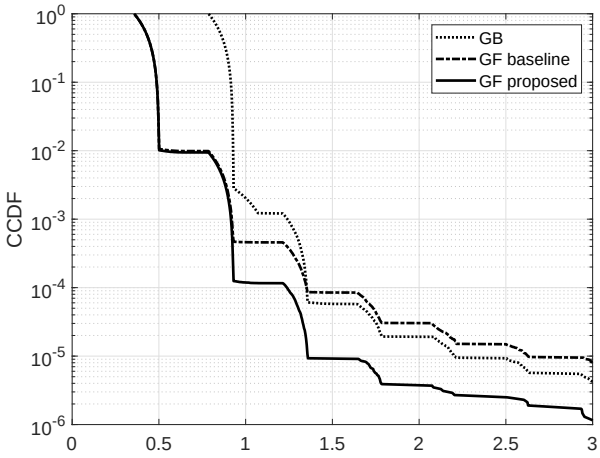


Fig. III.1: Latency CCDF of the GF baseline transmission scheme from Paper C, the enhanced GF transmission scheme with the RRM enhancements from Paper D and Paper E, and the GB transmission scheme with the enhanced scheduler. The aggregated URLLC load is 1.3 Mbps. Two receive antenna per base station (BS) is used.

receive antennas, it would be expected that the GB scheme, would surpass the GF schemes, but only when the latency requirement is relaxed to 1.4 ms and for much higher URLLC aggregated loads (5-10 Mbps).

5 Recommendations and follow-up studies

In this part, RRM mechanisms to enhance the network performance when serving sporadic uplink URLLC traffic with GF transmissions on shared radio resources has been proposed and evaluated. Based on the main findings the following recommendations are made:

- Uplink power control with full path loss compensation should be used. The target receive power density should be selected depending on at least; the aggregated offered load, scenario, number of receive antennas and transmission scheme.
- Applying the joint resource allocation and MCS selection scheme based on average device coverage conditions. With the scheme, a gain of 90% URLLC capacity is observed to support up to 0.5 Mbps aggregated URLLC load per cell.
- Multiple receive antennas is a very effective way to increase the URLLC capacity gains. More than two receive antennas are recommended.
- GF transmission scheme with the proposed RRM enhancements should be considered over GB transmission schemes for sporadic uplink URLLC traffic with latency requirements below 1.4 ms.

Further studies should focus on further techniques to increase the diversity order as a technique to increase the URLLC capacity for uplink GF URLLC. Especially techniques which may be of benefit to devices located near the cell-edge should be considered.

References

[1] 3GPP TS 38.213 v15.3.0, "Physical layer procedures for control (Release 15)," Sep. 2018.

[2] 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.

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

[4] 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.

Paper C

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
IEEE Wireless Communications and Networking Conference (WCNC), 2018

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:

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

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.

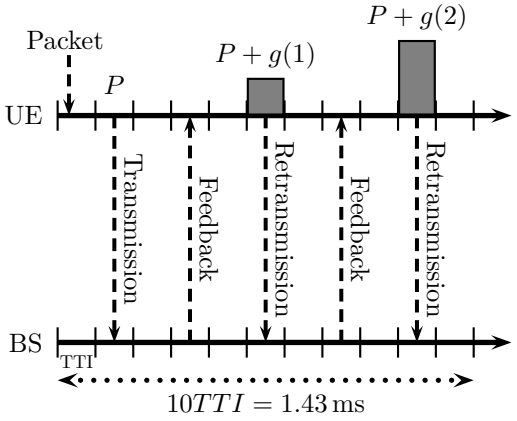


Fig. C.1: URLLC Uplink Grant-Free Transmission with Reactive HARQ and Power Boosting for the retransmissions. P is the transmit power without power boosting and $g()$ indicates the requested power boost.

The considered transmission scheme with power boosting is illustrated in Fig. C.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 (C.3). The considered uplink power control algorithm considered in this study then simplifies from (C.1) to the following:

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

where $g()$ is defined as:

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

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. C.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 (C.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 C.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 C.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. C.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. C.2a is with full path-loss compensation ($\alpha = 1$) and Fig. C.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.

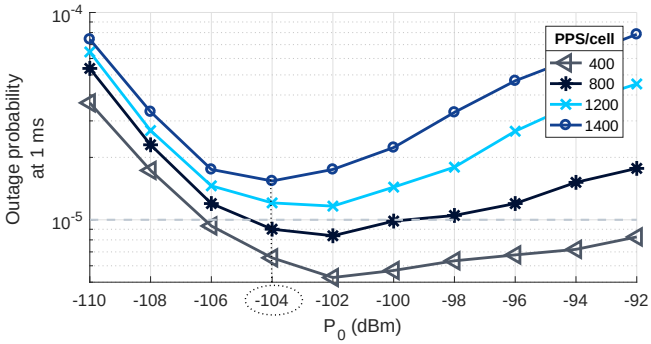
The comparison of fractional and full path-loss compensation is done in two different ranges of P_0 found by an initial sampling of a large P_0 range. It was found that $\alpha = 0.8$ provided the best performance for $-90\,\text{dBm} \leq P_0 \leq -72\,\text{dBm}$, while for $\alpha = 1$ the best range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

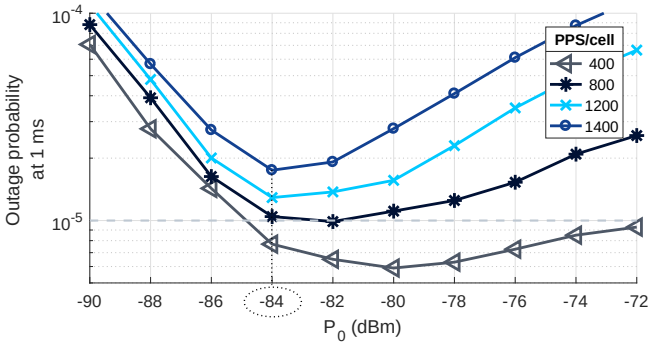
Comparing Fig. C.2a and Fig. C.2b it can be noted that the outage is slightly more sensitive to the P_0 setting for fractional path-loss compensation than for full path-loss compensation. This is due to the higher penalty to cell edge devices caused by fractional path-loss compensation, so operating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

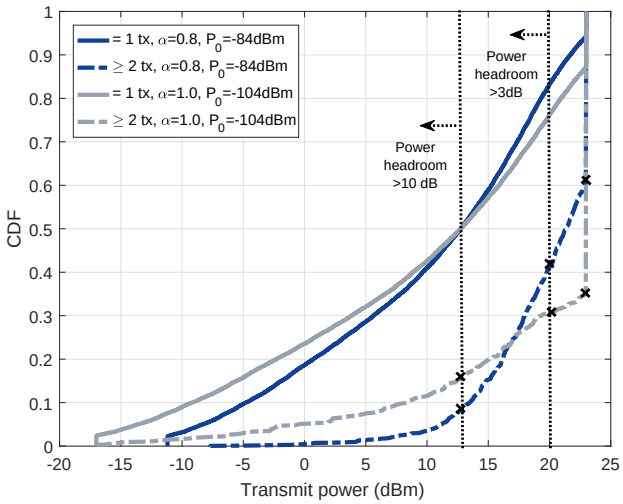


Fig. C.3: CDF of the transmit power according number of required transmissions and power control setting (load of 800 PPS/cell).

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 C.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. C.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. C.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

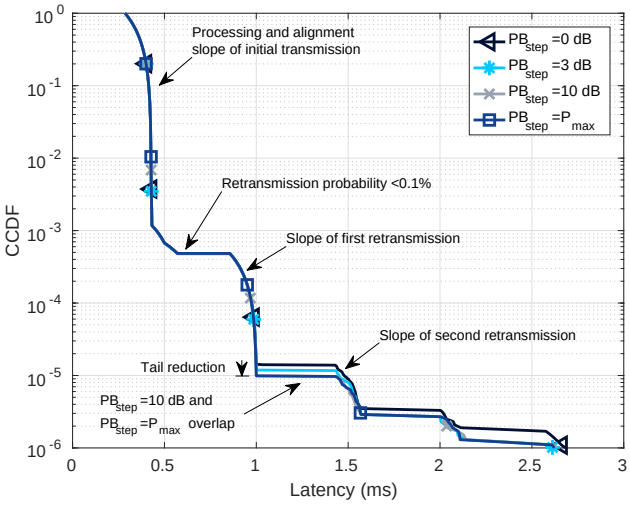


Fig. C.4: Latency CCDFs with 1200 PPS/cell.

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. C.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.

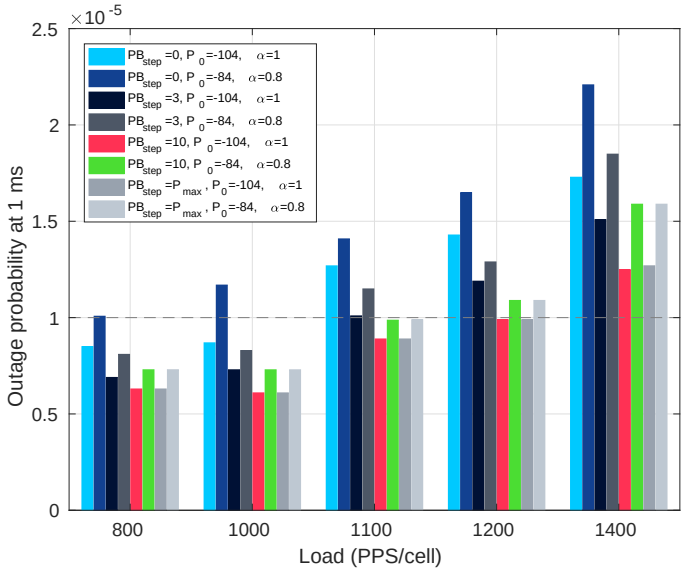


Fig. C.5: Outage at 1 ms for different power control configurations.

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 D

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 GlobeCom Workshops, 2018

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

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

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. D.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. D.2 illustrates examples of GF transmissions and their overlap which can occur using the configuration illustrated in Fig. D.1. Fully overlapping transmissions can occur for transmissions using the same MCS whereas transmissions using different MCS can partially overlap.

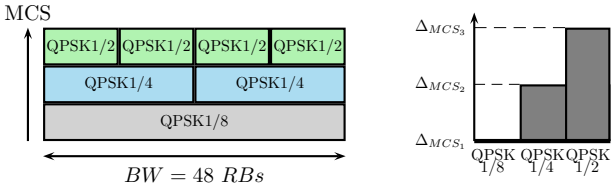


Fig. D.1: Example configuration of MCS, corresponding power spectral density offsets and frequency allocations for grant-free transmissions

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,

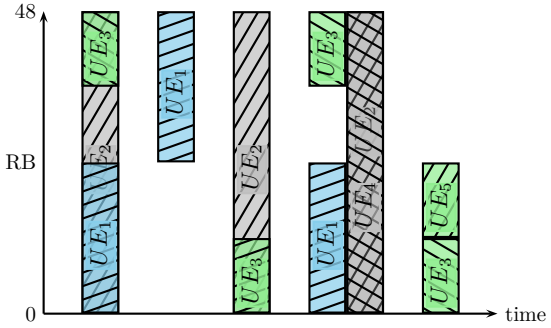


Fig. D.2: Example resource allocations for grant-free transmissions from five UEs using the example configuration from Fig. D.1

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 (D.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

$$\Delta_{MCS_n}[\text{dB}] = 10log_{10}(k), \tag{D.2}$$

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. D.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

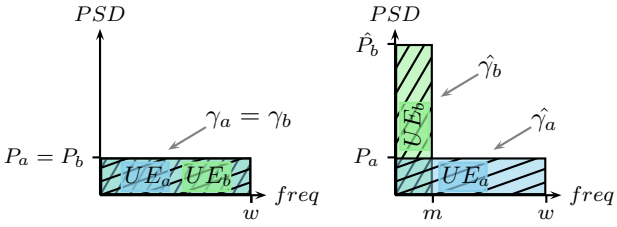


Fig. D.3: Two fully overlapping transmissions (left) versus two partial overlapping transmissions (right)

$$\hat{\gamma}_a = \frac{w-m}{w} \cdot \frac{P_a}{N_0} + \frac{m}{w} \cdot \frac{P_a}{N_0 + \hat{P}_b}. \tag{D.3}$$

An evaluation of the SINR gain $\hat{\gamma}_a/\gamma_a$ using (D.3) is shown in Fig. D.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 (D.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].

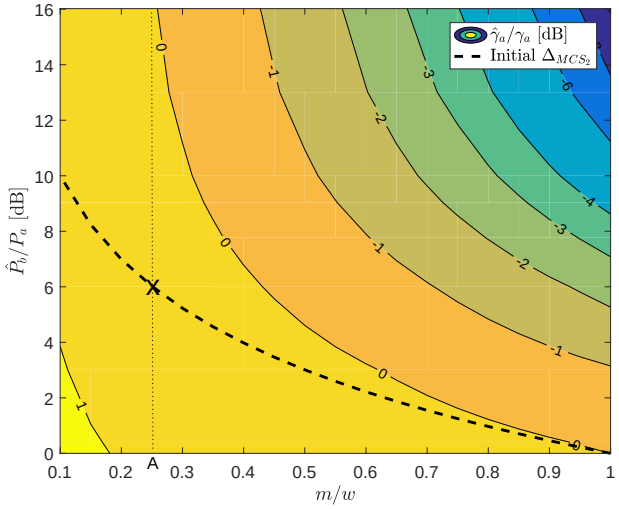


Fig. D.4: SINR gain $\hat{\gamma}_a/\gamma_a$ in dB of UE_a using the MCS_1 as a function of m/w and $\hat{P}_b/P_a$ ratios

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 D.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 D.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 (D.2).

Fig. D.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. D.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.

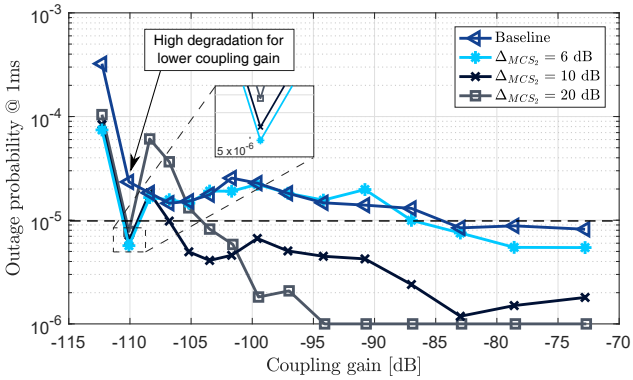


Fig. D.5: Outage probability in coupling gain intervals with $\approx 6\,\%$ of all transmission latency samples per interval. $L = 486.4\,\text{kbps/cell}$.

The reliability statistics per coupling gain interval in Fig. D.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. D.6, for both the baseline and the considered scheme with $\Delta_{MCS_2} = \{6\,\text{dB}, 10\,\text{dB}, 20\,\text{dB}\}$. The

staircase 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.

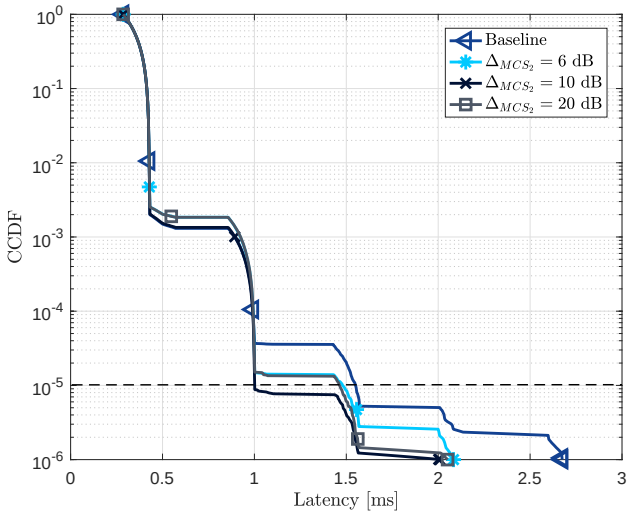


Fig. D.6: Complementary Cumulative Distribution Function (CCDF) of the latency with different MCSs configurations for $L = 486.4\,\text{kbps/cell}$

Fig. D.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. D.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.

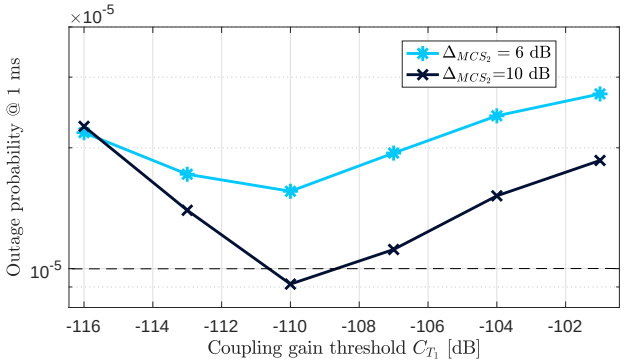


Fig. D.7: Outage probability at 1 ms versus coupling-gain threshold C_{T_1}. UEs with $C > C_{T_1}$ apply MCS_2 with a power offset Δ_{MCS_2}, otherwise MCS_1 is applied. $L = 486.4$ kbps/cell.

