Lean Beam Management for New Radio

Nawanit Kumar

Department of Electrical and Information Technology
Lund University

Supervisors: Stefan Höst, LTH
Thomas Svensson and Joachim Ramkull, Ericsson Lund

Examiner: Maria Kihl

August 19, 2020

Popular Science Summary

In this modern epoch, demand and growth of high-speed mobile broadband has been surging at never seen before pace. Hence primarily, it is need for tetherless communication with high data rates which has led to need of migration from Fourth generation to Fifth generation of mobile communication rather than voice calls. This leads to need for large radio bandwidth. Large bandwidth further leads to use of millimeter wave spectrum to achieve better power efficiency. Need for better bandwidth efficiency leads to use of phased array antenna modules or massive multiple input multiple output antenna system. The base station has multiple antennas in Fifth generation New Radio. Despite availability of many antennas, the base station is equipped only with a fixed set of steering vectors. These are called beams in New Radio. Each beam is pointing in a certain direction in space. To cover multiple user equipment in a single transmission is almost impossible when the radiation is beam shaped unless they are in very close proximity.

Fifth generation New Radio is the new radio access technology developed by the 3rd Generation Partnership Project for mobile communication network. It is the global standard for the air interface of Fifth generation networks. Analog beam-forming is a technique used widely in Fifth generation New Radio Access Network to counter propagation effects between transmitter and receiver. In analog beam-forming, connected devices are individually tracked so that the communication with each device takes place on the best beam. This is called beam management. Beam management comes with an additional cost. For radio access network to track the device and use the best beam, devices must perform regular measurements called channel sounding and estimation and report them to the base station of the mobile network system. Hence, these measurements consume air interface resources which are costly and scarce given the high traffic volume and could otherwise be used for data. The challenge increases with increase in the number of connected users in a cell.

The thesis aims to investigate current implementation of downlink beam management in New Radio high-band and propose better suggestions to minimise these measurements while still achieving decent throughput and coverage with reduced interference.

Abstract

The use of millimetre-wave frequencies in Fifth generation New Radio has lead to high path loss due to radio propagation environment. Simultaneously, high data rates is a must for current products due to high market demand. This requirement accounts for efficient tracking of the best beam for the user equipment to stay connected to it. Current products at Ericsson use such methods of beam management. These methods call for regular measurements of beams which add up to overhead. Overhead reduction is an area that is being continuously researched. Through this thesis, two algorithms are proposed with the aim to reduce overhead. First, by increasing the periodicity of measurements. Second, by reducing the number of narrow beam measurements. These solutions are simulated by comparing overall system throughput values from a selection of beams. Theoretical calculations have also been done. Both these results are compared to see the effectiveness of the proposed solutions. Different use cases are taken into consideration such as stationary and/or mobile users. In review, it can be clearly seen that increasing periodicity of channel sounding is a good way to decrease measurements and use the resources for data instead. However, reducing narrow beam measurements was not found to be an overall feasible algorithm as the system performance was poor for uplink leading to degradation in downlink. Thus, increasing periodicity for channel estimation in downlink can be applied to New Radio beam management scenarios in millimetre-wave high-band to yield better average cell downlink throughput as compared to current products.

Acknowledgements

Table of Contents

List of Figures

List of Tables

List of Acronyms

3GPP	Third Generation Partnership Project
3D-UMa	Three Dimesional Urban Macrocell
4G	Fourth Generation
5G	Fifth Generation
ABF	Analog Beamforming
ARQ	Automatic Repeat-reQuest
AWGN	Additive White Gaussian Noise
BLER	Block Error Ratio
BS	Base Station
CA	Carrier Aggregation
CC	Component Carrier
CDM	Coding and Modulation Scheme
CE	Control Element
CH	Channel Hardening
CORESET	Control Resource Set
CSI	Channel State Information
CSI-IM	CSI Interference Measurement
CSI-RS	CSI Reference Signal
CQI	Channel Quality Indicator
DCI	Downlink Control Information
DFT	Discrete Fourier Transform
DL	Downlink
DM-RS	Demodulation Reference Signal
EKF	Extended Kalman Filter
eMBB	enhanced MBB
eNB	eNodeB
EN-DC	E-UTRA NR Dual-Connectivity
EPA	Extended Pedestrian A Model
EPC	Evolved Packet Core

E-UTRA	Evolved UTRA
FDD	Frequency Division Duplex
FDM	Frequency-Domain Sharing
FR1	Frequency Range 1
FR2	Frequency Range 1
FTP	File Transfer Protocol
gNB	gNodeB
gNodeB	generalized NodeB
GSM	Global System for Mobile Communication
H-ARQ	Hybrid ARQ
IEEE	Institute of Electrical and Electronics Engineers
IMT-2020	International Mobile Telecommunications 2020
IoT	Internet of Things
ITU-R	ITU-Radio Sector
ITU	International Telecommunications Union
KPI	Key Performance Index
L1-RSRP	Layer 1 RSRP
LA	Link adaptation
LTE	Long Term Evolution
LOS	Line Of Sight
MAC	Medium Access Control
MAC-CE	MAC Control Element
MBB	Mobile Broadband
MCS	Modulation and Coding Scheme
MIMO	Multiple Input Multiple Output
ML	Machine Learning
MTC	Machine Type Communication
mMTC	massive MTC
MU-MIMO	Multi-User MIMO
mmWave	Millimeter Wave
NLOS	Non LOS
NOMA	Non-Orthogonal Multiple Access
NR	New Radio
NSA	Non Standalone
NZP CSI-RS	Non-zero-power CSI-RS
OFDM	Orthogonal Frequency-Division Multiplexing
PAAM	Phased Antenna Array Modules
PBCH	Physical Broadcast Channel
PCCH	Paging Control Channel
PDCCH	Physical Downlink Control Channel
PDSCH	Physical Downlink Shared Channel
PHY	Physical Layer
PRACH	Physical Random-Access Channel

PRB	Physical Resource Block
PT-RS	Phase Tracking Reference Signal
PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel
PMI	Precoder Matrix Indicator
QAM	Quadrature Amplitude Modulation
QCL	Quasi Co-Location
QoS	Quality-of-Service
RACH	Random Access Channel
RAN	Radio Access Network
RAT	Radio Access Technology
RB	Resource Block
RE	Resource Element
RI	Rank Indicator
RIM	Remote Interference Management
RIM-RS	RIM Reference Signal
RLC	Radio-Link Control
RRC	Radio Resource Control
RRM	Radio Resource Management
RS	Reference Symbol
RSRP	Reference Signal Received Power
RV	Redundancy Version
SNR	Signal-to-Noise Ratio
SRS	Sounding Reference Signal
SSB	Synchronization Signal Block
SDAP	Service Data Application Protocol
TBS	Transport Block Size
TCI	Transmission Configuration Indication
TCP	Transmission Control Protocol
TDD	Time-Division Duplex
TDM	Time Domain Sharing
TRP	Transmission Reception Point
TRS	Tracking Reference Signal
TTI	Transmission Time Interval
UDP	User Datagram Protocol
UE	User Equipment
UL	Uplink
UMTS	Universal Mobile Telecommunication System
URLLC	Ultra-reliable low-latency communication
UTRA	Universal Terrestrial Radio Access
ZP CSI-RS	Zero-power CSI-RS

