Tampere University

Syed Hassan Uz Zaman Gillani

TRADEOFF BETWEEN LATENCY AND THROUGHPUT IN 5G NETWORKS WITH HYBRID-ARQ RETRANSMISSIONS

Master Thesis
Faculty of Information
Technology and Communication
May 2020

ABSTRACT

Syed Hassan Uz Zaman Gillani: Trade-off Between Latency and Throughput In 5G Networks with Hybrid-ARQ Retransmissions
Master's Thesis
Tampere University
Faculty of Information Technology and Communication
May 2020

Ultra-reliable Low Latency Communications (URLLC) is one of the key enabling technology in Fifth Generation New Radio (5G-NR), which promises to provide reliability and ultra-low latency communication link for different mission critical and remote operations. To make it possible different retransmission schemes and error correction techniques are used in 5G-NR to minimize the delay and increase the reliability of the data packet transmission under variant wireless channel conditions.

The main purpose of this thesis is to find the trade-off between throughput and latency in the network by simulating different Modulation and Coding Schemes (MCSs) under varying channel conditions. The Cyclic Prefix Orthogonal Frequency Division Multiplexing (CP-OFDM) waveform is used for downlink data transmission. The waveform is propagated through different noisy channel conditions. The throughput and delay, in terms of required retransmissions, are then estimated on the successful packet reception by the receiver and the subframe retransmissions by hybrid-ARQ during the simulation respectively. By executing various MCSs different graphs are generated and results are concluded. Based on these conclusions robust modulation schemes with coding rates are suggested for the critical applications, whose requirements are very strict in terms of reliability and a threshold level of data rate.

In addition, a literature study is done on the 5G-NR radio interference architecture, waveform, data channels, transport channel process and the working mechanism of retransmission protocols for the data transmissions. Thus, this flourish the basic understanding of the overall process involved throughout the simulations.

Key words: Hybrid-ARQ, 5G-NR, Retransmission protocols, URLLC, MCS, Throughput, Delay.

PREFACE

First of all, I would like to thank Almighty Allah for his blessing and guidance, which has enabled me to believe in myself and complete my master's degree studies.

This thesis titled "Trade-off Between Latency and Throughput In 5G Networks with Hybrid-ARQ Retransmissions" is supervised by Dr. Jukka Talvitie for the completion of master's degree in faculty of Information Technology, Communication systems and Networks at Tampere University.

As a student of wireless communication, I would like to thank Dr. Jukka Talvitie for choosing me as a thesis candidate under his supervision. I acknowledge the guidance and valuable feedback from my professor during whole thesis, which enabled me to acquire sufficient knowledge about the fifth-generation mobile networks. While doing the thesis I have learned different research methodologies and scientific writing. I would also like to thank Dr. Toni Levanen, my thesis co-supervisor for his review and suggestion regarding the improvement of this thesis.

Finally, I would like to thank to my family members and friends for their motivational support and unconditional love for me. Special thanks to my father and mother, who always believe in me and remember me in prayer for every time.

Tampere, 22 May 2020

Syed Hassan Uz Zaman Gillani

CONTENTS

1 INTRODUCTION 1
1.1 Motivation and Scope of Thesis 2
2 FIFTH-GENERATION NEW RADIO (5G-NR) 3
2.1 Evolution of 5G 3
2.2 Fifth Generation (5G-NR) 5
2.2.1 Standardization for 5G 5
2.2.2 Key Enabler Technologies for 5G 6
2.2.3 High frequency and Spectrum for 5G 7
2.3 Radio Interface Architecture 8
2.3.1 Quality of Service 8
2.3.2 Radio Protocol Stack 9
2.4 Modulation Scheme 11
2.4.1 Frame Structure 12
2.4.2 Duplex Scheme 13
2.5 Low-latency Support 14
2.6 Max Throughput Support 16
3 DATA CHANNELS AND RETRANSMISSION PROTOCOLS 18
3.1 Data Channels in (5G-NR) 18
3.2 Transport Channel Process 19
3.3 Retransmission Protocols 22
3.3.1 Overview 22
3.4 Hybrid-ARQ Protocol 23
3.4.1 Soft Combining in Hybrid-ARQ 24
3.4.2 Downlink Hybrid-ARQ 26
3.4.3 Uplink Hybrid-ARQ 27
3.4.4 Uplink Acknowledgements Timing 28
3.4.5 Multiplexing of Hybrid-ARQ Acknowledgements 29
3.5 RLC layer 31
3.5.1 Retransmissions in RLC layer 33
3.6 PDCP 35
3.6.1 Retransmission in PDCP Layer 35
3.7 Why do we need multi-layer retransmission protocol? 36
4 SYSTEM MODEL AND SIMULATION RESULTS 37
4.1 Model Overview 37
4.2 Working Principle 39
4.3 Parameters 40
4.4 Simulation Findings 42
4.4.1 Throughput Analysis 42
4.4.2 Delay Analysis 43
4.4.3 Delay Vs Throughput at – 4dB 45
4.4.4 Delay Vs Throughput at 0dB 46
4.4.5 Delay Vs Throughput at 6dB 47
4.4.6 Delay Vs Throughput at 10dB 48
5 CONCLUSIONS 49
REFERENCES 51

LIST OF FIGURES

Figure 2-1. *Evolution of Mobile communication* 4
Figure 2-2. *Evolution of LTE and NR jointly providing the overall 5G radio-access solution [13]* 4
Figure 2-3. *3GPP Timeline* 5
Figure 2-4. *Key enabler technologies for 5G NR [16]* 6
Figure 2-5. *Spectrum Allocation for NR* 7
Figure 2-6. *Different combination of core networks and radio-access technologies [22]* 8
Figure 2-7. *QoS flows and radio bearers during a PDU session [15]* 9
Figure 2-8. *User-plane and Control-plane protocol stack* 11
Figure 2-9. *Time domain frame structure for NR [28]* 12
Figure 2-10. *Frequency domain frame structure for NR [29]* 13
Figure 2-11. *Different duplex schemes* 14
Figure 2-12. *Decoupling transmissions from slot boundaries to achieve low latency [15]* 15
Figure 3-1. *Mapping between logical, transport and physical channels* 19
Figure 3-2. *Selection of base graph for the LDPC code [15]* 20
Figure 3-3. *Overall transport process for the downlink and uplink [29]* 21
Figure 3-4. *Multiple hybrid-ARQ processes.* 23
Figure 3-5. *Codeblock-group retransmission.[15]* 24
Figure 3-6. *Example of incremental redundancy [15]* 26
Figure 3-7. *Illustration of per-CBG retransmission [15]* 27
Figure 3-8. *Determining the acknowledgement timing [15]* 28
Figure 3-9. *Example of semistatic hybrid-ARQ acknowledgement codebook [15]* 30
Figure 3-10. *Example of dynamic hybrid-ARQ acknowledgement codebook [15]* 31
Figure 3-11. *MAC and RLC.* 32
Figure 3-12. *SDU delivery in acknowledged mode [15]* 34
Figure 3-13. *Retransmission of missing PDUs.* 35
Figure 4-1. *Downlink transmission of PDSCH* 38
Figure 4-2. *Overall process block diagram* 39
Figure 4-3. *Throughput Vs SNR* 42
Figure 4-4. *Delay Vs SNR* 44
Figure 4-5. *Delay Vs Throughput at -4dB* 45
Figure 4-6. *Delay Vs Throughput at 0dB* 46
Figure 4-7. *Delay Vs Throughput at 6dB* 47
Figure 4-8. *Delay Vs Throughput at 10dB* 48

LIST OF TABLES

Table 1. *Frequency ranges* 7
Table 2. *Different numerology defined in NR [26]* 12
Table 3. *Use cases for 5G and their requirement in term of URLLC [30]* 14
Table 4. *Minimum processing time for feedback signal [15]* 16
Table 5. *Different parameters set for the whole simulations* 41

LIST OF SYMBOLS AND ABBREVIATIONS

1G	First Generation
2G	Second Generation
3G	Third Generation
3GPP	Third Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
ACKs	Acknowledgements
AM	Acknowledged Mode
AMPS	Advanced Mobile Phone System
ARQ	Automatic Repeat Request
BCCH	Broadcast Control Channel
BCH	Broadcast Channel
BG	Base Graph
CBGFI	Codeblock Group Flush Indicator
CBGs	Codeblock Groups
CBGTI	Codeblock Group Transmitter Indicator
CC	Chase Combining
CCCH	Common Control Channel
cDAI	Counter Downlink Assignment Index
CSI-RS	Channel State Information Reference Signal
CN	Core Network
CP	Cyclic Prefix
CRC	Cyclic Redundancy Check
DAI	Downlink Assignment Index
D-AMPS	Digital- Advanced Mobile Phone System
DCCH	Dedicated Control Channel
DCI	Downlink Control Information
DFT	Discrete Fourier Transform
DL-SCH	Downlink Shared Channel
DTCH	Dedicated Traffic Channel
eMBB	Enhanced Mobile Broadband
eNB	Evolved Node B
EPC	Evolved Packet Core
FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
FEC	Forward Error Correction
FFT	Fast Fourier Transform
FR	Frequency Range
Gbps	Gigabits per second
GHz	Giga Hertz
gNB	Next Generation Node B
GSM	Global System for Mobile Communication
HD	High Definition
HSPA	High Speed Packet Access
hybrid-ARQ	Hybrid Automatic Repeat Request
IMT	International Mobile Telecommunication
IoT	Internet of Things
IR	Incremental Redundancy
ITU-R	International Telecommunication Union Radiocommunication
LDPC	Low-Density Parity-Check

LTE	Long Term Evolution
MCS	Modulation and Coding Scheme
MAC	Medium Access Control
Mbps	Megabits per second
MHz	Mega Hertz
MIMO	Multiple Input Multiple Output
mm	millimeter
mMTC	Massive Machine-Type Communications
MRC	Maximum Ratio Combining
ms	milliseconds
NAS	Non-Access stratum
NGMN	Next Generation Mobile Network
NMT	Nordic Mobile Telephony
NR	New Radio
OFDM	Orthogonal Frequency Division Multiplexing
PBCH	Physical Broadcast Channel
PCCH	Paging Control Channel
PCH	Paging Channel
PDC	Personal Digital Cellular
PDCCH	Physical Downlink Control Channel
PDCP	Physical Data Convergence Protocol
PDSCH	Physical Downlink Shared Channel
PDU	Packet Data Unit
PHY	Physical Layer
PRACH	Physical Random-Access Channel
PUCCH	Physical Uplink Control Channel
QAM	Quadrature Amplitude Modulation
QFI	Quality of Service Flow Identifier
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RACH	Random-Access Channel
RAN	Radio Access Network
RAT	Radio Access Technology
RBs	Resources Blocks
RE	Resource Element
RLC	Radio Link Control
ROCH	Robust Header Compression
RRC	Radio Resource Control
RV	Redundancy version
SDAP	Service Data Application Protocol
SDN	Software Define Network
TACS	Total Access Communication System
TCP	Transmission Control Protocol
TD-SCDMA	Time Division Synchronous Code Division Multiple Access
tDAI	Total Downlink Assignment Index
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TF	Transport Format
TM	Transport Mode
TTI	Transmission Time Interval
UE	User Equipment
UL-SCH	Uplink Shared Channel
UM	Unacknowledged Mode
UPF	User Plane Functionality
UCI	Uplink Control Information

URLLC	Ultra-Reliable Low-Latency Communications
V2V	Vehicle to Vehicle

1 INTRODUCTION

During the last decades, the use of wireless technology has been exponentially expanding, and there is a trend of increasing demand for high-speed data rates and reliable connectivity. At present, we have advanced communication infrastructure that can support above mentioned requirements, but this infrastructure is not able to support all emerging technologies and applications. These applications include mission critical operations with high demanding throughputs, reliable connectivity even in worst situations, more extensive internet of things (IoT), industrial revolution into massive machine connectivity, remote surgeries with extreme low latency, vehicle to vehicle (V2V) communications and high data rate demand for high quality videos and virtual reality applications.

To address the requirement for these applications, Fifth Generation (5G) New Radio (NR) was approved in the end of 2017 with the first specification Release 15 by the Third-Generation Partnership Project (3GPP). The 5G promises to achieve high data rate peaks of 10 Gigabits per second (Gbps) which is 10 times more than the current wireless system deployed. It will enable to handle the communication between a large number of connected devices with low power consumption. It will provide ultra-low latency for remote critical mission with higher reliability. The three key enabler technologies for the 5G are Enhanced Mobile Broadband (eMBB), massive Machine-Type Communications (mMTC) and Ultra-Reliable Low-Latency Communications (URLLC). [1, 2]

The most challenging part for designing a wireless communication system is to eliminate the presence of the fading. The multipath propagation and signal variation with respect to time, due to different environmental obstacles or mobility of the receiver causes the phenomena of fading. This effect degrades the throughput and reliability of the signal power and gives rise to the errors in the data packet transmission [3]. To mitigate the fading effect different algorithms and retransmission schemes are implemented in the transmitter and receiver. In such schemes, different versions of signals can be transmitted with redundancy and then combined in the receiver to get the original signal. Every signal is transmitted with an individual fading path to overcome the errors and increase the signal reception power.

By using the different retransmission schemes is a standard approach to increase the diversity gain for the transmitted signals. The most common schemes used in the communications systems are repetitive and parallel. In a repetitive scheme, the transmitter sends the data packets in an available fading channel and the receiver tries to decode the data depending on the communication protocol implemented. While in the parallel retransmission scheme, the transmitted packets are placed differently and joined together in a way that, in receiver, these packets with different information are combined to retrieve all the data.

Most retransmission protocols follow these schemes. The Automatic Repeat Request (ARQ) is the simplest retransmission protocol and uses repetition coding scheme [4, 5]. The transmitters in ARQ, retransmit the errored packets until it receives the positive acknowledgement from the receiver. Further protocols which use a retransmission scheme is Hybrid Automatic Repeat Request (hybrid-ARQ) [6]. The hybrid-ARQ uses the same repetition coding scheme but it also uses forward error correction. Meaning that in the receiver's side, it buffers the errored packets and then by retransmissions it combines the different versions of the data packet to decode the information. There are several Hybrid-ARQ protocols and are classified according to how the packets are combined by the receiver. For example, the receiver performs Maximum Ratio Combining (MRC) on the received signal and decodes the information from it then this scheme is called Chase Combining (CC) hybrid-ARQ [7-9]. By using these schemes, the probability of successful transmission increases but the overall throughput of the system decreases. Several retransmission protocols implement parallel coding schemes to reduce losses. These protocols use Incremental Redundancy Schemes (IR) [6, 10, 11]. In this scheme the packets are transmitted in parallel and every individual retransmitted packet contains different information from others. These packet bits are combined in the receiver with previous transmission attempts of the same packet. The use of these retransmission protocols is adopted by several systems, like 3GPP, 3GPP LTE-IOT and Microwave Access (WiMAX). [12]

1.1 Motivation and Scope of Thesis

In this thesis, a literature study has been conducted on different retransmission protocols used in 5G New Radio (5G-NR). One of the key enabler technology of 5G NR is the URLLC, which promises to deliver ultra-low latency and high reliability for different critical application. The use cases for URLLC are industrial automation, intelligent transportation and remote health care.

So, to provide the optimal throughput and maximum reliability for these applications in context of retransmissions, different scenarios are observed in the thesis. As a parameter, different Modulation and Coding Schemes (MCSs) are used to analyse throughput and delay by applying the hybrid-ARQ protocol.

2 FIFTH-GENERATION NEW RADIO (5G-NR)

This chapter covers a brief overview of Fifth-generation new radio (5G-NR) cellular network which includes evolution of 5G, radio interface architecture, modulation scheme, low-latency support and maximum throughput support.

2.1 Evolution of 5G

For the last 4 decades, there has been a revolutionary change in the mobile communication. These changes have been developed step by step from First Generation (1G) to 5G as illustrated in Figure 2-1.

The 1G mobile communication technology was evolved around 1980 by the Nordic Mobile Telephony (NMT) established within Nordic countries, Advanced Mobile Phone System (AMPS) in North America and Total Access Communication System (TACS) used in the United Kingdom. The basic key enabler technology was based on an analog communication system which only facilities the voice services for limited users.

In the early 1990s, the Second Generation (2G) of mobile communication known as Global System for Mobile communication (GSM) was introduced. Within Japan, Digital AMPS (D-AMPS), Personal Digital Cellular (PDC) contributed to the development of 2G and further in America they used IS-95/CDMA and IS-136/TDMA technologies [13]. In this generation the communication between the radio links started to use new digital transmission, which provided better quality of voice calls with limited data services. Gradually the second-generation technology was spread from Europe to other parts of the world and was widely used by mobile telephony companies.

The Third Generation (3G) of mobile communication emerged in early 2000. The main improvement in this generation was high-quality mobile broadband. Further evolution of 3G is known as High Speed Packet Access (HSPA) which enabled fast wireless internet access, which is still currently widely used. [14]. The 3G introduced the unpaired spectrum based on china-developed Time Division Synchronous Code Division Multiple Access (TD-SCDMA), which used Time Division Duplex (TDD). While earlier mobile communication was designed to work in paired spectrum (separate spectrum for network to device and device to network links) which were based on Frequency-Division Duplex (FDD). [15]

The first release of technical specifications of the fourth-generation mobile communications, referred as Long-Term Evolution (LTE), was announced in 2009. It follows the footstep of HSPA, providing achievable higher data rate of broadband with increased efficiency to the end-users by using OFDM (Orthogonal Frequency Division Multiplexing) based scheme for transmission. It also uses a wider spectrum and advanced multi-antenna technologies. In addition, LTE supports both FDD and TDD, operating in paired and unpaired spectrum, within one common radio-access technology. In addition to conventional broadband data communications, LTE considers various supplementary use cases, such as low-cost devices with very long battery life, mMTC applications and low latency in air-interface. So, the evolution of LTE will be able to support a wide range of the use cases envisioned for 5G. [15]

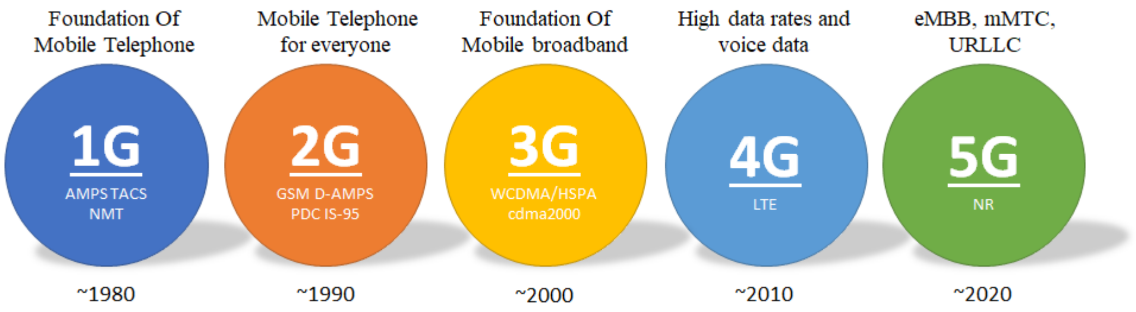

Figure 2-1. *Evolution of Mobile communication*

The NR technology has remodelled the envision technologies of LTE. There is tight interconnection between the evolution of LTE and NR as shown in Figure 2-2. As the LTE has been deployed for more than 10 years and the new more demanding technical requirements are not possibly satisfied. Therefore, the 3GPP held a workshop in fall of 2015 to meet the new requirement and specification for advanced technologies. As a result, new radio-access technology was found, nowadays known as NR. In the end of 2017, the first version of NR specification was published and initial design of 5G was commercially deployed in 2018. [15]

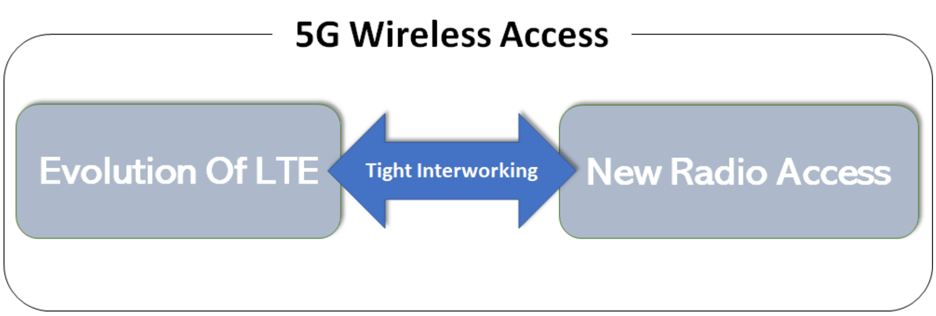

Figure 2-2. *Evolution of LTE and NR jointly providing the overall 5G radio-access solution [13]*

2.2 Fifth Generation (5G-NR)

5G stands for Fifth Generation of wireless cellular technology, which is totally based on digital communication. 5G is often used in a wider context but we mostly refer to 5G NR, which is the new radio access technology (RAT) standardized by the 3GPP.