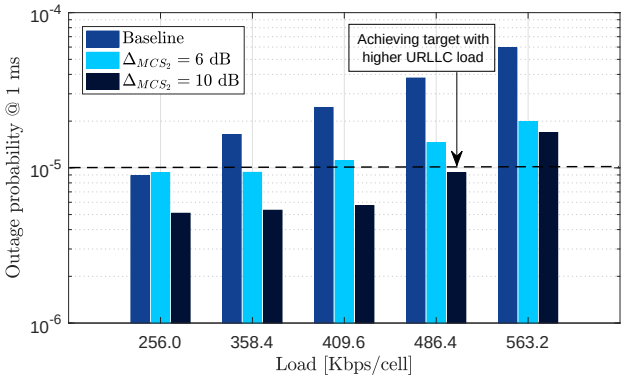


Fig. D.8: Outage probability at 1 ms as a function of offered load

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 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 journal paper has been submitted to the
IEEE Transactions of Vehicular Technilogy, 2019

At the time of creation of this document, the paper was still under peer-review and has therefore not been included in this public version. The reader is kindly referred to the publication channel for a copy of the paper.

Part IV

Diversity and multi-cell reception

Diversity and multi-cell reception

This part of the thesis investigates the potential of diversity and multi-cell reception for uplink grant-free (GF) Ultra-Reliable Low-Latency Communications (URLLC). This includes an investigation of how multi-cell reception in combination with other sources of diversity may enhance the uplink GF URLLC capacity. New multi-cell reception aware radio resource management (RRM) techniques are proposed to unleash the full potential of multi-cell reception.

1 Problems and solution space

One of the main limitations for GF transmission schemes is the generated intra-cell interference caused when using shared radio resources. This interference is particular harmful for devices with high path loss which are unable to transmit with sufficiently high power to meet the desired uplink received power density. This means that their average received signal at their serving cell is weaker than the received signal from devices with a lower path loss. This was observed in Paper D and E (see Fig. 7 from Paper E). These devices with high path loss are as a result particular challenged in reaching the URLLC reliability requirements.

Diversity is a technique to improve transmission reliability and is considered essential to satisfy the challenging URLLC requirements [1]. Diversity can be obtained in the frequency, time and spatial domains [2]. For transmissions over a frequency selective fading channel, frequency diversity can be achieved with a wide-band transmission [3]. Time diversity is challenging to achieve for URLLC use cases with low mobility, as the channel coherence time can be longer than the URLLC latency requirements. However, the interference conditions vary much faster, such that transmission diversity can be obtained e.g. through hybrid automatic repeat request (HARQ) retransmissions. Spatial diversity can be achieved by applying multiple receive antennas

at the receiver and enables the receiver to distinguish simultaneous transmissions. Spatial diversity can also be achieved with multiple physically separate receivers, which therefore may experience different fading and interference conditions. Macro-diversity reception is one such technique where multiple macro base station (BS) simultaneously receive a transmission from the same device. These cells might be deployed at the same site, or at different sites. Macro-diversity reception can be referred to as multi-cell reception.

Multi-cell reception is known from Wideband Code Division Multiple Access (WCDMA) [4] as a part of the soft- and softer-handover mechanism between cells. In Long Term Evolution (LTE) multi-cell reception techniques was standardized in the coordinated multipoint (CoMP) framework in Release 11, with the purpose of increasing coverage and network throughput [5, 6]. Uplink CoMP includes both coordinated scheduling and joint processing, where only the latter is relevant for sporadic GF traffic. Joint processing can be based on the exchange of: complex IQ samples, coded bits (soft bit representations) or decoded bits [7]. Combining based on the coded and decoded bits are sometimes referred to as soft and selection (or hard) combining respectively. The selection of the exchanged source for joint processing requires insights into the trade-off between combining complexity, backhaul capacity usage and the gains it can achieve in terms of URLLC capacity.

While multi-cell reception has been well studied in the open literature [7–11], it remains to be well understood how to best benefit from multi-cell reception in a fifth generation (5G) New Radio (NR) setting with sporadic uplink GF URLLC traffic and what the benefits are when multi-cell reception is combined with other sources of diversity such as multiple receive antennas and HARQ. Further, it remains to be understood how multi-cell reception aware RRM techniques may enhance the URLLC capacity by utilizing the estimated multi-cell reception signal to interference-and-noise ratio (SINR) enhancements.

Potential RRM enhancements could exploit the enhanced signal quality achieved with multi-cell combining. Promising techniques for dynamically scheduled traffic are presented in [11, 12] but are relying on dynamic adaptation of modulation and coding scheme (MCS) and transmission power. One of the techniques exploits the SINR enhancement to reduce the device transmission power in order to reduce the generated interference. Another increase the MCS to improve device spectral efficiency and achieve interference diversity. It remains to be fully understood if comparable techniques can be developed for sporadic uplink GF URLLC traffic.

2 Objectives

The research objectives of this part of the thesis are:

- Quantify the achievable URLLC capacity gain by transmission diversity (retransmissions) and additional spatial diversity from multiple receive antennas per BS.
- Evaluate the potential of multi-cell reception to enhance the achievable network URLLC capacity for sporadic GF URLLC on shared radio resources.
- Quantify the trade-off between URLLC capacity and backhaul throughput requirements for multi-cell reception using selection and soft combining techniques.
- Study RRM enhancements for GF URLLC in the presence of multi-cell reception.

The methodology to address the research objectives has lead to the following sub-objectives:

- Design and implement in the system-level simulator;
  - A multi-cell reception framework to enable device specific cell selection including a cell selection criterion, with a maximum number of assisting cell per device.
  - Multi-cell reception combining based on coded bits, decoded bits, and a hybrid which use coded bits for co-located assisting cells and decoded for non co-located assisting cells. These are referred to as chase-, selection- and hybrid-combining.
  - Long-term statistics of the enhanced signal quality (measured as the SINR) by multi-cell reception.
  - Multi-cell reception aware RRM enhancement based on closed loop power control.
  - Multi-cell reception aware RRM enhancement based on MCS selection.
  - Detailed multi-cell reception statistics.
- Conduct a sensitivity study on the multi-cell reception parameters and combining options, disclosing the impact on the URLLC capacity and the backhaul demand.
- Conduct a sensitivity study on the choice of SINR margin and the achievable URLLC capacity for the two multi-cell reception aware RRM enhancements.

3 Included Articles

The findings of this part are included in the following article:

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

This journal paper evaluates the achievable URLLC capacity enhancements with multi-cell reception based on exchange of coded bits, decoded bits and a hybrid of the two. These options are captured in three combining schemes denoted as; chase-, selection and hybrid-combining. Two novel multi-cell reception aware RRM techniques are proposed and evaluated; one based on closed loop power control and another based on an MCS selection scheme. The performance is evaluated using advanced and highly detailed system level simulations, and use detailed modeling of the main latency and reliability influencing radio access network (RAN) medium-access-control layer (MAC) and physical layer (PHY) layer mechanisms, such as device measurement procedures and explicit interference calculations and linear receivers. The URLLC capacity gains with multi-cell reception are evaluated in combination with two and four receive antennas per BS as well with and without HARQ retransmissions.

4 Main Findings

The main findings from this part of the thesis are:

High URLLC capacity gains with transmission and antenna diversity

Fig. IV.1 shows the achievable single-cell reception URLLC capacity for each configuration combination listed in Table IV.1. For each combination, the power density target is empirically optimized to obtain the achievable URLLC capacity. It shows that the URLLC capacity can be increased by using four in-

Table IV.1: Combination labels (inspired by Table 2 in Paper F)

Combination	BS receive antennas	Retransmissions
A	Two	Disabled
B	Two	Enabled
C	Four	Disabled
D	Four	Enabled

stead of two receive antennas by 3100% (combination A to C) without HARQ retransmissions and 880% when HARQ retransmissions is used (combination

B to D). The identified gain with retransmissions is comparable, but slightly larger than observed in Paper E. This is explained by the use of different uplink power control parameters. The gain of applying retransmissions is observed to be 900% when two receive antennas is used (A to B) and 206% when four receive antennas is used (C to D).

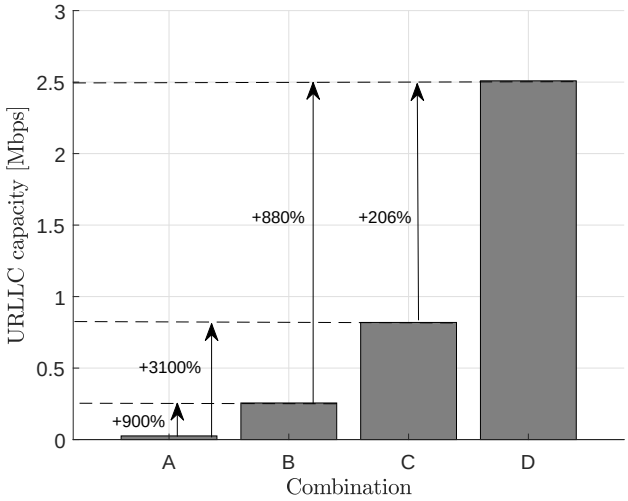


Fig. IV.1: Maximum achievable URLLC capacity with single-cell reception.

Fig. IV.2 shows the SINR cumulative distribution function (CDF) with single-cell reception only, fixed power density target of $P_0 = -98$ dBm and for the aggregated load per cell of 0.768 Mbps and 2 Mbps. In combination B and D, only one retransmission is used. It is observed that increasing the number of receive antennas from two to four (A to C and B to D) the average SINR is increased by 5-6 dB and the 10^{-5} quantile SINR is increased by 9-10 dB. These observations can be explained by the increased diversity order (visible by the slope of the distribution) and the increased received energy. The usage of HARQ (comparing A to B and C to D) is observed to enhance the SINR at the low quantiles ($< 10^{-3}$). This is because a retransmission is only triggered upon a failed initial transmission, which in this case has a reliability of around $1 - 10^{-3}$. The SINR impact of increasing the load, is observed to be less significant with retransmissions and four receive antennas. These observed benefits of retransmissions and HARQ retransmissions to cope with increased load supports the observed significant URLLC capacity gains shown in Fig. IV.1.

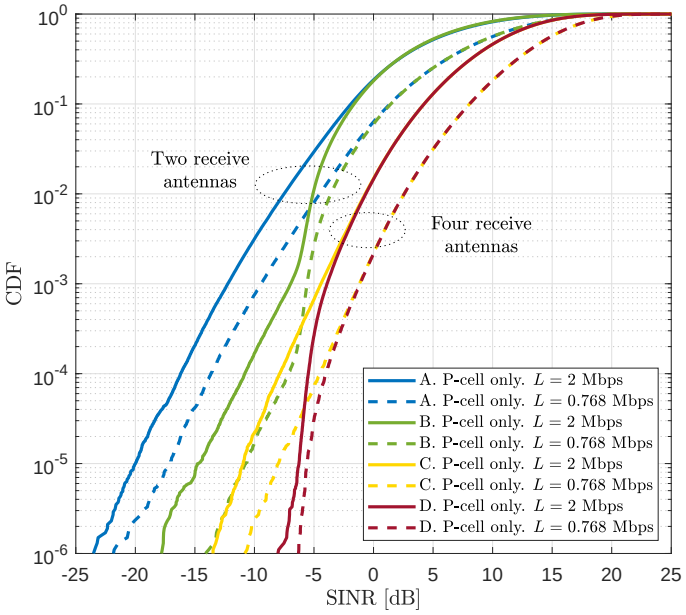


Fig. IV.2: Post-processing SINR for the combinations in Table IV.1. $P_0 = -98$ dBm is used for all combinations with and without multi-cell reception to aid comparison.

Multi-cell reception can dramatically improve URLLC capacity

Fig. 7 from Paper F from Paper F, shows the achievable URLLC capacity using multi-cell reception on top of the achieved URLLC capacity with single-cell configurations shown in Fig. IV.1. Between 200-440% URLLC capacity gains when BSs is equipped with two receive antennas and 20-50% gains are observed when BSs are equipped with four receive antennas and depending on whether retransmissions are enabled. This difference can be explained by the observations from Fig. IV.2, which showed the higher order of diversity changed the SINR distribution particular at the low quantiles. Introducing multi-cell reception which further increases the SINR at the tail, therefore has the strongest impact when the diversity order (from transmission or antenna diversity) is low for the single-cell equivalent. Selection- and hybrid-combining provides similar URLLC capacity gains. Chase-combining provides URLLC capacity gains between 7-22% when compared to selection- and hybrid-combining. However, chase-combining also needs about 10 times higher backhaul throughput when compared to selection- and hybrid-combining.

Multi-cell aware RRM mechanisms

Multi-cell aware RRM techniques are found to be capable of further enhancing the URLLC capacity on top of multi-cell reception. Two techniques have been proposed; closed-loop uplink power control and MCS selection. The former reduces the transmission power if the estimated SINR is above a predefined threshold. The latter increase the MCS-order when the SINR is above another predefined threshold. The adaptation, performed per device, has been based on SINR estimates from uplink transmissions and averaged over a 2 second sliding window to average out the presence of strong interference from overlapping GF transmissions. The threshold values are found using sensitivity studies. The URLLC capacity gain is observed to be up to 90% on top of multi-cell reception with chase-combining, when two receive antennas is used.

5 Recommendations and follow-up studies

In this part of the thesis we have studied the potential of multi-cell reception and multi-cell aware RRM for sporadic uplink GF traffic. Based on our findings the following recommendations are given:

- Use four instead of two receive antennas. This is shown to provide dramatic URLLC capacity gains with no latency increase and does not require additional backhaul usage.
- Support multi-cell reception with the simple selection-combining technique. Multi-cell reception with selection-combining is shown to be capable of providing URLLC capacity gains of 200-440% with two receive antennas per BS and 20-50% with four receive antennas per BS.
- Support multi-cell reception with the soft-combining multi-cell combining technique for networks with high backhaul capacity. Multi-cell reception with soft-combining has been shown to be capable of achieving an additional 7-22% URLLC capacity gain on top of selection-combining, but with the cost of 10 times higher backhaul throughput.
- The use of multi-cell aware RRM techniques such as the proposed closed loop uplink power control based technique and MCS selection scheme are recommended as they have demonstrated to be capable of unleashes the full URLLC capacity gain with multi-cell reception.

This thesis has so far addressed networks where only the URLLC service class is supported. However, with the 5G NR target to support services, it also becomes relevant to research how to efficiently support heterogeneous service classes, where URLLC is one of them, simultaneous over the same carrier without violating the challenging URLLC requirements.

References

[1] 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.

[2] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[3] 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.

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

[5] 3GPP TR 36.819 v11.2.0, "Coordinated multi-point operation for LTE physical layer aspects," Sep. 2013.

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

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

[8] 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.

[9] 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.

[10] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana, "Coordinated multipoint transmission and reception in LTE-advanced: deployment scenarios and operational challenges," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 148–155, Feb. 2012.

[11] 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.

[12] 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.

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 submitted to
IEEE Access, 2019

At the time of creation of this document, the paper was still under peer-review and has therefore not been included in this public version. The reader is kindly referred to the publication channel for a copy of the paper.

Part V

Multiplexing of URLLC and eMBB services

Multiplexing of URLLC and eMBB services

This part of the dissertation focus on how to efficiently multiplex enhanced Mobile Broadband (eMBB) and grant-free (GF) Ultra-Reliable Low-Latency Communications (URLLC) services on the same uplink carrier. More specifically, it is studied how radio resource management (RRM) can be used to allow the URLLC service to satisfy its requirements while keeping the penalty payed by the eMBB service to a minimum.

1 Problems and solution space

In fifth generation (5G) New Radio (NR), multiple service classes have to be supported on the same carrier and hence the question of how to efficiently support URLLC and eMBB on the same carrier, while satisfying the heterogeneous service requirements, has been raised in the study phase for 3rd Generation Partnership Project (3GPP) Release 16 [1]. Particular the challenge of multiplexing URLLC and eMBB has gained interest [2–4].

URLLC is characterized by small packet transmission with very strict latency and reliability requirements. The eMBB service is characterized by high data rates traffic with no strict latency requirements as the goal is to maximize throughput. Multiplexing services within the same carrier can be done in the time domain, frequency domain, power domain and the spatial domain. Fig. V.1 illustrates time (a), frequency (b) and spatial and/or power domain (c) multiplexing of short packet URLLC and eMBB.

Frequency domain multiplexing can be achieved by allocating the two service classes in separate bandwidth parts. For sporadic GF URLLC traffic, transmission opportunities should be available in every transmission time interval (TTI) to minimize frame alignment and queuing latency. This means that frequency domain multiplexing with sporadic GF URLLC needs semi-statically configured radio resources, which effectively means separate bandwidth parts for each service class. The drawback of splitting the carrier into

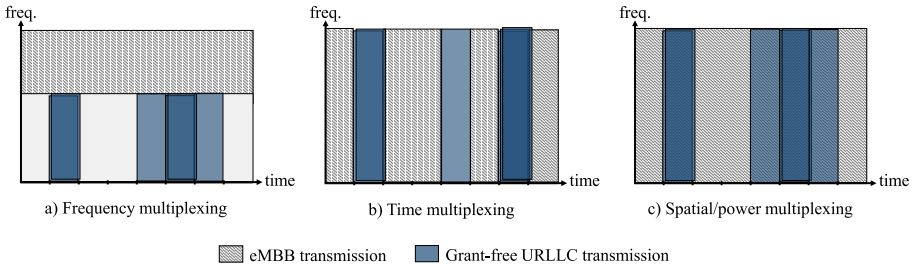


Fig. V.1: Illustration of multiplexing options of URLLC and eMBB.

two parts is a reduced frequency diversity and limited spectral efficiency for eMBB in high signal to noise ratio (SNR) conditions. This can partly be avoided when the grant-based (GB) URLLC transmission scheme is considered, such that the two service classes can be dynamically multiplexed in the frequency domain.