Introduction

1.1 Background

5G New Radio (NR) Radio Access Network (RAN) is expected to achieve very high data rates of more than 1 Gigabits per second (Gbps) with Enhanced Mobile Broadband (eMBB) according to Third Generation Partnership Project (3GPP) [14]. Looking at Ericsson's mobility report, mobile data traffic has grown by 49 percent between quarter 4 for 2018 and 2019 [34] to reach 6.3 billion mobile broadband subscriptions globally. A total of 49 million subscriptions were added in just one quarter. 5G subscriptions are expected to reach 190 million by end of 2020 [35]. This rapid traffic growth is caused by rising number of smartphone subscriptions and increasing average data volume per subscription. This is primarily due to increase in video content usage. Rise in subscriptions is also due to increase in number of connected devices through Internet of Things (IoT). This demand leads to need for large radio bandwidth which further leads to use of millimetre wave (mmWave) spectrum for better power efficiency. Due to physics of nature of waves, transmitting a signal in all directions and/or at relatively wide angles is possible for mid and low frequency range. In these cases, a single transmission would cover a lot of user equipment (UE) simultaneously without using massive antenna array. Need for better bandwidth efficiency leads to use of phased antenna array modules (PAAM) or massive multiple input multiple output (mMIMO) antenna system which further results in radiation being a beam. Beam is used for high frequency to cover multiple UE. However, mmWaves experience great propagation loss, especially for larger distances between transmitter and receiver and even due to the atmospheric absorption by oxygen molecules and water vapor [33]. Channel conditions and radio link quality also changes significantly for mobile users. In order to counter this difficulty, a very sophisticated technique of managing/controlling the beam to cover multiple devices scattered in all directions is required. This control mechanism called beamforming should be generic enough to adapt to different scenarios.

Beamforming concentrates transmitted signal in receiver's direction which improves received signal power. Analog Beamforming (ABF) is such a technique used widely in 5G NR RAN to counter propagation effects between transmitter and receiver [32]. In ABF, connected devices are individually tracked so that the communication with each device happens on the best beam. This is called beam

management. Beam Management is a collection of techniques such as initial beam establishment, beam refinement and beam tracking to refine directional links between transmitter and receiver beams [29].

1.2   Objective

The aim of the thesis is to investigate the current standards of beam management and beam tracking measurements applied in the industry as baseline solutions. Comprehensive study of recent ongoing work in this domain is done and based on the cumulative knowledge of the above, new mechanisms of reducing measurements of beam tracking are to be proposed and simulated. These proposed solutions are compared with baseline solutions for their effectiveness.

1.3   Thesis Formulation

Beam management involves selecting the best beam for connection establishment between the UE and the network or base station (BS). This is typically done by measuring Channel State Information Reference Signals (CSI-RS) of the candidate beams [1]. More specifically, Reference Signal Received Power (RSRP) values of the candidate beams are measured and the one with highest value is selected as the best beam. Ideally, this power will be highest where best possible channel quality is experienced. The standards on these channel reference signals is left to be independently designed by the network provider by 3GPP [1]. Thus, these measurements for best beam occur with a specific periodicity which is determined by the network provider alone.

Minimising the number of reference measurements is vital so that majority of time-domain resources are used for data and control information. This is even more crucial in dense deployments for reducing power consumption. However, trying to optimise and minimise all the parameters associated with beam management lies outside the scope of the thesis. The problem formulation of the thesis was limited to investigating effect of periodicity increase of CSI-RS signals and reducing narrow beam measurements on overall system performance and throughput.

1.4   Previous Work

Informative and diverse literature was explored as the thesis progressed which developed robust knowledge about the project and current standards. The area of beam management for 5G is an on-going field of research. Some of the previous works have been included in the references section.

Basic knowledge about beam forming can be can be found in [20][21]. Basic overall knowledge of NR can be found in textbook [28]. A lot of information stated in the thesis report is obtained by thorough reading of the book. Detailed tutorial on beam management for 5G can be found in [29].

Non-Orthogonal Multiple Access (NOMA) is a topic that is gaining popularity. NOMA is a proposal by 3GPP Long-term Evolution Advanced (3GPP-LTE-A) [19]

to address increase in user numbers in 5G. NOMA can provide better bandwidth efficiency over conventional orthogonal multiple-access (OMA) techniques. It has also been combined with reducing the overhead required for channel state information (CSI). Performance of downlink NOMA using partial CSI at transmitter has been investigated by Yang et al. More specifically, outage probability of NOMA by assuming either imperfect CSI or second-order-statistics-based CSI has been explored [22]. With knowledge of statistical CSI, Shi et al. [23] explored outage of NOMA. Cui et al. [24] studied best decoding order and best power allocation of users in downlink NOMA system assuming average CSI at the BS. Yang et al. [25] used single-bit feedback CSI from each UE to BS to study outage performance for downlink. NOMA is found to be a good alternative in case of partial CSI measurements.

Channel Hardening (CH) effect where channel variations decrease due to use of narrow beams causing correlated measurements by UE have been explored in [26]. Beams are used to send channel sounding reference signals in NR. Energy consumption increases when more beams are used. Channel Hardening (CH) is defined as decrease in channel variations with narrow beams under certain channel conditions. This results in correlated measurements. Thus, number of measurements can be decreased in such cases. A method for CH detection and measurement adaptation is proposed and results show reduction of measurements up-to eight times.

Performance of 5G NR downlink data channel for UE on the cell edge through inter-cell-interference from neighbouring cells is investigated in two ways in [30] Specifically, Non-zero-power CSI-RS (NZP CSI-RS) based method and the CSI interference measurement (CSI-IM) method have been investigated. IMs are essential in efficient link adaptation (LA) techniques. They provide essential information on channel quality indicator (CQI) calculations at UE. Effect of IM methods on 5G NR LA and throughput is investigated. System overhead and performance evaluation have been compared. It is concluded in the paper that NZP CSI-RS method achieves more accurate IM than CSI-IM in non-precoded interference scenario. On the other hand, CSI-IM method is better in precoded interference measurements.

CSI for mMIMO planar antenna array is studied and a hybrid CSI feedback mechanism is proposed in combination of beamformed CSI-RS and non-beamformed CSI-RS transmission in [41] Planar antenna array system has been used for the measurements. Non-beamformed CSI-RS codeword in the code-book for feedback has been associated with a beamformed CSI-RS. Results show accurate CSI feedback can be achieved without large CSI-RS overhead and maintaining CSI-RS coverage.

Remote interference problem in 5G NR MACRO deployment of remote aggressor base-stations where uplink reception is interfered by the downlink in time division duplexing is covered in [40]. This problem degrades the performance of MU grouping. Uplink reception is effected by downlink transmissions in time division duplex (TDD) networks under specific atmospheric conditions. Remote Interference Management (RIM) is suggested using reference signals. RIM reference signals (RIM-RS) based on CSI-RS for 5G NR and LTE are proposed. Results show that LTE RIM-RS is better when number of interfering BSs are low and

NR RIM-RS is better with more BSs. Additionally, NR RIM-RS proves to give smaller overhead and can be frequency multiplexed with the physical downlink shared channel (PDSCH).

