Enabling Ultra-Reliable Low-Latency Communications in 5G Networks

Hamidreza Shariatmadari
Enabling Ultra-Reliable Low-Latency Communications in 5G Networks

Hamidreza Shariatmadari

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Electrical Engineering, at a public examination held at the lecture hall AS1 of the school on 1.2.2019 at 12:00.

Aalto University
School of Electrical Engineering
Department of Communications and Networking
Department of Communications and Networking
Supervising professor
Professor Riku Jäntti, Aalto University, Finland

Thesis advisor
Dr. Sassan Iraji, Intel, Germany

Preliminary examiners
Professor Tapani Ristaniemi, University of Jyväskylä, Finland
Dr. Yulin Hu, RWTH Aachen University, Germany

Opponent
Professor James Gross, Royal Institute of Technology (KTH), Sweden
Abstract

The fifth generation (5G) of cellular networks aims at providing connectivity for a large number of applications. To achieve this goal, 5G has been designed considering three main service categories, known as enhanced mobile broadband (eMBB), massive machine-type communications (mMTC), and ultra-reliable low-latency communications (URLLC). The target for URLLC support in 5G is to deliver small packets within a short time (up to 1 ms) with high probability of success (at least 99.999%). The support of URLLC, along with other services, brings new challenges that should be addressed. This thesis studies some of the challenges and proposes practical solutions to them.

The research in this thesis considers the resource allocations for hybrid automatic repeat request (HARQ) transmission schemes for the efficient support of URLLC. The HARQ-based transmissions can achieve a high transmission rate while ensuring the communication reliability. It is, however, essential to consider different types of errors for resource allocations. This motivates employing a new link adaptation scheme, which considers jointly the errors of the data and the control channels.

Time allocation is another concern for ensuring the end-to-end communication latency for two users operating in cellular mode. Conventionally, the time budget is divided equally and assigned to the uplink and downlink. This approach is not efficient for URLLC due to the stringent latency budget. Instead, two novel time allocation schemes are proposed that outperform the conventional approach. The communication between two users can be facilitated by device-to-device (D2D) communication when they are in close vicinity. The performance of D2D can be enhanced by exploiting the diversity for the data retransmissions.

The design of the control channel is another important topic that is considered. The accuracy of control channels directly affects overall communication reliability. Generally, higher accuracy can be achieved by allocating more radio resources to the control channels. However, this approach can significantly reduce the communication efficiency of URLLC, as the amount of resources for the control channels becomes comparable to that allocated to the data channels. Instead, non-trivial solutions can be exploited to improve the performance of control channels for URLLC support.

Blind-transmission schemes can reduce the communication latency and achieve high reliability without relying on a feedback channel. These are achieved by transmitting replicas of the data message using different resources in both time and frequency domains. The reliability of different types of blind-transmissions are evaluated.

Keywords 5G, Control Channels, Machine-Type Communications, Radio Resource Allocations, Ultra-Reliable Low-Latency Communication
Preface

The research work for this doctoral dissertation has been carried out in the Department of Communications and Networking (COMNET) of the School of Electrical Engineering at Aalto University, Espoo, Finland. The work was partially funded by the Finnish Funding Agency for Technology and Innovation (TEKES), the European Institute of Innovation and Technology (EIT Digital), Business Finland, and the Nokia foundation, for which I am very grateful to them.

First of all, I would like to express my sincere gratitude to my Professor, Riku Jäntti, for the giving me the opportunity to pursue the doctoral degree under his supervision. I really appreciate his guidance, help, and support in all matters. I would also like to thank Dr. Sassan Iraji for his instruction during my research work and also for his guidance in my life as a friend.

During my studies, I was involved in some projects in collaboration with other universities and companies, such as Aalborg University, KTH University, Ericsson, Orange, and Nokia Bell Labs. They were all great opportunities for me to work with many experienced people. I am thankful to everyone for their invaluable insights and contributions to my research work. In particular, I would like to thank Dr. Mikko A. Uusitalo and Dr. Zexian Li from Nokia Bell Labs for their guidance. I want to thank all the co-authors, specially, Prof. Tarik Taleb, Prof. Klaus Pedersen, Dr. Ruifeng Duan, Dr. Guillermo Pocovi, Dr. Andres Laya, Dr. Klaus Hugl, Dr. Rapeepat Ratasuk, and Dr. Amitava Ghosh. I would also like to thank Prof. Petar Popovski for giving me the opportunity to stay with his team at Aalborg University as a research visitor, and also for the fruitful discussions that we had. I would like to express my sincere gratitude to all my colleagues at Aalto University.

I am extremely thankful to the pre-examiners Prof. Tapani Ristaniemi and Dr. Yulin Hu for their invaluable comments and recommendations.
for improving the quality of this dissertation.

Finally, I would like to express my sincere gratitude to my family members, my parents and my sister, for their endless love. I am also very grateful to all my friends for their support.

Espoo, December 22, 2018,

Hamidreza Shariatmadari
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>i</td>
</tr>
<tr>
<td>Contents</td>
<td>iii</td>
</tr>
<tr>
<td>List of Publications</td>
<td>vii</td>
</tr>
<tr>
<td>Author's Contribution</td>
<td>ix</td>
</tr>
<tr>
<td>List of Abbreviations and Symbols</td>
<td>xi</td>
</tr>
<tr>
<td><strong>1. Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Background and Motivation</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Objective and Contribution of the Dissertation</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Dissertation Structure</td>
<td>3</td>
</tr>
<tr>
<td><strong>2. 5G Overview</strong></td>
<td>5</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>5</td>
</tr>
<tr>
<td>2.2 Communication Services</td>
<td>5</td>
</tr>
<tr>
<td>2.2.1 eMBB</td>
<td>6</td>
</tr>
<tr>
<td>2.2.2 mMTC</td>
<td>6</td>
</tr>
<tr>
<td>2.2.3 URLLC</td>
<td>6</td>
</tr>
<tr>
<td>2.3 5G NR</td>
<td>7</td>
</tr>
<tr>
<td>2.3.1 Spectrum</td>
<td>7</td>
</tr>
<tr>
<td>2.3.2 Massive MIMO</td>
<td>8</td>
</tr>
<tr>
<td>2.3.3 Waveform</td>
<td>8</td>
</tr>
<tr>
<td>2.3.4 Channel Coding</td>
<td>8</td>
</tr>
<tr>
<td>2.3.5 Slot Structure</td>
<td>8</td>
</tr>
<tr>
<td><strong>3. Radio Resource Allocations</strong></td>
<td>11</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>11</td>
</tr>
<tr>
<td>3.2 Finite-block Length Bounds</td>
<td>11</td>
</tr>
</tbody>
</table>
3.3 Resource Allocations for Single-link Operation ........ 12
  3.3.1 Non-adaptive Retransmission ................ 13
  3.3.2 Adaptive Retransmission .................... 15
3.4 Time and Resource Allocations for Cellular-mode Operation 18
  3.4.1 Fixed-transmission Rounds ................... 19
  3.4.2 Adaptive-transmission Rounds ............... 22
3.5 Link Adaptation in Cellular Systems .................. 24
  3.5.1 Optimal Link Adaptation .................... 25
  3.5.2 Proposed Link Adaptation Scheme ............. 26
3.6 Mode Selections .................................... 28
  3.6.1 Cellular Mode ................................ 29
  3.6.2 D2D Mode .................................... 31
  3.6.3 Hybrid Mode .................................. 32
  3.6.4 Results ....................................... 32
3.7 Discussion ......................................... 33

4. Control Channel Enhancements 35
  4.1 Introduction .................................... 35
  4.2 Control Channel Errors and Reliability Requirements . 35
    4.2.1 Uplink Transmissions ....................... 36
    4.2.2 Downlink Transmissions ..................... 38
  4.3 Scheduling Request ................................ 40
    4.3.1 Quality of Service (QoS) Based SR ........... 41
    4.3.2 Group-based SR ................................ 41
  4.4 Resource Grant .................................... 42
    4.4.1 Supporting Higher Aggregation Levels ....... 43
    4.4.2 Semi-persistent Scheduling and Fast Uplink Access . 43
    4.4.3 In-resource Control Signaling ................ 43
    4.4.4 Joint Data and Control Coding ............... 44
    4.4.5 DCI Repetition with DTX Detection .......... 44
  4.5 ACK/NACK Feedback ................................ 45
    4.5.1 Early ACK/NACK Transmission ............... 46
    4.5.2 Multi-bit ACK/NACK ......................... 46
    4.5.3 Asymmetric ACK/NACK Detection ............. 46
  4.6 CQI Report ....................................... 49
    4.6.1 Delay-based Detection ....................... 51
    4.6.2 Early Termination ......................... 51
  4.7 Discussion ....................................... 52
5. Blind Downlink Transmission Schemes 55
   5.1 Introduction ............................ 55
   5.2 HARQ Transmissions ........................ 55
   5.3 Blind Transmissions ........................ 56
      5.3.1 Normal DCI Transmission ............... 56
      5.3.2 DCI Bundling ............................ 57
      5.3.3 Repeated DCI Bundling ................. 58
   5.4 Results .................................. 58
   5.5 Discussion ............................... 60

6. Conclusions and Future Research Directions 61

References 65

Errata 69

Publications 71
This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.


V Hamidreza Shariatmadari, Ruifeng Duan, Zexian Li, Sassan Iraji, Mikko A. Uusitalo, and Riku Jänni. Analysis of transmis-


Author’s Contribution

Publication I: “Machine-type communications: current status and future perspectives toward 5G systems”

The author reviewed the literature and organized the manuscript.

Publication II: “Resource allocations for ultra-reliable low-latency communications”

The author formulated the problems and derived the optimizations.

Publication III: “Optimized transmission and resource allocation strategies for ultra-reliable communications”

The author developed the resource-allocation strategies and compared their performances.

Publication IV: “Link adaptation design for ultra-reliable communications”

The author developed the link adaptation algorithm and evaluated its performance.
Publication V: “Analysis of transmission modes for ultra-reliable communications”

The author developed the transmission schemes and evaluated their performances.

Publication VI: “Control channel enhancements for ultra-reliable low-latency communications”

The author derived the reliability requirements for the control channels and developed mechanisms to enhance the performance of the control channels.

Publication VII: “Fifth-generation control channel design; Achieving ultrareliable low-latency communications”

The author organized the manuscript and developed solutions for improving the design of the control channels.

Publication VIII: “Achieving ultra-reliable low-latency communications: Challenges and envisioned system enhancements”

The author contributed in organizing the manuscript and writing the control channel section.

Publication IX: “Asymmetric ACK/NACK detection for ultra-reliable low-latency communications”

The author developed the optimization problem and performed the simulations.

Publication X: “Statistical analysis of downlink transmissions for ultra-reliable low-latency communications”

The author derived the analytical expressions for the reliability of the communication schemes and performed the simulations.
# List of Abbreviations and Symbols

## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3GPP</td>
<td>Third Generation Partnership Project</td>
</tr>
<tr>
<td>4G</td>
<td>4th generation</td>
</tr>
<tr>
<td>5G</td>
<td>5th generation</td>
</tr>
<tr>
<td>ACK</td>
<td>Acknowledgment</td>
</tr>
<tr>
<td>ARQ</td>
<td>Automatic repeat request</td>
</tr>
<tr>
<td>BER</td>
<td>Bit error rate</td>
</tr>
<tr>
<td>BLER</td>
<td>Block error rate</td>
</tr>
<tr>
<td>BPSK</td>
<td>Binary phase shift keying</td>
</tr>
<tr>
<td>CCE</td>
<td>Control channel element</td>
</tr>
<tr>
<td>CP</td>
<td>Cyclic prefix</td>
</tr>
<tr>
<td>CQI</td>
<td>Channel quality indicator</td>
</tr>
<tr>
<td>CRC</td>
<td>Cyclic redundancy check</td>
</tr>
<tr>
<td>CSI</td>
<td>Channel state information</td>
</tr>
<tr>
<td>D2D</td>
<td>Device-to-device</td>
</tr>
<tr>
<td>DCI</td>
<td>Downlink control information</td>
</tr>
<tr>
<td>DFT</td>
<td>Discrete Fourier transform</td>
</tr>
<tr>
<td>DTX</td>
<td>Discontinuous transmission</td>
</tr>
<tr>
<td>eMBB</td>
<td>Enhanced mobile broadband</td>
</tr>
<tr>
<td>FDD</td>
<td>Frequency division duplex</td>
</tr>
<tr>
<td>gNB</td>
<td>5G base station</td>
</tr>
<tr>
<td>HARQ</td>
<td>Hybrid automatic repeat request</td>
</tr>
<tr>
<td>ICI</td>
<td>Inter-carrier interference</td>
</tr>
<tr>
<td>IR</td>
<td>Incremental redundancy</td>
</tr>
<tr>
<td>ISI</td>
<td>Inter-symbol interference</td>
</tr>
<tr>
<td>IoT</td>
<td>Internet of Things</td>
</tr>
<tr>
<td>LDPC</td>
<td>Low-density parity-check</td>
</tr>
</tbody>
</table>
List of Abbreviations and Symbols

LLR  Log-likelihood ratio
LTE  Long Term Evolution
M2M  Machine-to-machine
MAC  Medium access control
MBB  Mobile broadband
MCS  Modulation and coding scheme
M-MIMO  Massive MIMO
MIMO  Multiple input multiple output
mMTC  Massive machine-type communications
MTC  Machine-type communications
NACK  Negative ACK
NB-IoT  Narrow-band Internet of Things
NR  New radio
OFDM  Orthogonal frequency division multiplexing
PDCCH  Physical downlink control channel
PDU  Protocol data unit
PHY  Physical
PDCCH  Physical downlink control channel
PDSCH  Physical downlink shared channel
PUCCH  Physical uplink control channel
PUSCH  Physical uplink shared channel
QoS  Quality of service
RE  Resource element
REG  Resource element group
RLC  Radio link control
RS  Reference signal
RTR  Remaining transmission rounds
RV  Redundancy version
SCS  Subcarrier spacing
SINR  Signal-to-interference-plus-noise ratio
SPS  Semi-persistent scheduling
SR  Scheduling request
SRS  Sounding reference signal
TDD  Time division duplex
TSN  Time-sensitive networking
TTI  Transmission time interval
UDN  Ultra-dense networks
UE  User equipment
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>URLLC</td>
<td>Ultra-reliable low-latency communications</td>
</tr>
<tr>
<td>VoIP</td>
<td>Voice over Internet protocol</td>
</tr>
<tr>
<td>VoLTE</td>
<td>Voice over LTE</td>
</tr>
</tbody>
</table>
Symbols