2.2.1 Standardization for 5G

The International Telecommunication Union Radiocommunication Sector (ITU-R) is responsible for the approval of 5G Standards. The International Mobile Telecommunication IMT-2020 in ITU is known for the standards of 5G, it encompasses pre-commercial activities and 5G trials, to assist in evaluating the candidate technologies and frequency bands.

The 3GPP follows the process of standardization in parallel with ITU-R timeline. The first release of 5G specification by 3GPP was published in mid-2017, when the 3GPP technical specifications groups agreed on a detailed work plan for Release 15. Every year roughly one 3GPP publication is released. These releases conduct study on both RAN and core aspects of NR. For NR, the radio aspect in 3GPP is defined in 38-series.

The first specification of Releases 15 is for non-standalone NR operation, in which the NR devices will rely on LTE for mobility and initial access. While the final Release 15 will also support standalone operation, as illustrated in Figure 2-3.

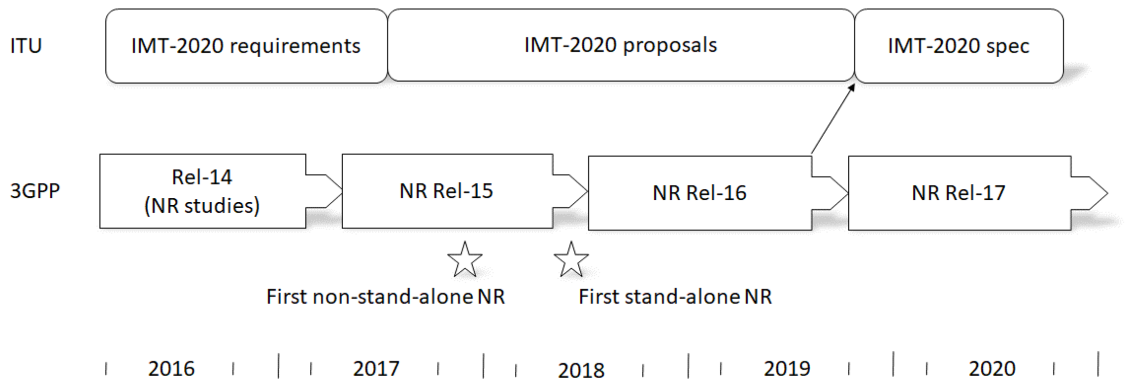

Figure 2-3. *3GPP Timeline*

2.2.2 Key Enabler Technologies for 5G

The fifth-generation mobile communication incorporates wide range of use cases, with higher peak throughput, high spectral efficiency, new spectrum band and low latency. The 5G network will not only flourish the mobile industry, it will also serve new business models and target new services. Later the vertical industries will also collaborate with the 5G development [16]. To fulfil the requirement for theses verticals the ITU-R [17], Next Generation Mobile Network (NGMN) Alliance [18] and METIS-I project [19] have considered the following main three 5G service type and show in Figure 2-4:

1. Enhanced Mobile Broadband (eMBB), provides extremely high data rates for high user density and very high capacity. The Enhanced Mobile Broadband will be more concentrated on human-centric communication and reliable access over large coverage areas.
2. Massive Machine-Type Communications (mMTC), as the name suggests is pure machine centric communication and will require to connect tens of billions of network-enabled devices, for example, in the context of Internet of Things (IOT). These devices will have a long battery life with remote deployment and very low cost.
3. Ultra-reliable low-latency Communications (URLLC), can be characterized as human-centric or machine-centric. It obligates ultra-low latency, high reliability and availability. An example for URLLC human-centric scenarios are tactile internet and augmented reality. Correspondingly, examples for machine-centric use cases are remote medical surgery, vehicle-to-vehicle communication, connecting smart grids and wireless control of industrial equipment.

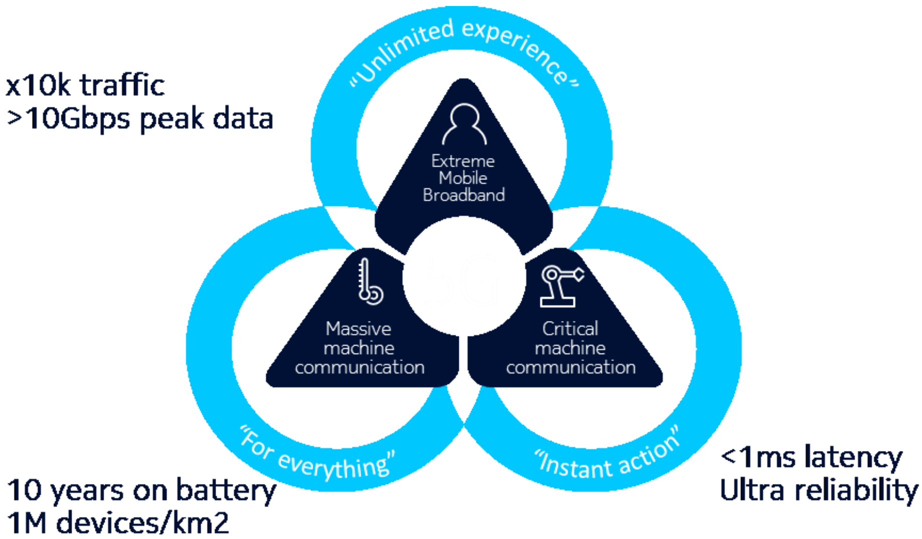

Figure 2-4. *Key enabler technologies for 5G NR [16]*

2.2.3 High frequency and Spectrum for 5G

As there is an increased demand for high data rate, the 5G mobile system introduces new spectrum for higher-frequency bands (above 6 GHz). Higher frequency bands are the key enabler for the 5G and there is abundant spectrum available to support ultra-high data rate. The frequency ranges from 30-300 GHZ are known as millimetre (mm) waves, but not necessarily all mm waves are suitable for mobile communication because of specific propagation characteristics. Above 6 GHz the candidate spectrum is affected by multiple channel propagation challenges, such as path loss, doppler effect and absorption by the atmosphere. Although these factors can be compensated by using new advanced technology like multi antenna transmission and receiving, similar to the beam centric design in NR. In 3GPP release 15 the NR frequencies are divided in two separate ranges as shown in Table 1 [20] . The Figure 2-5 illustrates the different frequency allocated for 5G NR.

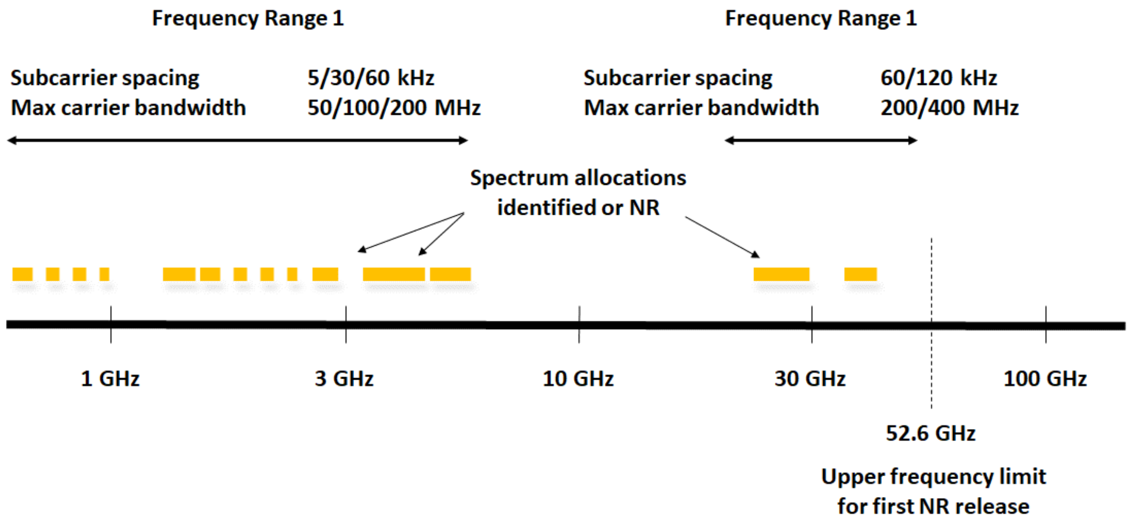

Figure 2-5. *Spectrum Allocation for NR*

Table 1. *Frequency ranges*

Frequency Range Designation	Corresponding Frequency Range
FR1	410 MHz – 7125 MHz
FR2	24250 MHz – 52600 MHz

2.3 Radio Interface Architecture

In 5G NR, the overall network architecture for Core Network (CN) and the Radio-Access Networks (RAN) were revisited. The core network of 5G function includes authentication, end to end connection setup, policing and charging. However, it is not responsible for any radio access network functionality. The Radio access network is entirely responsible for the radio related functions, that includes radio-resource management, retransmissions, scheduling, coding and various multi antenna schemes. Splitting these network functionalities is beneficial to connect different radio-access technologies to the same serving core network.

The radio-access network of the NR can be deployed to the legacy LTE (Long Term Evolution) core network known as Evolved Packed Core (EPC). This type of deployment is known as non-stand alone, in which EPC of LTE is providing the primary function like paging, mobility and set-up to the radio access network of NR shown in Figure 2-6 . In later release of NR, the 5G core network will directly connect to the radio network of NR known as standalone operation. [15, 21]

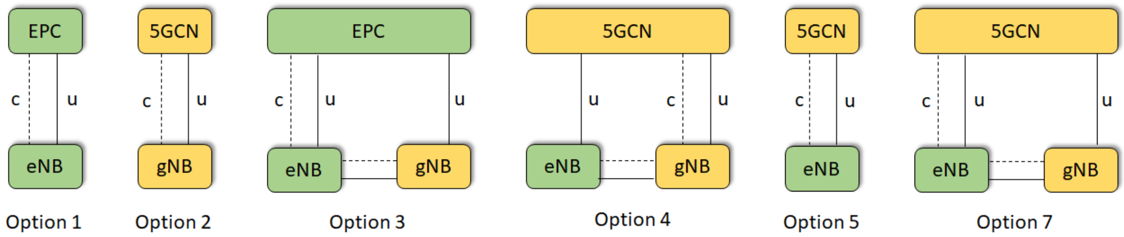

Figure 2-6. *Different combination of core networks and radio-access technologies [22]*

2.3.1 Quality of Service

The feature of Quality of Service (QoS) has been already considered in earlier mobile network generations, but the NR enhances the framework of QoS. Whereas in LTE only the core network is aware of QoS, in NR the QoS awareness is extended to the radio-access network. Due to this enhancement of QoS in radio-access, one of the architecture systems of 5G known as Network slicing is achievable and is an essential part of it.

The Figure 2-7 shows the QoS process in NR. When a User Equipment (UE) is connected to a gNB, one or more Packet Data Unit (PDU) sessions start. Each session has one or more QoS flows and data radio bearers. In core networks, these IP packets are mapped according to the QoS requirement like delay or required data rate as a part of User Plane Function (UPF) functionality. To assist uplink QoS each packet is marked with QoS Flow Identifier (QFI). While in radio-access network only mapping of QoS and data radio bearer is done. Not only one to one mapping of QoS-flow-to-radio-bearer is possible, but there can also be a multiple QoS flows mapping to the same data radio bearer.[15]

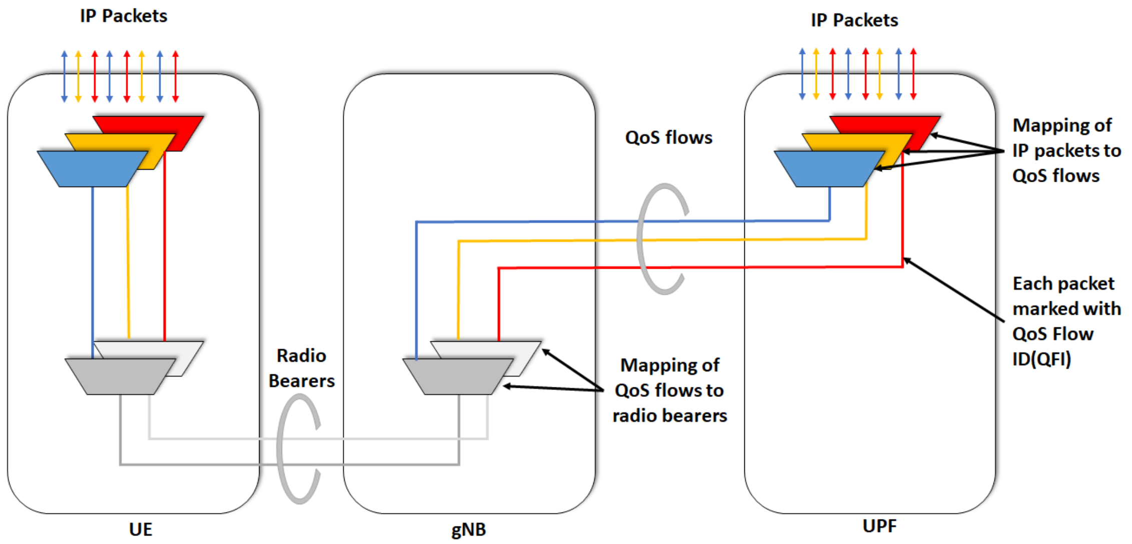

Figure 2-7. *QoS flows and radio bearers during a PDU session [15]*

2.3.2 Radio Protocol Stack

To study the radio-access network of NR, the radio protocol stack is divided into two parts that are user plane and control plane. The Figure 2-8 shows the user-plane and control plane protocol stack. The user plane carries the user data traffic while the control plane carries the control signalling traffic from the core network to the radio network. [15, 23]

The user plane protocol comprises of the Service Data Application Protocol (SDAP), Packet Data Convergence Protocol (PDCP), Radio-Link Control (RLC), Medium-Access Control (MAC) and Physical Layer (PHY).

Although the user data protocol of NR is similar to LTE it has some differences as well. The NR introduces a new Service Data Application Protocol (SDAP), which is responsible for the QoS handling in radio-access networks. If the NR is connected to the EPC then the functionality of SDAP is not in use. The different user data protocols are described in the following section.

- The Service Data Adaptation Protocol (SDPA) functionality is for mapping between QoS flow from the 5G core network and data radio bearer. It also marks the IP packets with QFI for uplink and downlink. However, when operating in non-standalone mode, the SDAP is not used.

- Packet Data Convergence Protocol (PDCP) performs compression of the IP header with ciphering and integrity protection. It deals with duplicate removal, retransmissions and in sequence delivery in the case of handover. Also, duplication and routing can be provided by the PDCP in case of dual connectivity with split bearers.
- Radio-Link Control (RLC) performs the segmentation and retransmission handling. In the form of RLC channels, it provides service to the PDCP. One RLC entity is configured per device. To reduce the delay it does not support the in-sequence delivery of data to higher protocol layers.
- Medium-Access Control (MAC) is responsible for the hybrid-ARQ retransmissions, multiplexing of logical channels, and scheduling related functions. In the form of logical channels, it provides services to the RLC layers. It also handles different numerologies defined in NR. The MAC protocol structure has been modified to support the low latency and high throughput in NR compared to LTE. The MAC layer is discussed in detail in chapter 3.
- Physical Layer (PHY) performs multi-antenna mapping, coding/decoding, modulation/ demodulation and mapping of signals to the physical time and frequency resource. It provides services to the MAC layer in the form of a physical channel.

The control plane protocols are responsible for mobility, connection setup and security. It includes Non-Access stratum (NAS), Radio Resource Control (RRC), PDCP, RLC, MAC and PHY.

- The Non-Access stratum (NAS) is a functional layer between the device and AMF in the core network. It provides security, authentication, assign IP address to the devices and also paging procedures.
- Radio Resource Control (RRC) is located in the gNB and responsible for resource management, broadcast of system information, transmission of paging messages, connection management, mobility functions, measurement configuration and reporting.