Joint grouping and scheduling scheme using CSI-RS to tackle limited sounding reference signals (SRS) scenario which restricts channel information acquired by BS is proposed in [42] In TDD mMIMO systems, multi-user (MU) grouping is used to enhance spectral efficiency. But MU grouping performance is degraded by limited SRS. A joint grouping and scheduling using CSI-RS has been proposed in the paper. Non-precoded and beam formed CSI-RS is used to measure channel accuracy in the beginning stage and then grouping algorithm is implemented. Low-correlation users with acceptable channel accuracy are grouped together. The measurements have been done for low and medium speed UE. Results show that low throughput UE performance cases can be enhanced by this method.

Considering previous mentioned papers, it is evident that innovative research has been on-going in the field of CSI. However, thorough investigation in reducing overhead using change in periodicity or using less beams has not been carried. Hence, the thesis will test proposed algorithms to see for their effectiveness.

1.5   Limitations

Some cases that could not be potentially exploited due to limitations of time and scope of the thesis have been stated in this section.

One limitation is that the thesis is limited to Ericsson's internal simulator and theoretical calculations. However, real time data traffic is highly unpredictable. No real time real world simulations were done. Thus analysing and predicting real-time output and behaviour of the algorithms still remains unexplored.

Also, the simulations are run for one cell with one BS. The scenario of handovers between multiple BSs were not exploited in the thesis.

Another limitation is that the Modulation and Coding System (MCS) implemented in the simulator is not entirely known. Some parameters were used to include incomplete measurements taken by UE on cell boundary to prevent statistics from turning skewed. However, the behaviour of measurements and thus reliability of results when UE is at cell boundary may not be completely accurate.

1.6   Layout

The thesis report comprises of several chapters. First is the current chapter of Introduction that introduces the thesis to the readers. The second chapter is Evolution of 5G which describes the current evolving standards of 5G. Third is Current standards and proposed solution which gives a description of the current beam management standards and the solution proposed for the thesis. Fourth is Theoretical calculations which shows desired ideal results to be expected. Fifth chapter is Methodology which describes the simulator and various parameters used for running simulations. Sixth chapter is Results which gives the obtained results along with respective plots. Last chapter is Future Scope which gives a view of potential future work that could grow from the thesis.

Evolution of 5G

This chapter describes the transition from 4G to 5G in depth covering the topics that are vital for the thesis alone.

2.1 5G NR

We are now progressing in the Fifth generation of mobile communication in the current year of 2020. The journey ranging over 40 years witnessed systems based on voice services made available to ordinary people in the first-generation to wide range of use cases far beyond voice and data alone for future mobile communication in the fifth-generation at present. The recommendations about various use cases in 5G started formally in the year 2015 by International Telecommunication Union (ITU)[14].

5G NR features many new and improved 4G LTE technologies such as massive MIMO (mMIMO), millimeter waves (mmWave) and beam management [27]. The term 5G represents new 5G radio-access technology (RAT) covering a wide range of new services. An array of such scenarios with improved performance was introduced in the International Mobile Telecommunication (IMT) 2020 shown in figure 2.1. These are broadly classified as following use cases [14]:
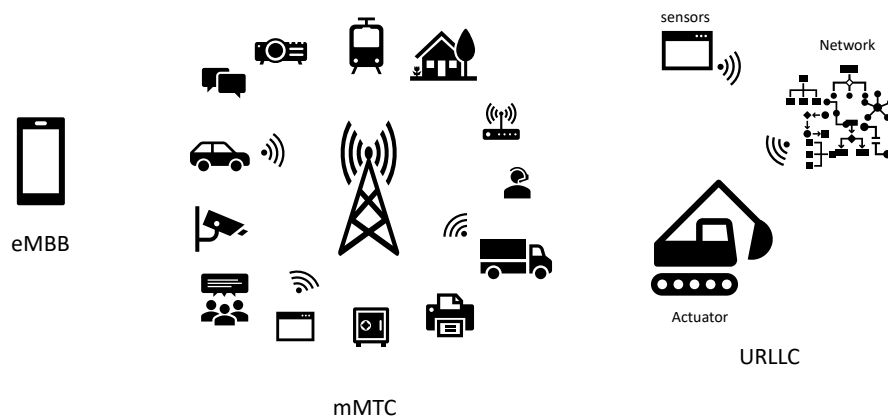
Figure 2.1: IMT-2020 new use cases

5

- Enhanced mobile broadband (eMBB): eMBB is the continuation of the current mobile broadband services enhancing human-centric scenarios of larger data volumes, user experiences and data rates.

- Massive machine-type communication (mMTC): mMTC is focused on machine-centric applications ranging of large volume of low cost connected devices like remote sensors and actuators. These devices would have sparse transmissions thus consuming very less energy corresponding to larger battery lives. These are not delay sensitive and thus do not require large data rates.

- Ultra-reliable and low-latency communication (URLLC): URLLC covers both human and machine-centric communication for very low latency and extremely high availability with high data rates. Examples are tactile internet or vehicle-to-vehicle communication for traffic safety and factory automation.

We will try to focus on the evolution of 5G from LTE and its salient features in the following sections.

2.2   Coexistence

Till today, a device could only be connected to one RAT at a time. For example, either to Global System for Mobile Communications (GSM) or Universal Mobile Telecommunications System (UMTS) or LTE. 3GPP Release 15 [12] [13] focuses on 5G NR complimenting LTE. This is possible due to the higher frequency bands used. This is referred to as Evolved UMTS Terrestrial Radio Access (EUTRA) NR dual node connectivity (EN-DC) or Multi-RAT Dual Connectivity (MR-DC). Dual connectivity allows device being connected to more than one cell and results in inter-connectivity between LTE and 5G NR. The cells can belong to different RATs as is the case for NR-LTE dual connectivity non-standalone operation (NSA) as depicted in figure 2.2.

NR RAN can connect to the legacy LTE core network which is called Evolved Packet Core (EPC). This mode of operation is called NSA. Here connection set-up and paging is under EPC while NR helps increase data-rate and capacity. Standalone connection with 5G Core Network (5GCN) exists for NR in low and mid-band frequency range. But for high-band frequency range , NSA is the current mode of operation.

In NSA, control-plane procedures like initial access and mobility are handled by LTE EPC which is linked to eNode BS (eNB). eNB can be viewed as BS for LTE. NR in NSA is used for handling user-plane data. gNode BS (gNB) and eNB are inter-connected as shown in figure 2.2. gNB is a 5G term for Base Transceiver Station (BTS) or network equipment that transmits and receives wireless communications between UE and a mobile network.

Although gNB is viewed as BS for NR, it is important to understand the real implementation of gNB is like a logical node than a physical node. In particular, Transmission Reception Point (TRP) is used for communication between gNB and UE. TRP is a transmission point of a gNB. TRP can also be viewed as an antenna array consisting of one or more antenna elements located at a specific geographical

location to be used by the network as shown in figure as shown in figure 2.2. This thesis deals with NR high-band, working will be in NSA operation mode.