$\mathcal{N}$ Gaussian distribution
$C$ Shannon capacity
$E$ BLER
$G$ Channel use gain
$k$ Message size
$l$ Number of DCI transmissions
$L$ Maximum number of data transmissions
$m_1$ The maximum number of transmission rounds in uplink
$m_2$ The maximum number of transmission rounds in downlink
$M$ Total number of transmission rounds in uplink and downlink
$MCS_i$ $i$th MCS
$n$ Blocklength
$n_1$ Channel uses for the initial data transmission
$n_2$ Channel uses for the data retransmission
$\vec{n}$ Vector of channel uses for the transmission rounds
$\vec{n}_1$ Vector of channel uses for the uplink transmission rounds
$\vec{n}_2$ Vector of channel uses for the downlink transmission rounds
$N$ Maximum number of data transmissions
$\overline{N}$ Average channel uses
$\overline{N}_{\text{Cellular}}$ Average channel uses in the cellular mode
$\overline{N}_{\text{Hybrid}}$ Average channel uses in the hybrid mode
$p_{\text{D}}(i)$ Residual BLER after $i$th data transmission round
$P_1$ Residual BLER using data from initial transmission round
$P_{1,2}$ Residual BLER using data from initial transmission and retransmission rounds
$P_{1,i}$ Residual BLER after initial transmission using $MCS_i$
$P_2$ Residual BLER using data from retransmission round
$P_{2,j}$ Residual BLER after retransmission using $MCS_j$
$P_{2D}$ Residual BLER when a DTX is detected
$P_{2N}$ Residual BLER when a NACK is detected
$P_{\text{AN}}$ Probability of decoding an ACK as a NACK
$P_{\text{NA}}$ Probability of decoding a NACK as an ACK
$P_{\text{success}}$ Success probability of message delivery
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P^\text{Blind(a)}_{\text{success}}$</td>
<td>Success probability of message delivery for blind transmissions with normal DCI transmissions</td>
</tr>
<tr>
<td>$P^\text{Blind(b)}_{\text{success}}$</td>
<td>Success probability of message delivery for blind transmissions with DCI bundling</td>
</tr>
<tr>
<td>$P^\text{Blind(c)}_{\text{success}}$</td>
<td>Success probability of message delivery for blind transmissions with repeated DCI bundling</td>
</tr>
<tr>
<td>$P^\text{DL}_{\text{success}}$</td>
<td>Success probability of message delivery in downlink</td>
</tr>
<tr>
<td>$P^\text{UL}_{\text{success}}$</td>
<td>Success probability of message delivery in uplink</td>
</tr>
<tr>
<td>$P^\text{HARQ}_{\text{success}}$</td>
<td>Success probability of message delivery for HARQ transmissions</td>
</tr>
<tr>
<td>$P^\text{Hybrid}_{\text{success}}$</td>
<td>Success probability of message delivery in hybrid mode</td>
</tr>
<tr>
<td>$Q$</td>
<td>$Q$-function</td>
</tr>
<tr>
<td>$r_i$</td>
<td>Transmission rate for $MCS_i$</td>
</tr>
<tr>
<td>$s$</td>
<td>Employed MCS</td>
</tr>
<tr>
<td>$T$</td>
<td>Expected throughput</td>
</tr>
<tr>
<td>$V$</td>
<td>Channel dispersion</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>SINR</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>BLER target for the overall communication</td>
</tr>
<tr>
<td>$\epsilon_1$</td>
<td>BLER target for the uplink communication</td>
</tr>
<tr>
<td>$\epsilon_2$</td>
<td>BLER target for the downlink communication</td>
</tr>
<tr>
<td>$\epsilon_A$</td>
<td>Error probability of decoding an ACK</td>
</tr>
<tr>
<td>$\epsilon_{A,N}$</td>
<td>Error probability of decoding ACK/NACK signals</td>
</tr>
<tr>
<td>$\epsilon_{AN}$</td>
<td>Error probability of decoding an ACK as a NACK</td>
</tr>
<tr>
<td>$\epsilon_{A,N,D}$</td>
<td>Error probability of decoding ACK/NACK and DTX signals</td>
</tr>
<tr>
<td>$\epsilon_1$</td>
<td>Error probability of decoding a DCI</td>
</tr>
<tr>
<td>$\epsilon_D$</td>
<td>Error probability of decoding a message</td>
</tr>
<tr>
<td>$\epsilon_{DA}$</td>
<td>Error probability of decoding a DTX as an ACK</td>
</tr>
<tr>
<td>$\epsilon_{DN}$</td>
<td>Error probability of decoding a DTX as a NACK</td>
</tr>
<tr>
<td>$\epsilon_N$</td>
<td>Error probability of decoding a NACK</td>
</tr>
<tr>
<td>$\epsilon_{NA}$</td>
<td>Error probability of decoding a NACK as an ACK</td>
</tr>
<tr>
<td>$\epsilon_{ND}$</td>
<td>Error probability of decoding a NACK as a DTX</td>
</tr>
<tr>
<td>$\epsilon_{SR}$</td>
<td>Error probability of decoding a scheduling request</td>
</tr>
<tr>
<td>$\epsilon_{RG}$</td>
<td>Error probability of decoding a resource grant</td>
</tr>
<tr>
<td>$\nu$</td>
<td>Threshold for distinguishing ACK/NACK signals</td>
</tr>
<tr>
<td>$\xi$</td>
<td>BLER target for MCS selection</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Overall communication reliability target</td>
</tr>
</tbody>
</table>
1. Introduction

1.1 Background and Motivation

Conventional cellular systems, such as the fourth generation (4G) of cellular systems and their predecessors, were designed to mainly support human-centric applications, such as voice calls and mobile broadband (MBB) connectivity. These technologies are capable of offering high-speed connectivity for a limited number of users. Fifth generation (5G) cellular systems have been introduced with the intention of not only addressing the problems of the previous generations but also providing connectivity for machine-centric applications. Thus, 5G will deal with a massive number of applications that have diverse and sometimes very stringent communication requirements. To achieve these goals, the third generation partnership project (3GPP), which is responsible for cellular system standardizations, has specified three services categories for 5G. These categories are known as enhanced MBB (eMBB), massive machine-type communications (mMTC), and ultra-reliable low-latency communications (URLLC) [1], [2]. These services can fulfill objectives of International Mobile Telecommunications for 2020 and beyond (IMT-2020) and enable emerging applications in various domains [3].

URLLC is a new communication service with stringent requirements in terms of availability, reliability, and latency [4]. This service is essential for supporting mission-critical applications, such as industrial automation, augmented reality, remote surgery, and intelligent transport systems [5]. The initial target for supporting URLLC in 5G is providing communication reliability of 99.999% with a latency of less than 1 ms for small data payloads. However, more stringent communication requirements may be considered for URLLC in the future.
The 3GPP has considered a flexible network design with new features for 5G in order to accommodate different services, including URLLC. While these features bring flexibility for the services, there are still some challenges that should be addressed. For instance, resource allocation is an important concern for supporting URLLC, ensuring the communication efficiency while meeting the latency and reliability requirements. Control channel design is the other issue, since the overhead of the control channel can become more than the size of the data payload. Another challenge is the ubiquitous support of URLLC, as reliable communication links might not be available in some wireless scenarios. In addition, the support of URLLC should not degrade the performance of the other services that operate using the same network infrastructure.

1.2 Objective and Contribution of the Dissertation

The main objective of this dissertation is the efficient support of URLLC in the future cellular systems, such as 5G. This dissertation addresses some of the main challenges of supporting URLLC in cellular systems. In addition, it proposes new solutions and enhancements for achieving better URLLC support. The main contributions of the dissertation can be summarized as follows

1. Optimal resource allocations for hybrid automatic repeat request (HARQ) schemes, over a single link, are considered when the feedback channel is prone to errors. The communication efficiency is evaluated. In addition, the allocations of resources are considered for two users communicating in cellular mode. Two different approaches are proposed to efficiently allocate the time budget between the uplink and downlink. In addition, a simple link adaption scheme is proposed, which can be easily adopted in 5G. The communication efficiency of different communication scenarios, such as cellular, device-to-device (D2D), and hybrid modes, are compared.

2. Comprehensive communication models are considered in order to evaluate the effects of the errors from the data and control channels on the overall communication reliability. Using these models, the reliability requirements for the control information are identified in order to support a communication service. Furthermore, various enhancements are pro-
posed to improve the efficiency of control channels for URLLC support.

3. Blind transmission schemes are considered to reduce the communication latency and improve the reliability. Their reliability performances are compared with that which can be achieved using HARQ-based transmissions.

1.3 Dissertation Structure

The structure of the dissertation is as follows. Chapter 2 provides an overview of 5G. In particular, the requirements of the generic communication services are explained. The main features of 5G are introduced. Chapter 3 considers resource allocations for HARQ-based transmissions in order to support URLLC in different communication scenarios. In addition, a link adaption scheme is proposed for supporting URLLC in practical wireless systems. Chapter 4 presents comprehensive communication models, considering different error sources from the data and control channels. In this regard, the reliability requirements for the control information are identified in order to support URLLC. Various enhancements are proposed for control channels to support URLLC more efficiently. Chapter 5 evaluates the reliability performance of blind data transmission schemes, which do not require employing a feedback channel. Finally, Section 6 concludes the dissertation and describes some relevant future work.
Introduction
2. 5G Overview

2.1 Introduction

5G mobile communication systems are expected to provide connectivity for a wide range of applications. This is an important concern for mobile operators in order to scale up the number of connected devices and increase their revenue. In order to achieve this goal, the 3GPP has identified the general communication requirements for supporting the envisioned applications [6]. Accordingly, 5G has been designed to meet the identified requirements, and also other needs that might be imposed by the future applications. The 3GPP adopted the name new radio (NR) for 5G radio access network and standardized new features for it in release 15. These features bring flexibility for supporting different applications and also network deployment. This chapter first introduces the main generic services considered for 5G networks. Then, the main features of the 5G NR have been highlighted.

2.2 Communication Services

The 5G NR should bring support for different applications with a wide range of communication requirements. These requirements are associated to three generic service classes, known as, eMBB, mMTC, and URLLC. These service classes have been considered by the 3GPP as the main drivers for designing 5G.
2.2.1 eMBB

eMBB is a communication service that fulfills the need for high data rate transmissions. This is essential for many human-oriented applications, such as video streaming, virtual reliability, and online gaming. Indeed, the demand for an increased data rate was the main motivation for developing the earlier generations of cellular systems, including 3G and 4G. 5G aims at offering a peak data rate of 20 Gbps and ensuring a decent data rate (50-100 Mbps) of ubiquity for unfavorable conditions; e.g., in densely-populated areas and under high mobility. In addition, the target for user data traffic latency is 4 ms.

2.2.2 mMTC

mMTC refers to the communication service for connecting a large number of devices. This service can foster the realization of the Internet of Things (IoT) by providing efficient Internet connectivity. Massive connectivity is required for many applications involved with environmental observations, traffic monitoring, asset tracking, resource utility management, and the smart grid, to mention a few. These applications may need additional communication requirements on top of the massive connectivity. For instance, some of the devices are battery-powered and should run for a long time, e.g., a couple of years, without the need to change the batteries. Some other devices, such as smart meters, might be deployed in locations with high penetration loss. Hence, they need the extended network coverage for their operations. In addition, a low-cost transceiver is essential for realizing many IoT applications. Some of these features were partially addressed in Long Term Evolution (LTE) releases 13 and 14, by introducing new UE categories, such as UE CAT-0, Cat-M1, and NB-IoT [PI], [7].

2.2.3 URLLC

URLLC is a new communication service that is considered for the cellular systems. This service is essential for supporting many emerging mission-critical applications, such as industrial automation, remote surgery, and vehicular safety communications. URLLC has stringent requirements on availability, latency, and reliability [4]. In other words, the communication link should be available almost all the time to carry data in a
short period of time and with high reliability. The 3GPP aims at providing URLLC for short data payloads (up to 32 Bytes) with the user plane latency of 1 ms (measured at layer 2 and layer 3) and the reliability of $1 - 10^{-5}$ (99.999%). To ensure the latency target, a symmetric latency budget of 0.5 ms is specified for both uplink and downlink. Data must be transported at the higher layers fast and error-free with high levels of integrity, authenticity, and confidentiality [8]. In addition, the target for the control plane latency is specified as 10 ms, which is the time that it takes the user equipment (UE) starts transmitting data from a battery-efficient state. A zero interruption time is desired, ensuring that the UE can always exchange user plane data even with mobility. While these requirements are satisfactory for many mission-critical applications, there are some other applications that have more stringent requirements as well as additional needs. For instance, there is a great interest in supporting the time-sensitive networking (TSN) in 5G network. Supporting TSN imposes additional requirements, such as accurate synchronization and very low jitter [9]. Hence, the 3GPP may specify more stringent requirements for URLLC service in later 5G releases.

2.3 5G NR

The 5G NR has been designed to be flexible and forward-compatible. These enable introducing new services in the future along with the supported services. The 3GPP has standardized new features for the 5G NR in release 15. This section presents the main features of the 5G NR, highlighting the relevant enhancements that facilitate supporting URLLC.