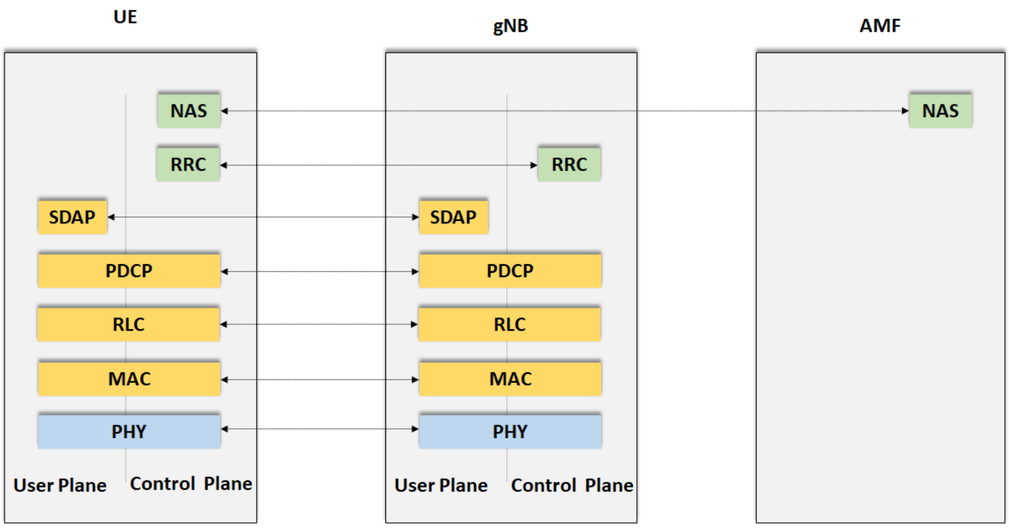

Figure 2-8. *User-plane and Control-plane protocol stack*

2.4 Modulation Scheme

In March 2016, the 3GPP proposed Orthogonal Frequency Division Multiplexing (OFDM) as a modulation scheme for the 5G NR because of its various significant features which remain dominant against any other waveform [24]. In OFDM, a channel is divided into many narrow subcarriers. The spacing between these subcarriers are such that they are orthogonal to each other and in ideal scenarios won't interfere even without the presence of guard bands. Because of this, OFDM has a very high spectral efficiency which meets the requirements of extreme high data rates. It is robust in time and frequency domain which allow it to select different subcarrier spacing and cyclic prefix (CP) according to channel conditions. In multiple-input and multiple-output (MIMO) antenna technology, which is a key enabler for NR, it is essential to provide high spectral efficiency and a greater coverage (via beamforming). And it can be achieved by using OFDM technology. [25]

The other important aspect of OFDM in NR is the selection of numerology (subcarrier spacing) and the cyclic prefix (CP). Unlike in LTE, the 5G NR has a flexible numerology multiple regarding subcarrier spacing and allows to support a wide range of frequency range from sub-1 GHz to 100 GHz according to different use case and deployment scenarios. In LTE, increasing bandwidth means increases in the subcarriers which lead to use of increased Fast Fourier Transform (FFT) size and create computational complexity. While in NR, using higher subcarrier spacing can result in lower FFT size. The Table 2 shows different numerology options for the NR, which can support up to 400 MHz with 120kHz subcarrier spacing. The parameter μ in Table 2 reflects the multiplication factor of the subcarrier spacing, as shown in the 2nd column title of the table.

***Table 2.** Different numerology defined in NR [26]*

μ	$\Delta f = 2^{\mu}.15$ kHz	Max BW (MHz)	Cyclic Prefix
0	15	50	Normal
1	30	100	Normal
2	60	200	Normal, Extended
3	120	400	Normal
4	240	-	Normal

2.4.1 Frame Structure

The structure for the NR transmission in time domain is organized in frames as shown in Figure 2-9. The time duration of one frame is 10ms. Each frame is further divided into 10 subframes with time duration of 1ms. One subframe is further divided into slots consisting of 14 OFDM symbols each and the duration of each slot is in milliseconds depending on the used numerology. For example, for the 15 kHz subcarrier spacing one slot duration is 1ms and for 30 kHz subcarrier spacing the one slot duration is 0.5ms. A slot can contain all uplink, all downlink or at least one downlink or uplink part. [27]

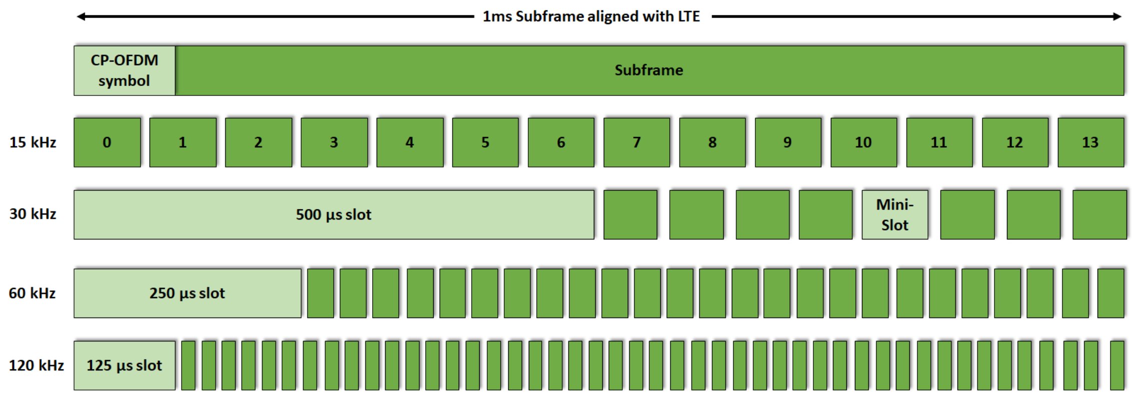

***Figure 2-9.** Time domain frame structure for NR [28]*

The frequency domain structure of transmission schemes for the NR is defined in a similar way to LTE. The physical resources are represented in a resource grid. A Resource Element (RE) is the smallest unit of the resource grid and is formed up from one OFDM symbol in time domain and one subcarrier in frequency domain. While in NR the resource block differs from LTE. The resource block is one dimensional measure spanning a frequency domain and consists of consecutive subcarriers. The reason for defining the resource block in frequency domain only is the flexibility in time duration for different transmissions. The Figure 2-10 illustrates the frame structure of NR in frequency domain.

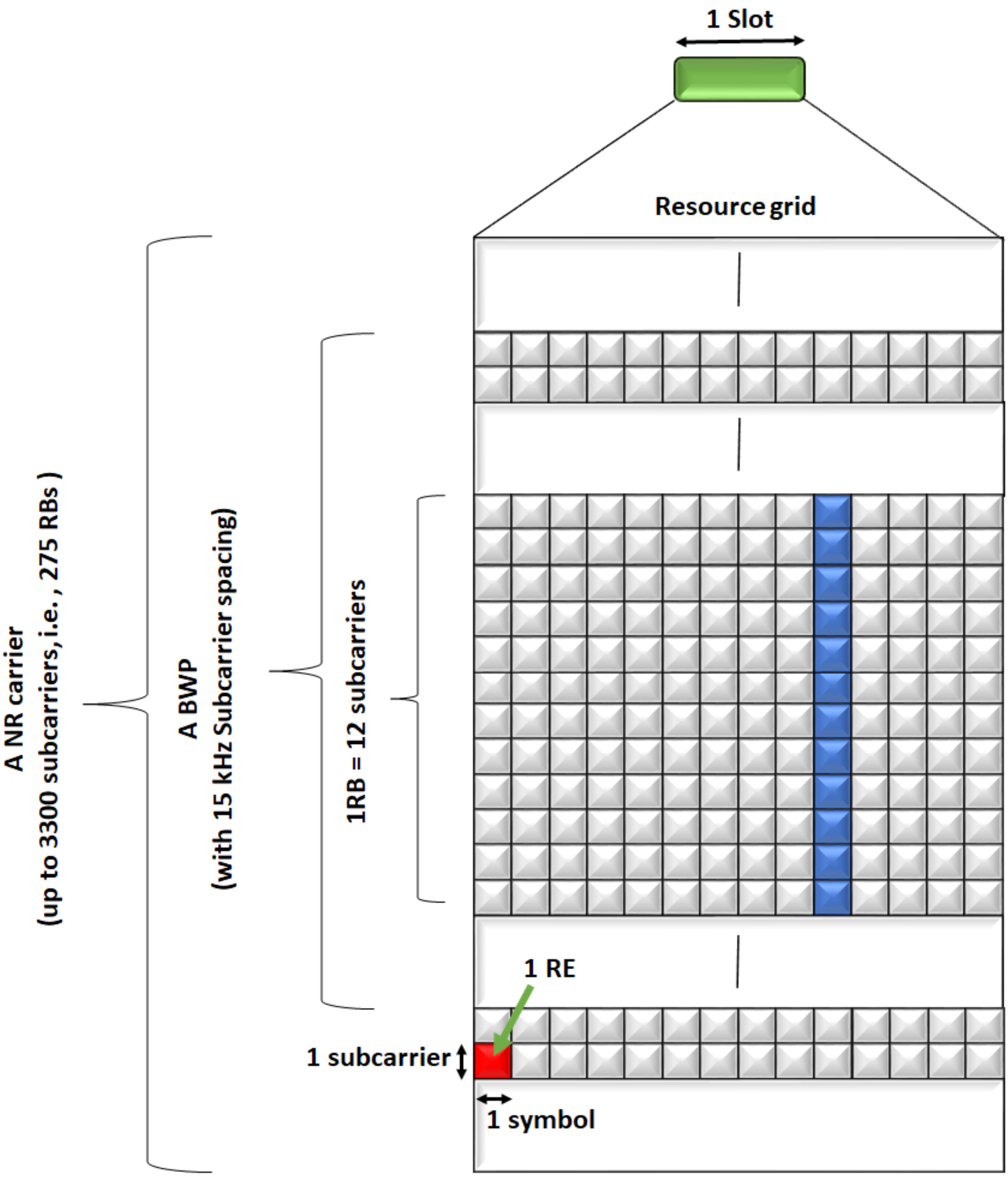

Figure 2-10. *Frequency domain frame structure for NR [29]*

2.4.2 Duplex Scheme

In NR both the FDD and TDD transmissions are supported. The deployment of the duplex scheme depends on the spectrum allocation. At higher frequencies, TDD is preferred because of the unpaired allocation of spectrum. At lower frequencies, FDD is preferred because of the paired allocation of spectrum. Furthermore, NR also supports dynamic TDD, in which the same carrier frequency is shared in time domain both in uplink and downlink. Due to this enhancement it can handle rapid variation of the traffic. The Figure 2-11 shows different duplex schemes used in NR.

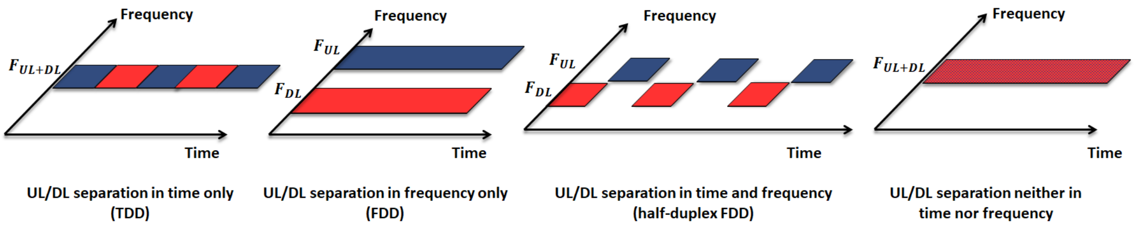

Figure 2-11. *Different duplex schemes*

2.5 Low-latency Support

The URLLC is one of the main service categories in 5G NR defined by ITU. The design of URLLC is the most challenging because of the requirement of low latency and ultra-high reliability. The use cases for the URLLC are, for example, in factory automation, smart city automation, remote surgery, tactile interaction and self-driving vehicles. The following Table 3 shows some envisioned use cases for 5G and their requirement in term of URLLC in 3GPP. [30]

Table 3. *Use cases for 5G and their requirement in term of URLLC [30]*

Scenario	End-to-end latency	User Experienced data rate
Discrete automation (motion control)	1 ms	1 Mbps up to 10 Mbps
Discrete automation	10 ms	10 Mbps
Process automation	50 ms	1 Mbps up to 100 Mbps
Process automation (monitoring)	50 ms	1 Mbps
Electricity distribution (medium voltage)	25 ms	10 Mbps
Electricity distribution (high voltage)	5 ms	10 Mbps
Intelligent transport systems (infrastructure backhaul)	10 ms	10 Mbps
Tactile interaction	0.5 ms	[Low]
Remote control	5 ms	[From low to 10 Mbps]

The end to end latency in the 4G network was significantly improved because the backbone of the network uses the best effort delivery mechanism compared to 3G and the latency ranges from 20-100ms. Still this latency is not suitable for mission critical application. For this reason, a new wireless radio interface and back bone network should emerge. In 5G NR the new concept of network slicing and Software Define Network (SDN) has been discussed. Utilizing these structures, URLLC services can be provided to the private customers by creating dedicated virtual slice connections. [31]

To support the low latency transmission in the RAN, 5G NR introduces the use of mini-slot (non-slot-based scheduling) transmission. As described in the frame structure section 2.4, the NR uses different numerology allowing to change the duration of the slot. The use of higher subcarrier spacing results in shorter slot duration. So, by decoupling the transmission duration from the slot duration, data transmission is not restricted to the slot boundaries and reduces the latency as illustrated in Figure 2-12 [15]. A slot can contain all uplink, downlink or at least one downlink or uplink part. Above 6 GHz, the length of the mini slot is defined as 1-2 symbols. [22]

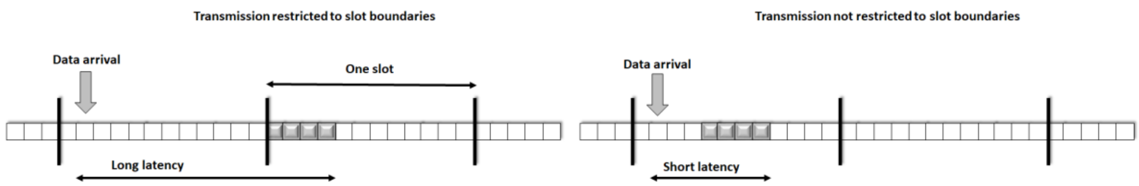

Figure 2-12. *Decoupling transmissions from slot boundaries to achieve low latency [15]*

The processing time in NR is very low compared to LTE. For example, the feedback time for the hybrid-ARQ acknowledged from the device is much faster than the last downlink it received. Table 4 shows the minimum baseline of processing time for all devices. The processing time capability is calculated per subcarrier spacing. As the duration of the slot get shorter when using the higher subcarrier spacing. The higher layer Radio Link Control (RLC) and Medium Access Control (MAC) in NR also supports low latency. The header structure of these protocols does not know the amount of data transmission while processing. Opposite in LTE, the MAC and RLC protocols know the amount of data to transmit while processing before which make challenges for the support of low latency applications. [15]

Table 4. *Minimum processing time for feedback signal [15]*

DM-RS Configuration	Device Capability	Subcarrier spacing				LTE Rel 8
		15 kHz	30 kHz	60 kHz	120 kHz	
Front-loaded	Baseline	0.57 ms	0.36 ms	0.30 ms	0.18 ms	2.3 ms
	Aggressive	0.18-0.29 ms	0.08-0.17 ms			
Additional	Baseline	0.92 ms	0.46 ms	0.36 ms	0.21 ms	
	Aggressive	0.85ms	0.4ms			

2.6 Max Throughput Support

The most common evaluation factor over different mobile generation is the achieved data rate. The advancement in the mobile internet and their services, such as 4k video streaming, multiplayer online games, multimedia sharing and virtual reality for mobile users, increased the demand for high data rates. Currently High Definition (HD) video streaming which requires 8-15 Mbps is supported but for the 4K Ultra-HD the data requirement is more than 25 Mbps. To achieve these high data requirements, the 5G NR is expected to support 10 Gbps which is 100-fold improvement over LTE. These high data rates will be acquired by using new advanced techniques like massive MIMO, millimetre wave communication and wireless software-defined networking etc [32]. In 3GPP, the maximum throughput of 5G NR in downlink and uplink can be calculated by the following formula depending on the user equipment UE.[33]

$$data\ rate(in\ Mbps) = 10^{-6} * \sum_{j=1}^{J} \left\{ v_{Layers}^{(j)} * Q_m^{(j)} * f^{(j)} * R_{max} * \frac{N_{PRB}^{BW(j),u} * 12}{T_s^u} * \left(1 - OH^{(j)}\right) \right\}$$

Where $\boldsymbol{j}$ is the number of aggregated component carriers in a band. $\boldsymbol{R_{max}}$ = (948/1024) is a max coding rate for NR. $\boldsymbol{v_{Layers}^{(j)}}$ defines the maximum number of MIMO-Layers PDSCH for downlink and max number MIMO-Layers CB-PUSCH for uplink. $\boldsymbol{u}$ is the numerology. $\boldsymbol{f^{(j)}}$ is the scaling factor given by higher layer parameters and the values can be 1,0.8,0.75 and 0.4. $\boldsymbol{T_s^u}$ is the average OFDM symbol duration in a subframe for numerology u, i.e. $\boldsymbol{T_s^u} = \frac{10^{-3}}{14*2^u}$. Note that normal cyclic prefix is assumed. $\boldsymbol{Q_m^{(j)}}$ is the maximum supported modulation order given by higher layer parameters. $\boldsymbol{N_{PRB}^{BW(j)}}$ is the maximum RB allocation in bandwidth $BW^{(j)}$ with numerology $\boldsymbol{u}$. $\boldsymbol{OH^{(j)}}$ is the overhead and can have the following values.[33]

- Frequency range FR1 for DL = 0.14.
- Frequency range FR2 for DL = 0.18.
- Frequency range FR1 for UL = 0.08.
- Frequency range FR2 for UL = 0.10.

3 DATA CHANNELS AND RETRANSMISSION PROTOCOLS

This chapter provides the detailed description of the uplink and downlink data channels, transport channel process, physical layer functionalities and retransmission protocol with reference to different layers. To understand the retransmission protocols, a literature study is conducted on the overall transport channel process. In this thesis, the system model and the corresponding computer simulation model, follows the same transport channel process.

3.1 Data Channels in (5G-NR)

To send the data packet over 5G NR radio access technology, every data packet is organised in a logical way. These data packets are clearly marked and have defined formats and positions. There are three main types of data channels used for mobile communication. Logical channel, transport channel and physical channel as shown in Figure 3-1.

The logical channel lies between the RLC and MAC layer. It is defined as the type of information it carries and can be categorized into two channels, control channel and traffic channel. In NR the logical channels are specified as Broadcast Control Channel (BCCH), Paging Control Channel (PCCH), Common Control Channel (CCCH), Dedicated Control Channel (DCCH) and Dedicated Traffic Channel (DTCH)[15].