Figure 2.2: NR NSA network

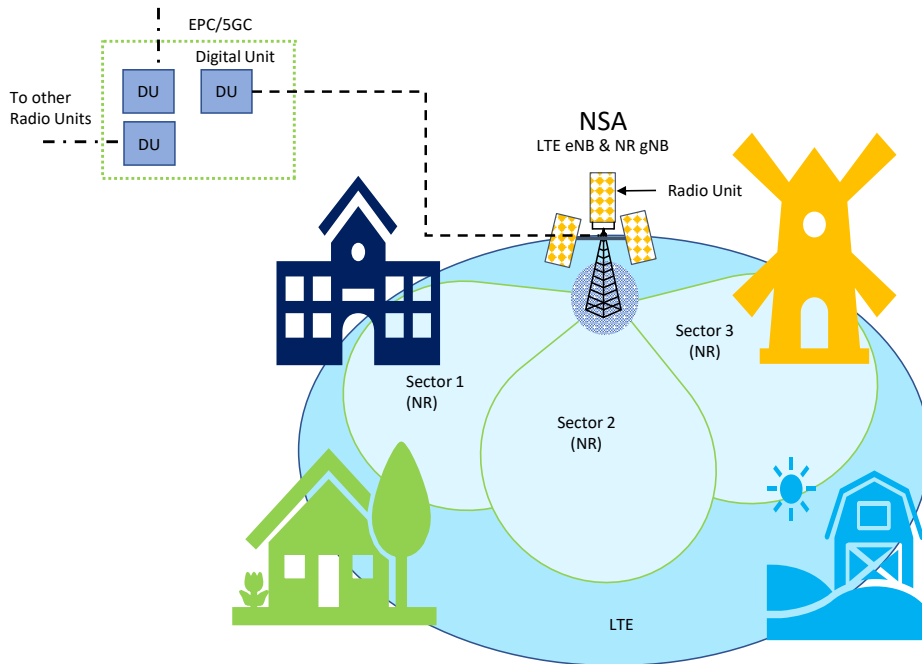## 2.3   NR v/s LTE

Some of the benefits of new design principles in NR associated with this thesis are enlisted below and discussed in detail in the subsequent sections:

- Enhanced transmission scheme

- Dynamic duplex scheme

- Beam-centred beam-forming and mMIMO for data transmission and control-plane processes

- Ultra-lean design boosting performance, reducing interference and energy conservation

2.4   Enhanced transmission scheme

Spectrum flexibility for NR in Release 15 by 3GPP [1]:

Very wide range of spectrum is supported in NR which is a big step ahead of LTE.

- Frequency range1(FR1): All existing and new bands below 6 Gigahertz (GHz).

- Frequency range2(FR2): New bands in and above the range of 24.25526 GHz.

Carrier bandwidths ranging up-to 400 Megahertz (MHz) are supported in NR which is way above 20 MHz supported in LTE. The thesis deals with NR high-band RF spectrum (24-40 GHz) and thus FR2. Current products at Ericsson have RF spectrum allocation as 39 GHz and 28 GHz in high-band range. Carrier bandwidth is determined by the operator according to spectrum allocation. Current high-band products support two different carrier bandwidths of 100 MHz and 50 MHz.

OFDM

Orthogonal frequency-division multiplexing (OFDM) was selected as the de facto transmission scheme for NR. This is due to a lot of features like ability to support both time and frequency domain when defining channels, robustness, decent receiver complexity and support for spatial multiplexing. Compared to LTE where Discrete Fourier Transform (DFT)-precoded OFDM is used, NR implements OFDM for uplink and downlink with an option of DFT-precoding for uplink in certain cases. This vital shift is due to the drawbacks of DFT-precoding version such as increase in complexity of MIMO receivers, a must need for contiguous allocations in frequency domain, loss of symmetry in uplink and downlink and implementation limitation to only one transmission layer while OFDM can support up-to four layers.

Numerology

Numerology refers to sub-carrier spacing between OFDM sub-carriers [18]. It is calculated mathematically as [4]

$$\Delta f = 2^{\mu} \cdot 15000 \text{Hz} \tag{2.1}$$

Different values of equation 2.1 are as follows:

- Δf: OFDM sub-carrier spacing taking values 15/30/60/120/240 kHz corresponding the respective μ

- μ: Numerology number taking values 0/1/2/3/4

A single sub-carrier spacing (corresponding to a single numerology) of 15 Kilohertz (kHz) and a cyclic prefix of approximately 4.7 microseconds (μs) is used in LTE. This is due to the network design of LTE with mostly macro-networks using carrier frequencies in a few GHz range. NR frequency of operation ranges from sub 3 GHz to mmWave of over 25 GHz. Due to physics, it is impossible for a

single numerology to cover this wide range without sacrificing efficiency and/or performance. Thus, NR supports flexible sub-carrier spacing corresponding to a flexible numerology. Maximum carrier bandwidths of 50/100/200/400 MHz for sub-carrier spacing of 15/30/60/120 kHz respectively is supported in NR as shown in figure 2.4. This flexibility supports the required network design of NR with wide range deployment scenarios and large cells.

The wide spectrum allocations in NR also lead to change in cyclic prefix duration of OFDM symbols. Both large and small numerology corresponding to large and small sub-carrier spacing have their advantages. A small sub-carrier spacing gives workable overhead and relatively long cyclic prefix in time domain. On the other hand, a larger sub-carrier spacing tackles the increased phase noise generated at higher carrier frequencies. This thesis in NR high-band used numerology 3 with a sub-carrier spacing of 120 kHz as shown in figure 2.3. Comparing sub-carriers supported in NR with LTE, it can be seen that up to 3300 sub-carriers are supported in NR as compared to 1200 in LTE (for 20 MHz spectrum). Note: Carrier aggregation (CA) is used for larger bandwidth support.

Time-frequency resources

NR resource grid is a collection of resource blocks in frequency domain and slots in time domain as shown in figure 2.3. Resource grid is formed with Resource Elements (REs). RE is the smallest unit of the resource grid. It constitutes one sub-carrier in frequency domain and one OFDM symbol in time domain.
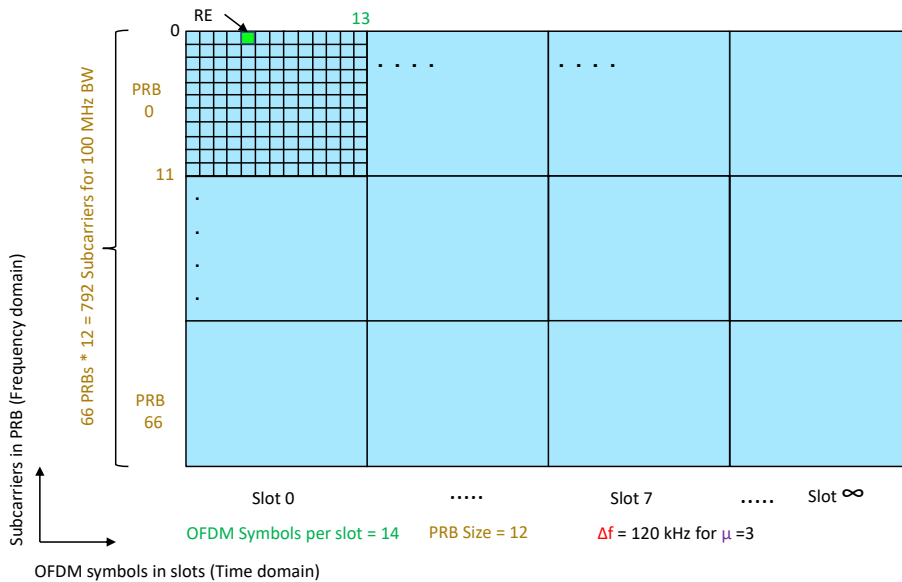