Time domain multiplexing of GF URLLC and eMBB is challenged by the URLLC requirements. GF URLLC opportunities on pre-configured radio resources are needed to minimize the impact from frame alignment and queuing. The unpredictability of sporadic URLLC traffic and the latency requirement can mean that the periodicity between GF transmission opportunities needs to be increased which reduce the pool of available radio resources for eMBB. Dynamical scheduling such as GB URLLC can be used to allocate resource on-demand. However, the latency due to waiting for the completion of a large eMBB transmission can be unacceptable for URLLC. This has lead to the concept of "pause-resume" currently being studied in 3GPP [2]. However, for strict URLLC latencies, the "pause-resume" is still insufficient as it is impacted by the time used to signal a transmitting eMBB device.

Spatial domain multiplexing aims to exploit the spatial domain to support simultaneous overlaying transmissions. Multiplexing in the spatial domain is therefore not forcing a bandwidth reduction or latency increase when multiplexing the services. Instead, it introduces intra-cell inter-service interference, which is a limiting factor for the achievable URLLC and eMBB capacity. Spatial domain multiplexing has been achieved with multi user (MU)-multiple-input multiple-output (MIMO) [5], using a-priori device pairing and dynamic selection of transmission precoding, which is suitable for sporadic uplink GF URLLC. Mechanisms capable of managing the mutually generated interference between URLLC and eMBB such that they can fulfill their service class requirements when multiplexed in the spatial domain is therefore particular interesting.

Additionally to spatial domain multiplexing, power domain multiplexing can be used to multiplex service classes based on their received power. Power domain multiplexing relies on advanced receiver capabilities such as succes-

sive interference cancellation (SIC). The drawback of advanced receivers is the high complexity and the need for user pairing [6].

Service dependent uplink power control is a RRM mechanism which can be used to manage the inter-service interference [7]. In Part II, the use of full path loss compensating open loop power control is recommended for GF URLLC. For service multiplexing, service differentiated uplink power control can be used to increase the URLLC signal to interference-and-noise ratio (SINR) for increasing the reliability, but at the cost of eMBB SINR and throughput as a result. An increased SINR can for example be achieved by either increasing the received power density target or decreasing the eMBB received power density target.

2 Objectives

The multiplexing options present a capacity trade-off between the eMBB and URLLC service classes. Spatial multiplexing of URLLC and eMBB has the potential to satisfy the URLLC latency requirement by allowing GF transmission opportunities in every TTI and the reliability by not constraining the bandwidth. However, the URLLC reliability will depend on the eMBB interfering power. In that case, a more favorable trade-off could be achieved by frequency domain multiplexing, as it removes the intra-cell inter-service interference but also limits the bandwidth used by the services.
The research objectives of this part of the dissertation are summarized as:

- Quantify the feasibility of using service differentiated uplink power control to manage the trade-off between the achievable URLLC and eMBB capacity when the service classes are spatial domain multiplexed.
- Evaluate the use of advanced receiver capabilities to enhance the trade-off between URLLC and eMBB capacity when the service classes are spatial domain multiplexed.
- Study the trade-off between the achievable URLLC capacity and the eMBB capacity for frequency and spatial multiplexing of the service classes.

3 Included Articles

The main findings presented in this part of the dissertation are presented in the following articles:

Paper G. System Level Analysis of eMBB and Grant-Free URLLC Multiplexing in Uplink

In this article the feasibility of spatial domain multiplexing full-buffer eMBB and sporadic GF URLLC is evaluated. Service differentiated uplink power control is used to manage the trade-off between URLLC and eMBB capacity. System level simulations of a multi-user multi-cell network including the main performance influencing factors are used to provide realistic performance figures in line with the previously presented papers in this dissertation.

Paper H. On the Multiplexing of Broadband Traffic and Grant-Free Ultra-Reliable Communication in Uplink

In this article the achievable URLLC and eMBB capacity trade-off for spatial and frequency domain multiplexing strategies is studied. An analytical framework is developed to study the two multiplexing options. The framework assume a single-cell network, but captures the effect of Rayleigh fading, minimum mean square error (MMSE) reception with different numbers of receive antennas, service differentiated ideal uplink power control, carrier bandwidth sharing ratio and simple SIC modeling. The evaluation is conducted in two steps: First, for an eMBB and URLLC traffic load and the average SNR, the required minimum post-combining SINR which ensures that the target reliability is fulfilled, is found by numerical means. Then, from the required SINR, the corresponding achievable capacities for eMBB and URLLC are calculated.

4 Main Findings

Carrier sharing with overlaying transmissions is feasible with service differentiated uplink power control

The findings presented in Paper G clearly show that fulfilling the URLLC service requirements in the presence of a full-buffer eMBB devices is only feasible when service differentiated uplink power control is used to reduce the eMBB received power density target. It is observed that spatial domain multiplexing can be used to support simultaneous URLLC and eMBB services, if the eMBB received power density target is reduced by at least 5 dB compared to URLLC, the base station (BS) are equipped with at least four receive antennas and the URLLC aggregated traffic load is below 0.26 Mbps. This load corresponds to only 10% of the maximum URLLC capacity observed in Paper F. In the presence of multiple eMBB streams per cell the impact on URLLC is severe and only a relaxed URLLC reliability require-

ment can be supported. These observations are confirmed in Paper H, which shows that the spatial multiplexing compared to frequency multiplexing, is only favorable when the average URLLC SNR is high (>10 dB on average) and when the URLLC aggregated load is below 50% of what can be achieved without the presence of eMBB. This difference in the identified maximum sustained URLLC load between Paper G (multi-cell urban macro scenario) and Paper H (single-cell) is due to the different scenarios which pose different interference and coverage conditions. When these conditions are not met, splitting the carrier for frequency multiplexing is shown to provide a better capacity trade-off for both eMBB and URLLC.

Advanced receivers improves the capacity trade-off between URLLC and eMBB

From Paper H, it is found that the use of a SIC capable receiver is particular beneficial for the eMBB capacity, such that a higher eMBB capacity can be reached with an unchanged achievable URLLC capacity. The greatest impact from using SIC on top of an MMSE-interference rejection combining (IRC) receiver, is observed when the average SNR for both eMBB and URLLC is high (>10 dB). That means that in a low SNR scenarios, such as an urban macro scenario, the benefit of using advanced receivers is expected to be limited.

5 Recommendations and follow-up studies

Based on the main findings presented in this part, the following recommendations are provided:

- Service differentiated uplink power control with full path loss compensation for both URLLC and eMBB should be supported. URLLC receive power density should be higher than for eMBB. It is noted that a this recommendation is in line with the preliminary conclusion from discussions in the 3GPP 5G NR study on inter-UE multiplexing and such functionality has recently been recommended for specification in Release 16 [3, 8].
- Spatial multiplexing of URLLC and eMBB service classes with a linear MMSE-IRC receiver is recommended. Service differentiated uplink power control is needed to reduce the eMBB received power density by at least 5 dB compared to URLLC in order to enable the URLLC service to fulfill its requirements. The combination with advanced receivers should be considered to enhance the eMBB capacity.

- When multiple eMBB spatial streams needs to be served or the BS are equipped with less than four receive antennas or service differentiated uplink power control is not used, frequency domain multiplexing is recommended over spatial domain multiplexing.

References

[1] RP-181477, "New SID on Physical Layer Enhancements for NR URLLC," Jun. 2019.

[2] ——, "SID on Physical Layer Enhancements for NR URLLC," Jun. 2018.

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

[4] 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

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

[6] Z. Wu, K. Lu, C. Jiang, and X. Shao, "Comprehensive Study and Comparison on 5G NOMA Schemes," *IEEE Access*, vol. 6, pp. 18 511–18 519, Mar. 2018.

[7] 3GPP TS 38.213 v15.3.0, "Physical layer procedures for control (Release 15)," Sep. 2018.

[8] 3GPP TSG RAN WG1, "RAN1 Chairman's Notes," Feb. 2019.

Paper G

System Level Analysis of eMBB and Grant-Free URLLC Multiplexing in Uplink

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

The paper has been accepted for publication in
IEEE Vehicular Technology Conference (VTC) Spring, 2019

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.

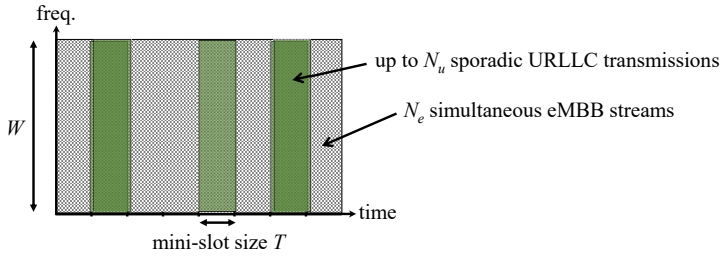


Fig. G.1: Overlaying eMBB and grant-free URLLC allocations in a cell.

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}.

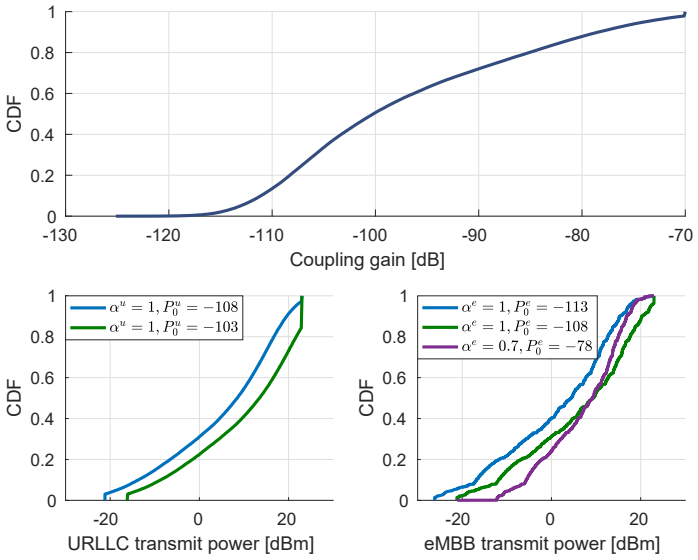
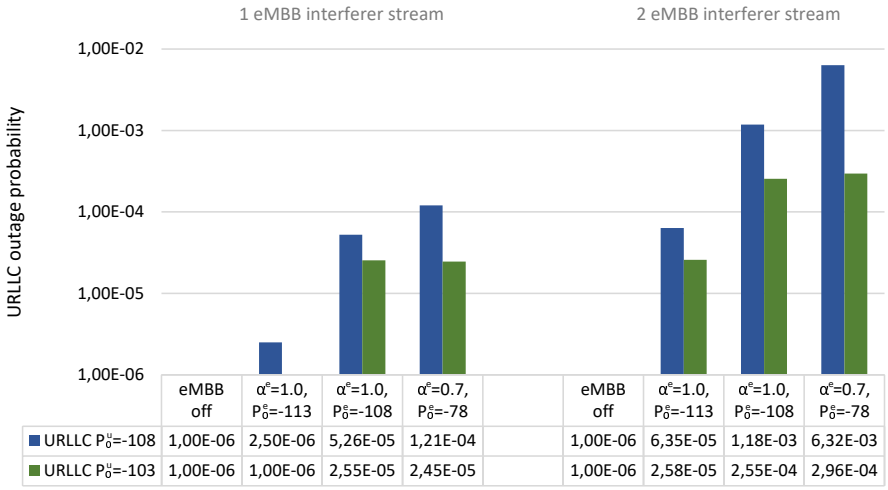


Fig. G.2: Coupling gain distribution in evaluated urban macro scenario outdoor (top). Transmit power distribution for URLLC UEs (bottom left), and eMBB UEs (bottom right).

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



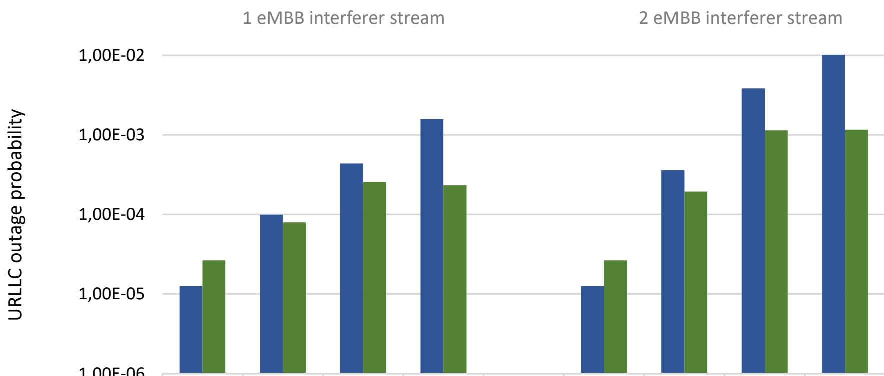
	eMBB off	α^e=1.0, P_0^e=-113	α^e=1.0, P_0^e=-108	α^e=0.7, P_0^e=-78	eMBB off	α^e=1.0, P_0^e=-113	α^e=1.0, P_0^e=-108	α^e=0.7, P_0^e=-78
URLLC P_0^u=-108	1,00E-06	2,50E-06	5,26E-05	1,21E-04	1,00E-06	6,35E-05	1,18E-03	6,32E-03
URLLC P_0^u=-103	1,00E-06	1,00E-06	2,55E-05	2,45E-05	1,00E-06	2,58E-05	2,55E-04	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.

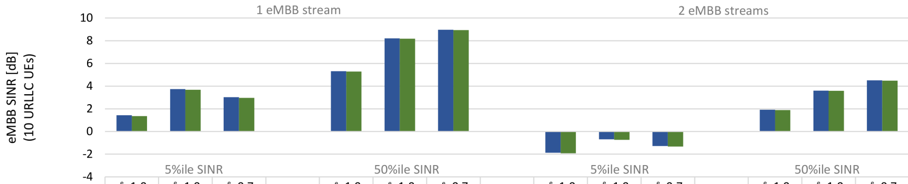
Fig.G.5 and Fig.G.6 shows the impact on the eMBB SINR for the different power control settings. For the lower URLLC load there is little difference on eMBB performance for the different URLLC P_0^u settings. As expected, the eMBB SINR is low in the case of a low P_0^e. And from full to fractional path-loss 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.



	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	1,25E-05	9,96E-05	4,37E-04	1,57E-03		1,25E-05	3,60E-04	3,85E-03	1,16E-02
URLLC P_0^u=-103	2,65E-05	7,96E-05	2,56E-04	2,33E-04		2,65E-05	1,94E-04	1,14E-03	1,17E-03

Fig. G.4: Outage probability of grant-free URLLC for $L = 768$ kbps.



	5%ile SINR α^e=1.0, P_0^e=-113	α^e=1.0, P_0^e=-108	α^e=0.7, P_0^e=-78		50%ile SINR α^e=1.0, P_0^e=-113	α^e=1.0, P_0^e=-108	α^e=0.7, P_0^e=-78		5%ile SINR α^e=1.0, P_0^e=-113	α^e=1.0, P_0^e=-108	α^e=0.7, P_0^e=-78		50%ile SINR 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	1,43	3,73	3,02		5,31	8,21	8,96		-1,87	-0,69	-1,28		1,91	3,6	4,51
URLLC P_0^u=-103	1,36	3,67	2,97		5,28	8,18	8,93		-1,92	-0,73	-1,32		1,88	3,58	4,48

Fig. G.5: eMBB SINR with grant-free URLLC load of $L = 25.6$ kbps.

	5%ile SINR α^e=1.0, P_0^e=-113	α^e=1.0, P_0^e=-108	α^e=0.7, P_0^e=-78		50%ile SINR α^e=1.0, P_0^e=-113	α^e=1.0, P_0^e=-108	α^e=0.7, P_0^e=-78		5%ile SINR α^e=1.0, P_0^e=-113	α^e=1.0, P_0^e=-108	α^e=0.7, P_0^e=-78		50%ile SINR 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	-1,69	1,43	1,01		3,06	6,21	6,9		-3,66	-1,66	-2,48		0,4	2,47	3,38
URLLC P_0^u=-103	-3,36	-0,1	-0,28		2,19	5,36	6,04		-5,06	-2,75	-3,41		-0,3	1,82	2,63

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 accepted for publication in
IEEE Vehicular Technology Conference (VTC) Spring, 2019

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.

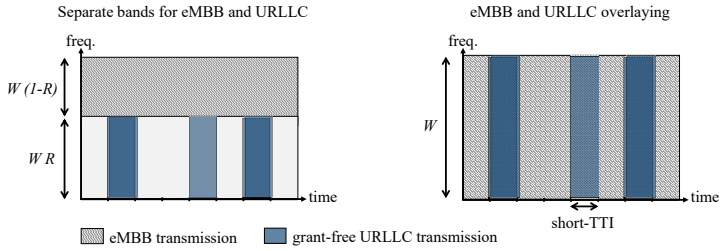


Fig. H.1: Separate bands vs. overlaying transmissions for eMBB and URLLC.

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.