2.3.1 Spectrum

The 5G NR is expected to operate over the licensed and unlicensed spectrum from sub-1 GHz up to 100 GHz. This ensures sufficient spectrum in order to satisfy the increasing demands of eMBB traffic. Generally, the sub-1 GHz frequency bands provide a wide area and deep indoor coverage, the 1-6 GHz frequency bands offer a compromised performance between the coverage and the capacity, and above 6 GHz frequency bands offer high data rates for specific uses. The high frequency bands also enable employing massive MIMO (M-MIMO) antenna systems with beamforming and ultra-dense networks (UDN).
2.3.2 Massive MIMO

Utilizing the higher frequency bands enables realizing the M-MIMO. In practice, the antenna size is reciprocal to the frequency. Hence, a large number of antennas can be accommodated at a base station when the system operates over the high frequency bands. This enables beamforming and simultaneously streaming data to different users, resulting in better spectrum utilization and higher capacity.

2.3.3 Waveform

The 5G NR adopts a cyclic prefix orthogonal frequency division multiplexing (CP-OFDM) waveform for both uplink and downlink transmissions [10]. The unified waveform simplifies the radio transceiver design and facilitates the support of device-to-device (D2D) communications and relay nodes. In addition, the discrete Fourier transform (DFT) spread OFDM is applicable in uplink for coverage enhancement. Other operations, such as windowing and filtering, can be applied on top of the CP-OFDM at the transmitter in order to enhance the spectrum confinement [11], [12].

2.3.4 Channel Coding

The 5G NR employs low-density parity-check (LDPC) code as the channel coding scheme for data channels. The employed LDPC is accompanied with the rate-compatible structure that permits the achievement of low code rates. Further, the repetition permits the achievement of code rates below 1/3. These features allow achieving high coding gains without suffering from the error floor that are essential for reliable communications. In addition, polar code is utilized for layer 1 and layer 2 control signaling, except for very short messages. Polar code can achieve the Shannon capacity for a wide range of channel conditions with reasonable decoding complexity. It also exhibits good performance for short block lengths, e.g., the general control signals.

2.3.5 Slot Structure

The 5G NR considers a new slot-structure design, applicable to both time division duplex (TDD) and frequency division duplex (FDD) transmissions. Some of the features of the new slot structure include scalable
subcarrier spacing (SCS), dynamic transmission time interval (TTI), and fast HARQ feedback. These features enable efficiently support different services, including URLLC.

The slot structure supports scalable SCS in order to encounter a wide range of spectrum. The SCS can be configured as 15, 30, 60, 120, and 240 kHz. The higher SCS can be adopted for the high carrier frequencies to better mitigate the inter-carrier interference (ICI), caused by the Doppler effect. In addition, multiple CP lengths are envisaged for each SCS value in order to adapt to various levels of inter-symbol interference (ISI) for different carrier frequencies and mobility [10].

A slot is comprised of 7 or 14 OFDM symbols for the SCS numerologies of 15, 30, and 60 kHz, while it is comprised of 14 OFDM symbols for the SCS numerologies of 120 and 240 kHz. Consequentially, the slot duration varies according to the employed SCS, ranging from 0.125 ms to 1 ms. There are three different symbol types: downlink, uplink, and flexible. The UE is assumed to receive information over the downlink and flexible symbols, while it shall transmit over uplink and flexible symbols [13]. The flexible symbols are configured according to the network preference.

The slot structure is accompanied with a wide range of configurations. Figure 2.1 represents the envisioned slot configurations according to the symbol type orders. The abundance of slot configurations brings flexibility for scheduling the users, the scheduling being either based on the slot or mini-slot. A mini-slot is formed from 1-13 OFDM symbols. The mini-slot scheduling is a promising feature for achieving low-latency communications, particularly, over low frequency bands as the slot durations are typically large. In addition, it allows performing the transmission immediately without waiting for the start of a new slot.
(a) A slot containing only downlink symbols.

(b) A slot containing only uplink symbols.

(c) A slot containing only flexible symbols.

(d) A slot containing downlink and flexible symbols.

(e) A slot containing downlink and flexible symbols.

(f) A slot containing flexible and uplink symbols.

(g) A slot containing flexible and uplink symbols.

(h) A slot containing downlink, flexible, and uplink symbols.

(i) Flexible slot structure.

Figure 2.1. Representation of slot configurations in 5G NR.
3. Radio Resource Allocations

3.1 Introduction

Resource management is an important part of cellular systems and encompasses several techniques that include scheduling, resource allocation, link adaptation, power allocation, and interference coordination [14]. In part, efficient resource allocation is an important concern for supporting URLLC. The main objective is to support the communication service while efficiently utilizing the limited radio resources in both time and frequency domains [15]. This directly affects the performance of the system and the number of users that can be served simultaneously. This chapter considers the resource allocations for HARQ-based transmission schemes, which are widely implemented in wireless systems. In addition, the efficiency of various transmission modes are analyzed.

3.2 Finite-block Length Bounds

Generally, wireless systems employ channel-coding schemes for the information carried over the wireless channels in order to achieve robustness against the noise and interference. A channel-coding scheme is applied to the input message containing $k$ information bits and forms a coded packet spanning over $n$ channel uses. The coding rate is defined as the ratio of the number of input bits to the number of channel uses, i.e., $R = k/n$ [PII]. The coding rate and achievable reliability are important criteria for evaluating the performance of the coding scheme that affects communication efficiency. Shannon determined the highest achievable rate for delivering large message sizes. This bound is generally utilized as an approximation of the coding rate achieved by the employed coding schemes in
cellular systems, such as the cyclic redundancy check (CRC) and Turbo coding schemes that are implemented in LTE. This allows evaluating the communication efficiency much easier. The Shannon capacity, however, cannot be applied to URLLC. This is due to the fact that URLLC considers short message sizes and does not meet the conditions for the Shannon bound. Instead, there are bounds for finite-block lengths that can be considered for evaluating URLLC. Such bounds can be approximated by [16], [17], [18], [19]

\[ R^*(n, \epsilon, \gamma) = C(\gamma) - \sqrt{\frac{V(\gamma)}{n}} Q^{-1}(\epsilon) + O\left(\frac{\log n}{n}\right), \]  

(3.1)

where \( n \) is the length of the block that is utilized for carrying \( k \) bits of information, \( \epsilon \) is the block error probability, and \( \gamma \) is the signal-to-interference-plus-noise ratio (SINR) at the receiver. In addition, \( C(\gamma) \) is the Shannon capacity, \( V(\gamma) \) is the channel dispersion\(^1\), and \( Q(.) \) denotes the Gaussian Q-function. In (3.1) the notation \( f(x) = O(g(x)) \) means that \( \limsup_{x \to \infty} |f(x)/g(x)| < \infty \). The Shannon capacity and the channel dispersion for an additive white Gaussian noise (AWGN) channel are given by [19]

\[ C(\gamma) = \frac{1}{2} \log(1 + \gamma), \]  

(3.2)

\[ V(\gamma) = \frac{\gamma - 1}{2 (\gamma + 1)^2} \log^2(e). \]  

(3.3)

In this dissertation, it is assumed that the employed channel coding schemes can achieve the maximum achievable rate, following the approximation (3.1). The approximation is tight for sufficiently large values of \( n \). In this case, the error probability of decoding a block carrying \( k \) bits of information when transmitted by a blocklength of \( n \) channels is

\[ E(n, k, \gamma) = Q\left(\frac{nC(\gamma) - k}{\sqrt{nV(\gamma)}}\right) \approx \epsilon. \]  

(3.4)

### 3.3 Resource Allocations for Single-link Operation

This section considers the resource allocations for incremental redundancy (IR) HARQ data transmissions. The communication is performed utilizing data and feedback channels. The data channel carries the actual data information, while the feedback channel carries the acknowledgment (ACK) and negative ACK (NACK) signals. In order to meet the

\(^1\)It is a measure of the stochastic variability of the channel relative to a deterministic channel with the same capacity [19].
latency constraint, only two data transmission rounds are envisaged, i.e, having only one data retransmission opportunity. A receiver should be able to retrieve the message successfully with the probability of $\rho = 1 - \epsilon$, where $\epsilon$ is the remaining block error rate (BLER) after all possible transmission rounds. The receiver sends either an ACK or a NACK after the first data transmission, in order to indicate the success or failure in decoding the message. However, the feedback signal might be decoded wrongly. Two error types are considered: 1) the erroneous decoding of an ACK as a NACK, and 2) the erroneous decoding of a NACK as an ACK. The former error leads to triggering the unnecessary data retransmission, which does not sacrifice the reliability of retrieving the message while wasting the data radio resources. The latter error results in terminating the essential data retransmission, which reduces the communication reliability.

It is assumed that the error rates of decoding the ACK and NACK signals are $\epsilon_A$ and $\epsilon_N$, respectively, and are known at the transmitter. The transmitter also has full knowledge of the data channel quality prior to the data transmission, which can be achieved by reporting channel state information (CSI) by the transmitter. In addition, the time-invariant fading channel is assumed, in which the quality of the data channel remains the same for both transmission rounds. The transmitter should perform the resource allocations for the data transmission rounds, considering the errors of the data and control channels.

The HARQ-based transmissions can adopt non-adaptive and adaptive retransmission scheme. The non-adaptive retransmission scheme entails performing the data transmission rounds with the same data size. While, the adaptive retransmission scheme allows performing the data transmission rounds with different data sizes. The rest of this section considers the resource allocation problems separately for these two schemes [PII].

### 3.3.1 Non-adaptive Retransmission

In the non-adaptive retransmission scheme, all the transmission rounds for a single message delivery are allocated with the same amount of radio resources. This scheme is mainly utilized in LTE uplink data transmissions [14]. For instance, the transmitter can send a different redundancy version (RV) of the encoded message in each transmission/retransmission round, while all the versions occupy the same amount of resources. The data retransmission continues until an ACK is detected or the maximum number of data transmission rounds is achieved.
Assume that the transmitter utilizes \( n \) channel uses for each data transmission/retransmission round for delivering \( k \) bits of information. For the case of having only one retransmission opportunity, the expected utilized channel uses can be expressed as

\[
N(n, k, \gamma) = n\{1 + E(n, k, \gamma)(1 - \epsilon_N) + (1 - E(n, k, \gamma))\epsilon_A\}. \tag{3.5}
\]

Accordingly, the success probability of retrieving the message by the receiver is

\[
P_{\text{success}} = \left(1 - E(n, k, \gamma)\right) + E(n, k, \gamma)(1 - \epsilon_N)\left(1 - \frac{E(2n, k, \gamma)}{E(n, k, \gamma)}\right) = 1 - E(2n, k, \gamma) - \epsilon_N(E(n, k, \gamma) - E(2n, k, \gamma)). \tag{3.6}
\]

The performance can be optimized by minimizing the expected channel uses, while ensuring the reliability target. The optimization is expressed as

\[
\min_n N(n, k, \gamma),
\]

subject to:

\[
P_{\text{success}} \geq 1 - \epsilon. \tag{3.7}
\]

A closed-form solution for this optimization problem is not easy to be derived. Here, only numerical results are presented. However, a close approximation for this solution can be found in [PII].

Figure 3.1 illustrates the BLER for the initial transmission round for carrying 80 and 800 bits of information and the overall reliability targets corresponding to the BLER of \( 10^{-5} \) and \( 10^{-7} \). It is assumed that the error probabilities of the ACK and NACK signals are the same, i.e, \( \epsilon_{A/N} = \epsilon_A = \epsilon_N \). The figure indicates that the BLER for the initial transmission round does not change significantly for different message sizes, while it changes according to the reliability target and the feedback error rate. For a given reliability target, two regions can be defined, which are separated with the dashed line in the figure. On the left side of that, the BLER for the initial transmission round is almost constant regardless of the feedback error rate. This is due to the fact that the retransmission can be triggered properly. Hence, the feedback channel is identified as reliable in this region. On the contrary, the BLER for the initial transmission round changes on the right side of the dashed line. In this region, it is not possible to only rely on the feedback channel to trigger the retransmission round due to the high feedback error rate. Hence, the initial transmission should be performed more robustly. The feedback channel is identified as unreliable in this region.
3.3.2 Adaptive Retransmission

In the adaptive retransmission scheme, the amounts of allocated radio resources for the transmission rounds can be different. This brings more flexibility for the resource allocations compared to the non-adaptive retransmission scheme. The adaptive retransmission scheme is employed particularly for LTE downlink data transmissions [14].

Assume that the transmitter intends to deliver $k$ bits of information to the receiver. The message should be delivered with a maximum of two transmission attempts, meeting the reliability target. The transmitter performs the initial data transmission utilizing $n_1$ channel uses, while it performs the data retransmission utilizing $n_2$ channel uses in case it detects a NACK. Accordingly, the expected channel uses can be expressed as

$$N(n_1, n_2, k, \gamma) = n_1 + n_2 \{E(n_1, k, \gamma) (1 - \epsilon_N) + (1 - E(n_1, k, \gamma)) \epsilon_A\}. \quad (3.8)$$
Figure 3.2. The normalized expected of channel uses for non-adaptive retransmission with respect to single-round data transmission, where $k = 80$ bits and $\epsilon = 10^{-5}$ [PII].

The success probability of retrieving the message by the receiver is

$$P_{\text{success}} = \left(1 - E(n_1, k, \gamma)\right) + E(n_1, k, \gamma) \left(1 - \frac{E(n_1 + n_2, k, \gamma)}{E(n_1, k, \gamma)}\right)$$

$$= 1 - E(n_1 + n_2, k, \gamma) - \epsilon N \left(E(n_1, k, \gamma) - E(n_1 + n_2, k, \gamma)\right).$$

(3.9)

The optimization problem can be formed such that the expected channel uses is minimized while the reliability target is met. This leads to the following optimization problem

$$\text{minimize } N(n_1, n_2, k, \epsilon),$$

subject to:

$$P_{\text{success}} \geq 1 - \epsilon.$$  

(3.10)

A closed-form solution for this optimization is difficult to be derived. Here, the numerical results are presented. However, a close approximation for the solution can be found in [PII].