Figure 2.3: NR Time-Frequency architecture for 100 MHz bandwidth and $\mu = 3$

Time-domain architecture

NR defines slots in the time domain consisting of OFDM symbols. NR frames similar to LTE in time domain, are of length 10 milliseconds (ms). These are further divided in 10 sub-frames of length 1 ms as can be seen in figure 2.4. Sub-frames are further divided into slots which have 14 OFDM symbols each as seen in figures 2.3 and 2.4. Thus, duration of a slot in milliseconds depends on the length of an OFDM symbol which further varies with the sub-carrier spacing or the numerology. This is a key difference between LTE and NR.
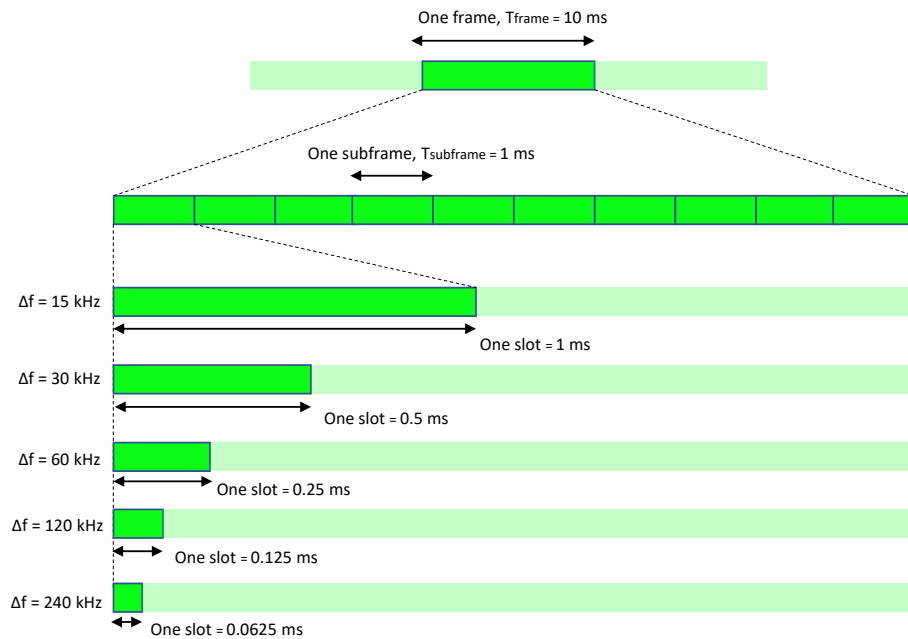
Figure 2.4: Time domain architecture in NR

Another difference between LTE and NR is that NR supports mini-slots transmissions. The necessary OFDM symbols needed to deliver the payload are used instead of entire slot duration. This is beneficial for latency-critical cases and exploring unlicensed spectra. But, the most important benefit is for supporting ABF. At mmWave frequencies with large bandwidths, less OFDM symbols would serve the purpose of payload delivery. This is a benefit when ABF only supports one beam at a time and leads to time-multiplexing of multiple UE.

Frequency-domain architecture

NR defines resource blocs in frequency domain as shown in figure 2.3. OFDM subcarriers carry parts of transmitted data and form the smallest physical resource or RE in NR. A combination of such REs is called a Physical Resource blocks (PRB). In NR, 12 such consecutive REs form one resource block. This is different from

LTE where two dimensional definition of PRB exists. One is of 12 sub-carriers in the frequency domain and another of one slot in the time domain. The definition of NR resource block exists only for the frequency domain to provide flexibility in time duration for different transmissions which was not the case in the original LTE release.

Antenna ports

As per 3GPP, an antenna port is defined such that the channel over which a symbol on the antenna port is conveyed can be inferred from the channel over which another symbol on the same antenna port is conveyed [4]. Similar to LTE, an antenna port in NR is defined so that it corresponds to one channel. Two symbols transmitted from the same antenna port assume the same channel of propagation. Thus, a specific antenna port whose identity is known to the receiving UE can be used for all downlink transmissions related to that UE.

Antenna port is an abstract idea and does not mean a physical antenna in reality. UE analyses multiple transmissions based on Beamforming Function or Mapping Function or Spatial Filter. Figure 2.5 shows input data going into spatial filter which forms a specific beam and transmit the data using this beam. The shape and direction of a beam is based on the filter in use. Four different beams are formed using same set of physical antennas with the use of different spatial filters. Using same spatial filter forms the same beam with same direction, shape and power. Thus, if two different physical antennas transmit signals using same spatial filter, UE will consider them travelling over a single channel. Overall transmission will then be assumed from a single antenna port and same for the two signals. Similarly, same physical antenna can transmit two signals with different spatial filters which are unknown to the receiving device. UE will consider them as travelling over different channels. Overall transmission will then be assumed from two different antenna ports. Figure 2.11 shows the stage where multi-antenna pre-coding is done for data mapping . This is carried using Spatial filters.
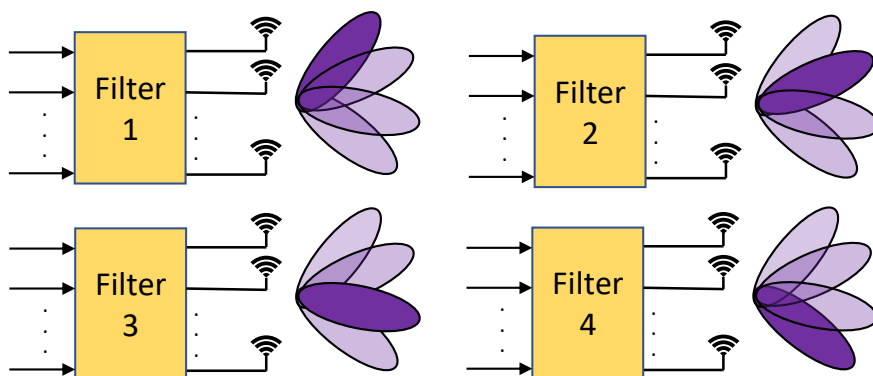

Figure 2.5: Antenna-port structure in NR

In the case of downlink, each antenna port also corresponds to a specific reference signal. Detailed channel sounding information related to the antenna port can be taken by the UE using these reference signals such as CSI-RS and Synchronisation Signal Block (SSB). This feature is also exploited for the thesis. Figure 2.6 shows two different spatial filters for different CSI-RS. gNB maps two different CSI-RS such that they are beam-formed in different directions. UE will see them as two CSI-RS transmitted over two different channels. In reality as shown in the figure, they are are transmitted from the same physical antennas and propagating via same physical channels.
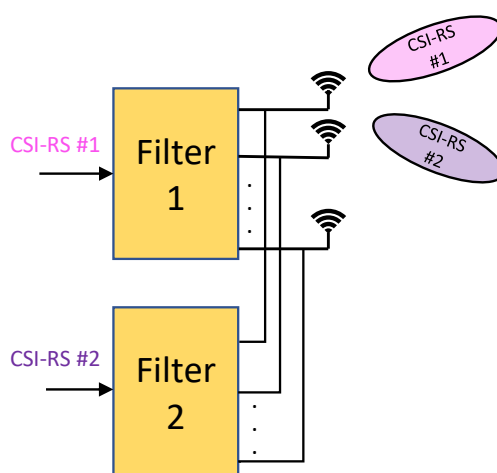
Figure 2.6: CSI-RS Antenna-port mapping

Large scale multi-path properties such as Doppler shift, average delay spread and gain experienced for two different channels relating to different transmitting signals from two different antenna ports may have common properties. In such cases, the antenna ports said to be in Quasi Co-location (QCL). This knowledge can be used by the receiving UE for setting parameters for channel estimation in NR multi-antenna scheme. If two signals have spatial QCL, they are assumed to be transmitted from the same place and in the same beam. Thus, UE can assume a receiver beam direction to be best for multiple signals if it verified for initial signal. For example, initially UE can confirm that a receiver beam direction is best for reference signals like CSI-RS used for channel sounding. QCL principle can then be applied and the same beam direction can be assumed to be ideal for data reception using downlink data channels. This feature is very important and also explored in the thesis.