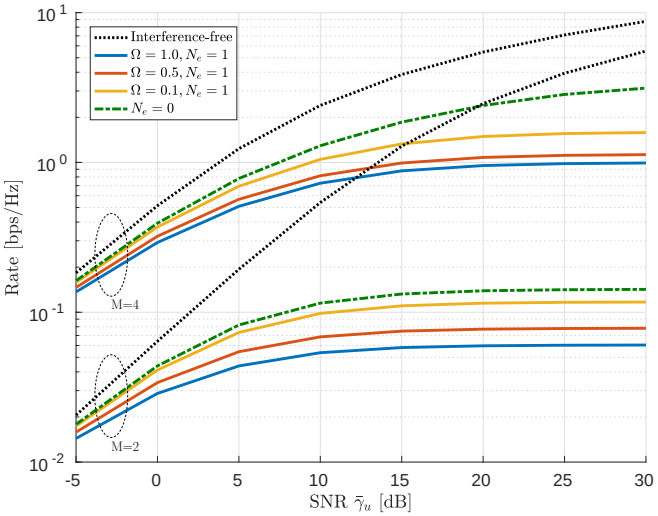


Fig. H.2: Achievable rates for URLLC overlaying one eMBB stream with different Ω, considering $N_u = 50$, $\lambda = 10^{-2}$, and MMSE with 2 and 4 antennas. For the interference-free curve it is assumed dedicated resources.

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.

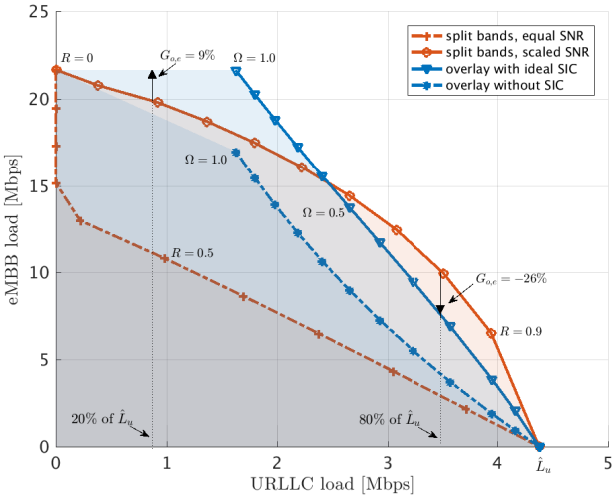


Fig. H.3: Achievable loads for URLLC and eMBB considering different receive strategies and low average SNR $\bar{\gamma}_u = 0$ dB. $W = 10$ MHz, $D = 256$ bits, $N_u = 50$, $N_e = 2$ and $M = 4$.

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

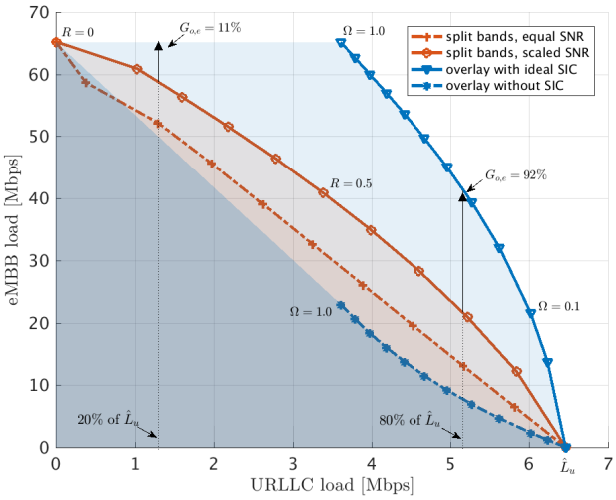


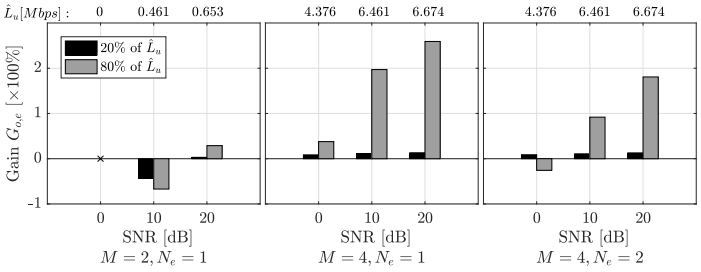
Fig. H.4: Achievable loads for URLLC and eMBB considering different receive strategies and moderate average SNR $\bar{\gamma}_u = 10\,\text{dB}$. $W = 10\,\text{MHz}$, $D = 256\,\text{bits}$, $N_u = 50$, $N_e = 2$ and $M = 4$.

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.

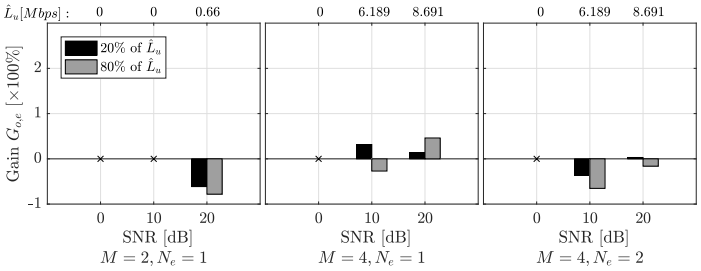
In many cases marked with "x", we note that no URLLC load can be supported. This is observed in most cases for $M = 2$ in low SNR scenarios, independent of the allocation scheme. As can be seen in Fig. H.5a and Fig. H.5c, for small packet size there is a significant gain of overlaying at high SNR, specially 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

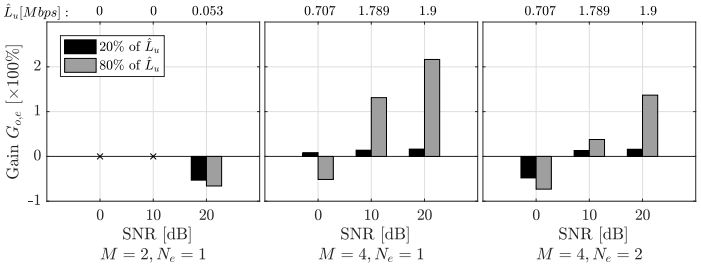
In many cases the allocation of separate bands for each service class shows to be more efficient, specially when SIC is not employed. In practice, it implies that the bandwidth needs to be reconfigured for all grant-free users whenever the target supported load changes. This results in additional control 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



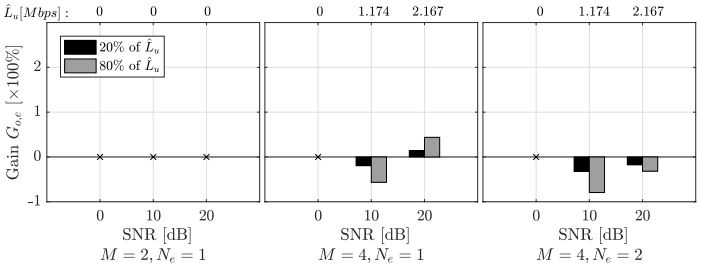
(a) $D = 256\,\text{bits}, \epsilon_u = 10^{-3}, T = 0.143\,\text{ms}$.



(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 VI

Conclusions

Conclusions

1 Summary of the main findings

Achieving efficient support for sporadic uplink Ultra-Reliable Low-Latency Communications (URLLC) requires rethinking and optimization of several components composing the state-of-the-art radio access network (RAN). In Part II of this dissertation, we have focused on identifying the feasible grant-based (GB) and grant-free (GF) transmission strategies. In Part III, radio resource management (RRM) techniques has been proposed which increase the supported URLLC capacity and hence spectral efficiency for particular GF transmission schemes. In Part IV, spectral efficiency enhancements through transmission, antenna and receiver diversity has been studied. Particular the technique of multi-cell reception has been proposed with tailored RRM mechanisms for uplink GF URLLC and has been demonstrated to provide significant spectral efficiency enhancements. Lastly, in Part V, multiplexing techniques of GF URLLC and enhanced Mobile Broadband (eMBB) service classes have been proposed.

With the enhancements proposed in Part III, it has been demonstrated that the most spectral efficient transmission scheme depends on the URLLC reliability and latency requirement and the chosen evaluation assumptions. It has been shown that a GF single-shot transmission can achieve the URLLC reliability requirements of 99.999% reliability with 0.5 ms latency. When relaxing the latency to 0.7 ms a factor of 2 spectral efficiency gain can be achieved with the GF repetition-based scheme. When relaxing the latency requirement further to 1 ms, the spectral efficiency can be additionally improved by a factor of 10 with the GF retransmission-based scheme. Relaxing the latency requirement to 1.4 ms, the GB transmission scheme is predicted to be capable of improving the spectral efficiency further by a factor of 10, when more receive antennas are deployed per base station (BS).

Rethinking of RRM techniques has been shown to be required for efficient sporadic uplink GF URLLC service support. One of these revisited techniques is uplink power control, where it has been demonstrated that tuning of the target receive power density and full path-loss compensation,

with the purpose to enhance the cell-edge reliability, has a significant effect on the achievable URLLC capacity. A novel RRM technique which combines resource allocation and an modulation and coding scheme (MCS) selection strategy has been proposed. The proposed technique has been shown to reduce the probability of overlapping GF transmissions and improve the URLLC capacity by 90%. Further, semi-static transmission parameter adaptation strategies have been shown to be beneficial for sporadic GF URLLC traffic, as they avoid basing the adaptation on only the latest GF transmission and its experienced signal to interference-and-noise ratio (SINR). Long-term and wide-band measured SINR has been demonstrated to be a good indicator for adapting the GF transmission MCS and the device specific uplink power control parameters.

Diversity and combination of multiple diversity sources have been proven to be a key enabler to increase the URLLC reliability and capacity. The use of retransmission can increase the URLLC capacity between 900% and 206% with two and four receive antennas respectively, and the gains of increasing the number of receive antennas is observed to be 3100% and 880% without and with retransmissions enabled respectively. Multi-cell reception has been demonstrated to provide URLLC capacity gains of 200-440% and 20-50% for two and four receive antennas per BS and depending on the use of retransmissions. That is even when a simple selection (can also be referred to as hard) multi-cell combining technique is used. An additional 7-22% capacity gain can be achieved when a soft multi-cell combining technique is used. Multi-cell reception aware RRM mechanisms based on long-term SINR measurements and slow adaptation of either closed loop power control parameters or MCS has demonstrated its capabilities to unleash an additional 50% URLLC capacity.

Efficient multiplexing of URLLC and eMBB, is one of the envisioned targets for fifth generation (5G) New Radio (NR). Multiplexing yields a service capacity trade-off, e.g. the maximum capacity of one service is accomplished when the other is absent. Service differentiated uplink power control is found to be an key enabler for spatial domain multiplexing of eMBB and URLLC. It is shown that with spatial domain multiplexing where the services are simultaneous using the same radio resources is a good option when the URLLC signal to noise ratio (SNR) is high and the eMBB SNR is low (at least 5 dB lower received power density target than used for URLLC) and the required URLLC load is low (such as 0.26 Mbps). When these conditions are not met, full separation of services in the frequency domain has shown to achieve a better service capacity trade-off.

In this dissertation, one of the objectives has been to explore what can be achieved in terms of spectral efficiency when the latency requirement is in the region of 1 ms. In Fig. VI.1 the relation between achievable URLLC spectral efficiency and latency, where the 99.999% reliability is satisfied, is illustrated

based on our main findings. The spectral efficiency values is provided relative to an eMBB requirement from the International Mobile Telecommunications for 2020 and beyond (IMT-2020) requirements [1]. Results obtained with two receive antennas is illustrated with black markers and results obtained with four receive antennas is illustrated with gray markers. Light gray is used to mark results obtained with multi-cell reception. A solid line indicate a confident estimation of the relation between spectral efficiency and latency and a dashed line indicate our best guess. It is observed that reducing the latency requirements to enable the URLLC services, has a cost in terms of spectral efficiency. This cost is particular steep when moving below 1 ms latency where only GF repetitions-based and GF single-shot transmission schemes have sufficiently low latency budgets. Further it is observed that our proposed RRM enhancements, diversity and multi-cell reception can provide clear spectral efficiency gains particular with very low latency requirements.

The intuitive explanation behind the shape of the latency and spectral efficiency relation, is best explained with the idea of degrees of freedom which are available to enhance the spectral efficiency, i.e. achieve the required reliability in the most efficiency way possible. At large latencies, the latency budget allows more complex scheduling algorithms, link adaptation, hybrid automatic repeat request (HARQ) retransmissions and higher layer radio link control (RLC) retransmissions. As the latency requirement is tightened, these degrees of freedom also decreases, at the cost of spectral efficiency. With that, it is essential to make proper system design choices for efficiency uplink URLLC support, particular at very low latency requirements.

1.1 Recommendations

Based on the main findings presented in this thesis, the following recommendations are provided:

- GF retransmission-based schemes are recommended when the latency requirement is in the order of 1 ms and GF repetition-based transmission schemes are recommended when the latency requirement is below 1 ms. GB transmission schemes are recommended when a relaxed latency requirement is considered with a queuing minimizing scheduler and when the BSs are deployed with at least four receive antennas.
- Open loop power control with full path loss compensation and careful tuning of the received power density target is highly recommended for uplink URLLC.
- Use GF transmission schemes with support for multiple MCS options along with a frequency domain structured resource allocation scheme, aiming to minimize the probability of fully overlapping GF transmissions.

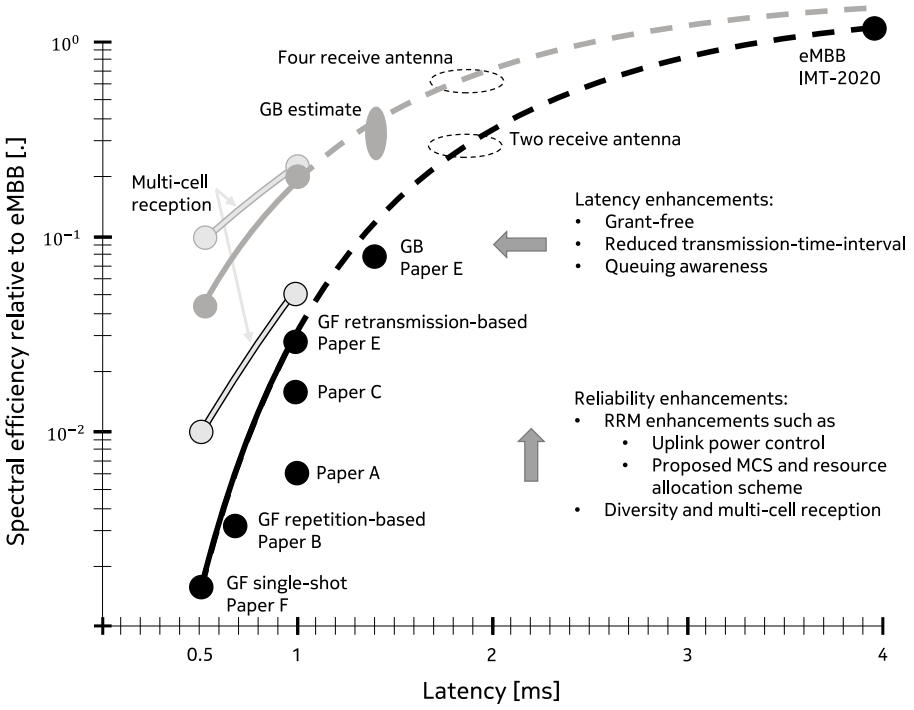


Fig. VI.1: Spectral efficiency as a function of latency with comparable reliability requirement.

- A linear minimum mean square error (MMSE) receiver with interference rejection combining (IRC) capabilities which is equipped with at least four receive antennas is recommended. While enhancing the radio coverage, the multiple receive antennas also enhance the capability to decode overlapping transmissions.
- Multi-cell reception is highly recommended with uplink GF URLLC. Significant URLLC spectral efficiency gains have been observed both when the receiver has either two and four receive antennas. A maximum of three receiving cells is recommended. Simply selection combining is recommended as it provides clear performance gains and can be used with a low capacity backhaul. Soft combining should be considered when the backhaul has a high capacity.
- Multiplexing of URLLC and eMBB by sharing radio resources is possible when service differentiated uplink power control is utilized and the URLLC SNR is high. When the URLLC capacity requirements are high or the URLLC SNR is low, multiplexing in the frequency domain is recommended instead.

2 Future work

This section provides a discussion on the work presented in this dissertation and looks forward on the remaining research challenges for uplink GF URLLC.

Techniques which enhance the spectral efficiency or decrease the latency has relevance for future work. One option to reduce the latency, is to reduce the transmission time interval (TTI) mini-slot length. In order to keep the full packet allocation in a TTI, the sub-carrier spacing (SCS) and bandwidth, needs to be increased correspondingly. For example, increasing the SCS and bandwidth by a factor of 8, means that the TTI length can be reduced equivalently. The results from Paper E and F were obtained with a SCS of 30 kHz, bandwidth of 10 MHz and TTI length of 0.143 ms, which by applying a factor of 8 becomes 240 kHz, 80 MHz and 0.079 ms respectively. With this example, a latency of 0.5 ms from Fig. VI.1 could ideally be achieved with the spectral efficiency equivalent to 4 ms latency. Equivalently, the spectral efficiency currently achieved with 1 ms latency can be achieved with a latency of only 0.125 ms. The challenge with increasing the SCS and bandwidth to reduce the TTI length is degraded coverage, which means that the technique is most suitable for smaller deployment scenarios which are not coverage limited.

Adaptation of transmission parameters based on the average coverage conditions has been proposed in Part III and evaluated for the GF transmission schemes based on retransmissions. A similar strategy could be considered for the GF repetition-based scheme, such that only devices which have power headroom to sustain a higher order MCS are configured to do so. Devices with poor channel conditions increase the number of used repetitions. Coverage-aware adaptation could also be extended to the GB transmission scheme. Applying such constraint for the GB scheme introduces a trade-off between queuing and coverage as a coverage aware scheduler would allocate more radio resources to coverage challenged devices and therefore increase the probability of having to postpone radio resource allocations to other devices.