The transport channel resides between the MAC and PHY. It is defined as how and with what characteristics the information is transmitted. The data on the transport layer is organized in the form of transport blocks with each Transmission Time Interval (TTI). A Transport Format (TF) is attached with each transport block which specify how the transport block will be transmitted over the radio interface. In TF, it contains the information about coding and modulation scheme, transport block size and antenna mapping. The transport channels in NR are defined as Broadcast Channel (BCH), Paging Channel (PCH), Downlink Shared Channel (DL-SCH) and Uplink Shared Channel (UL-SCH). There is also a Random-Access Channel (RACH) present but it does not carry any transport block. [15]

The physical channel is the set of time and frequency resources. It maps the transport block from a transport channel to corresponding physical resources. There are also physical channels which do not correspond to transport channels and are known as L1/L2 control channels. These are downlink control information (DCI) and uplink control information (UCI). The DCI contains the essential information for the device for decoding and proper reception of downlink data, whereas the UCI contains the information about the situation of the device for the scheduler and the hybrid-ARQ protocol. In NR, the physical channels are Physical Downlink Shared Channel (PDSCH), Physical Broadcast Channel (PBCH), Physical Downlink Control Channel (PDCCH), Physical Uplink Control Channel (PUCCH) and Physical Random-Access Channel (PRACH). [15]

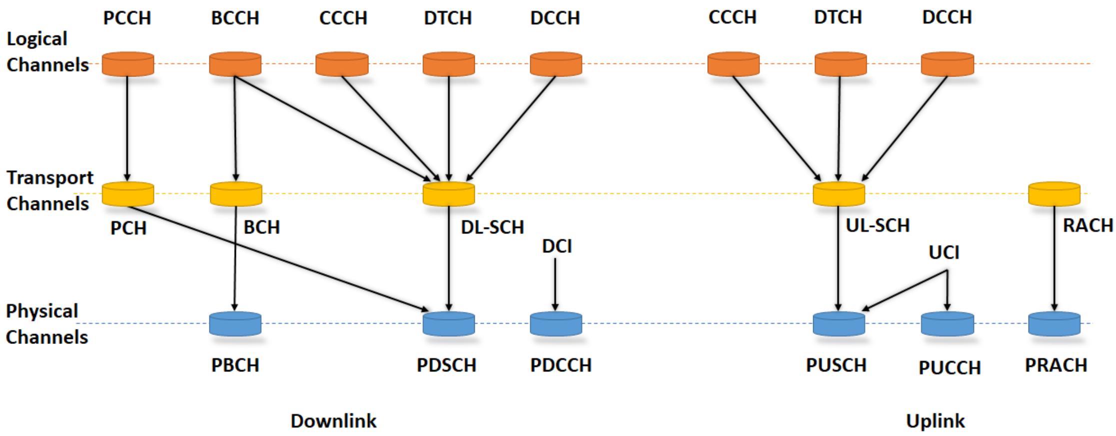

Figure 3-1 *Mapping between logical, transport and physical channels*

3.2 Transport Channel Process

As discussed in the previous section the transport channel resides between the MAC and physical layer. The three channels which are defined in NR for downlink are DL-SCH, PCH and BCH. In non-standalone operations the PCH and BCH are not used. For the uplink, only one transport channel is used known as UL-SCH. The overall transport channel process is the same for the downlink and uplink with little changes in the uplink.[15, 34]

In the transport channel the data from the logical channel is multiplexed into the transport blocks. These transport blocks are of dynamic size with each transmission time interval (TTI). For each component carrier up to two transport blocks are delivered to the physical layer. In case of spatial multiplexing two transport blocks are used with more than four layers, which is very useful in high signal to noise ratio in downlink transmission. [15, 34]

Firstly, a Cyclic Redundancy Check (CRC) is appended to the transport blocks for error detection. Followed by a Low-Density Parity-Check (LDPC) coder, which provides error correction for the data. In NR, the LDPC codes are defined in two base graphs (BG) shown in Figure 3-2. These base graphs are optimized for smaller and larger transport blocks sizes. If the payload load sizes are larger, which is a typical case with high data rates, then using code rates which are designed for low data rate is not efficient. The same is valid for low data rates which are essential to deliver better performance in rough conditions. The BG2 is designed for the coder rates 1/5 to 5/6 and BG1 from 1/3 to 22/24. The maximum high code rate can be 0.95 through puncturing, after this the device is not required to decode. The transport blocks are segmented into the code blocks. With each code block CRC is attached individually. Followed by LDPC codes which are individually rate matched. The rate matching and physical hybrid-ARQ are performed for the two reasons. To match the number of coded bits to assigned resources and generate different redundancy versions (RV) that are used in hybrid-ARQ protocols, which will be more discussed in upcoming sections. Finally, these code blocks are concatenated into codewords for the transmission. Simultaneously, up to two code words can be transmitted on the PDSCH. [15, 34]

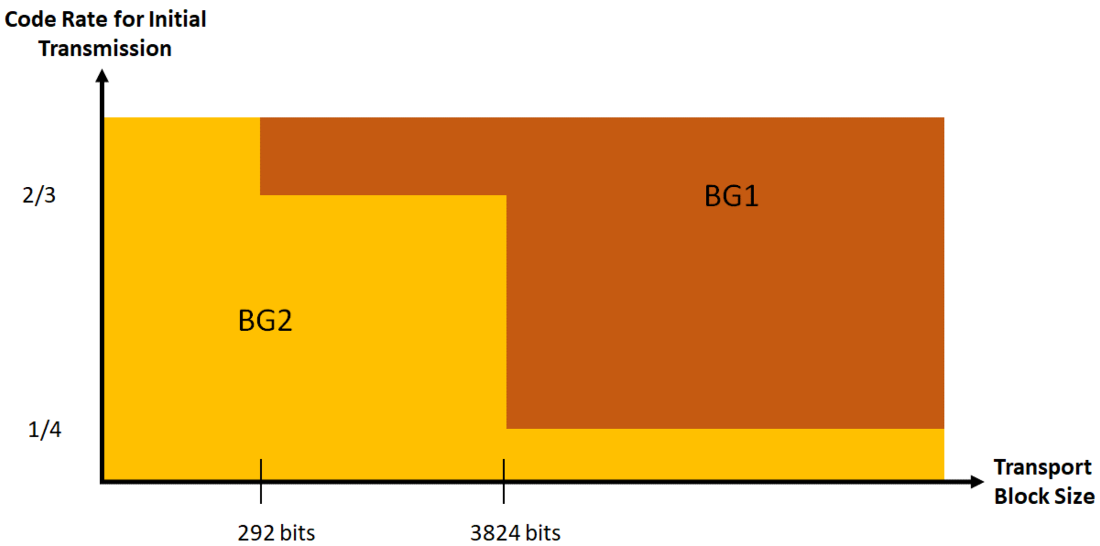

Figure 3-2. *Selection of base graph for the LDPC code [15]*

Each codeword contents are then scrambled and modulated. The scrambling helps the decoder at the receiver's side to suppress the interference form the targeted signal. The scrambling sequence depends on the device identity C-RNTI, in both uplink (PUSCH) and downlink (PDSCH). In NR, the supported modulation schemes are Quadrature Phase Shift Keying (QPSK), 16 Quadrature Amplitude Modulation (QAM), 64QAM and 256QAM. The scrambled bits are converted into the complex-valued symbols after modulation. These complex-valued symbols are then mapped into different MIMO layers. One codeword can be mapped up to four layers. After this, different layers are mapped to antenna ports and MIMO precoding or beamforming is performed. These layers are then mapped to RBs and used for transmission on physical layer PDSCH. [15, 34] The Figure 3-3 shows overall transport process for the downlink and uplink.

The uplink transport channel process is done via UL-SCH and PUSCH channels. The transmission of transport blocks of uplink is similar to the downlink. The CRC and LDPC codes are attached to the uplink transport block and code word is generated. Then the contents of the codewords are scrambled and modulated to form a complex-valued modulation symbol. In uplink the PDSCH only supports a single codeword which is mapped up to four layers. If a single layer transmission is required, an optional Discrete Fourier Transform (DFT) precoding can be used if enabled. Each Resource Blocks (RBs) are then mapped to the symbols after antenna port mapping. In NR, the mapping is done in frequency before time which helps in the early decoding at the receiver side. [34]

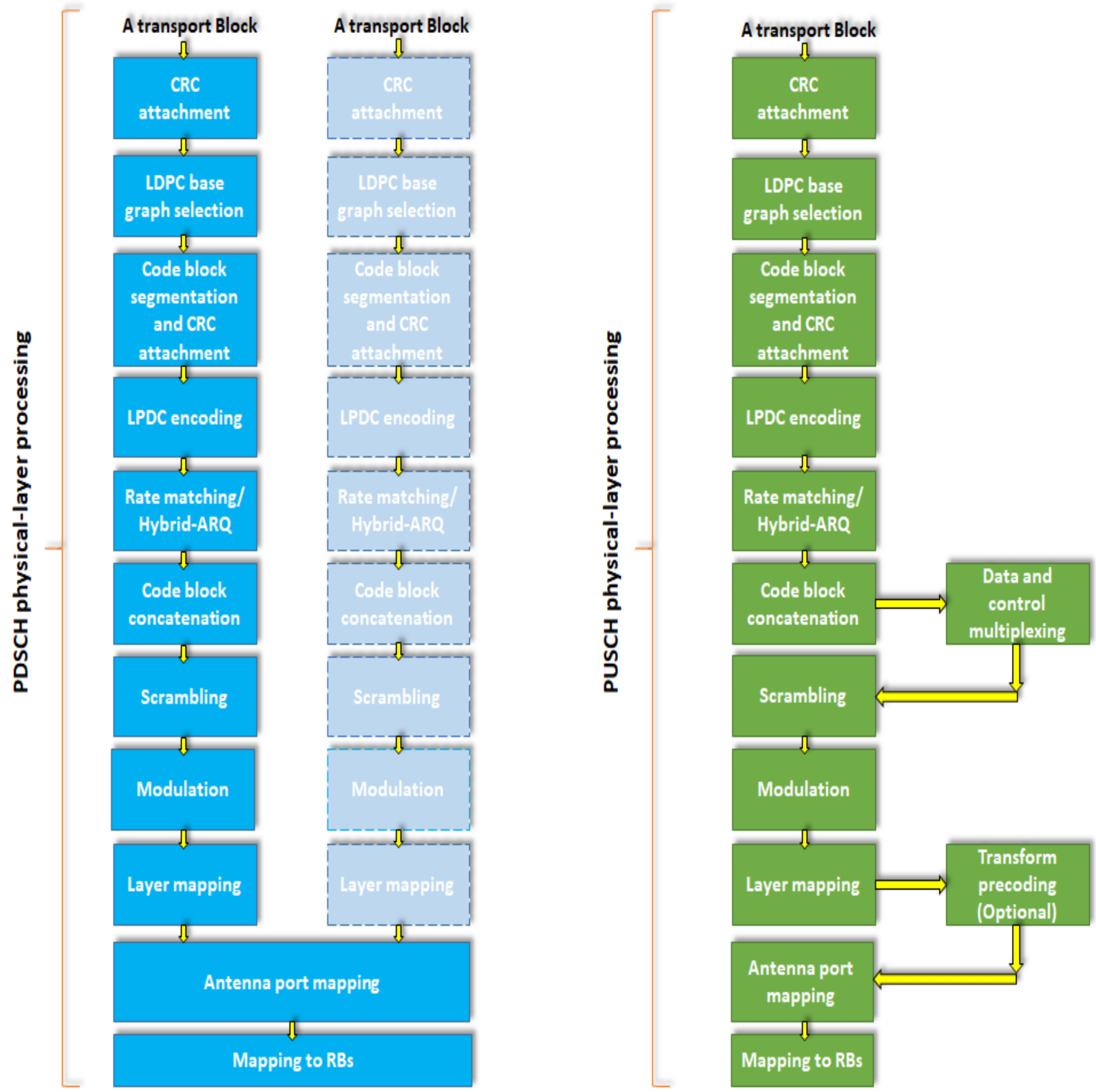

Figure 3-3. *Overall transport process for the downlink and uplink [29]*

3.3 Retransmission Protocols

Signals used in wireless communications travel through the air medium in which there are lot of obstacles like trees, building etc. Due to these obstacles, noise and unwanted interference, the signal may get distorted, faded or attenuated. These obstructions also effect the signal strength and introduce errors to the transmitted signal. To confront this problem every wireless system includes some type of Forward Error Correction (FEC) technique, which adds redundancy to the transmitted signal. When the signal is decoded at receiver, it will try to correct the errors. The pioneer work on FEC was done by the Shannon[35]. The NR uses LDPC coding as an error correction technique. Even using error correcting code, there will be still data bits with error in the receiver. This is because of the too high interference and noise level. Hybrid-ARQ is used, which is the combination of FEC and retransmission of erroneous data bit. It was firstly proposed by the Horstein and Wozencraft [36]. The hybrid-ARQ retransmission is commonly used in modern communication systems.

3.3.1 Overview

In NR, there are three layers in which retransmission protocol exists that are MAC, RLC and PDCP. The presence of multilevel retransmission structure is to have a trade-off between the reliability and fast feedback reports. In the MAC layer the retransmission process is very fast and after each transport block is received there is success or failure feedback to the transmitter. It is possible to get very low error probability of hybrid-ARQ feedback, but it comes with cost of transmission resources, such as signalling overhead and power. A feedback error of 0.1 - 1% is reasonable in many scenarios but if the requirement is ultra-reliable delivery with low latency, it is very critical to have less probability of feedback error rate. In such scenarios high-reliability feedback signalling might be required in order to achieve the performance targets. High data rates with Transmission Control Protocol (TCP) connection also requires close to error free delivery. When there is any error in the data flow, the TCP assumes it as a congestion and suddenly decreases the data rates. For example, if the data rate exceeds 100 Mbit/s the packet loss probability should be less than 10^{-5}. [37]

The RLC layer also transmits the feedback status reports but they are comparatively less frequent than hybrid-ARQ acknowledgements. For URLLC applications, it is suitable to use both fast retransmission by Hybrid-ARQ and reliable delivery provided by RLC. The retransmission in the PDCP layer also handles in-sequence delivery. These retransmissions mainly happen in case of inter gNB handovers.

3.4 Hybrid-ARQ Protocol

In NR, the retransmissions are primarily handled by the hybrid-ARQ protocol. Whenever erroneous packets are detected, the receiver requests a retransmission. The packet which was unable to be decoded due to error still includes some information. The receiver has some buffer memory in which that error packet is stored. So, when the retransmitted packet is received, it is combined with previous packets into a single packet by the error correction code, which increases the decoding performance. This whole process is known as hybrid-ARQ with soft combining. The hybrid-ARQ protocol belongs to the MAC layer but its soft combining functionality is also involved in the physical layer.

The hybrid-ARQ process remains similar in NR compared to LTE. It uses the same approach of stop-and-wait protocol on each single transport block. The working principle of the stop-and-wait protocol is that after each transport block transmission, the transmitter stops and waits for an acknowledgement from the receiver. The acknowledgement or feedback signal will be one bit positive or negative. However, in this whole stop and wait process the throughput is low. To overcome this deficiency, multiple stop and wait processes are used in parallel. When the first process starts and waits for acknowledgement, on that instant the second process starts to transmit data packets and these processes continue in parallel forming a hybrid-ARQ entity illustrated in Figure 3-4. One hybrid-ARQ entity is configured for one carrier.NR supports asynchronous hybrid-ARQ both in downlink and uplink with 16 processes. Whereas in LTE eight numbers of hybrid-ARQ processes are supported. The high number of hybrid-ARQ process then LTE is to counter the delay of certain front haul of remote radio heads and the shorter time for slots at higher frequencies. [15]

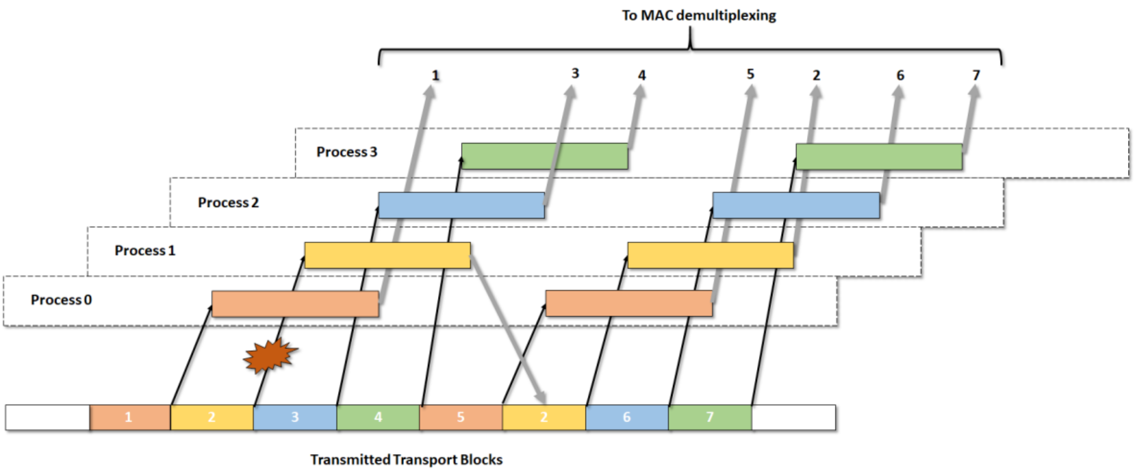

Figure 3-4. *Multiple hybrid-ARQ processes.*

The large transport blocks are divided into multiple Codeblocks (CBs). A 24 bits CRC is attached to each codeblocks. The errors can be detected on each codeblock and also on the overall transport block, as each codeblocks has its own CRC. A question arises, if some errors occur in the reception, is it beneficial to retransmit the whole transport block or just the erroneous codeblocks? As there can be several hundreds of codeblocks included in a large transport block, carrying several gigabits of data per second, it is better to retransmit only the erroneous codeblocks also illustrated in Figure 3-5.

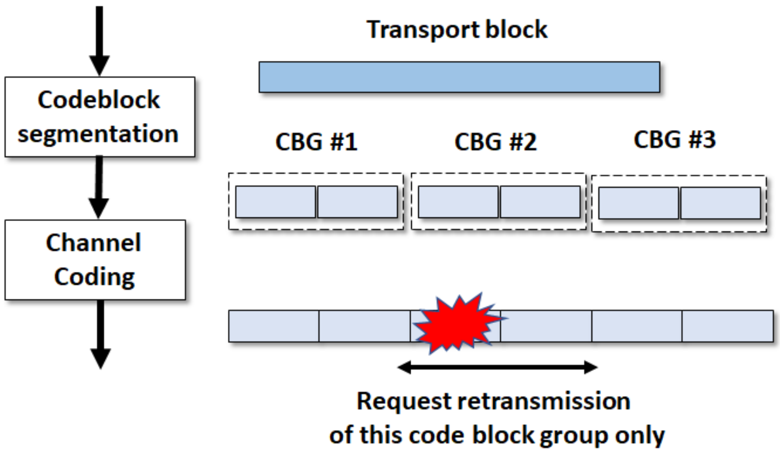

Figure 3-5 *Codeblock-group retransmission.[15]*

The retransmission of erroneous codeblocks individually will create a lot of overheads which introduces delay. This delay is originated from using Codeblock Groups (CBGs) for retransmission, which allows more effective resource utilization compared to retransmission of individually erroneous codeblocks. The codeblocks are configured in two, four, six or eight codeblocks per a codeblock group and depend upon transmission. The physical layer handles the functionality of CBGs retransmissions as per specification perspective.

3.4.1 Soft Combining in Hybrid-ARQ

Soft combining mechanism is used in hybrid-ARQ. In this process the receiver combines multiple versions of erroneous packets by retransmission in the buffer memory to get a single packet, which is more reliable than its constituent. The soft combining in NR is referred to Incremental Redundancy or Chase combining.

- The **Chase Combining** (CC) was firstly proposed by the Chase in 1985 [38]. This is a simple scheme, the transmitter retransmit again the whole packet which got errored. At the receiving side, the decoder combines the multiple copies of the retransmitted packet. In chase combining there is repetition of the coding as retransmitted packets are the identical copy of the original packet. Thus, the gain in accumulated received is E_b/N_0 for each packet.