Figure 3.3 illustrates the BLER for the initial transmission associated with the optimal resource allocations. The BLER is derived for delivering messages with the sizes of 80 and 800 bits. The overall communication reliabilities are associated with the BLER of $10^{-5}$ and $10^{-7}$. These results indicate that the BLER for the initial transmissions does not change significantly for different message sizes. However, it changes according to
Radio Resource Allocations

the overall communication target and the reliability of the feedback channel. For a given reliability target, two regions can be defined, which are divided by the dashed line. On the left side of the dashed line, the BLER for the initial transmission round is almost constant regardless of the feedback channel error. In this region, the BLER for the initial transmission is around 10%. This is due to the fact that the feedback channel can properly trigger the data retransmission, hence the initial transmission can be performed with a moderate reliability level. The feedback channel is considered as reliable in this region. On the other hand, the BLER for the initial transmission changes on the right side of the dashed line according to the error of the feedback channel. In this region, the feedback channel cannot properly trigger the retransmission round. Consequently, the initial transmission round should be performed more robustly to compensate the error of the feedback channel. The feedback channel is identified as unreliable in this region. The figure suggests that the state of the feedback channel in terms of the reliability/unreliability can be identified with respect to the tolerable communication error, i.e., $\epsilon$. For a given reliability target, a feedback channel is reliable if $\epsilon_{A/N} < 10\epsilon$, otherwise it is unreliable.

![Figure 3.3](image_url)

**Figure 3.3.** The BLER for the initial transmission round associated with the optimal resource allocations for adaptive retransmission scheme, where SINR=0 dB [PII].

Figure 3.4 illustrate the normalized channel uses of adaptive retransmission with respect to single-round data transmission for delivering
Radio Resource Allocations

$k = 80$ of information with the overall BLER of $10^{-5}$. The adaptive retransmission scheme utilizes significantly less radio resources compared to that with single-round data transmission, particularly, when the SINR is low and the feedback channel is reliable. In addition, the adaptive retransmission scheme has slightly better performance compared to the non-adaptive retransmission scheme (compare Figure 3.2 with Figure 3.4).

![Figure 3.4](image)

**Figure 3.4.** The normalized expected of channel uses for adaptive retransmission with respect to single-round data transmission, where $k = 80$ bits and $\epsilon = 10^{-5}$ [PII].

### 3.4 Time and Resource Allocations for Cellular-mode Operation

This section considers the time and resource allocations for supporting URLLC between two UEs operating in cellular mode. The base station receives the message from the transmitter UE and then forwards it to the receiver UE, applying encode-and-decode approach. Figure 3.5 illustrates the general model for cellular-mode operation. The time and resource allocations are considered in order to meet the latency and reliability requirements. The latency requirement can be expressed by the maximum number of transmission attempts for both uplink and downlink. It is assumed that the latency budget allows having $M$ transmissions attempts for both links, i.e, $m_1 + m_2 \leq M$, where $m_1$ and $m_2$ are the maximum number of transmission attempts for uplink and downlink, respectively.
The reliability requirement entails delivering the message with the success probability of $\rho$ within the latency budget, i.e., $(1 - \epsilon_1)(1 - \epsilon_2) \geq \rho$, where $\epsilon_1$ and $\epsilon_2$ are the failure probabilities of message delivery over the uplink and downlink, respectively.

![Diagram of communication model for cellular mode](image)

**Figure 3.5.** The communication model for cellular mode.

In the rest of this section, two different time allocation schemes are presented. The fixed time allocation scheme assigns fixed transmission time budgets for uplink and downlink, while, the adaptive time allocation scheme allows adaptively assigning time budgets for uplink and downlink. Both schemes are applicable to ARQ and HARQ transmissions. However, only the HARQ case is presented in this section. The ARQ case is covered in [PIII].

### 3.4.1 Fixed-transmission Rounds

In this approach, the time budgets for uplink and downlink data transmissions are fixed. This is equivalent to having a fixed number of transmission attempts for uplink and downlink, i.e., $m_1$ and $m_2$. Consider the adaptive retransmission scheme, in which a different number of channel uses can be allocated to each transmission attempt. For the link $i$ (1: uplink, 2: downlink) with the maximum number of $m_i$ transmission attempts, consider $\vec{n}_i = \{n_{i,1}, ..., n_{i,m_i}\}$, where $n_{i,j}$ denotes the channel uses in $j$th HARQ round. Assuming that the segmented data transmission can
meet the bound (3.4), the residual BLER after the \( j \)th HARQ round can be expressed as [20]

\[
\epsilon = E\left( \sum_{l=0}^{j} n_{i,l}, k, \gamma_i \right),
\]

(3.11)

where \( n_{i,0} = 0 \). This indicates that the residual BLER after the \( j \)th HARQ round depends on the amount of received information and not on the segmentation. The expected number of channel uses for this link can be expressed as

\[
\overline{N}_i(n_i, m_i, k, \gamma_i) = \sum_{j=1}^{m_i} n_{i,j} E\left( \sum_{l=0}^{j-1} n_{i,l}, k, \gamma_i \right).
\]

(3.12)

The time and resource allocations can be considered such that the sum of the expected channel uses in uplink and downlink is minimized, while the communication reliability is met. This can be achieved by considering the following optimization problem

\[
\arg\min_{\vec{n}_1, \vec{n}_2, m_1} \left\{ \overline{N}_1(n_1, m_1, k, \gamma_1) + \left( 1 - E\left( \sum_{l=0}^{m_1} n_{1,l}, k, \gamma_1 \right) \right) \overline{N}_2(n_2, M - m_1, k, \gamma_2) \right\}
\]

subject to

\[
1 - \left[ 1 - E\left( \sum_{l=0}^{m_1} n_{1,l}, k, \gamma_1 \right) \right] \left[ 1 - E\left( \sum_{l=0}^{M-m_1} n_{2,l}, k, \gamma_2 \right) \right] \leq \epsilon.
\]

The optimization problem involves determining the number of transmission rounds for the links and resource allocations for each transmission round. This can be solved numerically with an exhaustive search. Figure 3.6 represents the optimal transmission allocations between two links for the reliability targets corresponding to the BLER of \( 10^{-5} \) and \( 10^{-10} \). The message size is 100 bits and the maximum of four transmission rounds is envisaged, i.e., \( M = 4 \). Regions are indexed by pairs of \((i, j)\), where \( i \) and \( j \) represent the number of assigned data transmission rounds for uplink and downlink, respectively. The figure indicates that the same number of transmission rounds are assigned to the links when their SINRs are quite similar. In contrast, more transmissions rounds are assigned to the link with lower SINR level when the link qualities are different.

The optimization problem (3.13) determines three different parameters, which might considered to be complex. The optimization problem can hence be simplified by considering the same link reliability for both uplink
(a) $\epsilon = 10^{-5}$  
(b) $\epsilon = 10^{-10}$  

Figure 3.6. Optimal allocation of transmission attempts between uplink and downlink, where $M = 4$ [PIII].

and downlink. This can be achieved by setting $\epsilon_1 = \epsilon_2 = 1 - \sqrt{1 - \epsilon}$. Consequently, the simplified optimization becomes

$$
\arg\min_{m_1} \left\{ N_1(\vec{n}_1, m_1, k, \gamma_1) + \left( 1 - E(\sum_{l=0}^{m_1} n_{1,l}, k, \gamma_1) \right) N_2(\vec{n}_2, M - m_1, k, \gamma_2) \right\},
$$

(3.14)

where $\vec{n}_1$ and $\vec{n}_2$ are the vectors of the channel uses for uplink and downlink, respectively, and can be determined by

$$
\text{sum}(\vec{n}_1) = \min n : E(n, k, \gamma_1) \leq 1 - \sqrt{1 - \epsilon},
$$

(3.15)

$$
\vec{n}_1 = \arg\min_{\vec{n}} \left\{ N_1(\vec{n}, m_1, k, \gamma_1) \right\},
$$

$$
\text{sum}(\vec{n}_2) = \min n : E(n, k, \gamma_2) \leq 1 - \sqrt{1 - \epsilon},
$$

$$
\vec{n}_2 = \arg\min_{\vec{n}} \left\{ N_2(\vec{n}, M - m_1, k, \gamma_2) \right\}.
$$

The sum($\vec{n}$) indicates the sum of all the vector elements. For this optimization, $m_1$ is the main parameter that should be determined. For this purpose, the blocklength for each link is initially determined regardless of the number of transmission rounds. Then, the optimal data block segmentation is derived at discrete values of $m_1$ and $m_2$. Finally, the value of $m_1$ is determined such that the sum of the expected channel uses for both links is minimized. This approach separates the time and resource allocation problems, which reduces the computation complexity.

In order to evaluate the performance of the proposed time and resource allocation, the channel use gain can be defined. It indicates how much the employed scheme reduces the expected channel uses compared to the case of equal transmission allocations ($m_1 = m_2$). The channel use gain is
Radio Resource Allocations

defined as

\[
G = \frac{(\bar{N}_1 + \bar{N}_2)_{\text{equal transmission allocations}} - (\bar{N}_1 + \bar{N}_2)_{\text{fixed transmission allocations}}}{(\bar{N}_1 + \bar{N}_2)_{\text{equal transmission allocations}}}
\]

(3.16)

Figure 3.7 illustrates the channel use gains achieved by employing fixed transmission allocations, considering optimizations (3.13) and (3.14). A total of four transmission rounds are considered for the uplink and downlink, i.e., \( M = 4 \). Figure 3.7 indicates that the fixed transmission allocations achieve higher gain when the difference between SINR levels is high and the reliability target is stringent. In addition, both optimization problems achieve quite similar gains.

![Figure 3.7](image)

Figure 3.7. Channel use gain for fixed transmission allocations \((SINR_1 = 0 \text{ dB})\) [PIII].

3.4.2 Adaptive-transmission Rounds

The data transmissions in cellular mode occur in uplink and downlink successively. Hence, the resource allocation for downlink can be adaptively performed after successfully retrieving the message in uplink. Consider that the total of \( M \) transmission rounds is available for both uplink and downlink. If the message is delivered uplink utilizing \( j \) transmission rounds, the radio resources for downlink are allocated knowing that \( M - j \) transmission rounds are available. In this scheme, the number of transmissions for downlink and uplink can vary from 1 to \( M - 1 \). Consider \( \bar{n}_1 = \{n_{1,1}, ..., n_{1,M-1}\} \) where \( n_{1,j} \) denotes the number of channel
Radio Resource Allocations

uses in the \( j \)th HARQ round of uplink transmissions. The HARQ uplink transmissions are stopped if the message is correctly received or the maximum of \( M - 1 \) transmission rounds is reached. The resource allocations for downlink depend on the number of remaining transmission attempts. Consider \( \{ \vec{n}_1, \ldots, \vec{n}_{M-1} \} \) where \( \vec{n}_1^j = \{ n_{1,1}^j, \ldots, n_{1,j}^j \} \) represents the channel use segmentations for the case that a maximum of \( j \) transmission rounds can be realized. The following optimization can be considered to minimize the sum of the expected channel uses in uplink and downlink

\[
\arg\min_{\vec{n}_1, \vec{n}_2, \ldots, \vec{n}_{M-1}} \sum_{j=1}^{M-1} \left\{ \left[ 1 - E\left( \sum_{l=0}^{j} n_{1,l}, k; \gamma_1 \right) \right] n_{1,j} \right. \\
+ \left. \left[ E\left( \sum_{l=0}^{j-1} n_{1,l}, k, \gamma_1 \right) - E\left( \sum_{l=0}^{j} n_{1,l}, k, \gamma_1 \right) \right] N_2\left( \vec{n}_2^{M-j}, M - j, k, \gamma_2 \right) \right\}
\]

subject to

\[
1 - \sum_{j=1}^{M-1} \left[ 1 - E\left( \sum_{l=0}^{j} n_{1,l}, k, \gamma_1 \right) \right] \left[ 1 - E\left( \sum_{l=0}^{M-j} n_{2,l}^{M-j}, k, \gamma_2 \right) \right] \leq \epsilon.
\]

The computational complexity of this optimization is high as there are \( M \) vectors that should be determined. Hence, a simplified optimization can be considered to separate the uplink and downlink resource allocations. For this purpose, the same BLER target of \( 1 - \sqrt{1 - \epsilon} \) can be considered for uplink and downlink, regardless of the remaining transmission rounds. This results in the following optimization

\[
\arg\min_{\vec{n}_1} \left\{ \left[ 1 - E\left( \sum_{l=0}^{j} n_{1,l}, k, \gamma_1 \right) \right] n_{1,j} \right. \\
+ \left. \left[ E\left( \sum_{l=0}^{j-1} n_{1,l}, k, \gamma_1 \right) - E\left( \sum_{l=0}^{j} n_{1,l}, k, \gamma_1 \right) \right] N_2\left( \vec{n}_2^{M-j}, M - j, k, \gamma_2 \right) \right\},
\]

where \( \text{sum}(\vec{n}_1) \) and \( \vec{n}_2^{M-j} \) can be determined as

\[
\text{sum}(\vec{n}_1) = \min n : E(n, k, \gamma_1) \leq 1 - \sqrt{1 - \epsilon}, \quad \text{(3.19)}
\]

\[
\text{sum}(\vec{n}_2^{M-j}) = \min n : E(n, k, \gamma_2) \leq 1 - \sqrt{1 - \epsilon},
\]

\[
\vec{n}_2^{M-j} = \arg\min_{\vec{n}} \left\{ N_2(\vec{n}, M - j, k, \gamma_2) \right\}.
\]

In this optimization, the segmentation of uplink transmissions, i.e., \( \vec{n}_1 \), is the main parameter that should be determined. While, the optimal segmentations for downlink are derived at a discrete value of the possible transmission rounds. Figure 3.8 illustrates the channel use gain achieved by the adaptive transmission allocation scheme, using optimization (3.17) and (3.18), compared to equal transmission allocation. The
Radio Resource Allocations

adaptive transmission allocation scheme achieves gain even when the qualities of the links are quite the same. However, the achieved gain is higher when the qualities of the links are different and the reliability target is stringent. In addition, both optimization problems achieve quite similar gains.