In a scenario where the number of URLLC devices is small compared to the number of BS antenna elements, the scenario approaches massive MIMO (mMIMO) territory, where channel hardening properties and beam forming techniques can be exploited [2]. In such scenario, a relaxation of the received power density target and smaller adaptation margins can be used, compared to those used in Paper D, E and F. The power density targets and adaptation margins are tuned to account for the probability of interference and degradation of fading. mMIMO should result in higher spectral efficiency as less interference is generated and more energy is collected per transmission due to the increased quantity of receive antennas. With mMIMO, spatial domain

multiplexing of URLLC and eMBB is also more attractive compared to using frequency domain multiplexing.

The dependence of the control channel (CCH), used to reconfigure transmission parameters, conduct scheduling procedures and deliver reception feedback, is an aspect which has not been covered in this dissertation, but have recently been studied in the open literature [3]. CCH unreliability both in the uplink and downlink affects the reliability of the scheduling procedure and would particularly harm the capacity achieved by the GB transmission schemes. The GF retransmission-based scheme only depends on the downlink CCH. For enhanced end-to-end (E2E) reliability estimation, the reliability of the CCH should be accounted for and could be used in novel RRM mechanisms designed for E2E reliability.

Uplink URLLC RRM techniques designed for 5G NR with periodic traffic sources enables predictability and hence facilitates techniques taking proactive actions. Such actions could be channel estimation or coordination between transmission opportunities, e.g. to manage interference or limit the number of simultaneous transmissions. On top, techniques such as preconfigured multi user (MU) multiple-input multiple-output (MIMO), interference coordination techniques, design of combinatorial codes [4] or dynamically adjust the transmission parameters based on the predicted interference, is also identified as relevant for future studies.

The network scenario also plays a role in RRM design for URLLC. The simulation based work presented in this dissertation use the urban macro network described in [5] and considers only devices deployed outdoors. Simultaneously serving indoor and outdoor URLLC devices proved infeasible in this scenario. An alternative deployment is an indoor hot-spot scenario, as is considered for Industry 4.0 [6]. This scenario is not coverage limited but interference limited, and in that case, coordination between cells with the purpose of achieving interference diversity as well as spatial diversity e.g. by using multi-cell reception, is strongly encouraged.

Further knowledge on the characteristics of the statistical confidence at the very low 10^{-5} quantile is desired for URLLC [7]. Throughout this dissertation the collected latency samples has been assumed independent and it has been assumed that the residual distribution around the 10^{-5} quantile follows an Gaussian distribution. Future work should focus on characterizing this relation more accurately to enhance statistical estimation at very low quantiles.

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] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency," *Foundations and Trends® in Signal Processing*, vol. Nov., no. 3-4, pp. 154–655, 2017. [Online]. Available: http://dx.doi.org/10.1561/2000000093

[3] 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.

[4] 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.

[5] 3GPP TR 38.913 v14.1.0, "Study on Scenarios and Requirements for Next Generation Access Technologies," Mar. 2017.

[6] 3GPP TS 22.186 v15.4.0, "Enhancement of 3GPP support for V2X scenarios," Sep. 2018.

[7] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and Low-Latency Wireless Communication: Tail, Risk, and Scale," *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1834–1853, Oct. 2018.

Part VII

Appendix

Appendix

Paper I

On the Performance of One Stage Massive Random-Access Protocols in 5G systems

Nurul H. Mahmood, Nuno Pratas, Thomas Jacobsen, Preben Mogensen

The paper has been published in
9th International Symposium on Turbo Codes and Iterative Information Processing (ISTC), 2016

Abstract

The next generation of cellular system is expected to experience a proliferation in the number of emerging use cases alongside supporting high speed mobile broadband services. Massive Machine Type Communication (mMTC) catering to a large number of low-data rate, low-cost devices is such an emerging use case. Smart utility meters, automated sensors in farms, and vehicle tracking nodes for logistics monitoring are all examples of emerging mMTC devices. Ensuring efficient mechanisms to access the wireless channel for such a massive number of densely deployed devices is a key challenge posed by mMTC applications. A framework for the comparative analysis of the one-stage massive access protocol with respect to important performance metrics for mMTC services is proposed in this paper. The proposed framework allows us to determine the scenarios where the relative simple one-stage protocol can sufficiently meet the desired performance requirements.

1 Introduction

The fifth generation wireless systems (5G) is expected to experience a proliferation in the number of emerging use cases categorized into several broad service groups such as: enhanced Mobile Broadband (eMBB) supporting an evolution of today's broadband traffic with an increased spectral efficiency, Ultra-Reliable Low Latency Communications (URLLC) where messages need to be transferred with high reliability and low latency, and massive Machine Type Communication (mMTC) catering to a large number of (generally) low-data rate, low-cost devices [1].

One of the main drivers behind mMTC services is the role of Machine Type Devices (MTD) as an enabler for the Internet of Things (IoT). Massive machine type communication is commonly characterized by a large number of MTDs associated with each base station [2], the possibility of asynchronous activation of a massive number of MTDs [3], low payload sizes and traffic asymmetry in which the uplink traffic dominates the downlink [4].

It is well documented that the RA procedure in Long Term Evolution (LTE) is not suitable for mMTC services [5–8]. The traditional radio access procedure in LTE consists of several stages; a contention-based stage with RA procedure and a non-contention based stage[1] for additional signalling and transmission of the uplink payload. Scheduling the payload data transmissions has the further benefit of allowing the network to allocate radio resources and hence achieve a higher spectral efficiency. However, with small data payloads, the additional uplink and downlink signalling overheads are

[1] Prior to the this stage is the all-important collision resolution phase, which makes this stage non-contention based.

sub-optimal in an mMTC context [8]. To overcome the limitations of the random access (RA) scheme in LTE for mMTC applications, the *one-stage* access protocol is being considered as an alternative.

The first one-stage access protocol was reported in the research community in [9], and has formed the basis of various other similar access protocols [10–12]. The *one-stage* access protocol, denotes the case where the MTD contends for access with its data payload. Upon the transmission of the MTD RA request with its payload, the eNodeB (eNB) acknowledges the reception via an ACK/NACK message. The access itself is performed over a time and frequency resource assigned for RA. Since the access is uncoordinated, it is possible that multiple MTDs will attempt to access the same network resource, resulting in a collision. In case the eNB is not able to decode the collided transmissions, each affected MTD will re-attempt access at a later RA resource.

Ideally, the *one-stage* protocol is the preferred access mechanism for mMTC services due to its simplicity and low overhead. However, the decision depends on two main factors: (1) the data payload size, and (2) the device density. Specifically, as the data payload increases the number of required network resources increases proportionally. When collisions occur and the base station is not able to decode the transmitted information, then these resources are wasted. Therefore, it is clear that for a high number of active MTDs and high payloads the RA scheme as in LTE is preferable. But, for a low number of active MTDs, fewer collisions will occur making the one-stage access protocol desirable.

In this paper our goal is to provide insights for the design of RA protocols for mMTC services in a 5G setting. Specifically, the performance of the one-stage access protocol considering a number of important performance metrics is investigated in depth by modelling the network as a random Poisson Point Process [13]. To cover the wide variation in the kinds of devices, MTDs capable of adjusting their transmission power to compensate for the path loss are considered alongside simple low-cost fixed transmit power nodes. The findings from this contribution can help to identify networks scenarios on a physical layer perspective under which the *one-stage* access protocol can sufficiently meet the desired service targets.

The rest of the paper is structured as follows: Section 2 details the system model and assumptions, the signal and the collision model, and the evaluated performance metrics. The analytical expressions for the considered performance metrics are derived in Section 3. Section 4 provides numerical and simulation results, and discussions. Concluding remarks are found in Section 5.

2 System Model

We consider a single cell of radius R in a multi-cell network. The eNB is located at the cell center with uniformly distributed M2M devices throughout the cells as shown in Figure I.1. The number and the location of the M2M devices are modelled according to an independent homogeneous PPP Φ_u with intensity $\lambda_u = \pi R^2 \tilde{\lambda_u}$, where $\tilde{\lambda_u}$ is the device density per meter squared. Specifically, for a given PPP, the number of points and their locations are random and they follow Poisson and uniform distributions, respectively [13]. We assume that a device randomly chooses to transmit with probability ρ at each RA opportunity.

Once a device decides to transmit, it will randomly choose a RA request preambles, i.e. a digital signature that the MTD transmits. We assume that there are M orthogonal pseudo-random RA preambles available for the transmitting MTDs to choose from. Information about the available RA preambles is periodically broadcast by the eNB.

Transmission with random transmission probability ρ results in independent thinning of the PPP Φ_u. Hence, the number of devices transmitting concurrently is again a PPP Φ_t of lower density $\lambda_t = \lambda_u \rho$ [13]. Due to the orthogonal nature of the RA request preambles, a transmitting device only experiences interference from other concurrently transmitting devices using the same preamble. As a result, the number of interfering devices is again a PPP Φ with intensity $\lambda = \frac{\lambda_t}{M}$, resulting from the thinning of Φ_t.

2.1 Signal Model

The transmission of a random device located at a distance d meters away from the eNB is analysed in this contribution. The desired signal of interest can be given as $S = P_T d^{-\alpha} g_d$, where P_T is the transmit power, α is the path loss exponent and g_d is the random channel fading power. The desired device experiences interference from all other concurrently transmitting devices that uses the same preamble, which form the PPP Φ. Thus, the sum interference I_s can be expressed as:

$$I_s = \sum_{x \in \Phi} P_T r_x^{-\alpha} g_x, \tag{I.1}$$

where r_x and g_x are the random distance and the channel fading power of the devices to the eNB. Since Φ is a stationary PPP, the PDF of the distance r_x is given by the Rayleigh distribution, i.e., $f_{r_s}(r) = 2\pi\lambda r \exp\left(-\pi\lambda r^2\right)$ [13], defined for $r \geq 1$. Note that, we explicitly assume $r \geq 1$ to avoid amplification of the signal for $0 < r < 1$, and the singularity at $r = 0$.

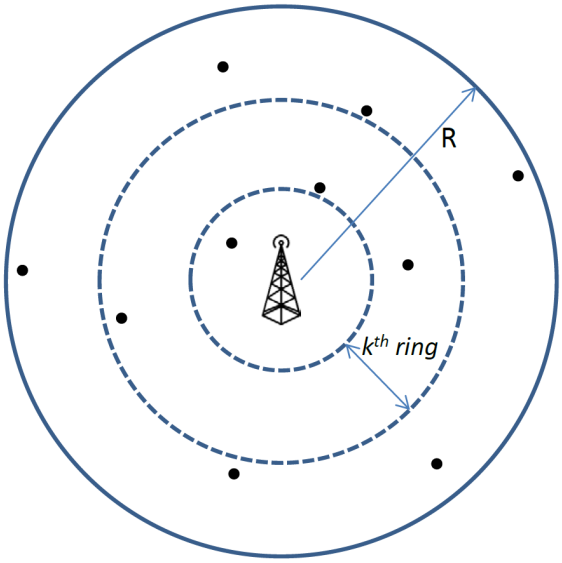


Fig. I.1: System Model showing the random distribution of the devices with the cell divided into multiple rings.

2.2 Collision Model

To cover the wide variation in the kinds of MTDs, we consider two different approaches in determining whether a random access request is successful, namely the *singleton* approach and the *SINR* approach, as detailed below.

Singleton Approach

Whenever an M2M device accesses the channel, it selects one of the M available preambles. A collision is said to occur if two or more devices transmit with the same RA request preamble on the same RA opportunity. Different RA request preambles can be detected by the eNB thanks to their orthogonality. The eNB replies with a RA response. However, a collision is not detected by the eNB if the two devices transmitting the same preamble are equidistant from the eNB, resulting in a false ACK. We use the term *blocking probability* to denote this scenario. In the event of a collision, the MTDs attempting transmission reattempts after a backoff time. The backoff period consists of a fixed backoff time of b_1 RA opportunities, followed by a random backoff time uniformly distributed in $[0, b_2]$ slots.

SINR Approach

Decaying of the signal power with distance is a fundamental feature of the wireless channel. The effects of channel propagation can be overcome with channel inversion power control, i.e., by transmitting with power $P_T = d^{\alpha}$, such that all transmissions are received with the same average received power at the eNB. However, this does not account for the power variation due to shadowing and fading. Furthermore, channel inversion may not always be possible due to maximum transmit power constraints limiting P_T to be below a certain level, and for some very simple low-cost MTDs.

An alternative approach to evaluating the success probability is through analysing the random signal-to-interference plus noise ratio (SINR). In this approach, all devices are assumed to transmit with a constant power P_T. A RA transmission can be considered successful if the SINR exceeds a given threshold γ_{th}. The SINR for the desired device's signal at the eNB, γ_s, is given by $\gamma_s = \frac{S}{I_s + N_0}$, where I_s is the sum interference from other devices as given by Eq. (I.1), and N_0 is the thermal noise power.

2.3 Performance Evaluation Metrics

The performance of the one-stage access protocol is evaluated in this contribution considering the following important evaluation metrics.

Channel Occupancy Rate

The channel occupancy rate, $p_{channel}$, is a general network evaluation criteria indicating the number of RA opportunities resulting in successful transmission. This is related to the singleton approach, and is given as the probability that there is exactly one active node in the PPP Φ. Specifically, $p_{channel} = \mathbb{P}[N = 1]$, where the Poisson distributed random number N is the cardinality of the PPP Φ.

Blocking Probability

To determine the blocking probability, we divide the cell into K equi-radius circular rings as shown in Figure I.1. The radius of each ring is chosen such that all nodes within the same ring arrive at the eNB with negligible time difference. The blocking probability is then given by the probability of having two or more devices attempting to simultaneously access the channel using the same preamble from the same ring.

Let $\mathcal{R}_k$ denote the k^{th} ring. The active MTDs in $\mathcal{R}_k$ choosing the same RA preamble then forms a PPP Φ_k with density $\lambda_k = \frac{\lambda\pi\left((Rk/K)^2 - (R(k-1)/K)^2\right)}{\pi R^2} = \frac{\lambda(2k-1)}{K^2}$. The blocking probability for MTDs in the k^{th} ring is therefore given

by $p_{block,k} = \mathbb{P}[N_k \geq 2]$, where N_k is the random number of nodes in the PPP Φ_k.

Success Probability

With the singleton approach, the random number of concurrently transmitting MTDs (irrespective of the selected RA preamble) is N_t, which is in fact the cardinality of the PPP Φ_t. A transmission attempt is said to be successful when the randomly selected RA preamble of an MTD attempting to transmit does not overlap with that of the other $N_t - 1$ co-transmitting MTDs. Mathematically, this can be expressed as $p_{single} = \mathbb{P}\left[N = 1 | N_t \geq 1\right]$.

On the other hand, the success probability for the SINR approach for a given target SINR γ_{th} is given as $p_{SINR} = \mathbb{P}[\gamma_s \geq \gamma_{th}]$, where γ_s is the desired SINR.

3 Performance Analysis Framework

The performance metrics of interest, as introduced in Section 2.3 are derived analytically in this section.

3.1 Channel Occupancy Rate and Blocking Probability

Owning to the PPP model assumption, the channel occupancy rate and the blocking probability are straightforwardly derived from the Poisson distribution as

$$\begin{aligned} p_{channel} &= \lambda \exp(-\lambda) \\ p_{block,k} &= 1 - \exp(-\lambda_k) - \lambda_k \exp(-\lambda_k), \end{aligned}$$

where $\lambda = \frac{\lambda_u \rho}{M}$ is the density of the random PPP Φ, with λ_u being the number of contending MTDs; and $\lambda_k = \frac{\lambda(2k-1)}{K^2}$.

3.2 Success Probability

Singleton Approach

The success probability with the singleton approach is given by $p_{single} = \mathbb{P}\left[N = 1 | N_t \geq 1\right]$. Using Baye's rule, p_{single} can be expressed as

$$p_{single} = \frac{\mathbb{P}\left[N = 1, N_t \geq 1\right]}{\mathbb{P}\left[N_t \geq 1\right]}.$$

We can evaluate $\mathbb{P}\left[N = 1, N_t \geq 1\right]$ by conditioning on N_t, followed by taking the expectation over its distribution; i.e., $\mathbb{P}\left[N = 1, N_t \geq 1\right] = \mathbb{E}_{N_t \geq 1}\left[\mathbb{P}\left[N = 1 | N_t\right]\right]$, where $\mathbb{E}[\cdot]$ is the expectation operator.

Let us consider a transmitting MTD and $N_t - 1$ potential colliding MTDs. Since the RA preambles are chosen randomly with uniform probability, the probability $\mathbb{P}[N = 1|N_t]$ is the probability that none of these $N_t - 1$ MTDs chooses the RA preambles selected by the desired MTD. Hence, $\mathbb{P}[N = 1|N_t] = \left(\frac{M-1}{M}\right)^{N_t-1}$. Following the Poisson distributed random variable N_t, the expectation over $\mathbb{P}[N = 1|N_t]$ evaluates to

$$\mathbb{E}_{N_t \geq 1}\left[\mathbb{P}[N = 1|N_t]\right] = \sum_{n \geq 1} \left(1 - \frac{1}{M}\right)^{n-1} \frac{\lambda_t^n \exp(-\lambda_t)}{n!}.$$

After some algebraic manipulations, the singleton success probability can be succinctly expressed as

$$p_{single} = \frac{\exp(-\lambda) - \exp(-\lambda_t)}{\left(1 - \frac{1}{M}\right)(1 - \exp(-\lambda_t))}, \tag{I.2}$$

where $\lambda_t = \lambda_u \rho$ is the density of the PPP Φ_t.