- In **Incremental Redundancy** (IR), the retransmitted packets are not the identical copy of the original packet. They are the progressive set of bits which are generated with different versions and represent the same original information of the packet [39, 40]. In NR, the rate matching function combines different coded bits with different redundancy versions as shown Figure 3-6. The incremental redundancy provides higher performance due to the coding gain while retransmissions. But compared to the chase combining, the incremental redundancy has more complexity in the UE because of the buffer requirement.

The rate matching serves two purposes, it assigns an appropriate number of coded bits to match the transmission resources and produces different redundancy versions for hybrid-ARQ. Rate matching is performed separately on each code block.

Let's assume that all the previous transmitted redundancy version of the data packet has been received by the receiver. And all these redundancy versions of data packets provide the same information then the order of RV is not that much critical. Though, there are some code structures where all redundancy version is not significant. This is the case for in LDPC coding in which the systematic bits are more significant than the parity bits. This means that in initial transmission there should be at least some parity bits with the systematic bits. So, for the retransmissions these parity bits are included and because of this reason the systematic bit are first inserted in the circular buffer as seen in Figure 3-6. The RV0 and RV3 in the circular buffer are self-decodable because of the placement in start. Also, the RV3 is placed after nine o'clock and most of the transmission part includes the systemics bits. By default, the order of redundancy version is 0,2,3,1 meaning that every second retransmission will be self-decodable. [15]

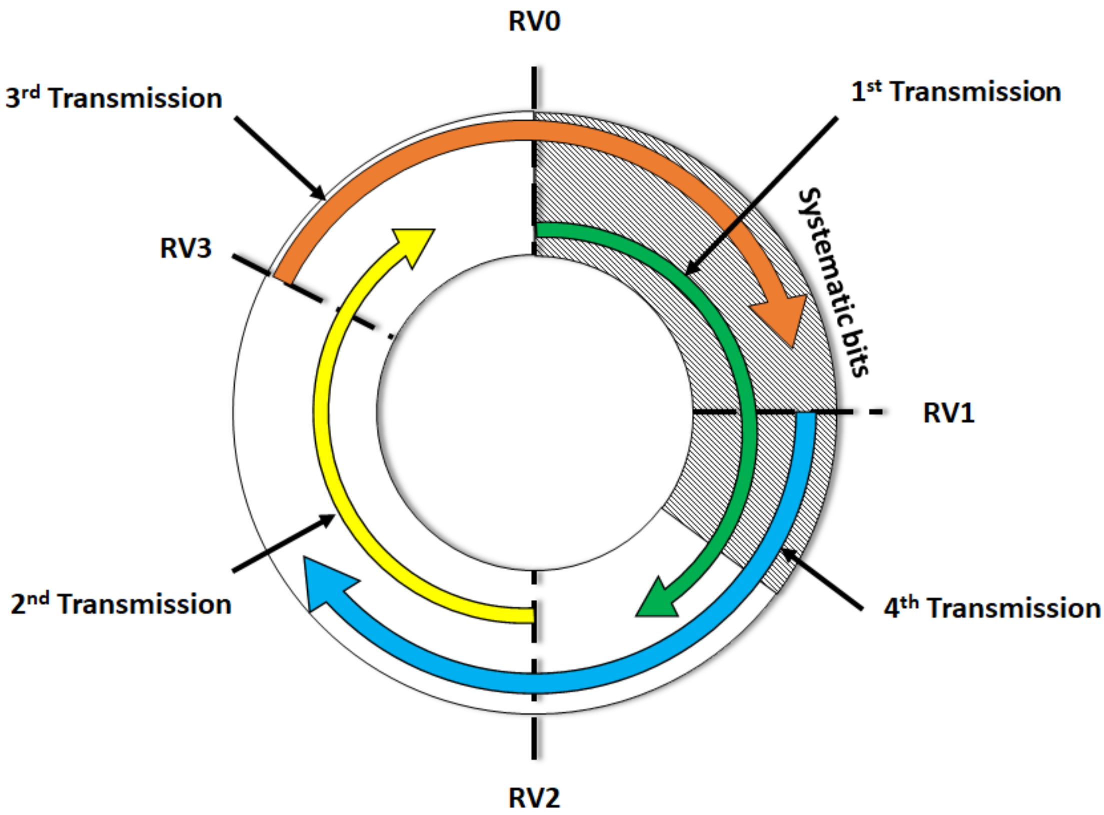

Figure 3-6. *Example of incremental redundancy [15]*

When the soft combining is operating, the receiver must know when to clear the buffer prior to decoding and the difference between the initial transmission (soft-buffer should be cleared) and receiving the retransmission. In the same way, the transmitter should also know whether to transmit new data or the erroneous retransmission data. To solve this, a new-data indicator is used both in uplink and downlink hybrid-ARQ, which is discussed more in the following sections.

3.4.2 Downlink Hybrid-ARQ

The retransmission in downlink is scheduled the same way as new data transmission meaning that the retransmissions may occur any time at any arbitrary frequency allocated in downlink cell bandwidth. The scheduling assignment includes hybrid-ARQ process number, controlling signals related to hybrid-ARQ and new-data indicator. When codeblock groups CBGs retransmission is configured then it also contains CBG Transmitter Indicator (CBGTI) and CBC Flush Indicator (CBGFI).

When the downlink connection is established, the receiver receives the scheduling assignment in Download Control Information (DCI) and decode it. Since the receiver should know that the transmission is new data, and thus the buffer should be cleared, or a retransmission, when the soft combining is performed. Therefore, with transport block a new-data indicator is included which is a single bit sequence number. Upon receiving the scheduled assignment in downlink, the receiver looks for the new-data indicator toggled with transport block, and determine whether the current transmission should be combined in the soft buffer or if the retransmission is requested or clear the buffer for new data. If codeblock group retransmission is configured, then additional two field new-data indicators are included in DCI that are CBGFI and CBGTI as shown in Figure 3-7. The CBGTI means that the CBGs are present in the transport block and it is a single bit. The CBGFI means that the CBGs indicated by the CBGTI will be flushed or soft combing will be performed. [15]

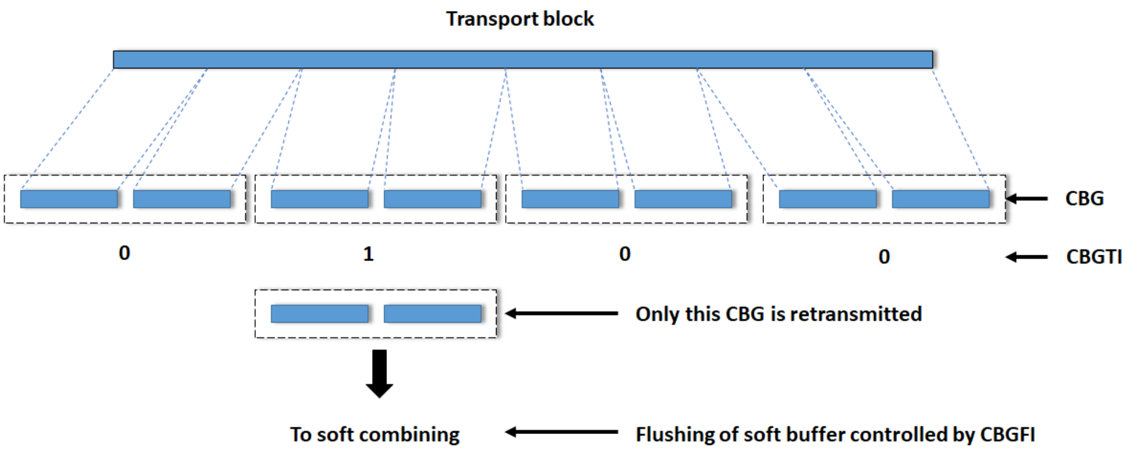

Figure 3-7. *Illustration of per-CBG retransmission [15]*

A positive acknowledgement is sent after successful decoding and if the decoding fails a negative acknowledgement is sent to the gNB in a feedback signal as an uplink control information.

3.4.3 Uplink Hybrid-ARQ

Both in uplink and downlink, asynchronous hybrid-ARQ is used. The uplink scheduling grant includes process number and information related to hybrid-ARQ as well as new-data indicator. And if CBGs is configured then it also contains CBGTI.

The new-data indicator is used to differentiate between retransmissions and new data. To request the new data, transport block is toggled with the new-data indicator. Otherwise, retransmission is requested for previous hybrid-ARQ process that the gNB can perform soft combining. The CBGFI is not used in uplink hybrid-ARQ because the soft buffer is placed in gNB and the scheduler makes the decision whether to clear the buffer or not. The CBGTI is used similar as in downlink, and if CBGs retransmission is configured, the CBGTI will retransmit the codeblock groups. [15]

3.4.4 Uplink Acknowledgements Timing

After receiving the downlink data the receiver sends acknowledgement back to the gNB. In LTE, the acknowledgement timing is fixed in specification. In FDD this is possible and the acknowledgement timing after the reception of data is 3ms. A similar approach can be implemented in semi static TDD for uplink and downlink in which half duplex operation is used. But for dynamic TDD it is not suitable to predefine the acknowledgement timing as the scheduler may dynamically change the transmission in downlink and uplink direction. In NR a new dynamic flexible scheme is introduced to adopt the control of acknowledgements.

The transmission of acknowledgement timing is controlled by the timing field present in the downlink DCI of hybrid-ARQ. In Figure 3-8 RCC-configured table with index and a three-bit field information provide the timing of hybrid-ARQ acknowledgements which should be used relatively upon receiving PDSCH. As from the Figure 3-8 three downlink data slots are scheduled before the acknowledgement in uplink. In downlink assignment, the DCI stores the RRC-configured table, where index 3 indicates receiving of the uplink acknowledgement of all three-downlink data at once by multiplexing. Also, from a latency perspective in NR the processing time for Hybrid-ARQ acknowledgement is much faster compared to the LTE discussed in Table 4.

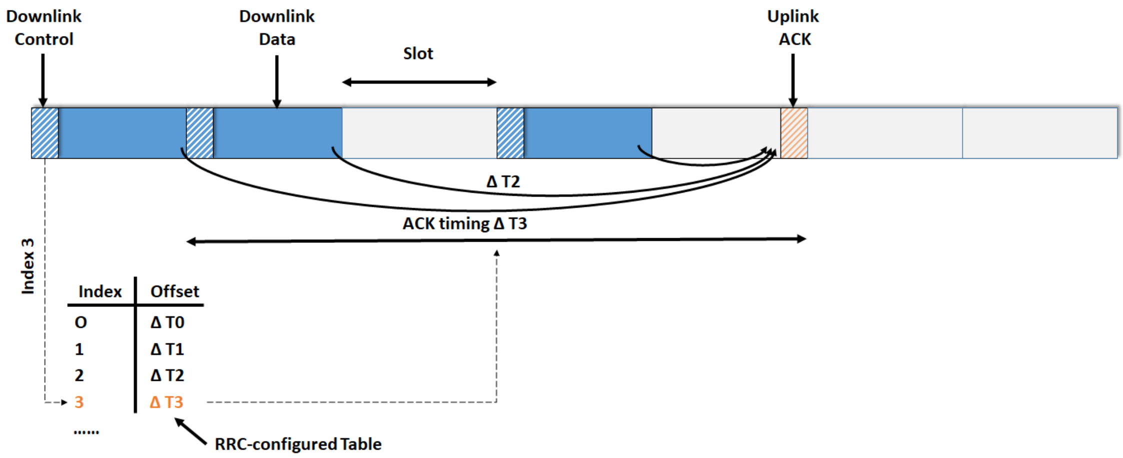

Figure 3-8. *Determining the acknowledgement timing [15]*

3.4.5 Multiplexing of Hybrid-ARQ Acknowledgements

In Figure 3-8, a single hybrid-ARQ acknowledgement is used in uplink transmission for the multiple transport blocks at the same time. Other examples like carrier aggregation and per-CBG retransmission, the UE must perform multiple acknowledgement at the same instant in uplink direction. So, the NR supports multiplexing of the acknowledgements in a one multi-bit acknowledgement message. It can be processed by a dynamic codebook or semi-static codebook with Radio Resources Control (RRC) configuration.

In a semi-static codebook, there are two-dimensional values including time domain dimension and a component -carrier dimension. These values are semi-statically configured, where the carrier domain values are the component of carriers containing transport blocks or CBGs and the time domain values are the maximum and minimum hybrid-ARQ acknowledgement timing. The Figure 3-9 illustrates an example of a semi-static codebook, where there are three carriers and a time span of one, two, three and four acknowledgement timing.

One carrier is with one transport block per slot, one carrier is with two transport blocks per slot and one carrier is with one transport block containing four CBGs configured per slot. As the codebook size is fixed, total number of bits for hybrid-ARQ report will be 7*4=28 bits. In the table, each entry shows the results of decoding either negative or positive acknowledgement for the equivalent transmissions. A negative acknowledgement is sent for those entries, whose corresponding transmission is not scheduled. This adds robustness to the system as for the missed downlink assignment a negative acknowledgement is sent to gNB which in return retransmits the missing CBG or transport block. [15]

The drawback of a semi-static codebook is that, for a huge number of carriers and codeblock groups, it is difficult to handle too many hybrid-ARQ reports simultaneously. To address this problem, NR also supports dynamic codebook which is used by default.

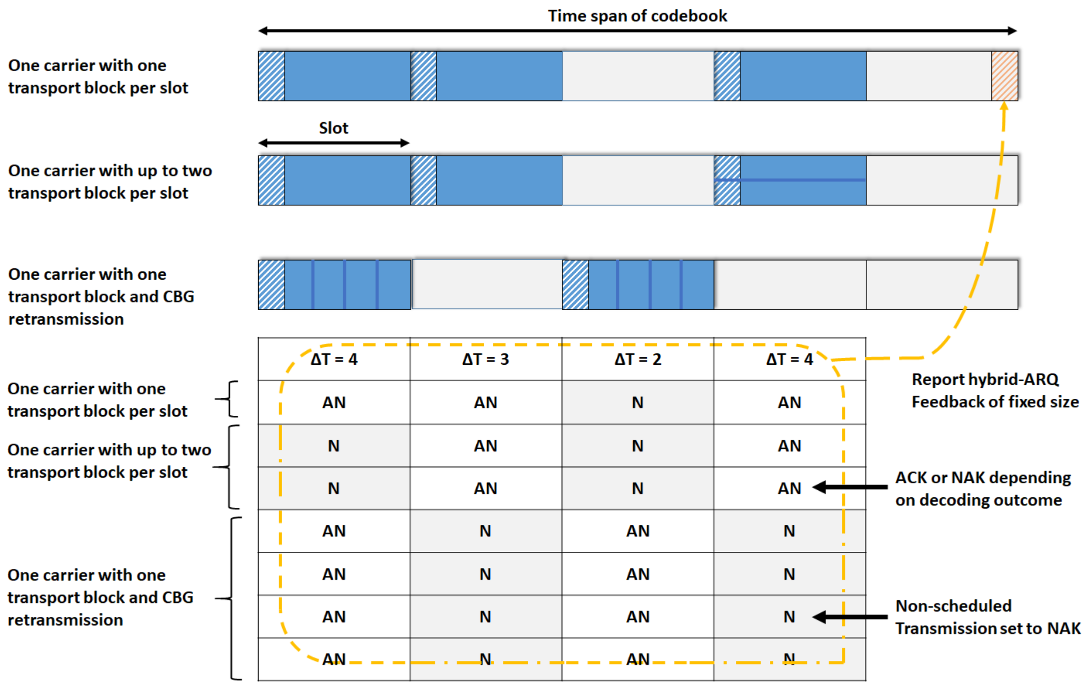

Figure 3-9. *Example of semistatic hybrid-ARQ acknowledgement codebook* [15]

In the dynamic codebook, the acknowledgements are transmitted for scheduled carriers only. Thus, the codebook size depends on the scheduled carriers and changes dynamically. In Figure 3-9 the table shows bold entries which are only added in the hybrid-ARQ acknowledgement report and the grey colour filled boxes same in the table are omitted as they correspond to the non-scheduled carriers. In result the size of acknowledgement messages is reduced.

Without any error in downlink control signalling the dynamic codebook would work in a straightforward manner. However, if there is any error caused in downlink control signal, there will be mismatching between the gNB and the device, which affects the feedback reporting timing of the hybrid-ARQ acknowledgement. Assume an example, where a device was scheduled in downlink transmission for two subsequent slots, but due to error it missed the PDCCH and hence scheduled assignment for the one slot. In return, the device will transmit the acknowledgement for one slot but the gNB will try to receive two slot acknowledgement which in result mismatches.

To address the error cases mentioned above, NR utilizes downlink assignment index (DAI) which is included in the DCI. The DAI field is divided into two parts, a Total DAI (tDAI) used in the carrier aggregation and Counter DAI (cDAI). The cDAI tells the number of downlinks scheduled transmissions up to the point DCI was received in the carrier first. The tDAI tells the total number of downlinks transmission across all carriers up to point in time. The cDAI and tDAI are represented in a decimal number consisting of two bits each with no limitation. In Figure 3-10 there are 17 scheduled transmissions and the corresponding acknowledgements are numbered as (0-16). If this is the case with semistatic codebook then it will require 28 entries irrespective of the number of transmissions.

The example shown in Figure 3-10 have 7 carriers and on carrier number 5, a transmission component got errored. If there was no DAI mechanism a mismatch codebook was generated between the gNB and device. From the DAI the device can compare the values of cDAI and tDAI, and conclude that transmission got errored. Also, the device has information that the gNB is expecting the 17 ACKs and a negative acknowledgement is sent. In scenarios where CBG retransmissions are configured for some carriers, the dynamic codebook is divided into two. One for the non-CBG carriers and one for the CBG carriers. The working principle is the same for codeblock groups as mentioned above.