![Figure 3.8. Channel use gain for fixed transmission allocations (\(SINR_1 = 0\) dB)](image)

### 3.5 Link Adaptation in Cellular Systems

Generally, cellular systems employ error control schemes, such as ARQ and HARQ, to achieve high spectral efficiency while meeting the communication reliability. In order to apply these schemes for URLLC, the errors of both data and control channel should be considered, since the errors of the control channels might be comparable to the communication tolerable error rate. The optimal resource allocations for the HARQ scheme are presented in Section 3.3. The results are derived assuming that the transmitter knows the SINR of the data channel and the error rates of the feedback signals. In addition, it is assumed that any arbitrary amount of radio resources can be assigned for the data channels. However, in practical wireless systems, the transmitter might not have full knowledge of the SINR. In addition, only a few transmission rates are available for the data transmissions. Hence, this section first considers the optimal link adaptation scheme, assuming that the exact SINR is known at the trans-
mitter. Then, a simple link adaptation scheme is presented, which can be utilized for downlink transmissions with a limited feedback overhead for reporting the channel quality.

In cellular systems, a base station employs link adaptation for the data channels by selecting a proper modulation and coding scheme (MCS). In downlink data transmissions, the MCS is mainly selected according to the reported channel quality indicator (CQI) by the UE. In LTE, the UE measures the reference signal (RS) and estimates the downlink SINR. It then selects a CQI value corresponding to an MCS that guarantees a BLER of 10% with a single transmission round [14]. The medium access control (MAC) layer employs a HARQ scheme, supporting up to three retransmission rounds. In case the MAC layer fails in delivering the message successfully, e.g., due to the errors of a feedback channel, radio link control (RLC) triggers the protocol data unit (PDU) retransmission. This procedure can achieve a satisfactory communication reliability with moderate latency. However, this cannot satisfy the URLLC requirements, due to the limited number of transmission opportunities. For this reason, the proposed link adaptation schemes consider the reliabilities of both data and control channels with a limited number of transmission rounds. In order to meet the stringent latency requirement of URLLC, only one retransmission round is envisaged [PIV]. The data transmission/retransmission can be performed employing one of the supported MCSs, which are represented as \{MCS_1, MCS_2, ..., MCS_M\}. The communication reliability target for each payload transmission is \(\rho = 1 - \epsilon\).

### 3.5.1 Optimal Link Adaptation

For optimal link adaptation, the base station needs to know the exact SINR at the receiver and also the error rates of the ACK/NACK signals. Assume that the base station employs \(MCS_i\) for the initial transmission and \(MCS_j\) for the retransmission when the NACK is detected. The expected throughput can be expressed as [21]

\[
T_{i,j} = r_i \{1 - P_{1,i}(\gamma)(1 - \epsilon_A)\} + \frac{1}{r_i + \frac{1}{r_j}} \{(1 - P_{1,i}(\gamma))\epsilon_A + P_{1,i}(\gamma)(1 - \epsilon_N)(1 - P_{2,j}(\gamma))\},
\]

where \(\gamma\) is the SINR at the receiver and \(r_i\) and \(r_j\) are the transmission rates for \(MCS_i\) and \(MCS_j\), respectively. \(P_{1,i}\) is the residual BLER after the initial transmission round and \(P_{2,j}\) is the residual BLER after the retransmission round. Note that \(P_{2,j}\) is independent from \(P_{1,i}\) if the ARQ
Radio Resource Allocations

scheme is implemented, i.e., \( P_{2,j}(\gamma) = P_{1,j}(\gamma) \). In case of implementing a HARQ scheme, the residual BLER of the retransmission conditionally depends on the BLER of the initial transmission; in this case, \( P_{2,j}(\gamma) \leq P_{b,j}(\gamma) \).

The optimal MCSs for the initial transmission and retransmission rounds can be selected in order to maximize the expected throughput while meeting the communication reliability target. This leads to the following optimization problem

\[
\begin{align*}
    s_1, s_2 &= \arg \max_{1 \leq i,j \leq M} T_{i,j} \\
    \text{subject to} \\
    1 - \{ (1 - P_{1,i}(\gamma)) + P_{1,i}(\gamma)(1 - \epsilon_N)(1 - P_{2,j}(\gamma)) \} &\leq \epsilon.
\end{align*}
\]

### 3.5.2 Proposed Link Adaptation Scheme

The optimal link adaptation scheme requires reporting the exact SINR to the transmitter, which can lead to a high signaling overhead over the feedback channel. To address this issue, a simple link adaptation scheme is proposed that can be implemented utilizing conventional CQI report [PIV]. For this scheme, a BLER target set is defined as \( \{\xi_1, \xi_2, \ldots, \xi_r\} \) whose components are arranged in decreasing order. The base station selects the BLER target values for the transmission rounds from this set. These BLER targets are delivered to the UE, and the UE reports the corresponding CQI values, i.e., \( s_1 \) and \( s_2 \). The base station then performs the transmissions accordingly. Figure 3.9 illustrates the communication model for the considered link adaptation scheme.

Motivated by the results from Section 3.3, the MCS selection can be performed by giving the preference to have the initial transmission with the highest possible transmission rate. However, the reliability of the transmission rounds should meet the overall reliability target. The BLER targets for the transmission rounds can be set as described in Algorithm 1. The base station delivers the BLER targets, i.e., \( \epsilon_1 \) and \( \epsilon_2 \), to the UE. For each BLER target, the UE needs to determine the highest CQI value that the corresponding MCS meets the BLER target. In case the available MCSs cannot meet the BLER targets, due to the low SINR levels, the UE reports the lower CQI values corresponding to the most robust MCS. The
CQI report at the UE is performed as follows

\[
s_1(\gamma) = \begin{cases} 
    MCS_i & \text{if } \exists i, j : P_{1,i}(\gamma) \leq e_1, P_{2,j}(\gamma) \leq e_2, \\
    MCS_1 & \text{others.}
\end{cases} 
\]

(3.22)

\[
s_2(\gamma) = \begin{cases} 
    MCS_j & \text{if } \exists i, j : P_{1,i}(\gamma) \leq e_1, P_{2,j}(\gamma) \leq e_2, \\
    MCS_1 & \text{others.}
\end{cases} 
\]

(3.23)

In order to evaluate the performance of the link adaptation schemes, they are applied to the LTE physical (PHY) layer framework. LTE defines 16 CQI values that are associated with different MCSs that have coding rates between 1/13 to 1, combined with 4-QAM, 16-QAM, and 64-QAM modulations [22]. The BLER target set is defined as \( \{10^{-1}, 10^{-2}, ..., 10^{-10}\} \). Figure 3.10 illustrates the BLER targets for achieving different communication reliability targets. This indicates that the initial transmission can be performed with the BLER of 10% regardless of the overall communication reliability target when the feedback channel is reliable. However, the retransmission should be performed robustly according to the overall communication reliability. As the reliability of the feedback channel reduces, it is not feasible to rely properly on the retransmission round for correcting the data transmission failure. Hence, the initial transmission round should be performed more robustly. Figure 3.11 compares the achievable throughput using the optimal and proposed link adaptation schemes. The
Algorithm 1 Target BLER selection

\begin{algorithmic}
\STATE \textbf{Initialization:} \textit{condition} $\leftarrow 0$, \textit{i} $\leftarrow 0$, \textit{j} $\leftarrow 0$
\WHILE {\textit{condition} $= 0$}
  \STATE \textit{i} $\leftarrow \textit{i} + 1$
  \WHILE {\textit{condition} $= 0$}
    \STATE \textit{j} $\leftarrow \textit{j} + 1$
    \IF {\(1 - \{(1 - \xi_i) + \xi_i(1 - \epsilon_N)(1 - \xi_j)\} \leq \epsilon\)}
      \STATE \textit{condition} $\leftarrow 1$
    \ENDIF
  \ENDWHILE
\ENDWHILE
\STATE \textit{e}_1 \leftarrow \xi_i$, \textit{e}_2 \leftarrow \xi_j
\end{algorithmic}

considered overall reliability targets correspond to the BLER of $\epsilon = 10^{-5}$ and $\epsilon = 10^{-10}$. The performance of the proposed link adaptation scheme is quite close to the one achieved by employing the optimal scheme. However, the differences become more as the error of the feedback channel increases.

### 3.6 Mode Selections

Conventionally, cellular-mode operation is only considered to provide connectivity between UEs. However, D2D communications have been considered as a more efficient mode for data transmissions between UEs that are located in close proximity to each other. This allows UEs to directly communicate without passing their data through a base station, which can increase the spectral efficiency and reduce the communication latency [23], [24]. This section compares the performance of cellular and D2D communication modes for supporting URLLC. In addition, a hybrid mode operation is proposed, which can be considered as a combination of D2D and cellular modes [PV]. These modes are illustrates in Figure 3.12. For the simplicity, the control signaling is not considered for these modes. In addition, only two transmission attempts are envisioned for each mode in order to meet the latency constraint.
Radio Resource Allocations

\[ \epsilon_{A,N} \]

\[ \begin{array}{cccccccc}
10^{-12} & 10^{-10} & 10^{-8} & 10^{-6} & 10^{-4} & 10^{-2} & 10^0 \\
10^{-10} & 10^{-8} & 10^{-6} & 10^{-4} & 10^{-2} & 10^0 & \\
10^{-8} & 10^{-6} & 10^{-4} & 10^{-2} & 10^0 & \\
10^{-6} & 10^{-4} & 10^{-2} & 10^0 & \\
10^{-4} & 10^{-2} & 10^0 & \\
10^{-2} & 10^0 & \\
10^0 & \\
\end{array} \]

\[ \epsilon = 1 \times 10^{-4} \]

\[ \epsilon = 1 \times 10^{-5} \]

\[ \epsilon = 1 \times 10^{-6} \]

\[ \epsilon = 1 \times 10^{-7} \]

\[ \epsilon = 1 \times 10^{-8} \]

\[ \epsilon = 1 \times 10^{-9} \]

\[ \epsilon = 1 \times 10^{-10} \]

(a) Initial data transmission

(b) Data retransmission

Figure 3.10. Target BLER for data transmission rounds [PIV].

3.6.1 Cellular Mode

In cellular mode, the base station receives messages from the transmitter UE and forwards them to the receiver UE. Considering the symmetric time allocations for uplink and downlink, each payload can be transmitted using two transmission attempts over each link. Denote \( n_{i,1} \) and \( n_{i,2} \) as the channel uses for the initial transmission and retransmission over the link \( i \) (1: uplink and 2: downlink). The expected channel uses over the link \( i \) is

\[ N(n_{i,1}, n_{i,2}, k, \gamma_i) = n_{i,1} + n_{i,2} E(n_{i,1}, k, \gamma_i), \tag{3.24} \]
where $\gamma_i$ is the SINR at the receiver over the link $i$. The base station forwards the message if it could retrieve the message from the transmitter UE. Hence, the sum of the expected channel uses in uplink and downlink is

$$
\bar{N}_{\text{Cellular}}(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2}, \gamma_1, \gamma_2) = \bar{N}(n_{1,1}, n_{1,2}, k; \gamma_1) + (1 - E(n_{1,1} + n_{1,2}, k; \gamma_1)) \bar{N}(n_{2,1}, n_{2,2}, k; \gamma_2).
$$

The optimal resource allocations can be derived to minimize the expected channel uses in uplink and downlink, while meeting the overall communi-
Radio Resource Allocations

Figure 3.12. Different transmission modes [PV].

The reliability target. This leads to the following optimization problem

\[
\arg\min_{n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2}} \mathcal{N}_{\text{Cellular}}(n_{1,1}, n_{1,2}, n_{2,1}, n_{2,2}, k; \gamma_1, \gamma_2)
\]

subject to

\[
1 - (1 - E(n_{1,1} + n_{1,2}, k; \gamma_1))(1 - E(n_{2,1} + n_{2,2}, k; \gamma_2)) \leq \epsilon. \tag{3.26}
\]

3.6.2 D2D Mode

In the D2D communication mode, two UEs directly communicate without passing their data through the base station. However, the base station provides radio resources for the D2D link. For this purpose, the receiver UE can report the estimated SINR, i.e., \(\gamma\), to the base station. Accordingly, the base station allocates radio resources for the data transmissions. Assume that each payload can be transmitted using two transmission attempts. The initial transmission is performed using \(n_1\) channel uses. If it is required, the retransmission is performed using \(n_2\) channel uses. The
optimal resource allocations can be derived as follows

\[ \arg\min_{n_1, n_2} N(n_1, n_2, k, \gamma) \]  
\[ \text{subject to} \quad E(n_1 + n_2, k, \gamma) \leq \epsilon. \]

### 3.6.3 Hybrid Mode

The hybrid mode can be considered as a combination of the cellular and D2D modes. In this mode, the transmitter UE performs the initial transmission in D2D mode, while the base station also tries to decode the message. However, the retransmission may be performed using the D2D link or downlink in the cellular mode. The downlink is selected if the base station has decoded the message successfully and it has a better channel quality compared to the D2D link, i.e., \( \gamma_2 \geq \gamma \). The expected channel uses for the hybrid mode can be expressed as

\[ N_{\text{Hybrid}}(n_1, n_2, n_2, k, \gamma, \gamma_1, \gamma_2) = n_1 + E(n_1, k, \gamma) \]
\[ \times [n_2 E(n_1, k, \gamma_1) + (1 - E(n_1, k, \gamma_1)) \min(n_2, n_1, 2)]. \]  

The success probability of delivering the message for this mode is

\[ P_{\text{Hybrid success}} = \begin{cases} 
1 - E(n_1 + n_2, k, \gamma) & \text{if } \gamma \geq \gamma_d, \\
1 - Q \left( \frac{n_1 C(\gamma) + n_2 C_2(\gamma_2) - k}{\sqrt{n_1 V(\gamma) + n_2 V(\gamma_2)}} \right) & \text{if } \gamma < \gamma_d.
\end{cases} \]  

The optimal resource allocations can be derived according to

\[ \arg\min_{n_1, n_2, n_2, 2} N_{\text{Hybrid}}(n_1, n_2, d_2, k, \gamma, \gamma_1, \gamma_2) \]
\[ \text{subject to} \quad 1 - P_{\text{success}}^{\text{Hybrid}} \leq \epsilon. \]  

### 3.6.4 Results

This section compares the performance of the presented transmission modes. The reliable communications is required for delivering 100 bits of information with an overall BLER of \( 10^{-10} \). Two communication scenarios are considered. In one case, \( \gamma_1 = 5 \text{ dB} \) and \( \gamma_2 = 10 \text{ dB} \), while in the other case, \( \gamma_1 = 10 \text{ dB} \) and \( \gamma_2 = 5 \text{ dB} \). The achievable average transmission rates for these cases are the same in cellular mode. For each case, the SINR of the D2D link is changed. This results in different average transmission rates for the D2D and hybrid modes. Figure 3.13 compares the
achieved transmission rates by considered modes. For a very low SINR of the D2D link, the cellular mode outperforms other operation modes. While for a high SINR of the D2D link, the D2D and hybrid modes provide better performance compared to cellular mode. The hybrid mode always outperforms the D2D mode by exploiting the transmission diversity.