SINR Approach

Using the definition of the SINR γ_s, the success probability for the SINR approach p_{SINR} can be re-expressed as $p_{SINR} = \mathbb{P}[\gamma_s \geq \gamma_{th}] = \mathbb{P}[S - \gamma_{th} I_s \geq \gamma_{th} N_0]$. Let us define the random variable $u \triangleq S - \gamma_{th} I_s$. Hence, p_{SINR} can be written as $p_{SINR} = \mathbb{P}[u \geq \gamma_{th} N_0] = 1 - F_u(\gamma_{th} N_0)$, where $F_u(\cdot)$ is the cumulative density function (CDF) of u.

Evaluating the CDF of u directly requires an expression for the probability distribution function (PDF) of I_s, which is not readily obtainable. We propose to circumvent this limitation by using the relationship between the CDF and the Laplace Transform (LT), which leads to the SINR success probability being derived as [14, Eq. 19]

$$p_{SINR} = 1 - \frac{1}{2\pi j} \oint \frac{\mathcal{M}_u(s)}{s} \exp(s\gamma_{th} N_0)\, ds, \tag{I.3}$$

where $\mathcal{M}_u(s)$ is the LT u defined as $\mathcal{M}_u(s) = \mathbb{E}[\exp(-su)]$. By the independence assumption between desired signal S and the sum interference I_s, we readily obtain $\mathcal{M}_u(s) = \mathcal{M}_S(s)\mathcal{M}_{I_s}(-\gamma_{th} s)$, where $\mathcal{M}_S(s)$ and $\mathcal{M}_{I_s}(s)$ are the LTs of S and I_s respectively.

Assuming Rayleigh fading channels, the desired channel power gain S is exponentially distributed with mean $P_T\, d^{-\alpha}$. The corresponding Laplace Transform of S is then $\mathcal{M}_S(s) = (1 + sP_t d^{-\alpha})^{-1}$ [15]. On the other hand, the

Laplace Transform of I_s is obtained as

$$\begin{aligned}\mathcal{M}_{I_s}(s) &= \mathbb{E}\left[\exp(-s\, I_s)\right] \\ &= \mathbb{E}_{\Phi,g_d}\left[\prod_{x\in\Phi}\exp\left(-s\, P_T r_x^{-\alpha} g_x\right)\right] \\ &= \mathbb{E}_{\Phi}\left[\prod_{x\in\Phi}\left(1+sP_T r_x^{-\alpha}\right)^{-1}\right] \\ &= \exp\left[-2\pi\lambda\int_1^{\infty}\left(1-\left(1+sP_T r_x^{-\alpha}\right)^{-1}\right) r_x\, dr_x\right] \\ &= \exp\left(-\pi\lambda\left({}_2F_1\left[1,\frac{-2}{\alpha},1-\frac{2}{\alpha},-sP_T\right]-1\right)\right). \quad \text{(I.4)}\end{aligned}$$

The first step in Eq. (I.4) follows from the independence assumption among the interference from the different sources; the second step is obtained by assuming a Rayleigh fading channel model (i.e., an exponentially distributed g_x); the third step is a result of invoking the probability generating functional of the PPP distribution [13]; and the final step is evaluated by carrying out the integration using the distribution of r as given in Section 2.1.

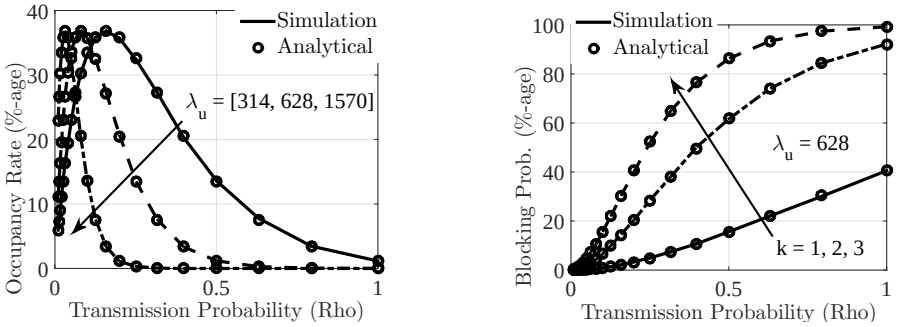
4 Numerical Results

We now present numerical results based on analytical expressions developed in Section 3 and investigate the impact of key system parameters on the performance. The simulations adopt the following parameters, unless stated otherwise: number of RA preambles $M = 50$, path loss exponent $\alpha = 3$, target SINR of 0 dB and transmit power $P_T = 0$ dBm. The channel occupancy rate and blocking probability of the network, and transmission success probabilities for the singleton and SINR approach through MATLAB based Monte-Carlo simulations. The presented results provide insights into the operation of the *one-stage* access protocol and its limitations with respect to the network parameters.

4.1 Channel Occupancy Rate and Blocking Probability

The channel occupancy rate for different transmission probabilities ρ with number of contending MTDs $\lambda_u \approx [315, 630, 1600]$ is presented in Figure I.2a; alongside the blocking probability for different transmission regions $k \in \{1 \dots K\}$ with $K = 3$ for $\lambda_u = 630$ MTDs in Figure I.2b. Analytical findings presented in Section 3 are found to closely match the simulation results. We can observe that the maximum channel occupancy rate is around 37%, which is in-line with the well known occupancy figures of CSMA-like RA

approaches. It is however interesting to note that a large number of transmissions results in false ACK (corresponding to the blocking probability) with high device density. This results from many devices at similar distances to the eNB having to choose the RA preamble from a set with limited number of options. In fact, this is one of the inherent limitations of the *one-stage* random access procedure. Improved access mechanisms are required to overcome such high blocking probabilities.



(a) Channel Occupancy Rate

(b) Blocking Probabilities

Fig. I.2: Simulated and Analytical Channel Occupancy Rate and Blocking Probability vs. transmission probability (ρ) with the singleton approach.

4.2 Success Probability - Singleton Approach

The success probability for an MTD attempting transmission with the singleton approach, as analysed in Section 3.2, is illustrated in Figure I.3. Contrary to the trends observed in Figure I.2, it is always better for an MTD to transmit with a lower probability ρ. This is because the success probability only accounts for the RA opportunities with transmission attempts, and hence idles slots are not accounted for. Therefore, a careful selection of the transmission probability ρ is an important step for the *one-stage* access mechanism, especially at massive MTD densities.

The cumulative success probability after the n^{th} retransmission attempt for $\rho = 0.1$ is shown in Figure I.4. The average number of re-transmissions required to achieve a target success probability can be extracted from such cumulative success probability curves.

4.3 Success Probability - SINR Approach

Complementary to the singleton approach, the success probabilities with the SINR approach for a target SINR $\gamma_{th} = 0$ dB for various transmission probabilities ρ and normalized device distance to the eNB with MTD density

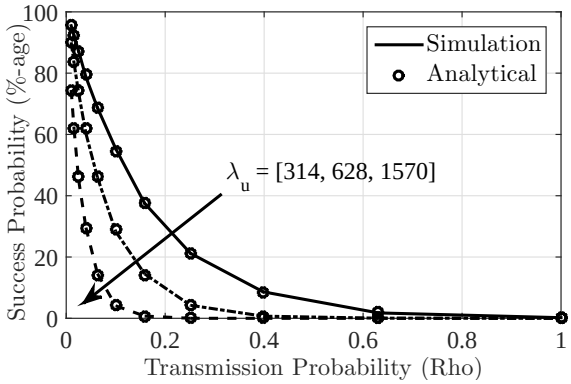


Fig. I.3: Success Probabilities vs. transmission probability ρ with the singleton approach and $\lambda_u = [314, 628, 1570]$ users.

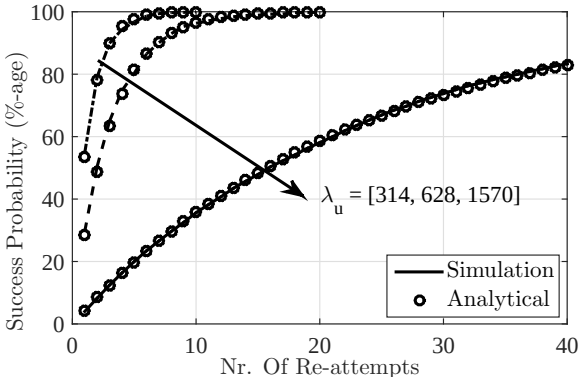


Fig. I.4: Cumumulative Success Probabilities after n^{th} re-attempt for $\rho = 0.1$ with the singleton approach and $\lambda_u = [314, 628, 1570]$ users.

$\lambda_u = 500$ MTDs is presented in Figure I.5. Since the same transmit power and the same transmission probability is considered for all devices, MTDs closer to the eNB experience a higher success probability compared to those at the cell edge in general.

Similar to the trends in Figure I.3, the success rate falls with increasing transmission probability ρ. However, devices closer to the eNB are much less affected compared to those further away. In fact MTDs closer to the cell edge experiences a very low success rate as a result of the overwhelming interference with $\rho = 0.5$. To overcome this limitation, we investigated the impact of having a variable transmission probability in Figure I.6. First we consider only a distance dependent transmission probability where $\rho \propto \frac{1}{d}$, i.e. the MTDs closer to the eNB transmit with a smaller probability compared to those further away. As a result, MTDs further away from the eNB are

less likely to be swarmed with strong interference from MTDs closer to the eNB. Such a distance dependent dynamic ρ is found to provide a success probability improvement of approximately 25% over having a fixed ρ.

In addition, significantly further improvement to the success probability is observed when we incorporate the network information in selecting ρ. For example, allowing $\rho \propto \frac{1}{d\lambda_u}$ results in more than 300% gain for cell edge MTDs, as shown in Figure I.6. It must be noted that additional signalling or autonomous *neighbour awareness protocols* are required for MTDs to estimate the device density λ_u. Furthermore, the cost of such increased success probabilities is an increased delays associated with having a lower ρ.

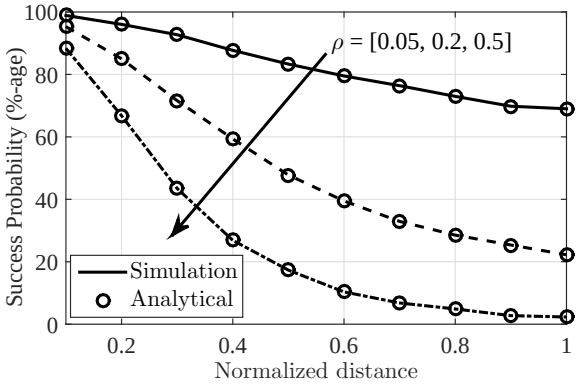


Fig. I.5: Simulated and Analytical Success Rate vs. normalized distance to the eNB with the SINR approach for $\rho = [0.05, 0.2, 0.5]$ and $\lambda_u = 500$ MTDs.

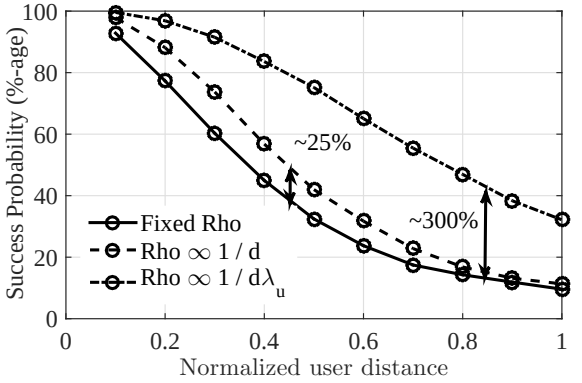


Fig. I.6: Success Rate vs. normalized user distance with a variable transmission probability for $\lambda_u = 1500$ MTDs.

Figure I.6 summarily demonstrates the potential of incorporating location and network awareness into the random access protocol. However, detailed

investigation into such awareness mechanisms is beyond the scope of this contribution, and is left for future studies.

5 Conclusion

The existing access procedure in 3GPP LTE needs to be revised to handle massive Machine Type Communication services envisioned for a 5G system. The *one-stage* random access procedure is a simple proposal aimed at addressing the mMTC access challenge. This paper presents a generic evaluation of the *one-stage* RA mechanism. Two different approaches are considered, namely: the *singleton approach* - where the transmission from all the devices arrive at the eNB with the same average received power; and the *SINR approach* - wherein all devices are assumed to transmit with a constant power. Various performance indicators encompassing both, the network and the device perspective, are evaluated. All analytically derived findings are validated via extensive system level Monte-Carlo simulations.

The evaluation framework presented in this paper has shown its potential in evaluating the advantages and the limitations of the *one-stage* RA protocol for large scale machine type communication. Furthermore, the potential of incorporating location and network awareness into the random access protocol is summarily demonstrated. As part of the future work, we would like to investigate possible enhancements that can improve the channel occupancy rate from the current maximum of $\sim 37\%$, and perform a comparative analysis of the *one-stage* RA procedure with other RA proposals for massive random access.

Acknowledgement

This work has been performed in the framework of the Horizon 2020 project FANTASTIC-5G (ICT-671660) receiving funds from the European Union. The authors would like to acknowledge the contributions of their colleagues in the project, although the views expressed in this contribution are those of the authors and do not necessarily represent the project.

References

[1] N. H. Mahmood *et al.*, "Radio resource management techniques for eMBB and mMTC services in 5G dense small cell scenarios," in *InProc. IEEE VTC-Fall*, Montreal, Canada, Sep. 2016, accepted.

[2] M. T. Islam, A. e. M. Taha, and S. Akl, "A survey of access management techniques in machine type communications," *IEEE Communications Magazine*, vol. 52, no. 4, pp. 74–81, Apr. 2014.

[3] G. C. Madueno *et al.*, "Assessment of LTE wireless access for monitoring of energy distribution in the smart grid," *IEEE JSAC*, vol. 34, no. 3, pp. 675–688, Mar. 2016.

[4] M. Z. Shafiq *et al.*, "Large-scale measurement and characterization of cellular machine-to-machine traffic," *IEEE/ACM Transactions on Networking*, vol. 21, no. 6, pp. 1960–1973, Dec. 2013.

[5] A. Biral *et al.*, "The challenges of M2M massive access in wireless cellular networks," *Digital Communications and Networks*, vol. 1, no. 1, pp. 1–19, 2015.

[6] A. Zanella *et al.*, "M2M massive wireless access: Challenges, research issues, and ways forward," in *Globecom Workshops (GC Workshops), 2013 IEEE*, Dec 2013, pp. 151–156.

[7] G. Madueno, S. Stefanovic, and P. Popovski, "Efficient LTE access with collision resolution for massive M2M communications," in *Globecom Workshops (GC Workshops), 2014*, Dec 2014, pp. 1433–1438.

[8] A. Laya, L. Alonso, and J. Alonso-Zarate, "Is the random access channel of LTE and LTE-A suitable for M2M communications? A survey of alternatives," *Communications Surveys Tutorials, IEEE*, vol. 16, 2014.

[9] C. Schlegel, R. Kempter, and P. Kota, "A novel random wireless packet multiple access method using CDMA," *IEEE Transactions on Wireless Communications*, vol. 5, no. 6, pp. 1362–1370, June 2006.

[10] "Deliverable D4.2 - 5GNOW final MAC/networking concepts," 5th Generation Non-orthogonal Waveform for Asynchronous Signalling (5GNOW), Tech. Rep., 2015.

[11] S. Saur, A. Weber, and G. Schreiber, "Radio access protocols and preamble design for machine type communications in 5G," in *Asilomar Conference on Signals, Systems, and Computers*, Nov 2015.

[12] D. Lin, G. Charbit, and I.-K. Fu, "Uplink contention based multiple access for 5G cellular IoT," in *Vehicular Technology Conference (VTC Fall), 2015 IEEE 82nd*, Sept 2015, pp. 1–5.

[13] M. Haenggi and R. K. Ganti, "Interference in large wireless networks," *Foundations and Trends in Networking*, vol. 3, no. 2, pp. 127–248, Feb. 2009.

[14] N. H. Mahmood, F. Yilmaz, M.-S. Alouini, and G. E. Øien, "Heterogeneous next-generation wireless network interference model-and its applications," *Transactions on Emerging Telecommunications Technologies*, vol. 25, no. 5, pp. 563–575, May 2014.

[15] M. K. Simon and M.-S. Alouini, *Digital Communication over Fading Channels*, 2nd ed. New Jersey, USA: John Wiley & Sons, Dec. 2005.