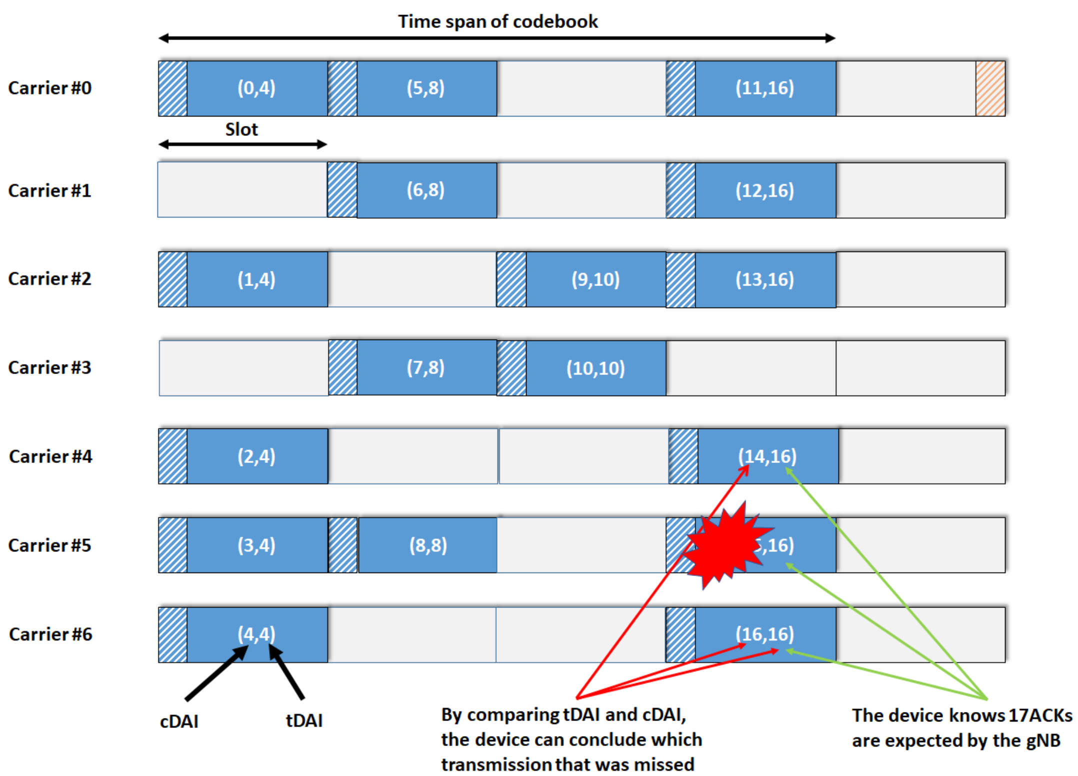

Figure 3-10. *Example of dynamic hybrid-ARQ acknowledgement codebook [15]*

3.5 RLC layer

The radio link control RLC protocol delivers RLC SDUs data to the corresponding RLC entity in the receiver via using MAC and physical layer functionality. The Figure 3-11 illustrates the working between the MAC and RLC, where a single transport channel is formed by multiplexing of the multiple logical channels. The RLC entity is configured per logical channel. The RLC is responsible for the segmentation of RLC SDUs, RLC re-transmission and duplicate removal. [15]

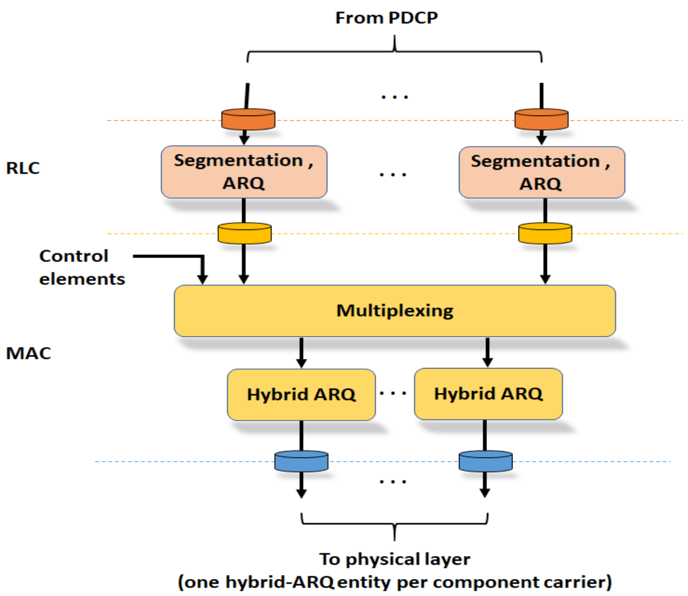

Figure 3-11. *MAC and RLC.*

To reduce latency in the NR, RLC layers do not support in sequence delivery or concatenation which was present in LTE. According to different service requirements: transfer of large data files with error free delivery and services like live video streaming which can bear some packet loss. So, the RLC operates in three different modes depending upon the service requirements that are Transparent mode (TM), Unacknowledged mode (UM) and Acknowledged mode (AM). [15]

- In Transparent mode (TM), the data is delivered from PDCP to MAC layer totally transparent, without any retransmission, reassembly/segmentation and duplication detection. This mode is used for control-plane broadcast data which are CCCH, BCCH and PCCH. The message is formatted in a size that it will deliver to all devices without any need of segmentation and can handle the variation in the channel conditions for error free delivery.
- In Unacknowledged mode (UM), the data packets are segmented without any retransmission support in the RLC layer. Applications like streaming video, VoIP, where error free delivery is not required UM is used.
- Acknowledged mode (AM) supports all the main RLC functionalities such as segmentation, retransmission of error packets and duplication detection. UL-SCH and DL-SCH are also operated in acknowledged mode.

3.5.1 Retransmissions in RLC layer

The RLC layer performs retransmission of missing packet data units PDUs in acknowledged mode. The missing PDUs are traced by the sequence number by the receiver and a retransmission is requested. The hybrid-ARQ handles most of the error by adding another layer of retransmission and adding more reliability.

To decrease the overall latency in the NR, the RLC layer does not support the in-sequence delivery of the packets. Due to this, a missing packet does not affect the later packet reception in retransmission and increases the performance of the RLC buffering memory. If a packet got errored by the temporary burst or interference, the RLC SDUs is not forwarded to the higher layer until all previous subsequent SDUs are received in correct order. In result a delay is generated, which is not desirable for the ultra-low latency applications.

The flow of data is bidirectional between the RLC peer entities in acknowledged mode. The transmitted PDUs are acknowledged by the receiver for the reception of more PDUs. A status report is used between the transmitter and the receiver to provide information about the missing PDUs. The header contains the sequence number which helps to keep the track of PDUs while transitioning. In acknowledged mode, both RLC entities the transmitter and receiver both have their respective windows. Only the transmission of PDUs are valid between these windows. The PDUs below windows have already been transmitted and acknowledged with the sequence number. In the receiver, if any duplicate PDUs is present it will be discarded and only SDUs will be transmitted to a higher layer.

The Figure 3-12 illustrates the mechanism of RLC working, there are two RLC entity nodes present, in which one is transmitter node and one is receiver node. As the transmission in RLC is bidirectional, so in the following example only transmission is discussed as the other direction is alike. In the transmitter the transmission window buffer includes the PDUs numbered from n to n+4, which are ready for transmission at time = t_0. In the receiver only PDUs n-1 and n-2 are acknowledged. Moreover, n is received in the receiver but not yet acknowledged. After receiving the PDUs till n, the SDUs are reassembled and delivered to higher layer PDCP.

At time = t_1, the PDUs n+1 and n+2 are transmitted, but in the receiver only n+2 PDU is received. The SDU of n+2 PDU is reassembled and delivered to the higher layer without waiting for the missing n+1 PDU. The reason for the missing n+1 PDUs may be hybrid-ARQ requested for the retransmission and the MAC layer did not deliver to the RLC layer. The transmission window remains the same as n and higher PDUs are not yet acknowledged by the receiver. As any errored or missing PDUs can be requested by the receiver.

The window in the receiver is not updated because of the missing n+1 PDU. At time = t_1, reassembly timer in a receiver starts and if the missing PDU is not received, the timer expires and receiver requests for the retransmission. For example, the n+1 PDU is received at time = t_2, by hybrid-ARQ retransmission protocol before the timer expires. The receiving window is advanced and the missing PDU is reassembled for higher layer delivery. The duplication detection mechanism is the responsibility of RLC, if another n+2 replica packet arrives then it will be discarded.

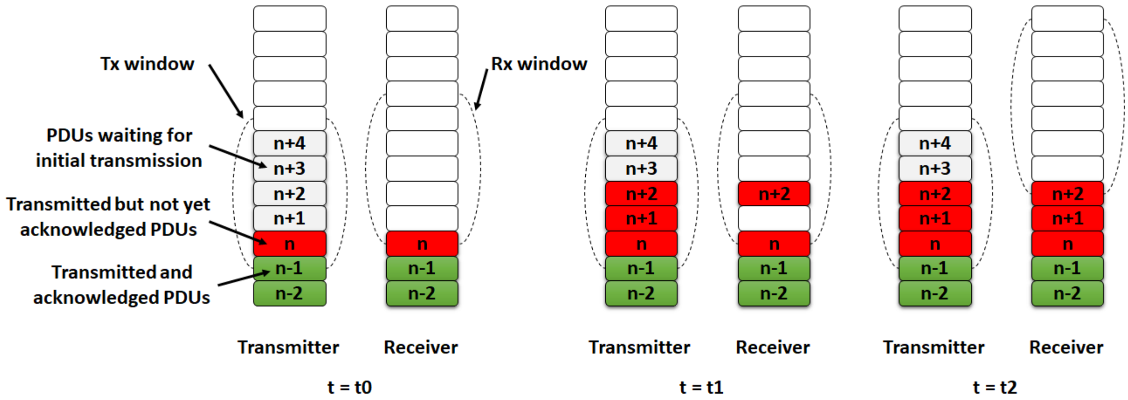

Figure 3-12. *SDU delivery in acknowledged mode [15]*

In Figure 3-13 at time = t_3, the transmission of n+3, n+4 and n+5 PDUs continues but only n+5 PDU is received. The n+3 and n+4 PDUs are missing and in this example the reassembly timer starts. When no missing PDUs are received, at time = t_4 ,the receiver sends a request to its peer entity in the form of control PDU including status report. To reduce the retransmission delays the control PDUs are of higher priority than the data PDUs. At time = t_5 , the receiver receives the status report and has all information till n+2 and all PDUs are received. The receiving window is advanced and this time missing PDUs (n+3 and n+4) are received correctly. Finally, at time = t_6, all PDUs are received and the transmitter sends a last RLC data PDUs setting a flag in the header to get the status report form the receiver. The receiver acknowledges all PDUs reception by transmitting the status report. The transmitting window advances and declares all PDUs reception correctly to the receiver.

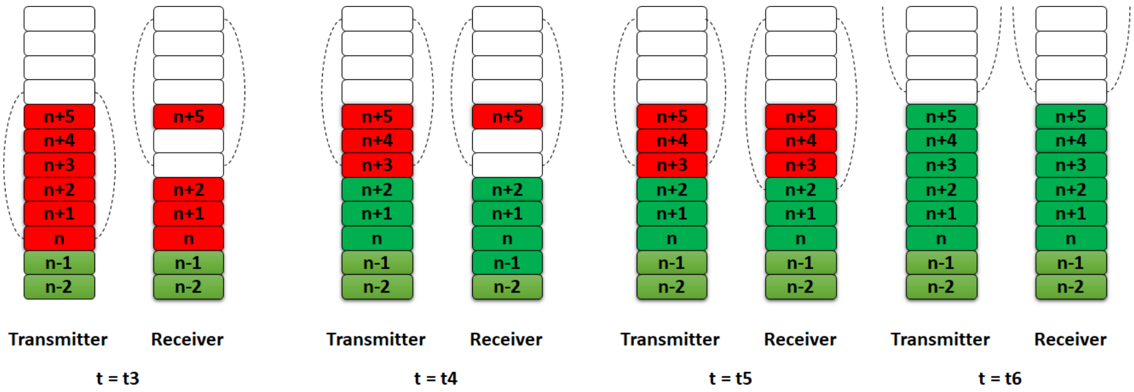

Figure 3-13. Retransmission of missing PDUs.

3.6 PDCP

The responsibilities of Packet Data Convergence Protocol (PDCP) are ciphering and integrity protection, header compression, routing and duplication for split bearers, and recording, retransmission and SDU discard.

Header compression is the primary function of the PDCP and decompression in the receiver. It reduces the number of bits transmitted in the air and increases the spectral efficiency. In NR, the Robust Header Compression (ROHC) algorithm is used for header compression. The integrity protection helps to ensure the data is sent from the authenticated and correct source. While the ciphering protects the data against the eavesdropping. In PDCP, the control plane and data plane both implement the ciphering and integrity protection.

PDCP also performs duplication and routing functionality. Some radio bearers are controlled by the master cell and some by secondary cell group in dual connectivity. So, the data flow between different bearers is done by routing functionality of PDCP. It also routes the data flow between the distributed unit (gNB-DU) and central unit (gNB-CU) in the scenario of split gNB. For duplication, the same data is delivered through two separate logical channels in case of carrier aggregation or dual connectivity. In receiver multiple copies of the same SDU are received and the duplicate data packets are discarded. This increases the diversity which meet high reliability requirements.[15]

3.6.1 Retransmission in PDCP Layer

A question arises why we need PDCP layer retransmission even though two retransmission protocols are working already in two layers, hybrid-ARQ in MAC layer and ARQ in RLC layer. The reason for retransmission in the PDCP layer is the inter-gNB handover. When the handover takes place, the undelivered downlink data packets are forwarded from old gNB to new gNB by PDCP. In this scenario the RLC status is lost and in new gNB a new RLC and hybrid-ARQ entity is initiated. The PDCP retransmission assures that no data packet is lost during the handover process. For the uplink, the retransmissions are handled by PDCP entity in the device for the not yet delivered packets to gNB upon handover as the hybrid-ARQ buffer is cleared.

To reduce the overall latency the in-sequence delivery is not support by the RLC layer. For some applications rapid transmission of packets are more important than in sequence delivery. However, if there is a need for in-sequence delivery, PDCP can be configured to support it. [15]

3.7 Why do we need multi-layer retransmission protocol?

The multi-layer retransmission protocol provides a trade-off between high reliability and low latency fast retransmission support. In the hybrid-ARQ, a fast retransmission is guaranteed by a feedback status report either successful or failure packets delivery to receiver. We can achieve a very low error probability for the hybrid-ARQ feedback but in result it will increase the transmission power cost. Thus, undermining the cost of transmitted power, a threshold level of Hybrid-ARQ retransmission feedback error rate is achieved which is around 10^{-1}. In result a residual error still remains in hybrid-ARQ which is not suitable for some use cases. Also, the RLC retransmissions status reports are comparatively infrequent and getting high reliability from RLC is relatively small. Therefore, the use of multi-layer retransmission protocol, hybrid-ARQ in MAC layer for fast transmission and ARQ in RLC layer for reliability, helps to get a short round-trip time and a modest feedback overhead.[41]

4 SYSTEM MODEL AND SIMULATION RESULTS

4.1 Model Overview

To study the overall throughput and latency in 5G-NR, we have to first understand the downlink data transmission in NR. The physical downlink shared control channel (PDSCH) is responsible for the transmission of data. In this thesis, throughput and delay of 5G NR is investigated by implementation of physical downlink shared channel (PDSCH) and downlink transport channel (DL-SCH) as defined in 3GPP. The propagation model used in simulation is cluster delay line (CDL) [42]. The Figure 4-1 illustrates the overall process for the downlink transmission of PDSCH.

The downlink shared channel (DL-SCH) is a transport channel and carries the user data. It also delivers the different system information block (SIB). The DL-SCH includes cyclic redundancy coding (CRC), code block segmentation, rate matching, hybrid-ARQ processing and code block concatenation.

For error detection CRC is attached to each transport block and get segmented. The CRC size depends upon the size of the transport block. Usually 16-bit CRC is attached to the transport block size of 3824 bits. If the transport size increases a 24-bit CRC is attached. Low density parity check (LDPC) channel coding is used for error correction in NR for its simplicity and better performance at higher code rate compared to turbo coding which was used in LTE. For transmission, a suitable number of coded bits are assigned to match the resources by the rate-matching block. The hybrid-ARQ process also performed with rate-matching block. When the retransmission is required for the erroneous blocks, in hybrid-ARQ different redundancy versions of bits are generated which represent the same information of missing block more detail discussed in chapter 3. Finally, these code blocks are concatenated into 1 or 2 codewords to form a sequence of coded bits.

The physical downlink shared channel (PDSCH) includes scrambling, modulation, layer mapping, multiantenna precoding and resource mapping. The bits from the hybrid-ARQ are scrambled with coded bit-level scrambling sequences. Modulation schemes supported in 5G NR are QPSK, 16QAM, 64QAM and 256 QAM for both uplink and downlink. The simulation includes the analysis of different MCSs. The scrambled bits are modulated into complex modulation symbols. Layer mapping converts the codewords into different transmission layers. A single codeword can

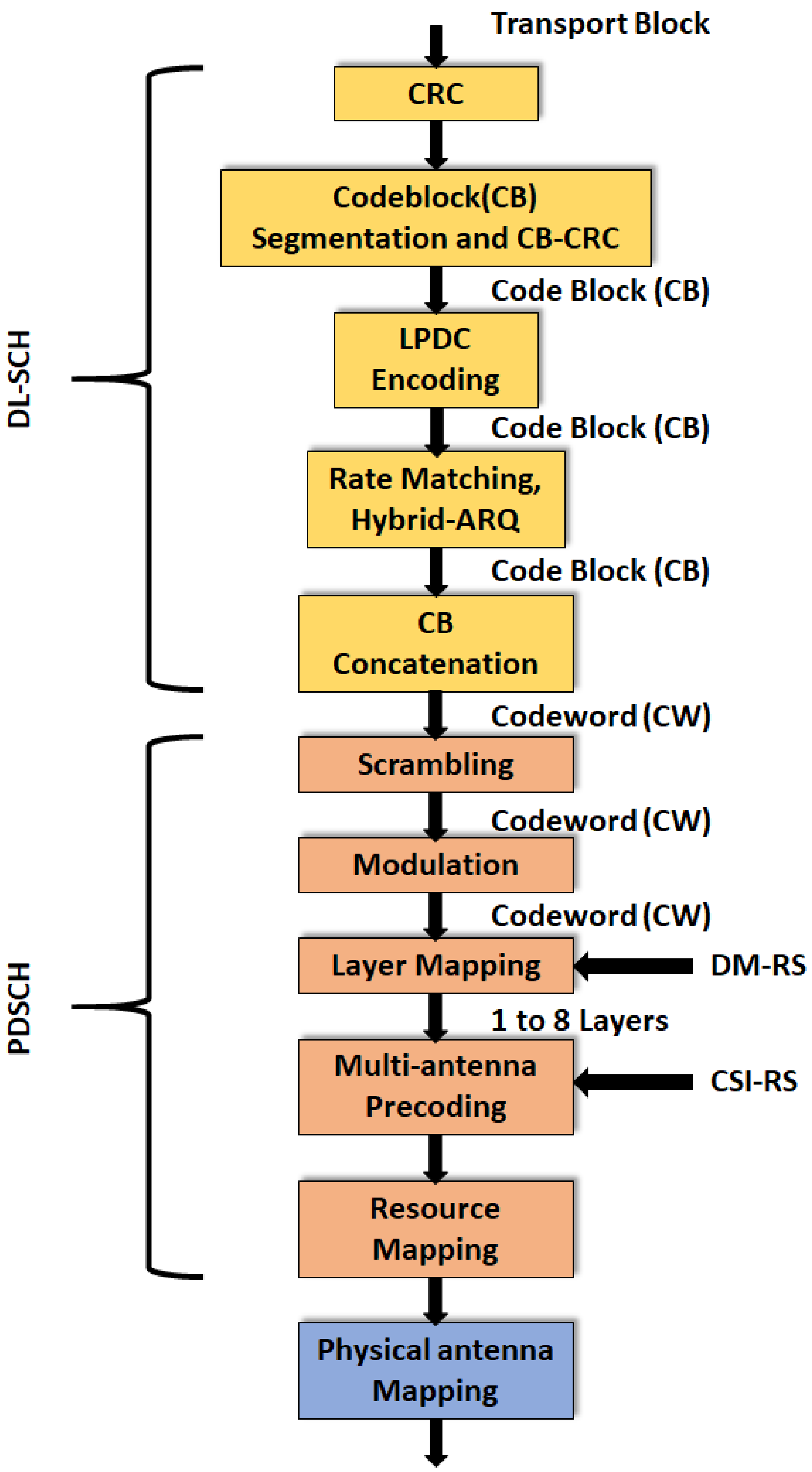

Figure 4-1. *Downlink transmission of PDSCH*

be mapped up to four layers and for two codewords than eight layers are mapped for transmission. After layer mapping these layers undergo a precoding. In precoding the layers are mapped to different antennas ports by multiplying with a precoder matrix. In case of beam forming, one layer is mapped to multiple antenna ports. For spatial multiplexing several layers are mapped to multiple antenna ports during the precoding step. The Demodulation Reference Signal (DM-RS) is also added with a precoder and is used for the estimation of channel. Firstly, in the resource mapping block, the PDSCH symbols are mapped to virtual resource blocks. Then these symbols are mapped to available resource elements of the resource block assigned by the MAC scheduler. The mapping avoids the reserved locations for DM-RS and Channel State Information Reference Signal (CSI-RS). Finally, these PDSCH symbols are transmitted from the gNB to the receiver UE through radio waves.