Figure 3.13. Achieved transmission rates with different operational modes.

3.7 Discussion

The use of HARQ-based transmissions for supporting URLLC requires considering a different source of errors from the data and control channels. The developed communication models mainly took into account the errors related to the data decoding and detection of the ACK/NACK signals, assuming the perfect channel estimation. For a better evaluation, the models should be extended in order to comprehend the channel estimation errors. In addition, the time gap between the channel measurement report and the data transmissions brings inaccuracy for the resource allocations. This type of error can also be considered for extending the communication model.

Results have shown that setting different BLER targets for the transmission rounds in HARQ-based transmissions can offer better performance for URLLC. This approach has been considered in 3GPP RAN1 meetings. It was agreed that the UE with URLLC support should be able
to report the CQI according to the BLER targets of $10^{-1}$ and $10^{-5}$. This brings flexibility for data transmissions, as the base station can configure the UE for each CQI report using one of the BLER targets.
4. Control Channel Enhancements

4.1 Introduction

Cellular systems define various physical channels in order to segregate different information types. Generally, the channels are categorized as the control and data channels, which carry control and user plane information, respectively. All these channels are prone to errors due to the noise and the interference in the system. In the conventional cellular systems, such as LTE, the higher levels are generally responsible for detecting and resolving the errors of the control channels. However, this approach is not applicable to the URLLC, due to the stringent latency requirement. In this regard, the errors of the control and data channels should be considered while designing the URLLC service. The errors of the control and data channels affect the overall communication reliability differently. Hence, these channels are associated with different reliability requirements in order to support a specific service. This chapter explains the control channel errors. It also provides comprehensive communication models for uplink and downlink data transmissions in order to support URLLC. These models help to identify the reliability requirements for the control channels. In addition, various enhancements are proposed for designing the control channels in order to support URLLC more efficiently.

4.2 Control Channel Errors and Reliability Requirements

The uplink and downlink data transmissions are performed differently in the cellular systems. In this section, the general communication models are considered for HARQ-based transmissions with a limited num-
ber of transmission attempts to meet the stringent latency constraint for URLLC [PVI]. It is assumed that the maximum of two transmission rounds is envisioned.

4.2.1 Uplink Transmissions

For uplink data transmissions, the base station, known as gNB in 5G, needs to know the channel quality of the UE before allocating the radio resources. For this purpose, the UE is configured to periodically transmit sounding reference signal (SRS), while the base station measures it and estimates the channel quality. In addition, the UE should inform the base station when it has some data for transmission. This is achieved by sending a scheduling request (SR) by the UE. The UE is provided with periodic resources for SR transmissions over the physical uplink control channel (PUCCH). When the base station identifies an SR, it allocates radio resources for the uplink data transmissions and instructs the UE to utilize them by sending an uplink resource grant (RG) as a part of the downlink control information (DCI) carried over the physical downlink control channel (PDCCH). Then, the UE utilizes the assigned radio resources over the physical uplink shared channel (PUSCH) for uplink transmissions. In case the base station cannot retrieve the message correctly, it triggers the data retransmission by sending a new uplink RG.

![Uplink data transmissions in cellular systems](image)

**Figure 4.1.** Uplink data transmissions in cellular systems.

Figure 4.1 illustrates the communication model for uplink data transmissions. The maximum of one data retransmission is assumed in order
to meet the latency requirement of URLLC. If the base station does not decode the SR from the UE, it does not grant radio resources for the data transmission. The UE identifies this failure when it does not receive any uplink RG. Hence, it should retransmit the SR. In this case, only one data transmission round is envisaged to compensate the delay caused by the SR retransmission. The UE might not decode the uplink RG correctly. In such case, it does not perform the uplink transmission. Consequently, the base station does not receive any signal from the UE and triggers the data retransmission by sending a new uplink RG. Considering all the error types, the probability of retrieving the message correctly within the delay budget can be expressed by

\[
P_{\text{UL success}} = (1 - \epsilon_{\text{SR}})(1 - \epsilon_{\text{RG}})\{(1 - P_1) + P_1(1 - \epsilon_{\text{RG}})(1 - P_{1,2})\}
\]

+ \(\epsilon_{\text{SR}} + \epsilon_{\text{RG}})(1 - \epsilon_{\text{SR}})(1 - \epsilon_{\text{RG}})(1 - P_2),\)

where \(\epsilon_{\text{SR}}\) is the probability of missing the RG by the base station, \(\epsilon_{\text{RG}}\) is the probability of missing the uplink RG. The error probability of retrieving the message using the initial data transmission round is \(P_1\). The error probability of retrieving the message is \(P_{1,2}\) when data from both initial transmission and retransmission rounds are available, and it is \(P_2\) when only the data from the retransmission round is available.

The reliability requirements for the control information depend on the overall communication reliability target and the performance of the data transmission/retransmission. The considered communication service can be supported only if \(\exists \epsilon_{\text{SR}}, \epsilon_{\text{RG}} \in [0,1] : P_{\text{UL success}} \geq 1 - \epsilon\). Figure 4.2 illustrates the reliability regions for the control information in order to meet the overall communication reliability of \(1 - 10^{-5}\). The initial transmission round is performed with three different BLER targets, i.e., \(P_1 = \{10, 1, 0.1\}\)%. It is assumed that \(P_1 = P_2\) and \(P_{1,2} = 10^{-5}\). The figure also represents trade-offs between the reliability requirements of the control information and the reliability of the initial data transmission. For instance, \(\epsilon_{\text{SR}}\) and \(\epsilon_{\text{RG}}\) should be less than \(10^{-4}\) when the BLER for the initial transmission round is \(10\%\). These reliability requirements are relaxed to around \(10^{-3}\) when the BLER for the initial transmission round is \(1\%\). However, this requires utilizing more radio resources for the data channel in order to achieve greater robustness for the initial data transmission.
4.2.2 Downlink Transmissions

For downlink data transmissions, a base station needs to know the channel quality of the receiver UE in order to employ a proper MCS for data transmissions. For this purpose, the base station transmits RS in downlink, while the UE measures it and estimates the SINR. The estimated SINR is mapped to one of the CQI values and then reported to the base station. Accordingly, the base station selects the proper MCS and allocates radio resources for the data transmissions. The base station informs the UE about the allocated radio resources by sending a downlink RG over the PDCCH. By decoding the downlink RG, the UE can monitor the data transmission over the physical downlink shared channel (PDSCH). The UE should report either an ACK or a NACK in order to indicate the success or failure in decoding the message. The ACK/NACK signals are also carried over the PUCCH. In case the UE cannot decode the downlink RG, it does not monitor the data transmission. Consequently, it does not report an ACK nor a NACK, which is known as a discontinuous transmission (DTX). The base station should trigger the data retransmission if it receives a NACK signal or detects a DTX.

Figure 4.3 illustrates the communication model for downlink data transmissions. If the UE decodes the RG for the initial data transmission, it can monitor the data transmission. The error probability of retrieving the message using the data from the initial transmission round is $P_1$. If the base station identifies a NACK, it assumes that the UE has received
the data information from the initial transmission, however, it could not decode the message correctly. Hence, it performs the data retransmission. The error probability of retrieving the message by the UE is $P_{1,2}$, considering the combined information from the initial transmission and retransmission rounds. In case the UE only receives the RG for the retransmission attempt, it might receive data information in different ways. If the base station correctly detects a DTX, it assumes that the UE has missed the initial data transmission. Hence, it can perform the data retransmission robustly to compensate the loss of data information in the initial transmission. In this case, the error probability of decoding the message is $P_{2D}$. The base station might erroneously decode a DTX as a NACK. In this case, the base station performs the data retransmission assuming that the UE has received some data information, while it could not decoded the message correctly. The error probability of decoding the message for this case is $P_{2N}$. Considering all the mentioned cases, the communication reliability is expressed by

$$P_{\text{DL success}} = (1 - \epsilon_{\text{RG}})(1 - P_1) + P_1(1 - \epsilon_{\text{NA}} - \epsilon_{\text{ND}})(1 - \epsilon_{\text{RG}})(1 - P_{1,2})$$

$$+ \epsilon_{\text{ND}}(1 - \epsilon_{\text{RG}})(1 - P_{2D}) + \epsilon_{\text{RG}}(1 - \epsilon_{\text{RG}})(\epsilon_{\text{DN}}(1 - P_{2N})$$

$$+ (1 - \epsilon_{\text{DN}} - \epsilon_{\text{DA}})(1 - P_{2D})),$$

where $\epsilon_{\text{RG}}$ is the probability of missing the downlink resource grant, $\epsilon_{\text{NA}}$ is the probability of the erroneous detection of a NACK as an ACK, $\epsilon_{\text{ND}}$ is the probability of the erroneous detection of a NACK as a DTX, $\epsilon_{\text{DN}}$ is the probability of the erroneous detection of a DTX as a NACK, and $\epsilon_{\text{DA}}$ is the probability of the erroneous detection of a DTX as an ACK.

The reliability requirements for the control information depend on the overall reliability target and the performance of the data transmission/retransmission. The considered services can be supported if $\exists \epsilon_{\text{RG}}, \epsilon_{\text{NA}}, \epsilon_{\text{DN}}, \epsilon_{\text{DA}} \in [0, 1] : P_{\text{DL success}} \geq 1 - \epsilon$. Figure 4.4 illustrates the reliability regions for the control information in order to meet the overall communication reliability of $1 - 10^{-5}$. The initial transmission round is performed with three different BLER targets, i.e., $P_1 = \{10, 1, 0.1\}$%. It is assumed that $P_{1,2} = 10^{-5}$ and $\epsilon_{\text{A,N,D}} = \epsilon_{\text{NA}} = \epsilon_{\text{ND}} = \epsilon_{\text{DN}}$. The figure indicates that there are trade-offs between the reliability requirements for the control information and the reliability of the data transmissions. For instance, the reliability requirements for the control channels, i.e., $\epsilon_{\text{RG}}$ and $\epsilon_{\text{A,N,D}}$, are relaxed when the initial transmission is performed more robustly. In addition, the higher reliability is required for the RG compared to the feedback channel, carrying ACK/NACK signals. This is due
to the fact that the RG is essential for both data transmission and retransmission. While, the feedback channel only affects the data retransmission round.

**Figure 4.4.** Reliability regions for the control information in downlink data transmissions [PVI].

### 4.3 Scheduling Request

The UE is assigned with periodic resources over the PUCCH to transmit the SR. In LTE, the SR is carried without employing any modulation, and the base station detects it applying energy detection. The periodicity
of the SR is configurable. The higher periodicity entails allocating more radio resources for the SR transmissions, while the UE can send an SR earlier. This section proposes some of the enhancements for SR transmissions.

### 4.3.1 Quality of Service (QoS) Based SR

The current SR does not carry any information regarding the communication requirements, such as reliability or latency, for delivering a message in uplink. This prevents performing the resource allocations efficiently. To address this issue, the UE can carry additional information regarding the communication requirements, which can form a QoS set, along the SR message [PVII]. This can be achieved by including a few extra bits in the SR or by assigning different SR sequences for the UE. The UE selects one of the QoS items and sends it along the SR. The base station schedules the UE according to the received QoS.

Consider a similar scenario to that described in Section 4.2.1. The UE is assigned with two SR sequences. The UE utilizes the first SR sequence when initially trying to send a request to be scheduled. If it does not receive the uplink RG in response, due to the missing of the SR by the base station, it retransmits the SR using the second sequence. If the base station detects the second SR sequence, it knows that the UE has only one transmission round for delivering the message data to meet the latency budget. Hence, it can assign radio resources for the data transmission such that a higher reliability is achieved. Figure 4.5 illustrates the reliability requirements for the control information for uplink data transmission using a QoS-based SR. It is assumed that the BLER after the initial data transmission is \( P_1 = \{10, 1, 0.1\} \% \) when the first SR sequence is detected. The BLER after the data retransmission is \( P_{1,2} = 10^{-5} \). In case the UE detects the second SR sequences, it performs the data transmission robustly, in which the BLER is \( P_2 = 10^{-5} \). The figure indicates that the proposed approach relaxes the reliability requirement of SR transmissions, also it can achieve more efficient scheduling.