Paper J

Generic Energy Evaluation Methodology for Machine Type Communication

Thomas Jacobsen, István Z. Kovács, Mads Lauridsen, Li Hongchao, Preben Mogensen, Tatiana K. Madsen

The paper has been published in
Multiple Access Conference (MACOM), 2016

Abstract

It is commonly accepted that the 3rd Generation Public Partnership Long Term Evolution (also known as 3GPP LTE) standard is likely to be unfit for future large scale machine type communication (MMTC). As a result, a new standard, LTE Narrow-Band Internet of Things (NB-IoT) and several radio protocol proposals are being developed. One of the main performance indicators for MMTC is the radio energy consumption. It is important to be able to evaluate the energy consumption of the new standard and the proposed protocols, therefore a generic energy consumption evaluation methodology tailored for MMTC devices is required. Such methodology is the contribution of this paper. It is developed by defining a generic radio transmission and describing the factors which affect the energy consumption. Special attention is put on the factors; power control, link-level performance and a radio power model with a non-constant power amplifier (PA) efficiency model intended for MMTC devices. The results show the impact of the factors and highlight first that applying a commonly used constant radio PA efficiency model can result in an overestimation of the battery life of up to 100 % *depending on the traffic scenario. It is also highlighted that combining power control, transmit repetitions and the radio power model opens for new methods to minimize the radio energy consumption.*

1 Introduction

The 3rd Generation Public Partnership (3GPP) Long Term Evolution (LTE) can become unfit for large scale (massive) machine type communication (MMTC) [8, 14, 16]. An example is the LTE access procedure which is known to become congested when serving a massive number of devices [13]. As a consequence new standards are being developed, like 3GPP Narrow-Band Internet of Things (NB-IoT) and LTE for MTC (LTE-M) [1–3]. In parallel with the standardization work on NB-IoT and LTE-M, there was and still is, significant research ongoing regarding MMTC protocols, as reported in [9, 11, 13, 15]. As it is important to be able to evaluate the new standards and the proposed protocols, a generic methodology is required which is applicable for all proposed protocols and new standards.

MMTC is generally characterized as communication with infrequent small payloads, in scenarios with high device density. The devices can be in challenging coverage conditions and have extreme battery life requirements. Therefore one of the main performance indicators for MMTC is the radio energy consumption [13].

The most common energy evaluation methodology is given in [3] and targets NB-IoT which is the state-of-the-art standard for MMTC. The common energy evaluation methodology does not include the energy consumption impact from challenging coverage conditions or the interference caused by a

high density of devices. Neither does it include a realistic model of the power amplifier (PA) energy efficiency. The authors of [12] show that assuming a constant PA efficiency is not valid for smartphones from 2013-2014. Even though older smartphone radios cannot be directly compared to the radio in MMTC devices, it seems unlikely that the efficiency of an MMTC radio PA will be constant as it is assumed in [3].

This paper presents a generic energy evaluation methodology tailored for MMTC. The methodology includes important features for MMTC, such as uplink power control to manage the level of interference occurring from a high density of devices, transmit repetitions to cope with challenging coverage conditions, and a radio power model intended for MMTC devices. The methodology is generic through its model of a transmission. The power model is based on [3], where we propose to use a non-constant PA efficiency model intended for MMTC radios, derived from empirical measurements on smartphones [12].

The paper is organized as follows. Section 2 presents the generic energy evaluation methodology along with the revised radio power model. In section 3 we apply the proposed methodology and demonstrate the impact of its input. The results and implications of the model is discussed in section 4. The paper is concluded in section 5 which also outlines the future work.

2 Generic Energy Evaluation Methodology

For the energy evaluation methodology to be applicable to the new standards and proposed protocols it needs to be generic. To achieve this we have identified the most important factors that affect the energy usage in a radio transmission. The identified factors are illustrated in Fig. J.1. These are channel aspects such as radio fading and interference, power control, link-level performance, power model and radio access configurations such as transmit repetitions. Modelling of specific protocols which utilize several radio transmissions can be done as a chain of radio transmission blocks. The following sections will describe these factors.

2.1 Radio Fading and Interference

MMTC devices can experience challenging radio coverage conditions [1–3] e.g. due to being located deep indoors.

One method to overcome the effect of a large path loss is by repeating transmissions in time [3]. When doing so, the receiver combines the received transmissions to increase the energy of the desired signal. The configuration of transmit repetitions is a part of the access configuration which also dictates when and how often to transmit. The number of transmit repetitions needs to

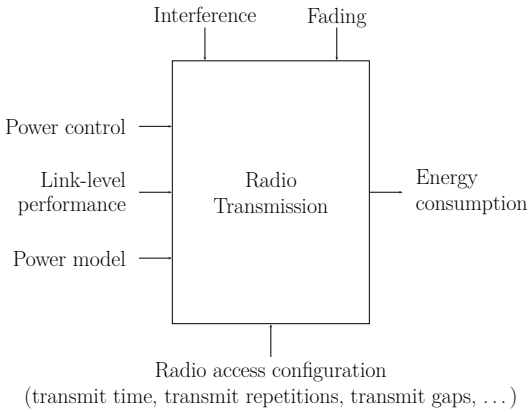


Fig. J.1: Characterization of the input and outputs of a radio transmission. The inputs are power control, link-level performance, power model and the radio access configurations. The transmission is affected by interference and fading. The output is the radio energy consumption

be taken into account by the power control to manage the level of interference. With the use of repetitions the quantity of devices active in a single time slot (TTI) depends on how many devices start their transmission and how many are already repeating their transmissions.

2.2 Power Control

In order to control the level of interference, power control is included in the methodology and for simplicity reasons open loop power control (OLPC) is chosen. OLPC aims to equalize the signal strength from the devices at the base station (BS) receiver. In the OLPC implementation, used in this energy evaluation methodology, it is chosen to use the number of simultaneous transmitting devices as the traffic intensity (M) and the path loss compensation factor (α) along with the target received signal strength at the BS (P_0) as the information which is broadcast to the transmitting devices. The number of active devices can be estimated by the BS through e.g. multi-user-detection techniques. The details on how this is done is out of the scope of this paper.

Each device uses the broadcast information to calculate which transmit power (Ptx) they should use. The transmit power is calculated using (J.1), where Ptx_{dBm} and $Ptx_{max,\mathrm{dBm}}$ is the transmit power and maximum transmit power in dBm and PL_{dB} is the path loss in dB. The devices estimate the path loss from the downlink reference received signal strength.

$$Ptx_{\mathrm{dBm}} = min(Ptx_{max,\mathrm{dBm}}, P_{0,\mathrm{dBm}} - PL_{\mathrm{dB}} \cdot \alpha) \qquad [\mathrm{dBm}] \qquad \text{(J.1)}$$

The target received power at the BS (P_0) is determined from the desired sig-

nal to interference and noise ratio (SINR) which dictates the target performance of the transmission. The target SINR is derived from link-level performance curves, simply referred to as performance curves in this paper and are described in further detail in section 2.3. The target SINR (γ_{SINR}) can be expressed as in (J.2). The interference is caused by all other transmitting devices ($M-1$). The noise power is denoted by N. Notice that (J.2) uses the linear version of PL_{dBm} and Ptx_{dBm} and implies that all devices use the same number of transmit repetitions (R) and that it is only valid when $R > 0$, $Ptx \leq Ptx_{max}$ and as long as the path loss can be compensated.

$$\gamma_{SINR} = \frac{(Ptx \cdot PL^{\alpha}) \cdot R}{(M-1) \cdot (Ptx \cdot PL^{\alpha}) \cdot R + N} \qquad [1] \qquad \text{(J.2)}$$

2.3 Link-Level Performance Curves

The SINR affects the receiver's probability of correctly decoding the transmitted message. By setting a target performance (e.g. 10 % successful decode probability) the corresponding target SINR (γ_{SINR}) can be found. The use of performance curves (successful decoding probability vs SINR) enable the evaluation methodology to support any coding and modulation, data type and multiple access technique with different multi user detection abilities. The corresponding SINR can be translated to SNR by considering the interference as noise.

An example of two performance curves (denoted curve A and B) are shown in Fig. J.2. They are generated by a link level simulator, which has been simulating LTE PRACH sequences [4] (Zadoff-Chu sequences) of two different lengths. The performance shows the probability of the eNB not successful decoding the PRACH sequence at various SNR. In the figure the target performance is set to 10 % error rate which translates to a target SINR (γ_{SINR}) of -18.9 dB and -11.0 dB for curve A and B respectively. The better SINR performance of A comes at the cost of taking more time to transmit. In this example, A requires 0.8 ms (without cyclic prefix which takes 0.103 ms) and B requires 0.134 ms to transmit. The pair of a performance curve and transmit time is denoted a mode through the rest of the paper.

2.4 Power Model

The radio power model used in this evaluation methodology originates from [3]. The model is updated with a PA efficiency model which is based on the work presented in [12] and modified to be used for MMTC radios. The radio power model from [3] utilizes four power states; power saving mode (PSM), receiving (RX), idle ($Idle$) and transmitting (TX). All states are included in the evaluation in order to include the energy impact of synchronization, configurations receptions, gaps between transmissions and receiving downlink

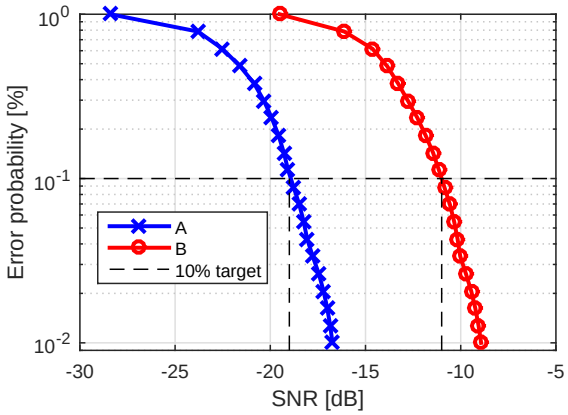


Fig. J.2: Example performance curves A and B.

traffic. A transmission with all four states (PSM, RX, $Idle$, and TX) is depicted in Fig. J.3. Notice this is an example of what occurs in each state and the order of the states.

The transmission starts with the radio sleeping for a certain period before waking up from power saving mode (PSM). The time spent in the PSM state (T_{PSM}) depends on the traffic model, sleep configurations, and whether uplink data is ready for transmission.

Once the radio is awake it will change to RX state where it will start acquiring downlink synchronization such that it is able to decode the broadcast channel (and control channel) to receive the configuration information. This information includes power control configurations, access configurations and scheduling grants (if the protocol utilize scheduled access). If downlink data are scheduled for the device, the radio will acquire the data in the RX state. The time spent in the RX state therefore depends on whether downlink payload is available for the device, the payload size, modulation and coding and the SINR. Devices in bad coverage can be assumed to spend more time to acquire synchronization compared to others with better coverage conditions. The power consumption in the RX state (P_{RX}) is assumed to be independent of the modulation and coding scheme, data rate and bandwidth which is a reasonable assumption according to [12].

If the radio has uplink data to transmit, it will change to the TX state. The time spent in the TX state (T_{TX}) depends on the configured number of transmit repetitions (R), gaps between the transmit repetitions, uplink modulation and coding scheme (UL MCS) and uplink payload size. The power drawn in the transmit state (P_{TX}) depends on the transmit power dictated by the power control and the efficiency of the radio PA which similarly depends on the transmit power [12]. P_{TX} is similar to P_{RX} assumed to be independent

of the modulation and coding scheme, data rate and bandwidth.

Time spend on waiting (e.g. for an opportunity to transmit uplink payload or in gaps between transmit repetitions) are spend in the *Idle* state where the radio maintains synchronization, as described in [3]. This is the main difference between *Idle* and *PSM*, where the radio in *PSM* is turned off such that synchronization cannot be maintained. The power draw in the *PSM* state (P_{PSM}) and the *Idle* state (P_{Idle}) are device specific.

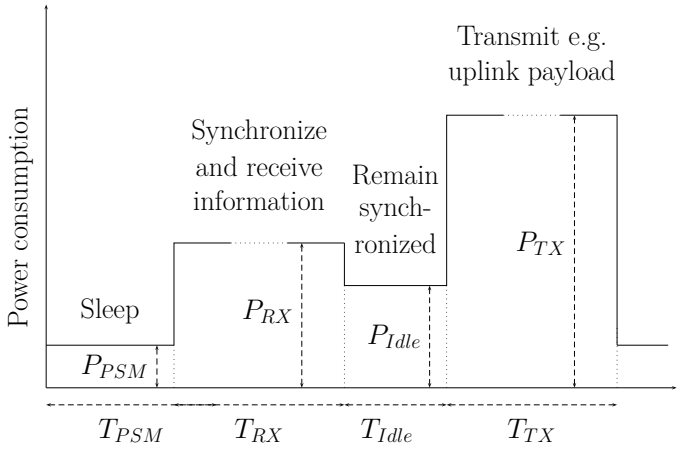


Fig. J.3: Power model states (*PSM*, *RX*, *Idle* and *TX*) in a power and time domain with examples of what happens in each state.

The energy consumption of a transmission can be calculated as the area below the line in Fig. J.3. The model proposed in this paper is given by (J.3). Notice that it does not consider ramp up or ramp down time in state transitions as in [12].

$$E_{tot} = P_{PSM} \cdot T_{PSM} + P_{RX} \cdot T_{RX} + P_{Idle} \cdot T_{Idle} + P_{TX}(Ptx_{\text{dBm}}) \cdot T_{TX} \quad \text{[J]} \tag{J.3}$$

The energy efficiency of the radio PA dictates the relation between the transmit power (Ptx) and the consumed power (P_{TX}). A common model from [3] assumes that the energy efficiency is constant at either 30 % or 40 % independent on the transmit power. A study conducted by [12] shows that this is not a valid assumption for LTE smartphones in 2013-2014.

The research done in [12] might not be directly applicable in terms of absolute power values in a power model intended for MMTC devices. Clearly the best fitting PA energy efficiency model would be derived from emperical measurements from a MMTC device using NB-IoT or LTE-M radios. But as

no such device is commercially available (to the best of the authors' knowledge) the work presented in [12] is used to create the PA efficiency model intended for MMTC radios.

One of the targets for NB-IoT and LTE-M is low cost [3]. If the PA should be cheap and the bandwidth is lower (e.g. from 20 MHz in LTE to 1.4 MHz in LTE-M or 200 kHz in NB-IoT), the PAs high-gain mode (described in [12]) used in smartphones PAs might not be needed and hence can be removed. This means that the maximum transmit power of the non-high-gain mode have to be extended from 10 dBm to 23 dBm ($Ptx_{max,\mathrm{dBm}} = 23\,\mathrm{dBm}$). Notice that $Ptx_{max,\mathrm{dBm}}$ should be considered a parameter in this energy evaluation methodology and hence can be set to another value. The corresponding efficiency at $Ptx_{max,\mathrm{dBm}}$ is scaled to be 40 %, such that this model matches the assumption used in [3] for the constant efficiency model. This means that both models have the same power consumption at $Ptx_{max,\mathrm{dBm}}$.

The resulting PA efficiency model proposed for MMTC devices, which has been derived from [12], is described in (J.4) and consists of two states; one where Ptx_{dBm} takes values from $-30\,\mathrm{dBm}$ to $0\,\mathrm{dBm}$ where the power consumption is constant as in [12], and one where Ptx_{dBm} takes values from 0 dBm to 23 dBm where the power consumption increases at a moderate rate. Please note that (J.4), relates quantities given in dBm to the power consumption in W, similar to the model presented in [12].

$$\begin{aligned} &P_{TX}(Ptx_{\mathrm{dBm}}) = P_{PA}(Ptx_{\mathrm{dBm}}) \\ &= \begin{cases} 0.0197 \cdot Ptx_{\mathrm{dBm}} + 0.0454, & \text{if } 0 < Ptx_{\mathrm{dBm}} \leq 23 \\ 0.0454, & \text{if } Ptx_{\mathrm{dBm}} \leq 0 \end{cases} \quad [\mathrm{W}] \end{aligned} \tag{J.4}$$

The radio PA efficiency model for MMTC is depicted in Fig. J.4 along with the commonly used constant PA efficiency model. The power consumption is given in relative values (in log scale) to the power consumption at $Ptx_{max,\mathrm{dBm}}$. This new model will result in higher energy consumption if transmit powers lower than maximum transmit power ($Ptx \leq Ptx_{max}$) is used, as it models a lower efficiency than the constant model for $Ptx < Ptx_{max}$.

3 Results

This section demonstrates the use of the proposed energy evaluation methodology described in the previous section. Assumptions and parameters used throughout this section are listed in table J.1.

The assumptions intend to model a MMTC device which on average spends 60 min in PSM between consecutive uplink transmissions. Before the device is ready to transmit its payload it has to perform synchronization and read the needed control channel. Then it initiates the transmissions and

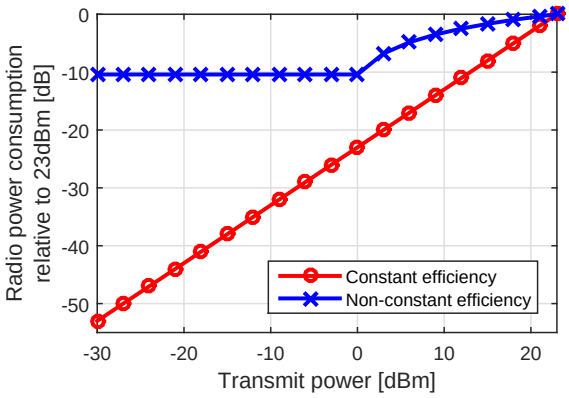


Fig. J.4: Radio PA power consumption ($P_{PA}(Ptx_{\mathrm{dBm}})$) for the proposed (non-constant efficiency) model for MMTC radios against the commonly used constant efficiency model (40 %).