2.5 Dynamic Duplex scheme

Frequency spectrum of operation is the deciding factor for the kind of duplex scheme to be used. NR supports paired allocations for FR1 lower-frequency bands using Frequency-Division Duplex (FDD) and unpaired spectrum allocations at FR2 higher-frequency bands using Time-Division Duplex (TDD) as illustrated in figure 2.7. This is achieved using one common frame structure for both schemes. This is different from LTE where two different structures are used. This frame structure supports half-duplex operations like TDD where simultaneous transmissions are not possible and full-duplex operation like FDD where simultaneous transmissions are possible.
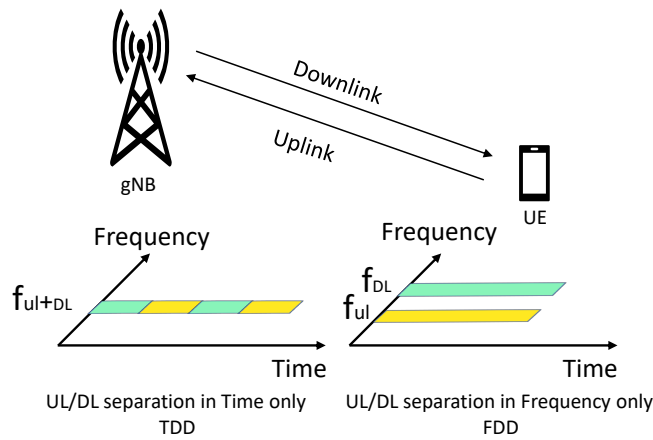
Figure 2.7: Duplex schemes

2.5.1 Dynamic TDD pattern

NR supports dynamic Physical Layer (PHY) TDD pattern. The procedure of setting up the TDD pattern was not standardized by 3GPP and was left for different vendors to regulate individually. The TDD pattern is also changed according to the network and traffic requirements. In TDD scheme, uplink and downlink transmissions occur at different time instants but use the same carrier frequency. It is the scheme used at higher frequencies and also the area of focus since the thesis is in NR high-band. TDD features same uplink-downlink allocation for all cells sharing same carrier frequency. The working of dynamic TDD allows scheduler to dynamically allocate uplink and downlink resources according to traffic variations.

The thesis was implemented in the current TDD pattern of 4+1 which means 4 slots for downlink and 1 slot for uplink as depicted in figures 2.12 and 2.13. The TDD pattern is divided into slots and each slots contain 14 OFDM symbols used for sending data, reference signals and acknowledgements. Typically, around 75 percent of the slots in entire TDD pattern are dedicated for downlink symbols, approximately 23 percent are used for uplink symbols and the remaining for gap symbols. This is covered in detail in a section below.

2.6   Multiple antenna transmission

Beam-forming and management

Use for a large number of antennas at higher frequencies is done to extend coverage area. This technique is called beamforming. Beamforming is of two different types, analog and digital. For higher frequencies, analog beam forming is used. Beams are shaped after digital-to-analog conversion in this technique. However, these transmissions can only be done in one particular direction at a given instant as it is done on carrier basis. To tackle this, beam-sweeping is used where multiple narrow transit beams are used to reach the coverage area. The same signal is repeated using OFDM symbols on each beam. Similarly the receiver beam-forming can only be done in one direction at a time.

Beam correspondence, as defined by 3GPP [17] is the process of assuming reciprocity. Best beam pair used for downlink transmission can turn out to be best for uplink also. Hence, if a suitable downlink beam pair is formed and used, explicit uplink beam management is not required. The same beam pair is assumed to be suitable for uplink. The same applies in the reverse order.

Radio products are customized to have a fixed number of wide beams. These wide beams further comprise of narrow beams which is also fixed. The thesis explores these beams for beam management.

Parts of beam-management

ABF beam-management comprises of following three steps:

- Beam establishment: This is the initial cell synchronization step which leads to establishment of a connection between gNB and UE.

- Beam adjustment: This is the refinement step that helps selection of the best beam from the set of available beams.

- Beam recovery: This is the recovery step in case of link failure and loss of connection between gNB and UE.

The thesis covers initial beam establishment and is primarily focused on beam adjustment in downlink. We will limit our discussions to these two. The process involves three steps called P1,P2 and P3 [15]. These processes are for downlink beam management in connected state for UE to have better downlink beam or data reception.

Beam establishment

This consists of selection of best beam pair for uplink and downlink transmission. Initially, SSB reference signal is transmitted on all wide beams by gNB in downlink transmission. UE measures all SSB RSRP values. UE then selects the best wide beam based on best SSB power value and reports it to gNB. After selection of best wide beam, UE needs to find the best narrow beam mapped to the current serving wide beam. gNB transmits CSI-RS reference signals on all narrow beams. Similar to earlier process, UE measures all CSI-RS Layer 1(L1)-RSRP values and reports

to gNB. Thus, gNB assigns the best narrow beam based on best power value to the UE. This initial random access transmission in downlink enables selection of best gNB transmitted beam or UE receive beam. This is called P1 process as shown in figure 2.8.
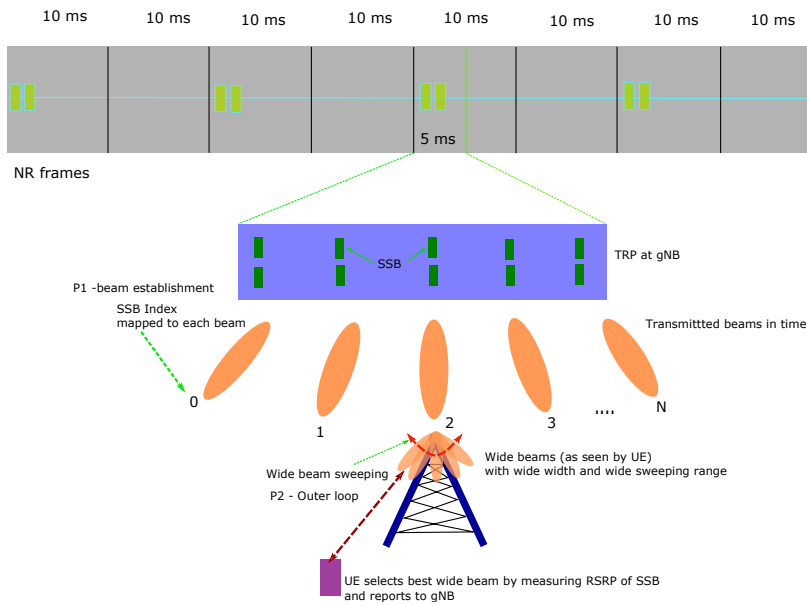

Figure 2.8: Beam establishment and refinement - P1 and Outer loop P2

After an established connection, the same SSB is further used by cell and UE for data communication in both uplink and downlink respectively. This is done using QCL principle stated earlier. It is not mandatory for the UE to be aware of download beam-forming and the specific beam used at the BS for transmission. However, NR is implemented with the feature of informing the UE by gNB about using the same beam for data transmission which was earlier used for referencing and initial beam establishment. This feature is called beam indication and is implemented using Transmission Configuration Indication (TCI) states. TCI includes reference signal information together with downlink data control channel information.