### 4.3.2 Group-based SR

A UE may need to transmit data for different services, e.g., URLLC and eMBB traffic. To multiplex different services efficiently, the radio resources assigned to the SR transmissions can be divided into different
groups. Each group is assigned to a specific traffic type. This allows guaranteeing different reliabilities for each SR group. For instance, the SR resources for URLLC can be available during each TTI for a limited number of UEs, while the SR resources for eMBB can be available with lower periodicity. In this way, the UE can send the SR for URLLC very rapidly. In addition, the SR transmissions for the eMBB traffic do not degrade the performance of SR detection for URLLC. Another advantage of group-based SR transmissions is its efficient scheduling. For instance, the base station can schedule the UEs with URLLC traffic with a short TTI in order to achieve low-latency, while it can schedule the UEs with the eMBB traffic with a longer TTI in order to reduce the signaling overhead.

### 4.4 Resource Grant

In scheduled-based transmissions, the base station sends an RG to inform the UE regarding the assigned radio resources for data transmissions/reception. The RG is carried by the mean of a DCI over the PDCCH. Missing the RG results in missing the data transmission/reception. This section proposes some enhancements for the RG transmissions.
4.4.1 Supporting Higher Aggregation Levels

LTE defines a control channel element (CCE) over the PDCCH. Each CCE consists of a nine-resource element group (REG) that contains four consecutive resource elements (REs). A DCI can be carried using a different number of consecutive CCEs, i.e., 1, 2, 4 or 8, which is called the aggregation level. A higher aggregation level achieves a better performance for decoding the DCI by the UE. The base station may employ several aggregation levels in order to deliver DCIs with different reliability performances. The 5G NR may consider supporting higher aggregation levels, e.g., 16 to achieve more robustness for delivering a DCI for URLLC.

4.4.2 Semi-persistent Scheduling and Fast Uplink Access

LTE supports semi-persistent scheduling (SPS) for periodic data traffic, such as voice over LTE (VoLTE) and voice over IP (VoIP). SPS enables reserving periodic radio resources for the initial data transmission/reception. The UE is informed regarding the resources during the SPS configuration procedure. After that, the UE can utilize these resources without receiving a DCI for them separately. In case the initial transmission/reception fails, the base station allocates additional resources for the retransmission and sends a DCI for that. The SPS can be utilized for URLLC with periodic traffic [25]. This scheme improves the overall communication reliability, as the UE can always utilize the resources for the initial transmission. In addition, the UE does not need to send an SR for uplink transmissions. Another similar method is a fast uplink access, which reserves periodic resources for the uplink data transmissions. However, the UE only utilizes these resources when ever it has some data for transmissions. This leads to better power consumption compared to SPS uplink transmissions for the UE.

4.4.3 In-resource Control Signaling

In LTE, the radio resources for the PDCCH and PDSCH are separated over the time domain. The PDCCH occupies the first symbol(s) of each downlink subframe. This design limits the number of DCI transmissions. In addition, this introduces a delay for scheduling the downlink transmissions, as the incoming data during a subframe is queued until the beginning of the next downlink subframe. One possible solution for these
issues is in-resource control signaling, which allows transmitting the DCI at any time during the downlink subframe [26]. The costs of this approach is that the UE needs to monitor larger resources for an incoming DCI, which increases the power consumption. In addition, the in-resource control signaling brings more flexibility for the resource allocations for the DCI transmissions.

4.4.4 Joint Data and Control Coding

Conventionally, data and control information in downlink is encoded separately. This approach simplifies the procedure of extracting the control and data information. The UE first decodes the control channel information. If the UE finds a DCI assigned for its downlink data transmissions, then it tries to decode the data message. However, this approach might not be efficient for URLLC with short message sizes. This is due to the fact that the encoding efficiency increases with the message size. To increase the communication efficiency, the control and data information can be encoded jointly [PVII]. However, this will increase the complexity of decoding procedure at the UE, which may lead to higher power consumption.

4.4.5 DCI Repetition with DTX Detection

Decoding a DCI is a prerequisite to transmit and receive data utilizing the assigned radio resources. In uplink, a UE does not transmit data when it misses the DCI. The base station may distinguish this event when it does not receive any signal from the UE, i.e., a DTX. When the base station distinguishes the DTX, it triggers the data retransmissions by transmitting a new DCI. With the conventional slot structure, in which a symbol is assigned either for uplink or downlink, the retransmission can be performed during the following slots. This introduces delay between detecting the DTX and providing the opportunity for the data retransmission. This delay can be significant with long TTI scheduling. Missing the DCI results in having fewer data transmission opportunities considering the delay budget, which can reduce the communication efficiency. For instance, Figure 4.6(a) illustrates a scenario in which the SR is detected by the base station and data should be delivered in uplink within a two-slot duration. The UE misses the DCI for the initial uplink transmission. The base station detects this event and triggers the data retransmission in...
the next slot. Since this is the last opportunity for delivering the data, the base station should allocate more radio resources for the data retransmission.

The flexible slot structure can be exploit to reduce the latency caused by missing the DCI [PVII]. For this goal, the flexible symbols are initially assigned for uplink data transmissions. However, when the base station identifies the DTX, it reconfigures part of the flexible symbols for the downlink in order to retransmit the DCI. This gives another opportunity to the UE for performing the uplink data transmission in the same slot. Figure 4.6(b) illustrates a similar scenario for uplink data transmission utilizing the flexible slot structure. The UE misses the initial DCI and the base station identifies this incident as it does not receive any signal over the assigned radio resources. The base station retransmits the DCI using the flexible symbols. The UE decodes the retransmitted DCI and starts performing the data transmission. In case the initial transmission fails, there is still another opportunity for performing the data retransmission in the next slot. This approach can achieve better communication efficiency, as more transmission opportunities are envisaged even if a DCI is missed.

![Figure 4.6](image-url)

(a) Conventional slot structure

(b) Flexible slot structure

**Figure 4.6.** Uplink data transmissions with an error in detecting the DCI [PVII].

### 4.5 ACK/NACK Feedback

The HARQ-based transmissions require utilizing a feedback channel in order to report the ACK/NACK signals. In LTE, the ACK/NACK signals
are carried over the PUCCH. These signals can be carried alone or along with the CQI. This section presents some of the envisaged solutions for improving the performance of the ACK/NACK transmissions.

4.5.1 Early ACK/NACK Transmission

One of the main issues of HARQ-based transmissions is the high processing time for decoding the data, which introduces delay before sending the ACK/NACK feedback. In LTE, the feedback is sent after at least 4 TTI, i.e., 4 ms. To reduce this delay, the UE can send an early ACK/NACK feedback based on the prediction of the success or failure in decoding the message even before the message decoding is completed [PVIII]. For instance, the prediction can be performed based on estimating the uncoded bit error rate (BER) from the log-likelihood ratios (LLRs) of the input bits at the decoder, and mapping such an estimate to the reference-coded BER curves [27].

4.5.2 Multi-bit ACK/NACK

In LTE, the UE reports a single bit feedback as the indication of success or failure in decoding the downlink data transmissions. In this regard, the base station does not know how close the UE’s decoder was to retrieve the message upon receiving a NACK signal from the UE. This may reduce the communication efficiency for supporting URLLC due to the limited number of retransmission attempts; i.e., the base station needs to perform the data retransmission very robustly. In order to improve the communication efficiency, multi-bit ACK/NACK feedback can be employed [PVIII]. This enables the UE to report the status of the decoder. Accordingly, the base station can adapt the redundancy of data retransmission.

4.5.3 Asymmetric ACK/NACK Detection

The performance of HARQ-based transmissions highly depend on the reliability of a feedback channel that carries ACK/NACK signals. The reliability of the feedback channel can be improved by allocating more radio resources, which results in high signaling overhead. Another approach is applying asymmetric signal detection in order to ensure higher reliability for the NACK signals compared to the ACK signals [PIX].

In order to assess the performance of asymmetric ACK/NACK detection, a simplified communication model is considered for downlink data trans-
missions. The maximum of one data retransmission is envisaged upon detecting a NACK signal. The ACK/NACK signals are carried using one-bit message over the control channel employing binary phase shift keying (BPSK) modulation. The received feedback signal at the base station can be expressed by

\[ y_f = h_f x + n, \]

where \( h_f \) is the channel coefficient known at the base station and \( x \in \{+1, -1\} \) denotes the ACK/NACK signals. The additive white Gaussian noise is modeled by \( n \sim \mathcal{N}(0, \sigma^2) \). The SINR for the feedback channel is \( \gamma_f = |h_f|^2/\sigma^2 \). Accordingly, the probability distribution of the received signal can be expressed as

\[ p(y|s) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-s)^2}{2\sigma^2}}, \]

where \( s = +A = |h_f| \) when an ACK is transmitted and \( s = -A = -|h_f| \) when a NACK is transmitted. The base station distinguishes the ACK/NACK signals by comparing the received signal with a threshold \( \nu \).

The received signal is distinguished as the ACK if the \( y > \nu \), otherwise as the NACK. There are two error types for decoding the feedback signals. One is the erroneous decoding of an ACK signal as a NACK signal and the other is the erroneous decoding of a NACK signal as an ACK signal. These errors are associated with the following probabilities

\[ P_{AN} = \Pr(\text{NACK}|\text{ACK}) = \int_{-\infty}^{\nu} p(y|A) dy = Q\left(\frac{A - \nu}{\sigma}\right), \]

\[ P_{NA} = \Pr(\text{ACK}|\text{NACK}) = \int_{\nu}^{\infty} p(y|-A) dy = Q\left(\frac{A + \nu}{\sigma}\right). \]

Figure 4.7 illustrates an example of incorrect detection of ACK/NACK signals. The symmetric signal detection is achieved by setting \( \nu = 0 \), which ensures \( P_{NA} = P_{AN} \). However, the asymmetric signal detection can be applied by setting \( \nu > 0 \) in order to achieve higher reliability for detecting the NACK signals compared to the ACK signals, i.e., \( P_{NA} < P_{AN} \).

The base station performs the initial data transmission using \( n_1 \) channel uses. The UE tries to decode the message and reports either an ACK or a NACK. The base station triggers the data retransmission if it identifies a NACK. The retransmission is performed using \( n_2 \) channel uses. The expected channel uses for the data transmissions can be expressed as

\[ N = n_1 + n_2 \{ E(n_1, k, \gamma)(1 - P_{NA}) + (1 - E(n_1, k, \gamma)) P_{AN} \}. \]

This indicates that the retransmission is triggered if 1) the UE sends a NACK and the base station decodes it correctly, or 2) the UE sends an
ACK but the base station wrongly decodes it as a NACK. The success probability of retrieving the message is

\[
P_{\text{success}} = \left(1 - E(n_1, k, \gamma)\right) + E(n_1, k, \gamma)(1 - P_{\text{NA}})\left(1 - \frac{E(n_1 + n_2, k, \gamma)}{E(n_1, k, \gamma)}\right)
\]

\[
= 1 - E(n_1, k, \gamma)P_{\text{NA}} - E(n_1 + n_2, k, \gamma)(1 - P_{\text{NA}}).
\]

(4.4)

In order to minimize the expected channel uses while meeting the overall reliability target, the following optimization can be considered

\[
\arg \min_{n_1, n_2, \nu} \overline{N}
\]

subject to

\[
P_{\text{success}} \geq 1 - \epsilon.
\]

(4.5)

Consider a scenario in which a transmitter intends to deliver \( k = 256 \) bits of information. The tolerable communication errors are \( 10^{-5} \) and \( 10^{-7} \). The SINR of the data channel is 0 dB. Figure 4.8 illustrates the error rates of the ACK and NACK signals when the symmetric and asymmetric signal detections are applied to the feedback channel. Figure 4.9 illustrates the BLER of the initial transmissions. The asymmetric signal detection imposes a higher reliability detection for the NACK signals compared to the ACK signals for moderate SINR levels of the feedback channel. Accordingly, the initial data transmission can be performed with a moderate BLER, which requires allocating less radio resources. Figure 4.10 illustrates the channel use gain for the data transmissions when
the asymmetric signal detection is applied compared to the symmetric
detection. The gain is achieved only for moderate SINR levels of the feed-
back channel. For low SINR levels, the feedback channel is very unre-
liable. Applying the asymmetric detection to ensure higher reliability
for the NACK detection is not helpful as it results in a high error rate
of ACK detection, which triggers the retransmission most of the time,
i.e., $P_{AN} \approx 1$. For the high SINR levels, the feedback channel is reliable
enough with the symmetric detection. Hence, there is no gain from em-
ploying asymmetric detection to provide higher reliability for the NACK
detection.

![Figure 4.8](image)

**Figure 4.8.** Probabilities of missing ACK and NACK signals for symmetric and asym-
metric detection [PIX].

### 4.6 CQI Report

The UE reports the CQI in order to enable link adaptation for downlink
data transmissions. LTE supports both periodic and aperiodic CQI re-
porting [14]. The periodic CQI is carried mainly by the PUCCH, while
the aperiodic CQI is carried along the data information over the PUSCH.
There are two error types concerning the CQI report that affect commu-
nication reliability. One error is related to the channel variation. The
base station employs the MCS according to the reported CQI. However,
the channel condition during the data transmissions might become worse
compared to the time that the UE measured and reported the CQI. The
other error is related to the CQI detection. A base station may erroneously decode a CQI value as a lower or higher value. The former case results in employing an MCS with a lower rate than is desired, which only reduces the communication efficiency. The latter case leads to employing an MCS with a higher rate than desired, which also reduces communication reliability. This section proposes some solutions regarding the errors related to the CQI report.

Figure 4.9. BLER for the initial data transmissions for symmetric and asymmetric ACK/NACK detections [PIX].