4.2 Working Principle

The Figure 4-2 illustrates the overall process chain used for the simulation, to measure the throughput and delay of the PDSCH of 5G NR. The transmitter model consists of DL-SCH, PDSCH, precoding and Cyclic Prefix Orthogonal Frequency Division Multiplexing (CP-OFDM). The receiver model includes CP-OFDM demodulation, channel estimation, PDSCH decoding and DL-SCH decoding. The modulated symbols generated by the gNB are transmitted via CDL fading channel with varied simulation parameters to the UE receiver. The whole simulation is performed on different SNR values from -20dB to 20dB. The throughput and retransmission of the PDSCH data is analysed per transmission instance. Different MCS are used in simulation for the comparison and analysis of the result on each SNR value.

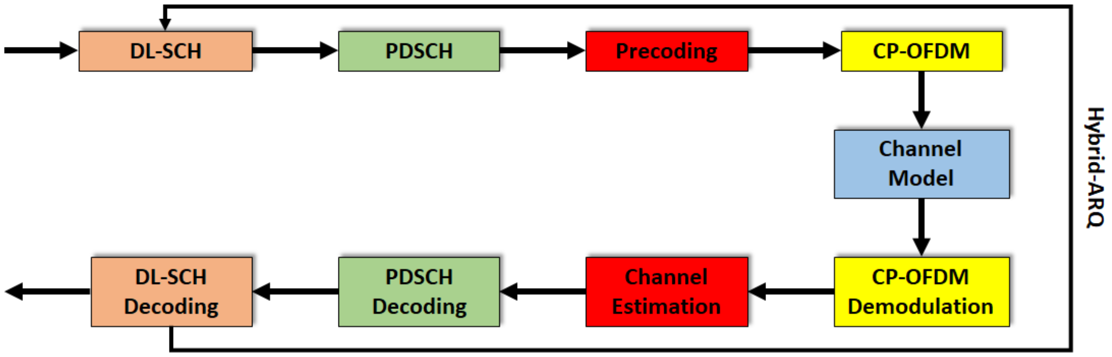

Figure 4-2. *Overall process block diagram*

The processing loop of the system works in a way that for each PDSCH transmission instance CRC is checked with a hybrid-ARQ process. If the CRC indicates a packet error due to low SNR values or poor channel conditions, a retransmission is requested for that instance. If no retransmission is required new data is generated by the gNB and hybrid-ARQ process is updated. The DL-SCH performs the channel coding and operates on the transport block provided as an input. A copy of the transport block is retained in the DL-SCH in case retransmission is required. The PDSCH converts these coded bits into complex modulation symbols. The precoding operation is applied to the resulting signal and generates a resource grid. This resource grid is then OFDM modulated to construct a waveform.

The signal is then propagated through CDL fading channel. For noisy channel modelling, Additive White Gaussian Noise (AWGN) is added. At each UE antenna SNR is defined per resource element.

The received signal is demodulated and synchronized with the reconstructed channel impulse response. The DM-RS is used for channel estimation. The recovered symbols are demodulated in PDSCH decoder. In PDSCH decoder, descrambling is performed to obtain an estimate of received codewords. The decoded soft bits are passed through the DL-SCH decoder which decodes the codewords. In return, the block CRC errors are exploited to determine the overall throughput of the system.

4.3 Parameters

To obtain the desired results from the simulations, different parameter configurations are used, as illustrated in Table 5.

For the simulation of overall PDSCH data transmission, certain parameters are set as a constant. The total number of frames for all the simulation results is 10 and the time duration for each individual frame is 10ms. As per 5G NR, the subcarrier spacing used is 30khz, and thus the time duration of one slot is defined as 0.5ms and the duration for subframe is 1ms. The signal is analysed over the SNR range from -20dB to 20dB. The number of used hybrid-ARQ process is 16. The packets are transmitted in slots, with initial transmission consisting of 16 parallel packets as discussed in section 3.4. If any packet is observed erroneous, a retransmission is requested by hybrid-ARQ. Different redundancy versions of retransmission data is then transmitted for a soft combing with retransmissions sequence of 0,2,3 and 1. In this thesis retransmissions are considered to introduce the delay in the system. After the reception, the throughput in Mbps is calculated for the successful decoded data packets received.

Table 5. *Different parameters set for the whole simulations*

Simulation length and SNR points	
Number of frames	10
SNR range	From -20 to 20 dB
Wave form type and numerology	
Number of resource blocks	51
Subcarrier spacing	30 kHz
Cyclic prefix	Normal
Channel Model	
Propagation channel type	CDL
Channel delay profile	CDL-C (Urban macro-cell model)
RMS delay spread	$300 * 10^{-9}$
Channel type	Non-line of sight
DL-SCH/PDSCH Parameters	
Hybrid-ARQ Process	16
Number of layers	1
Number of transmission antenna	8
Number of UE receive antenna	2
Number of Codewords	1
Code rates	0.478, 0.753, 0.925
Modulation Schemes	QPSK, 16 QAM, 64 QAM, 256 QAM
Waveform	CP-OFDM

4.4 Simulation Findings

This section includes the findings and results of the simulation of PDSCH transmission for the user data. At different SNR values, throughput and delay in terms of retransmission are investigated. Four modulation schemes which includes QPSK,16 QAM, 64 QAM and 256 QAM are analysed for finding the significant high data rates transmission with a smaller amount of delay in terms of low latency and reliability of packet delivery requirement. Another important parameter which is studied is different coding rates. In NR, LDPC is used as an error correction technique is used because of its robustness and reduced complexity compared to turbo coding as discussed in chapter 3. The coding rates used for simulations are 0.478,0.753 and 0.925. These coding rates are subject to increase the number of parity bits with the original set of data for error correction and error detection techniques as discussed in chapter 3. Thus, different MCSs are implemented for the PDSCH data transmission for downlink simulation to study the throughput and delay with different SNR values.

4.4.1 Throughput Analysis

The Figure 4-3 shows the graph between the throughput and SNR. In total twelve downlink data transmission schemes are analysed, which includes four modulation schemes with three different coding rates. Modulation schemes are presented by different colours, green for QPSK; blue for 16 QAM; red for 64 QAM and yellow for 256 QAM, while coding rates are presented by markers, circle for coding rate 0.478; cross for coding rate 0.753 and plus for coding rate 0.925. The y-axis shows the throughput calculated in Mbps and x-axis shows SNR in decibels (dB).

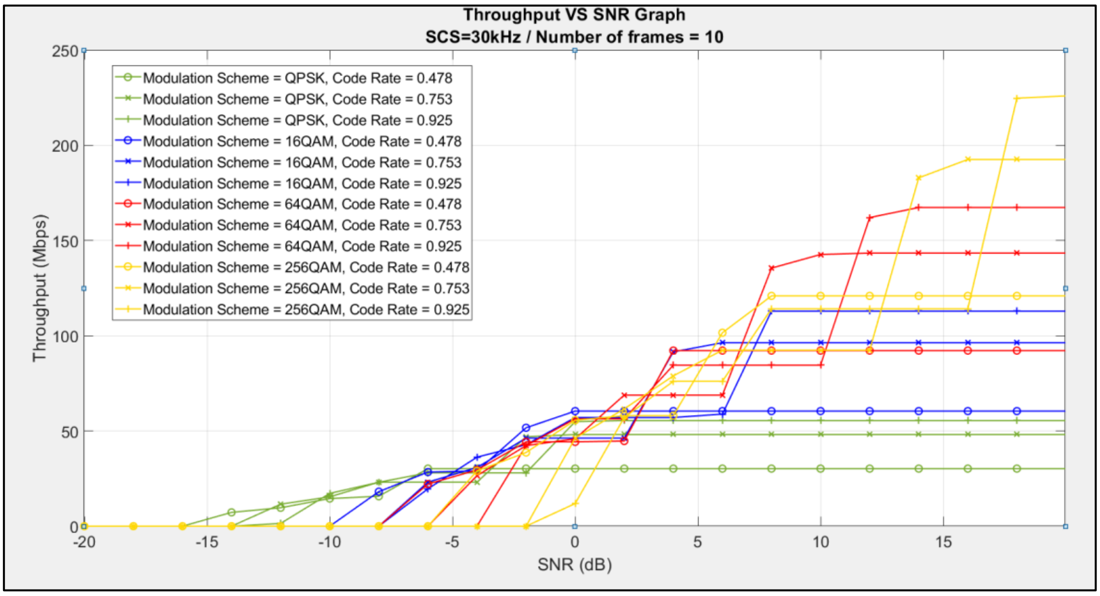

Figure 4-3. *Throughput Vs SNR*

The overall throughput for all modulation schemes gradually rises with the increase in the SNR values. The maximum throughput of 230 Mbps is observed for the 256 QAM with code rate 0.925 at 20 dB, because of the better signal quality and higher modulation scheme with less parity bits. The important observation from the graph is that when the SNR power is less, throughput of lower modulation schemes is better than the higher modulation schemes. Since the lower order modulation schemes are more robust to poor channel conditions

The QPSK modulation scheme remains reliable for almost all conditions. The 16 QAM and 64 QAM throughput gradually increases from -10dB and -5dB respectively. The maximum performance for the 256 QAM is proportional with the better SNR power. The low coding rate performs well for the low SNR value. Considering the different coding rates for the QPSK ranging from -15dB to 0dB, the code rate 0.478 remains steady with high throughput, while the code rate 0.753 and 0.927 dramatically increase when the signal power increases with respect to noise. This happens because an equal number of data bits are appended with parity bits, which help the receiver to detect the erroneous packets and decode it more effectively via error correction techniques. The use of high code rate is beneficial when the condition of the channel is better with high signal quality reception.

In 5G NR one of the key technologies is the URLLC, in which ultra-low latency and reliable communication is provided. The applications in which the throughput requirement is low, but the reliability and low latency is critical. So, according to the findings in the simulations provided a promising result for these applications, meaning that by using low modulation scheme with appropriate coding rate the communication link reliability remains optimal even in worst SNR conditions.

4.4.2 Delay Analysis

The delay is the most critical factor for URLLC. For designing a good communication link for URLLC, the delay is evaluated at high priorities. The Figure 4-4 illustrates the delay vs SNR graph. The x-axis represents the values of SNR in decibel dB and the y-axis represents the delay in terms of occurred retransmissions in total for erroneous packets within a subframe. The same simulation parameters and legends are used in delay analysis which were used for throughput analysis in section 4.4.

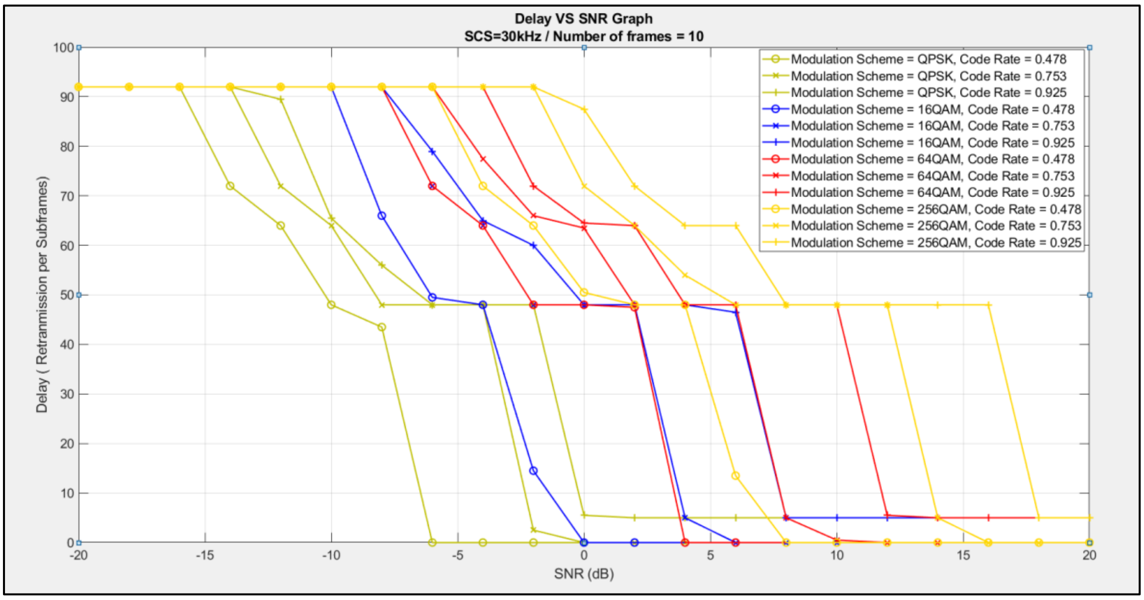

Figure 4-4. *Delay Vs SNR*

The maximum delay for all the modulation schemes and coding rate is 92 retransmissions per subframe from the range -20db to -15dB, which are the worst channel conditions for any radio packets transmissions. As expected, the most reliable communication in the worst conditions is the QPSK with coding rate of 0.470 and 0.753 with less number of retransmissions compared to other transmission schemes. Although the data throughput is less because of the use of a low modulation scheme. The most critical findings are between the range -4dB to 10dB, where different modulation schemes with different coding rates provide similar latency with different throughputs. For 256 QAM with higher coding rate, the delay factor remains non-zero even in good conditions because of low tolerance for noise and channel fading. When the high data rates are transmitted with higher modulation schemes, the packet error rate increases and the hybrid-ARQ retransmission are more frequent as seen from the results.

The study of the delay in terms of retransmissions is very significant for different applications in 5G NR. It provides the information about the using of different modulation schemes and coding rate under different channel conditions, which in result benefits to use maximum throughput with the same delay and reliability of packet delivery. In addition, it also helps to analyse the trade-off between used total transmission energy and number of retransmissions.

To get more objective results for the simulations the throughput and delay are compared at four different channel conditions -4dB, 0dB, 6dB and 10dB.

4.4.3 Delay Vs Throughput at – 4dB

The Figure 4-5 shows the comparison between the throughput (Mbps) and the delay (retransmission per subframes %) at -4dB. Modulation schemes are presented by different colours, green for QPSK; blue for 16 QAM; red for 64 QAM and yellow for 256 QAM, while coding rates are presented by markers, circle for coding rate 0.478; square for coding rate 0.753 and diamond for coding rate 0.925.

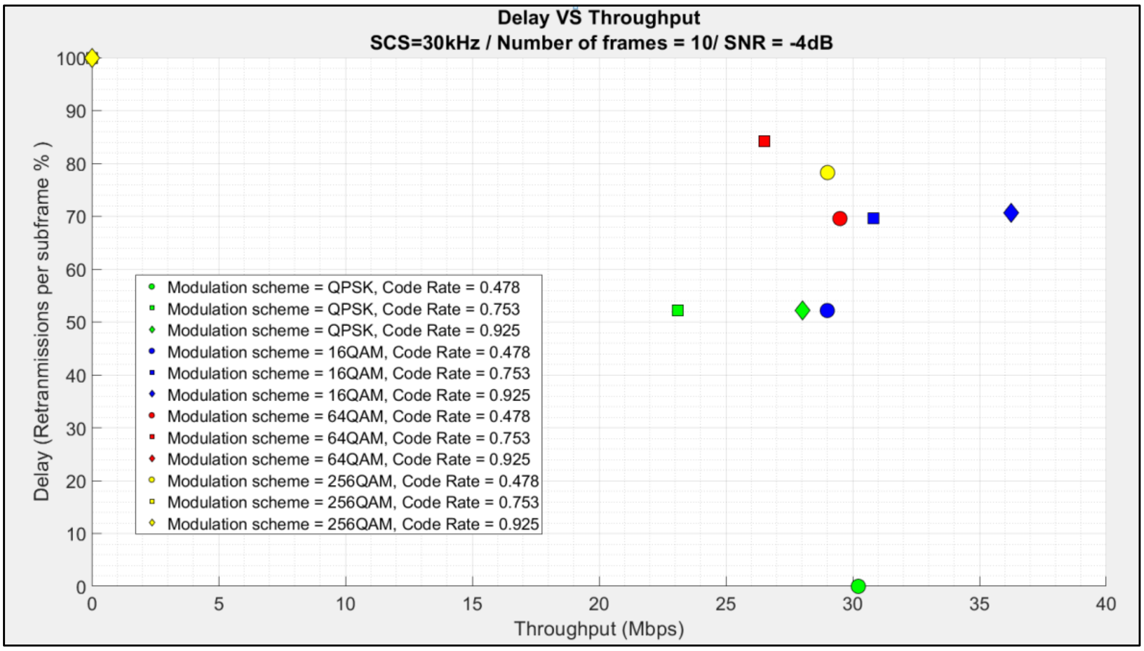

Figure 4-5. *Delay Vs Throughput at -4dB*

The maximum throughput is 36 Mbps but with a delay of 70% for 16QAM with a coding rate of 0.925. However, the most reliable and error free delivery is for the modulation scheme QPSK with a coding rate of 0.478 and the throughput is 30Mbps. The worst modulation scheme in this scenario is 256 QAM with a coding rate of 0.753 and 0.925.

So, in conclusion the most significant and reliable modulation scheme at -4dB or worst condition is QPSK with a low coding rate of 0.478. If the application is very critical and cannot resist any delay or errors and can operate at a low throughput, the QPSK modulation is suitable for that application.