Beam adjustment/refinement

After successful beam establishment, beam directions are constantly evaluated due to multi-path environment in both transmitter and receiver side. Further, beam refining is also done by reducing initial wider beam width used for initial establishment to narrow. The concept of reciprocity is used for beam refinement also. This leads to adjustments done for a beam pair in either downlink or uplink.

Downlink beam adjustment

Downlink beam adjustment aims to find or adjust the best beam pair for downlink transmission. It is done for both wide beams and narrow beams. It should be noted that at this stage, the device is now in connected state and is receiving downlink data transmitted by gNB. Beam adjustment can be done in two ways:

P2 process

P2 process involves UE measurement on different TRP transmitted beams to possibly change current serving beam. It is done using two loop approach and is the focus of the thesis:

- Outer loop approach: This process is for refining and selecting best wide beam. After P1 process, SSBs are beam-swept by UE over available wide beams for cell search and/or synchronisation as shown in figure 2.8. These measurements happen aperiodically and RSRP measurements are triggered. If a better wide beam with better SSB RSRP is reported, gNB transmits CSI-RSs on the all narrow beams mapped to the new best wide beam. UE measures all CSI-RS RSRP values and reports it to gNB. If new best narrow beam is reported, the gNB selects it and starts transmitting to the UE on this beam. Further, QCL info for spatial relation is updated in the UE via Medium Access Control- Control Element (MAC-CE). This is covered in detail in a section below.

- Inner loop approach: This process is for refining and selecting best narrow beam. After outer loop, CSI-RSs are beam-swept by UE over available narrow beams within serving wide beam as shown in figure 2.9. These measurements happen aperiodically and RSRP measurements are triggered. UE measures all CSI-RS RSRP values and reports it to gNB. If new best narrow beam is reported, the gNB selects it and starts transmitting to the UE on this beam. However, QCL info for spatial relation is not updated in the UE.

Thus, if the best receive beam is fixed for the UE, corresponding best transmitted beam by the BS is to be selected. Measurement report quality is of vital importance to beam measurement and will be discussed in detail in a section below.

P3 process

P3 process involves UE measurement on the same fixed TRP transmit beam by changing UE receive beam. Thus, if an best transmit beam is fixed for the BS, then corresponding best receive beam for the UE is to be selected. In contrast to previous method, various reference signals are transmitted from gNB on the same fixed serving beam but on different symbols. The receiving UE also assumes that all different reference signals are transmitted on the same serving beam. Beam sweeping is used on the UE side to sweep through several reference signals. UE measures these signals for best value internally but does not report anything to gNB unlike P1 and P2.
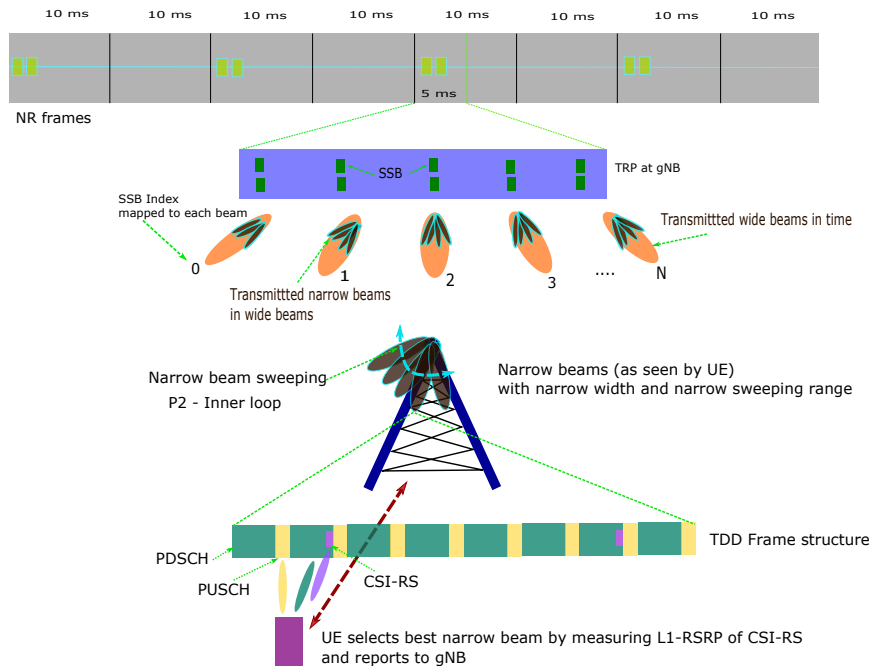
Figure 2.9: Beam refinement - Inner loop P2

Uplink beam adjustment

Uplink beam adjustment aims to find or adjust the best beam pair for uplink transmission. In this case, the device is transmitting uplink data to the BS. As mentioned earlier, beam reciprocity can be used to set up beam pair for uplink using the chosen one for downlink and vice-versa. This topic is not concerned with the thesis and we will limit our discussions to downlink beam adjustment.

Beam recovery

For ABF, the device assists in tracking and selection of an best receive beam for reception of data and control. With multi-antenna configurations with narrow beams in multi-path environment, tracking failure can occur. This happens when the current best serving beam gets blocked for a certain amount of time beyond a set threshold. Beam adjustment can not be used to re-establish the connection in that case. This would result in need for beam-recovery procedures. This consists of beam-failure detection based on reference signal power reports, new candidate-beam identification, recovery-request transmission and network response to the beam-recovery request. This topic is not concerned with the thesis and we will limit our discussions to downlink beam tracking.

2.7 Ultra-lean design

NR RAN can be broadly divided into control-plane and user-plane protocol architecture and the thesis deals with user-plane protocol.