Table J.1: Assumptions and parameters

Channel model	
Path loss (PL)	154 dB
Shadow fading	Eliminated by power control
Bandwidth	1.08 MHz
Noise	Noise figure (2 dB), thermal noise (-111 dBm)
Performance curve	
Target performance	10 % error rate
$\gamma SINR$ (from Fig. J.2)	Mode A (-18.9 dB), mode B (-11.0 dB)
Power model	
P_{TX}@23 dBm	500 mW
P_{other}	60 mW
T_{TX}	Mode A (0.903 ms), mode B (0.237 ms)
PSM	$P_{PSM} = 0.015$ mW, $T_{PSM} = 60$ min
$Idle$	$P_{Idle} = 3$ mW, $T_{Idle} = 2$ s
RX	$P_{RX} = 70$ mW, $T_{RX} = 0.36$ s
Power control	Uplink open loop with $\alpha = 1$
Traffic model	UL only. Poisson call inter-arrival
Deployment	Single cell. Path-loss compensated by PC
Access configuration	Transmission in all TTI. Consecutive repetitions

stays in TX until all transmit repetitions have been performed. The values used in this evaluation are inspired by NB-IoT and should be considered as example values only. The power consumption values are from [7]. The time to conduct synchronization is from [5] and set to 200 ms which is spend in RX. The total acquisition time of the control channel is from [6] and is set to 2 s which is spend in $Idle$ followed by 160 ms in RX.

3.1 Impact of Power Control

To demonstrate the impact of uplink power control on the energy consumption, lets first have a look at Fig. J.5 which shows the transmit power at different traffic intensities (M) with the number of transmit repetitions (R) ranging from 1 to 512 for performance curve A (right) and B (left). Note that all devices are using the same number of transmit repetitions. The white area is the outage region and is clearly seen to the right and in the lower right corner for both performance curves. The outage region is where (J.2) is not satisfied meaning that the power control requests a transmit power above the maximum allowed (Ptx_{max}) and the UL γ_{SINR} cannot be met. Notice the outage region appears at much lower traffic intensity for performance curve B than for A. The reason for this is its higher γ_{SINR} which forces the transmit power to be generally higher than when performance curve A is used.

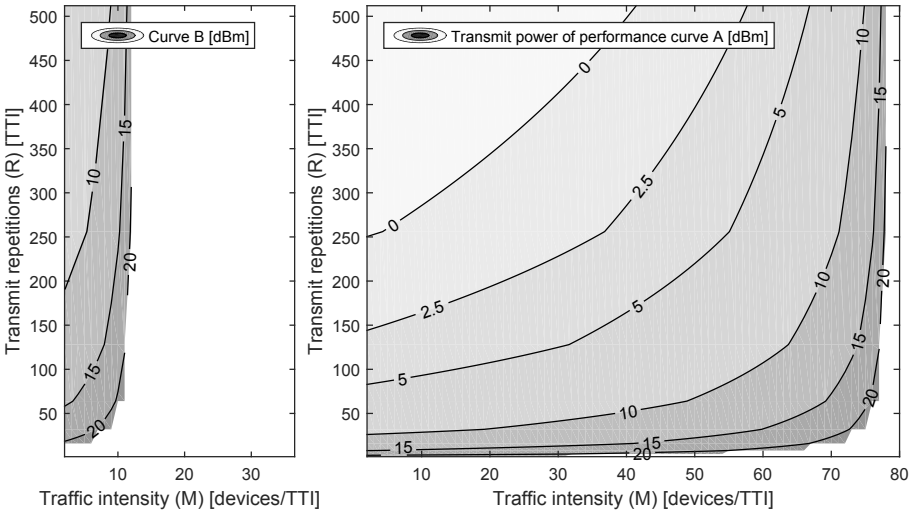


Fig. J.5: Contour plot of the radio transmit power for performance curve A (right) and B (left). The figure for performance curve A can be interpreted as if 50 devices are transmitting at the same time ($M = 50$) and 64 transmit repetitions are configured ($R = 64$), then each device should transmit with $Ptx_{\text{dBm}} = 10\,\text{dBm}$ to reach the performance target (γ_{SINR}).

The corresponding energy consumption to the transmit powers shown in Fig. J.5 is shown in Fig. J.6. The markers in Fig. J.6 at mode A (right) indicates the energy consumption if the transmit power is fixed to $Ptx_{\text{dBm}} = 10\,\text{dBm}$ meaning that power control is not utilized. Let's say that the traffic intensity at a given time is $M = 50$ devices / timeslot (TTI) and that $R = 64$ transmit repetitions is used. Then the traffic intensity increases to $M = 65$. To keep the target performance (γ_{SINR}) for the mode, in the case that power control is not utilized, meaning that the transmit power is fixed, the number of

transmit repetitions have to be increased. The result is an increase in energy consumption. However, if power control is utilized, the transmit power can be increased with the result of maintaining the energy consumption instead and keeping the target performance. This is illustrated in the figure as the two arrows from $M = 50$ to $M = 65$, one with power control (blue) and one with fixed transmit power (red). Note that it is the combination of the power model with the non-constant PA efficiency model for MMTC radios and power control that causes the non-linear energy consumption contour lines and enables new options to optimize the device energy consumption.

3.2 Impact of Performance Curves

Two modes are considered (A and B) each having a performance curve (A and B from Fig. J.2) and transmit time. The target performance error rate is the same (10 %), but the corresponding target SINR (γ_{SINR}) is different -18.9 dB for mode A and -11.0 dB for mode B (see also table J.1). The time it takes to transmit a single transmission is set to 0.903 ms and 0.237 ms.

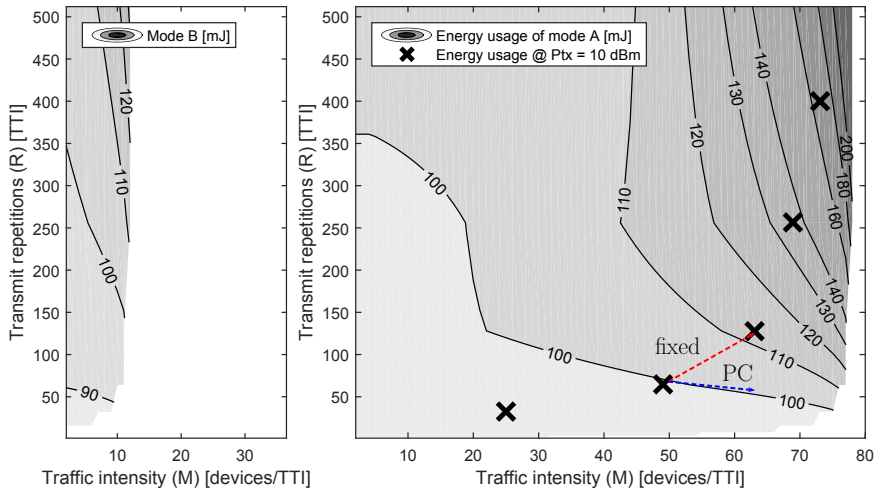


Fig. J.6: Contour plots of the device energy consumption for mode A (right) and B (left) with power control as a function of transmit repetitions (R) and traffic intensity (M). The energy consumption is per transmission. The crosses in mode A are reference energy usage if the transmit power is fixed to 10 dBm. The arrows indicate the possible options when power control is used (blue) and not used (red) to keep the target performance (γ_{SINR}) when M increases from $M = 50$ with $R = 64$ to $M = 65$.

The energy consumption of mode A and B are seen in Fig. J.6. Again note the significant difference in maximum supported traffic intensity, which is much lower in B at $M \leq 12$ than in A at $M \leq 78$, as also seen in Fig. J.5. The outage region appears much earlier with mode B due to the higher γ_{SINR}

which is close to 10 dB higher. To compare the energy consumption of the two modes in further detail, one method is to extract energy consumption values at the same traffic intensities and transmit repetitions. If the transmit repetitions is fixed to $R = 128$ it can be found that the energy consumption of B is slightly lower than A but only for $M \leq 10$ (e.g. at $M = 5$ the difference is 94.3 mJ against 97.4 mJ).

3.3 Impact of Power Model

Figure J.7 shows the effect of the radio power model with a non-constant PA efficiency on the energy consumption when mode A is used. Using mode A and B shows the same tendencies. The figure shows that the energy consumption ratios between the two radio power models range from 1 to 2 and have an average of 1.15 across traffic intensity (M) and transmit repetitions (R). This means that using the power model with a non-constant PA model intended for MMTC radios will estimate a overall higher power consumption. The exception to this is when $Ptx = Ptx_{max}$, as expected. This is where the ratio is 1.

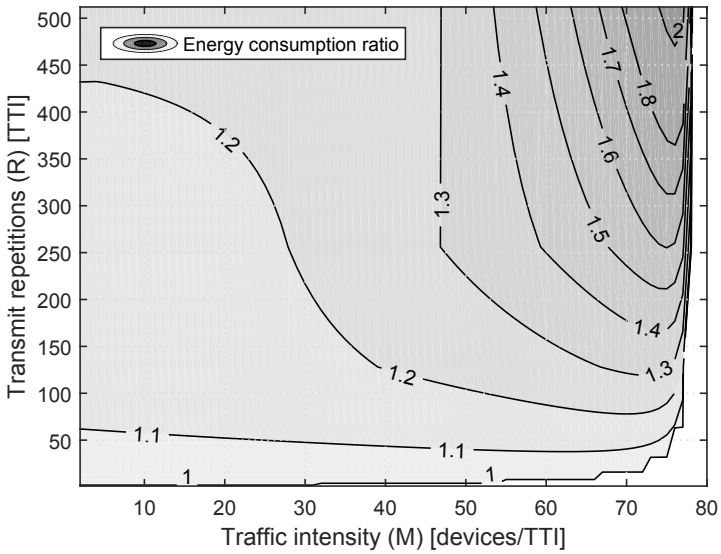


Fig. J.7: Energy consumption ratio per transmission of the radio power model with the non-constant PA efficiency intended for MMTC devices over the commonly used radio power model with a constant PA efficiency model. Mode A is used.

It should be noted that the energy consumption difference between the PA efficiency models only comes from the difference in energy consumption in TX state. But the energy consumption ratio depends on the energy consumption of the other states such as RX and $Idle$. For instance if the time

spend in RX is increased to $T_{RX} = 1\,\text{s}$ the average energy consumption ratio becomes 1.1 and the maximum ratio is 1.75. If T_{RX} is further increased to $T_{RX} = 2\,\text{s}$ the average ratio is 1.07 and the maximum ratio becomes 1.5.

4 Discussion

The results shown in this paper demonstrate the generic energy evaluation methodology and emphasize the impact of uplink power control, link-level performance and the power model. This section will discuss the results and the implications of the methodology.

One of the assumptions is the path-loss which is set to 154 dB in the results. The reason for selecting the path-loss as one value is that shadow fading is omitted as it is assumed that the power control is capable of compensating for this. Path-loss selected as a single value can also be interpreted as an upper bound of the shadow fading. Introducing fading as a random variable corresponds to introducing imperfect power control.

Perfect power control means, in this paper, that the devices are capable of doing perfect path-loss estimation and always have perfect knowledge of M and the path-loss. However in reality the power control might be inaccurate [10], in terms of e.g. path-loss estimation or knowledge of M. The effect of an imperfect power control will be that devices will select non-optimal transmit powers. Even if the average transmit power is the same as the optimal transmit power, the average amount of devices which fulfils the target performance will decrease. The overall consequence of a non-optimal transmit power is a lower outage capacity and a higher energy consumption.

The effect of using the non-constant PA efficiency model depends on the energy consumption of the transmit state compared to the other states (RX, $Idle$ and PSM) and the relative difference to the constant PA efficiency model. The relative difference of P_{TX} decreases as the transmit power increases (Fig. J.4). However, when the transmit power increases the impact of P_{TX} in E_{tot} (J.3) increases. So when the transmit power increases on average, the average ratio between the power models (Fig. J.7) will also increase as long that $Ptx < Ptx_{max}$. The maximum ratio will however not change if $Ptx = Ptx_{max}$ is already present.

Throughout the paper it is assumed that all devices uses the same number of transmit repetitions. Translated into a cellular deployment it will correspond to a group of devices which uses the same configuration, but are orthogonal to other groups and devices in the cell. If the path-loss is not the same for all devices in the group, the power control really proves its worth as it allows the devices to regulate themselves such that the received signal strengths after transmit repetitions at the BS receiver from the devices are still equally strong. When all devices in the group use the same number of trans-

mit repetitions, the resource usage will also be fixed. This can, however, be optimized if devices are configured depending on their coverage conditions.

The proposed methodology is device centric and focuses on the uplink transmissions. The example evaluation done in this paper is for one uplink transmission being transmitted with transmit repetitions. The example considers what happens when the device is sleeping, has synchronized, read the broadcast information and control channel, received downlink traffic, transmitted its uplink transmission and returned back to sleep. It is possible to use this methodology for a protocol consisting of multiple uplink and downlink transmissions. Downlink transmissions and their energy consumption impact are included as a parameter in the power model, where the most important parameter to change is how long time the device needs to be active. If the uplink transmissions utilize different modulation and coding scheme, multiple link-level performance curves will be needed.

The outcome of the proposed generic energy evaluation methodology is to make it easier to compare and evaluate standards and protocols for MMTC. The proposed methodology is simple, yet it includes important factors that affects the energy consumption. This means that evaluations performed with this methodology are more realistic than those done with the existing energy evaluation methodology from [3]. Further the outcomes of this paper can and should be used as input when new MMTC protocols and standards are being developed.

5 Conclusion and Outlook

This paper has presented a generic energy evaluation methodology tailored for MMTC. The methodology is demonstrated with special focus on three important factors which affect the energy consumption evaluation; power control, link-level performance and radio power model. The results presented in this paper provide important take-away messages:

- Using the commonly used radio power model with a constant PA energy efficiency instead of the radio power model with a non-constant PA model intended for MMTC radios can result in an overestimation of the battery life up to 100 % and on average 15 % across traffic intensity and transmit repetitions configurations. The proposed PA efficiency model does, however, need to be validated using a similar approach as used in [12] when NB-IoT or LTE-M devices become available.

- The combination of link-level performance, power control and the radio power model with a PA model intended for MMTC devices, provides options to optimize both access capacity and energy consumption.

Future work involves applying the proposed methodology on concrete protocols and help the development of future cellular MMTC solutions. These could be for example, new schemes and protocols such as one-stage and two-stage access protocols by [15]. For simplicity, in the presented evaluation, all devices are assumed to use the same radio access configuration. This can be generalized and interpreted as a group of devices within a larger set of MMTC devices. Our future work will focus on how cell radio resource management and higher layer protocol mechanisms can help minimizing the device energy consumption when several groups of devices are considered.

Acknowledgments

This work has been partly performed in the framework of the Horizon 2020 project FANTASTIC-5G (ICT-671660) receiving funds from the European Union. The authors would like to acknowledge the contributions of their colleagues in the project, although the views expressed in this contribution are those of the authors and do not necessarily represent the project.

References

[1] 3GPP: Study on Enhancements to Machine Type Communications and other Mobile Data Applications. TR 37.869 V12.0.0 (09 2013)

[2] 3GPP: Study on Enhancements to Machine Type Communications and other mobile data applications communications enhancements. TR 23.887 V12.0.0 (12 2013)

[3] 3GPP: Cellular System Support for Ultra Low Complexity and Low Throughput Internet of Things. TR 45.820 V2.1.0 (08 2015)

[4] 3GPP: E-UTRA: Physical channels and modulation. TS 36.211 V12.8.0 (12 2015)

[5] 3GPP: Narrowband LTE - Synchronization Channel Design and Performance. R1-156009 (10 2015)

[6] 3GPP: NB-PBCH design for NB-IoT. R1-160441 (02 2016)

[7] 3GPP: UE battery life evaluation for mMTC use cases. R1-165008 (05 2016)

[8] Biral, A., Centenaro, M., Zanella, A., Vangelista, L., Zorzi, M.: The challenges of M2M massive access in wireless cellular networks. Digital Communications and Networks 1(1), 1–19 (2015)

[9] Fantastic5G: Technical Results for Service Specific MultiNode/Multi-Antenna Solutions. Public Deliverable D4.1, H2020-ICT-2014-2 (06 2016)

[10] Holma, H., Toskala, A.: WCDMA for UMTS: HSPA Evolution and LTE. John Wiley & Sons, Inc. (2007)

[11] Jover, R., Murynets, I.: Connection-less communication of IoT devices over LTE mobile networks. In: Sensing, Communication, and Networking (SECON), 2015 12th Annual IEEE International Conference on. pp. 247–255 (06 2015)

[12] Lauridsen, M., Noël, L., Sørensen, T., Mogensen, P.: An Empirical LTE Smartphone Power Model with a View to Energy Efficiency Evolution. Intel Technology Journal 18(1), 172–193 (03 2014)

[13] Laya, A., Alonso, L., Alonso-Zarate, J.: Is the random access channel of LTE and LTE-A suitable for M2M communications? A survey of alternatives. Communications Surveys Tutorials, IEEE 16 (2014)

[14] Madueno, G., Stefanovic, S., Popovski, P.: Efficient LTE access with collision resolution for massive M2M communications. In: Globecom Workshops (GC Wkshps), 2014. pp. 1433–1438 (12 2014)

[15] Saur, S., Weber, A., Schreiber, G.: Radio access protocols and preamble design for machine type communications in 5G. In: IEEE 49th Asilomar Conference on Signals, Systems, and Computers (11 2015)

[16] Zanella, A., Zorzi, M., dos Santos, A., Popovski, P., Pratas, N., Stefanovic, C., Dekorsy, A., Bockelmann, C., Busropan, B., Norp, T.: M2M massive wireless access: Challenges, research issues, and ways forward. In: Globecom Workshops (GC Workshops), 2013 IEEE. pp. 151–156 (Dec 2013)

ISSN (online): 2446-1628
ISBN (online): 978-87-7210-448-5

AALBORG UNIVERSITY PRESS