Figure 4.10. Channel use gain for asymmetric ACK/NACK detection [PIX].
4.6.1 Delay-based Detection

The time gap between the CQI report and the data transmission degrades the accuracy of the link adaptation. This gap can be reduced by increasing the periodicity of the CQI report, which requires allocating more radio resources for the control channel. However, this approach is not efficient for the UEs generating sporadic traffic. Another solution is considering the time gap while selecting the MCS. In this way, a more robust MCS is selected when there is a long time gap between the received CQI and data transmission instances [28].

4.6.2 Early Termination

Data transmission may fail in case of employing an inappropriate MCS. In downlink, an inappropriate MCS might be selected due to the channel variation or erroneously decoding a CQI report. This results in selecting an MCS that achieves a lower BLER than it is desired. Conventionally, the base station may become aware that the employed MCS was not proper after finishing the data transmission by receiving feedback, e.g., a NACK, from the UE. This means that the base station performs data transmission that fails with a high probability, which is not desirable for URLLC.

The flexible slot configurations can be exploit to address the issue of inappropriate MCS selection [PVII]. The UE can identify the employed MCS from the DCI prior to the data transmission. It distinguishes whether the employed MCS can achieve the target BLER with the current channel condition. In case the employed MCS is not satisfactory, the UE terminates the data transmission by sending an early NACK along with the updated CQI. This early feedback can be transmitted over the assigned radio resources for the data transmission. When the base station detects the early NACK, it assigns new radio resources for the data transmission employing more robust MCS according to the updated CQI. Then, it sends a new DCI to the UE and transmits the data. The base station needs to extend the resource allocations over the frequency domain to compensate the shorter transmission time for delivering the data.

Figure 4.11(a) illustrates a communication scenario in which an inappropriate MCS is selected for the initial transmission with the conventional slot structure. The base station performs the initial transmission and the UE tries to decode the message. However, the UE fails in decod-
ing the message and sends a NACK along with the updated CQI at the end of the slot. The base station triggers the data retransmission in the next slot. The retransmission should be performed very robustly, since the initial transmission is performed with the inappropriate MCS and there is only one chance for the retransmission according to the latency budget. Figure 4.11(b) illustrates a similar scenario with the flexible slot structure. The base station performs the initial transmission with an inappropriate MCS. The UE distinguishes this from the DCI and terminates the data transmission by sending an early NACK along with the updated CQI. The base station reallocates radio resources for the downlink data transmission according to the new CQI and then sends a new DCI to the UE. The early data termination allows the base station to perform the initial transmission such that the desired BLER is achieved. In addition, the base station still has the opportunity for data retransmission, if the initial transmission fails. Hence, the early termination can improve the communication efficiency.

Figure 4.11. Downlink data transmissions employing an inappropriate MCS for the initial data transmission round [PVII].

4.7 Discussion

The performance of the data and control channels affects the communication reliability and the latency in different ways. In part, the control information suffers from different error types that should be considered for supporting URLLC. There are non-trivial solutions that improve the
performance of the control channels without necessarily utilizing more radio resources. Some of these solutions help in protecting the information signals that highly affect the communication reliability, while others allow detecting the errors of the control information at the early stages and enable taking immediate actions to compensate them.
5. Blind Downlink Transmission Schemes

5.1 Introduction

The HARQ transmissions can achieve high transmission rates by incrementally sending data segments [29]. As mentioned in the Section 3.5, it is not feasible for URLLC to rely on higher layers, such as RLC, in order to resolve the control channel error, due to the latency constraint. In this regard, the errors of data and control channels should be considered for supporting URLLC with HARQ scheme. As identified in Section 4.2, the URLLC requirements can be met with HARQ scheme when the feedback channel, carrying ACK/NACK signals, is reliable enough. Instead, blind transmissions can be utilized when a reliable feedback channel is not available. This is achieved by transmitting replicas of data information without relying on a feedback channel. The data transmission can be performed using radio resources spanned over time and/or frequency domains in order to exploit diversity gain [PX]. This chapter presents different types of blind transmission schemes. Their communication reliabilities are assessed statistically. To have a fair comparison, their performances are compared to the one achieved with the HARQ scheme.

5.2 HARQ Transmissions

Consider the HARQ scheme with the maximum of $L$ transmission rounds. Each data transmission is followed by a DCI, indicating the assigned radio resources for the data transmission. The UE can access the data transmission if it can decode the corresponding DCI correctly. The UE reports either an ACK or a NACK after each data reception indicating the success or failure in decoding the message. In case the UE does not decode
the DCI, it would not send any feedback, which might be detected by the base station as a DTX. Figure 5.1 illustrates the HARQ scheme with the maximum of four transmission rounds.

![Figure 5.1. HARQ-based downlink data transmissions.](image)

Assume that the error of decoding the DCI by the UE is $\epsilon_I$. In addition, the probability of erroneous detecting a NACK as an ACK is $\epsilon_{NA}$, the probability of wrong detecting a DTX as an ACK is $\epsilon_{DA}$. The residual BLER of decoding the message after $i$th transmission rounds is $p_D(i)$. The success probability of retrieving the message is

$$P_{\text{HARQ success}} = \sum_{i=1}^{L} (1 - \epsilon_I)^i (1 - \epsilon_{NA})^{i-1} (1 - p_D(i)) \sum_{j=0}^{L-i} \epsilon_D^j (1 - \epsilon_D)^j$$

$\times \{1 + (i - 1)j - \text{mod}(1 + ij, i(L - 2) + 1)(1 - \epsilon_I)(1 - \epsilon_{NA})\}.$

(5.1)

5.3 Blind Transmissions

In blind transmissions, a replica of a message is transmitted multiple times, using different radio resources. The UE can try to retrieve the message from a single received data unit, or by combining multiples of the received data units. Depending on the implementation of DCI, different blind transmission schemes are envisaged. This section presents some of those schemes.

5.3.1 Normal DCI Transmission

Blind transmissions are enabled by transmitting replicas of data information. Prior to each data transmission, a DCI is transmitted to inform the UE regarding the assigned radio resources that carry the actual data information. The UE can monitor the data transmission and try to decode the message if it decodes the corresponding DCI successfully. Figure 5.2 illustrates the considered scheme with four data transmissions. Assume that $L$ replicas of data are transmitted, using the same radio
resource size for all of them. The error probability of decoding a DCI is $\epsilon_I$. In addition, each data transmission is self-decodable. The error probability of decoding the message depends on the number of received data units. $p_D(i)$ is the error probability of decoding the message using $i$ data units. The success probability of decoding the message can be expressed as

$$P_{\text{Blind(a) success}} = \sum_{i=1}^{L} \binom{L}{i} (1 - \epsilon_I)^i \epsilon_I^{N-i}(1 - p_D(i)) .$$

(5.2)

5.3.2 DCI Bundling

Decoding a DCI is prerequisite for monitoring data transmissions, and consequently decoding the message. This motivates employing enhanced DCI transmission in order to improve the reliability of DCI detection. This can be achieved by defining a frequency hopping pattern and including a few extra bits in the DCI as the indication of remaining transmission rounds (RTRs). In this way, when the UE decodes a DCI, it can monitor the corresponding data transmission and all the remaining data transmissions. Another advantage of this scheme is the flexibility for the number of DCI transmissions. The number of DCI transmissions can be equal or less than the number of data transmissions. This allows saving radio resources over the control channel by reducing the number of DCI transmissions for the UE that has a good DCI-detection performance. Figure 5.3 illustrates the blind data transmissions with four data transmissions and three bundled DCI transmissions.

Figure 5.2. Blind downlink data transmissions with normal DCI.

Figure 5.3. Blind downlink data transmissions with DCI bundling.
Assume that the number of bundled DCI transmissions is $l$. The success probability of decoding the message can be expressed by

$$P_{\text{Blind(b)}}_{\text{success}} = \sum_{i=1}^{l} \epsilon_{i}^{i-1}(1 - \epsilon_{1})(1 - p_{D}(L - i + 1)). \quad (5.3)$$

### 5.3.3 Repeated DCI Bundling

To further improve the communication reliability of blind transmissions, a DCI can be transmitted multiple times before performing the actual data transmissions. This allows the UE to monitor all the data transmissions if it can decode one of the DCI transmissions. The DCI can be transmitted using different radio resources expanded in the frequency domain in order to benefit from the frequency diversity gain while not increasing the latency. This scheme also provides a flexibility for the number of DCI transmissions, as the number of DCI transmissions can be different from the number of data transmissions. Figure 5.4 illustrates the blind data transmissions with the repeated DCI bundling, performing four data transmissions and three DCI transmissions.

![Blind downlink data transmissions with repeated DCI bundling.](image)

Assume that the number of bundled DCI transmissions is $l$. The success probability of decoding the message can be expressed as

$$P_{\text{Blind(c)}}_{\text{success}} = (1 - \epsilon_{l}^{l})(1 - p_{D}(L)). \quad (5.4)$$

### 5.4 Results

In order to compare the performance of the blind transmissions, two different scenarios are considered. In the first scenario, two data transmission rounds are considered, i.e., $L = 2$. In addition, the residual BLER of decoding the message for data transmission rounds are $p_{D}(1) = 10^{-3}$
and $p_D(2) = 10^{-6}$. In the second scenario, there are four data transmission rounds, i.e., $L = 4$. The residual BLER of data transmissions are $p_D(1) = 10^{-1}$, $p_D(2) = 10^{-3}$, $p_D(3) = 10^{-5}$, and $p_D(4) = 10^{-7}$. Figure 5.5 compares the reliability performance of the HARQ and blind transmission schemes. The performance of the HARQ scheme highly depends on the reliability of NACK detection while, any type of blind transmission schemes outperforms the HARQ scheme. When the error probability of DCI detection is low, the number of DCI transmissions can be reduced while high overall reliability is achieved.

**Figure 5.5.** Uplink data transmissions with an error in detecting the DCI.
5.5 Discussion

Blind data transmissions can achieve high communication reliability when a reliable feedback channel is not available. They can also achieve lower latency compared to the HARQ-based scheme. There have been some discussions in 3GPP meetings to support these schemes in 5G NR. Generally, the communication efficiency of blind transmissions are less than those of HARQ-based scheme. Hence, the blind transmissions should be activated only for specific UEs that cannot achieve satisfactory performance with a HARQ-based scheme. This motivates developing admission control schemes to select the proper transmission scheme for each UE. To further enhance the reliability and latency analysis, the system models can be applied to the 5G NR framework. In addition, the reliability of DCI detection depends on the number of DCI transmissions and the employed aggregation level. Hence, further analysis is required to properly configure these parameters to achieve good performance.
6. Conclusions and Future Research Directions

URLLC has been emerged as a new service for 5G networks to support mission-critical applications. This enables devising applications in various domains, such as industrial automation, remote surgery, and vehicular safety communications. The 5G NR has introduced new features in release 15 in order to meet the stringent requirements of the URLLC, in terms of availability, reliability, and latency. More enhancements are expected in release 16, which bring a better support for URLLC with more stringent requirements.

The dissertation studied the resource allocations for HARQ-based transmissions, achieving high reliability with a limited number of transmission rounds that are typically considered for supporting URLLC. It was observed that the reliability of the feedback channel, carrying ACK/NACK signals, can significantly affect the communication efficiency. In addition, the resource allocations should be performed considering the reliabilities of the data and control channels. To achieve this in cellular systems, a simple link adaptation scheme was proposed. The dissertation considered two time allocation schemes for UEs operating in cellular mode. These schemes can achieve better performance compared to case that the time budget is equally divided between uplink and downlink. In addition to cellular mode, D2D can be utilized for UEs located in a close proximity. The base station provides the radio resources for the UEs, while the UEs directly communicate for exchanging data. This mode can achieve a better radio utilization. The performance can be further enhanced by considering the hybrid mode, which is a combination of cellular and D2D modes. In this mode, the initial transmission is performed over the D2D link, while transmission diversity can be exploited for the data retransmissions, i.e., using either the downlink or D2D links.

Comprehensive communication models were considered in order to as-
Conclusions and Future Research Directions

sess the effects of the data and control errors on the overall communication reliability. The models were utilized to determine the reliability requirements for the control information in order to support a specific service. It was observed that trade-offs exist between the reliability requirements for the control information and the performance of the data transmissions. In addition, different enhancements were proposed for the control channels to better support URLLC. In case a reliable feedback channel is not available, blind transmission schemes can be utilized. These schemes achieve high communication reliability by transmitting replicas of the message, which requires utilizing more radio resources.

In summary, the dissertation studied some of the problems related to the support of URLLC in cellular systems. It also presented some enhancements, particularly for the control channels, to achieve a better performance. Some of the research directions for further improvement are as follows.

- The communication models can be extended to more realistic communication scenarios, considering the channel fading, channel estimation error, and the channel report delay. These parameters can be considered, for instance, for resource allocation and link adaptation.

- The resource allocation optimization can be considered jointly for data and control channels. For instance, the periodicity of CSI report and the assignment of data resources can be configured according to the UE’s traffic parameters.

- A similar idea of asymmetric ACK/NACK detection can be implemented for detecting CQI values. This limits the rate of erroneous decoding a CQI value as a higher, which sacrifices the communication reliability.

- Multi-user scheduling is another important concern that should be addressed. The problem becomes more complex if the users have different communication requirements in terms of latency and reliability.

- 5G supports mini-slot scheduling in order to reduce the TTI. Employing the shorter TTI gives the opportunity for having more transmission opportunities, while increasing the signaling overhead. In this regard, the trade-off between the TTI and communication efficiency should be
investigated.
References


[6] 3GPP TR 38.913, “Technical specification group radio access network; study on scenarios and requirements for next generation access technologies; (Release 14).”


[22] 3GPP TS 136.201, “LTE; evolved universal terrestrial radio access (E-UTRA); LTE physical layer; general description,” Release 14, 2017.


Errata

Publication X

In equation (1), $\epsilon_N$ and $\epsilon_D$ should be replaced by $\epsilon_{NA}$ and $\epsilon_{DA}$, respectively.
Enabling Ultra-Reliable Low-Latency Communications in 5G Networks

Hamidreza Shariatmadari