4.4.4 Delay Vs Throughput at 0dB

At 0db, the 16QAM with a coding rate of 0.478 is the highest throughput of 60 Mbps with no delay meaning that all packets delivered at first go without any error and retransmissions as shown in Figure 4-6. Also, the performance of the QPSK modulation scheme is significant and more reliable in the same conditions with all coding rates. The use of higher modulation in this scenario is not valuable because of the more retransmission and errors.

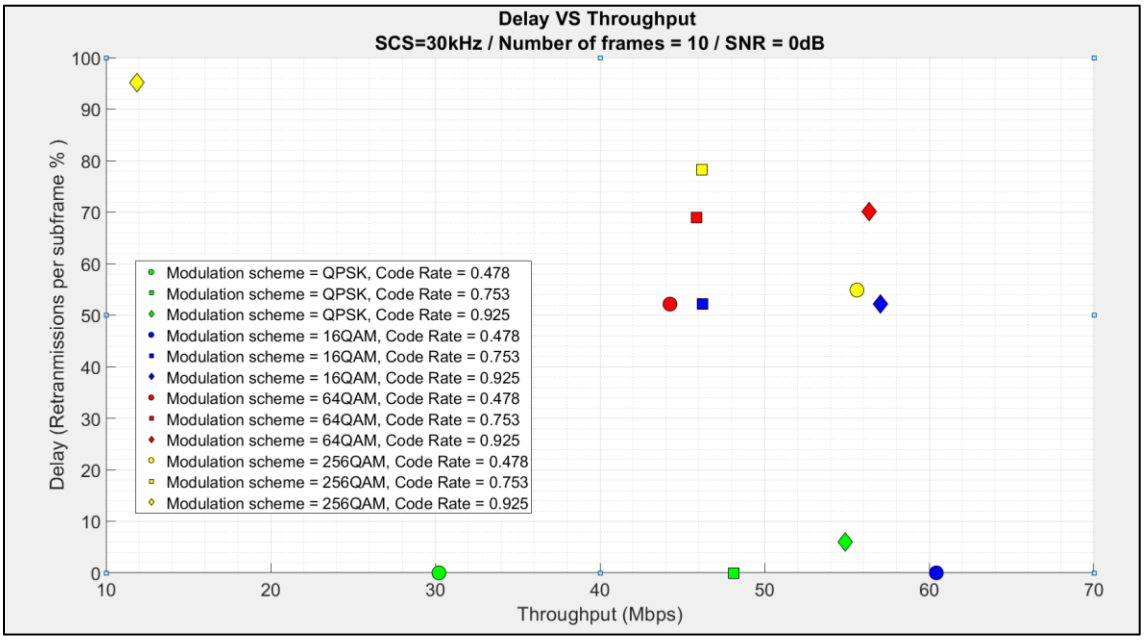

Figure 4-6. *Delay Vs Throughput at 0dB*

In this scenario, the applications which need high data rates with error free delivery can implement a 16QAM modulation scheme with low coding rate. Compared to the previous scenario the throughput is doubled and provides the same reliability of packet delivery. The 16 QAM,64 QAM and 256 QAM with coding rate of 0.925,0.925 and 0.478 respectively also provides good throughput but with the 50% of delay which might have a considerable effect in certain URLLC applications.

In conclusion the 16 QAM modulation scheme provides more throughput than the QPSK using low coding rate in current scenario with more maximum reliability.

4.4.5 Delay Vs Throughput at 6dB

As the channel condition gets better, the higher modulation schemes are performing better because of error free delivery and larger potential throughput capability compared to the lower order modulation schemes. The Figure 4-7 shows the delay vs throughput comparison at SNR value 6dB.

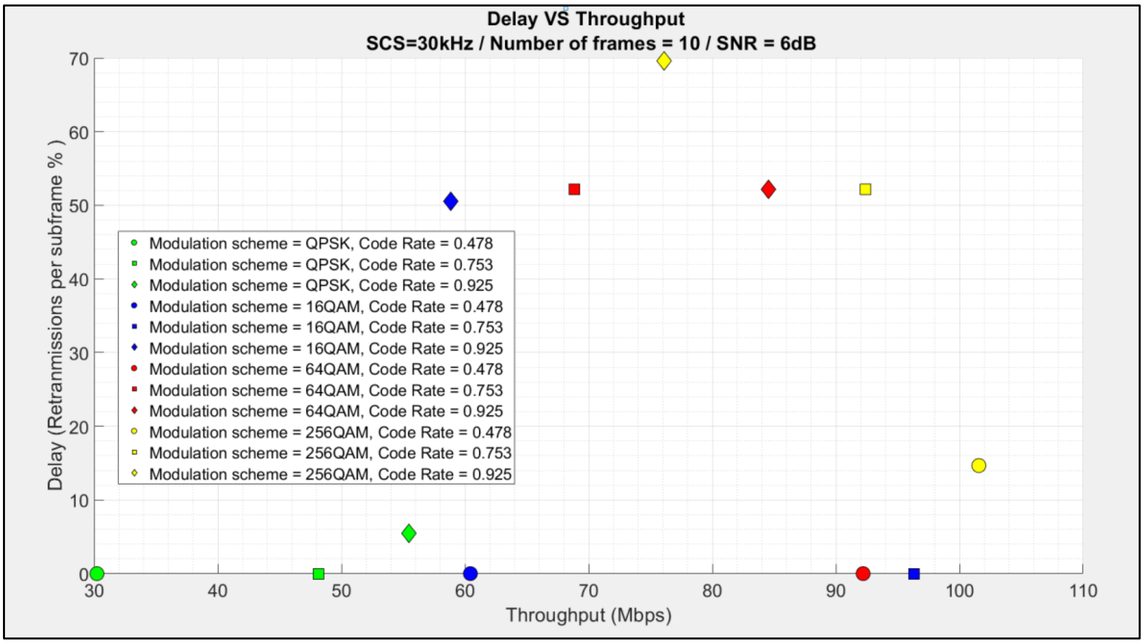

Figure 4-7. *Delay Vs Throughput at 6dB*

In this scenario, the 256 QAM modulation scheme with 0.478 has the highest throughput with value of 102 Mbps which is two times more than previous scenarios. But it also has some delays in terms of retransmissions. By further analysing the graph the modulation schemes are proportional to the channel conditions. At 6 dB, we have more reliable modulation schemes which are QPSK with coding rate of 0.478 and 0.753; 16 QAM with coding rate of 0.478 and 0.753 and last 64 QAM with coding rate of 0.478.

So, from the results it can be observed that the use of a high coding rate is not suitable for the given conditions. The medium and low coding rate are less vulnerable to the errors and hence provide better reliability of packet delivery reliability and throughput with all modulation schemes.

4.4.6 Delay Vs Throughput at 10dB

From the Figure 4-8, which shows the delay vs throughput comparison at SNR value 10d, it can be observed that almost all the modulation schemes perform at their maixmum throughput with no retransmissions and delay. The 64QAM with coding rate of 0.753 has highest throughput of 144 Mbps with no delay. As the channel conditions are getting better with more signal power, less errors and retranamissions occurs.

The applications which need more than 100 Mbps data rate with no latency due to retransmissions and reliable communication link in the given conditions, the 64 QAM with coding rate 0.753 is most suitable compared to other modulation schemes.

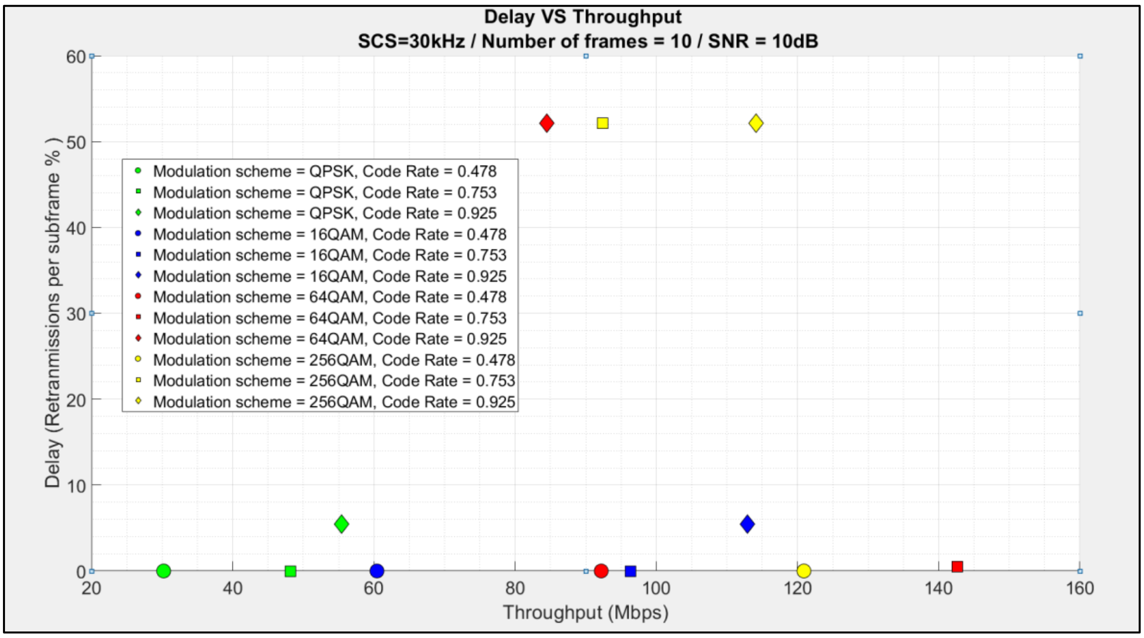

Figure 4-8. *Delay Vs Throughput at 10dB*

As the signal to noise ratio increases the higher modulation schemes with the higher coding rate have less delay and provide maximum throughput as can be seen in Figure 4-8. However, for the negative SNR values the higher modulation schemes had more retransmission and errors. For these conditions lower modulation schemes with low coding rate perform well and provide a reliable communication link and error free delivery.

5 CONCLUSIONS

In this thesis, throughput and latency are analysed for different URLLC applications requirements, which is one of the key enabler technologies in 5G NR, considering various channel conditions. Different MCSs are set as a parameter to evaluate the optimal reliability with maximum performance in downlink communication link of 5G NR. The throughput and delay are analysed by the successful reception of data packets at the receiver and the numbers subframe retransmissions by the hybrid-ARQ, which occurs due to variant channel conditions respectively.

In chapter 2, a brief overview is presented about the protocol stack, network architecture and modulation schemes used in 5G NR. The waveform used for the simulation is CP-OFDM with 30 kHz subcarrier spacing. The applications defined in 3GPP which require low latency are also discussed. In chapter 3, a detailed literature study is conducted on the data channels and downlink data transmission via PDSCH in 5G NR. The working principle and requirement of the hybrid-ARQ retransmission protocols in data transmission is also briefly described.

Chapter 4 includes the system parameters; overall function block used for the simulation and conclude the result. The most important outcome of the simulation is that, different robust MCSs provides maximum throughput with no delays and retransmissions which means high reliability link is established with respect to varying channel conditions. The low modulation schemes like QPSK with coding rate of 0.470 and 0.753 respectively are suggested for the critical and remote applications, which requires a threshold level of data rate with error free delivery in worst channel conditions. As predicated, when the channel conditions are good the performance of higher modulation schemes gets better and provides high data rates with low error deliveries. The throughput verses delay graphs provide critical results at different channel conditions, that some modulation schemes display the same latency but with various throughputs. Thus, these findings help us to determine the trade-off between maximum throughput of data transmission and the low latency.

As a future work, a further investigation is done by implementing higher subcarrier spacings and studying the impact of the short slot duration on the throughput and latency using the same retransmission protocol. In addition, defining average delays in time units, instead of number of retransmissions, is of considerable interest. Furthermore, the uplink transmission can be simulated, and a trade-off can be found between the maximum data rate and reliability.

REFERENCES

[1] M.R. Palattella, M. Dohler, A. Grieco, G. Rizzo, J. Torsner, T. Engel and L. Ladid, '"Internet of things in the 5G era: Enablers, architecture, and business models," *IEEE J.Select.Areas Commun.*, vol. 34, no. 3, pp. 510-527.

[2] S. Andreev, O. Galinina, A. Pyattaev, M. Gerasimenko, T. Tirronen, J. Torsner, J. Sachs, M. Dohler and Y. Koucheryavy, '"Understanding the IoT connectivity landscape: a contemporary M2M radio technology roadmap," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 32-40.

[3] A. Goldsmith, '"Wireless communications,", 2005.

[4] L. Le and E. Hossain, '"An analytical model for ARQ cooperative diversity in multi-hop wireless networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 5, pp. 1786-1791.

[5] M. Jabi, M. Benjillali, L. Szczecinski and F. Labeau, '"Energy efficiency of adaptive HARQ," *IEEE Trans.Commun.*, vol. 64, no. 2, pp. 818-831.

[6] A. Chelli, E. Zedini, M. Alouini, J. R. Barry and M. Pätzold, '"Performance and Delay Analysis of Hybrid ARQ With Incremental Redundancy Over Double Rayleigh Fading Channels," *IEEE Transactions on Wireless Communications*, vol. 13, no. 11, pp. 6245-6258.

[7] S. H. Kim, S. J. Lee and D. K. Sung, '"Low-Complexity Rate Selection of HARQ With Chase Combining in Rayleigh Block-Fading Channels," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 6, pp. 2818-2824.

[8] E. Dosti, U.L. Wijewardhana, H. Alves and M. Latva-aho, '"Ultra reliable communication via optimum power allocation for Type-I ARQ in finite Block-Length,", pp. 1-6.

[9] K. Wang and Z. Ding, '"Diversity integration in hybrid-ARQ with chase combining under partial CSI," *IEEE Trans.Commun.*, vol. 64, no. 6, pp. 2647-2659.

[10] A. Chelli and M. Alouini, '"On the performance of hybrid-ARQ with incremental redundancy and with code combining over relay channels," *IEEE transactions on wireless communications*, vol. 12, no. 8, pp. 3860-3871.

[11] C. Ji, D. Wang, N. Liu and X. You, '"On Power Allocation for Incremental Redundancy Hybrid ARQ," *IEEE Transactions on Wireless Communications*, vol. 14, no. 3, pp. 1506-1518.

[12] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski and A. Dekorsy, '"Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Communications Magazine*, vol. 54, no. 9, pp. 59-65.

[13] E. Dahlman, S. Parkvall and J. Skold, '"4G, LTE-Advanced Pro and The Road to 5G,", 2016.

[14] T. Chapman, E. Larsson, P. Von Wrycza, E. Dahlman, S. Parkvall and J. Skold, '"HSPA Evolution: The Fundamentals for Mobile Broadband,", 2014.

[15] E. Dahlman, S. Parkvall and J. Skold, '"5G NR: The Next Generation Wireless Access Technology,", 2018.

[16] S. E. Elayoubi, M. Fallgren, P. Spapis, G. Zimmermann, D. Martín-Sacristán, C. Yang, S. Jeux, P. Agyapong, L. Campoy, Y. Qi and S. Singh, '"5G service requirements and operational

use cases: Analysis and METIS II vision," *2016 European Conference on Networks and Communications (EuCNC)*, pp. 158-162.

[17] M. Series, '"IMT Vision–Framework and overall objectives of the future development of IMT for 2020 and beyond,".

[18] NGMN Alliance, '"NGMN 5G White paper,", February,.

[19] ICT -317669 METIS, Deliverable 1. 5 Version 1, '"Updated scenarios, requirements and KPIs for 5G mobile and wireless system with recommendations for future investigations,", April,.

[20] 3GPP and 1.5. Release, '"NR; User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone
," *Specification #: 38.101-1*, Janurary,.

[21] 3GPP and 1.4. Release, '"Study on new radio access technology: Radio access architecture and interfaces," *Specification # 38.801*, March,.

[22] 3GPP and 1.5. Release, '"Study on New Radio (NR) access technology," *Specification #: 38.912*, July,.

[23] 3GPP and 1.5. Release, '"NR; Overall description; Stage-2," *Specification #: 38.300*, Janurary,.

[24] 3GPP TSG RAN Meeting 71 and RP-160671, '"Study on NR New Radio Access Technology,", March,.

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

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

[27] 3GPP and 1.4. Release, '"Study on new radio access technology Physical layer aspects
," *Specification #: 38.802*, September,.

[28] A. Ghosh, '"5G New Radio (NR): physical layer overview and performance,".

[29] X. Lin, J. Li, R. Baldemair, J.T. Cheng, S. Parkvall, D.C. Larsson, H. Koorapaty, M. Frenne, S. Falahati and A. Grovlen, '"5G new radio: Unveiling the essentials of the next generation wireless access technology," *IEEE Communications Standards Magazine*, vol. 3, no. 3, pp. 30-37.

[30] Z. Li, M.A. Uusitalo, H. Shariatmadari and B. Singh, '"5g urllc: Design challenges and system concepts,", pp. 1-6.

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

[32] I.F. Akyildiz, S. Nie, S. Lin and M. Chandrasekaran, '"5G roadmap: 10 key enabling technologies," *Computer Networks*, vol. 106, pp. 17-48.

[33] 3GPP and 1.5. Release, '"NR; User Equipment (UE) radio access capabilities," *Specification #: 38.306*, January,.

[34] X. Lin, J. Li, R. Baldemair, J.T. Cheng, S. Parkvall, D.C. Larsson, H. Koorapaty, M. Frenne, S. Falahati and A. Grovlen, '"5G new radio: Unveiling the essentials of the next generation wireless access technology," *IEEE Communications Standards Magazine*, vol. 3, no. 3, pp. 30-37.

[35] C.E. Shannon, '"A mathematical theory of communication," *Bell system technical journal*, vol. 27, no. 3, pp. 379-423.

[36] J.M. Wozencraft and M. Horstein, '"Coding for two-way channels,".

[37] J. Padhye, V. Firoiu, D.F. Towsley and J.F. Kurose, '"Modeling TCP Reno performance: a simple model and its empirical validation," *IEEE/ACM transactions on Networking*, vol. 8, no. 2, pp. 133-145.

[38] D. Chase, '"Code Combining - A Maximum-Likelihood Decoding Approach for Combining an Arbitrary Number of Noisy Packets," *IEEE Transactions on Communications*, vol. 33, no. 5, pp. 385-393.

[39] M.B. Pursley and S.D. Sandberg, '"Incremental-Redundancy Transmission for Meteor-Burst Communications," *IEEE Trans Commun*, vol. 39, no. 5, pp. 689-702.

[40] S.B. Wicker and M.J. Bartz, '"Type-II hybrid-ARQ protocols using punctured MDS codes," *IEEE Trans.Commun.*, vol. 42, no. 234, pp. 1431-1440.

[41] C. Shi, A. Bergstrom, E. Eriksson, P. Frenger and A. A. Zaidi, '"Retransmission Schemes for 5G Radio Interface," *2016 IEEE Globecom Workshops (GC Wkshps)*, pp. 1-6.

[42] 3GPP and 1.6. Release, '"Study on channel model for frequencies from 0.5 to 100 GHz,", January,.