NR RAN user-plane protocol architecture

NR RAN downlink user-plane protocol layers consists of Service Data Application Protocol (SDAP), Packet Data Convergence Protocol (PDCP), Radio-Link Control (RLC), Medium-Access Control (MAC) and Physical Layer (PHY). Control signalling called L1/L2 control is used for downlink and uplink transport channels. As the name suggests, this control signalling is associated with information from the PHY (L1) and MAC (L2). The thesis is also focused on L1/L2 control signals and our discussion will be limited to these only.
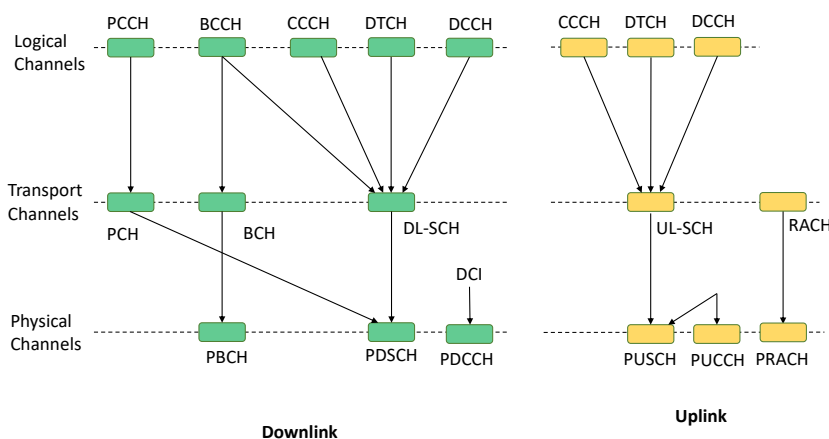
Figure 2.10: Logical, transport, and physical channels mapping

Medium-Access Control scheduling

The MAC layer serves RLC using logical channels [7] as shown in figure 2.10. MAC layer in NR is a vast improvement over LTE using a new header structure. It has different responsibilities such as handling data multiplexing functions for carrier aggregation, logical channel multiplexing, hybrid automatic repeat request (H-ARQ) re-transmissions, handling various numerology and scheduling-related functions. Our area of focus is scheduling.

NR RAN enables dynamic channel share of time-frequency resources among users. This is done using a scheduler which is a part of MAC layer. Scheduling regulations are left by 3GPP for the network providers to decide based on their needs. These scheduling strategies are such that they exploit multi-path channel variations. The gNB scheduler assigns independent uplink and downlink resources to both time and frequency domain and informs the devices about the scheduling decision. The scheduling is done on a dynamic basis and mostly one per slot.

The downlink scheduler dynamically decides the devices to which transmissions will be made. Channel-dependent scheduling in downlink is performed using CSI. This is the baseline mode of operation where CSI is reported to the gNB by the UE. This contains array of vital information like instantaneous downlink channel quality and multi-antenna processing needs in both time and frequency domain. The thesis is focused in CSI transmissions which is explained in detail in a section below.

Physical Layer

The PHY serves the MAC layer using transport channels as illustrated in in figure 2.10. It has different responsibilities such as mapping of transport channels to physical channels, channel-coding, modulation/demodulation schemes etc. as shown in figure 2.11. Our area of focus is multi-antenna processing and mapping of the signal to the appropriate physical time-frequency resources.

A physical channel is defined as a set of time-frequency resources shown in figures 2.12 and 2.13.
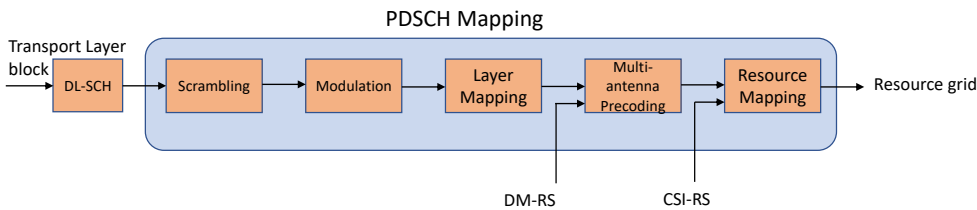

Figure 2.11: PDSCH mapping

Physical channels in NR

Different physical channels as defined by 3GPP [4] for 5G are stated below:

- Physical Downlink Shared Channel (PDSCH): It is the is prime physical channel used for transmission of data, paging, random-access responses and system information. It is the key component of a TDD framework. Figure 2.11 shows mapping of Transport layer data channel to PHY data channel PDSCH. These are finally mapped to PRB. Figures 2.12 and 2.13 show physical channels and reference symbols on TDD 4+1 pattern used in the thesis with 4 downlink PDSCH slots occuring is succession followed by one uplink PUSCH slot. The periodicity or occurrence of PDSCH slots is decided by the vendor and is usually high to facilitate maximum slots for data.

- Physical Uplink Shared Channel (PUSCH): Similar to PDSCH, it is the main physical channel used for transmission of data for uplink. It is limited to at most one per uplink component carrier (CC) per UE. The periodicity is less than PDSCH as more resources are dedicated for downlink than uplink due to traffic requirements as seen in figures 2.12 and 2.13. The slots containing PUSCH occur after PDSCH slots have been transmitted.
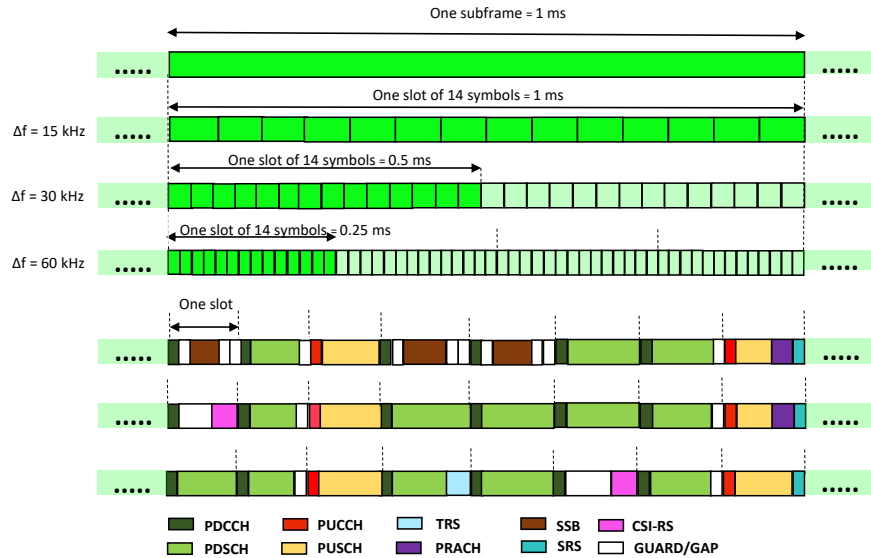
Figure 2.12: TDD 4+1 PHY data and control channel architecture

- Physical Broadcast Channel (PBCH): It has system information needed by UE to access the network.

- Physical Downlink Control Channel (PDCCH): It is used for downlink control information (DCI) like scheduling decisions. DCI is needed to gain scheduling grants for PUSCH transmission and/or for PDSCH reception as shown in figure 2.13. They are the first OFDM symbol in the slot used for PDSCH. Typically they occupy one symbol but sometimes can also consume two consecutive symbols based on configured CORESET.

- Physical Uplink Control Channel (PUCCH): It is used by UE to transmit H-ARQ acknowledgments. ARQ informs the gNB about reception of downlink transport block. It requests the network for resources on which uplink data is transmitted. It also helps downlink channel-dependent scheduling by transmitting channel-state reports as shown in figure 2.13. They are placed between PDSCH slot and following PUSCH slot as can be seen in figure 2.12. Usually few OFDM symbols are left vacant before PUCCH symbols as guard symbols where no transmissions take place. These gurad symbols allow for circuitry in gNB to switch from downlink to uplink, especially since a single gNB is handling multiple UEs. This ensures that uplink and downlink or BSs involved do not interfere with each other. Guard symbols are also used to support timing advance feature. Since distance exists between gNB and UE, it takes some time for transmitted signal to propagate. Due to this propagation delay, certain time gap exists between transmission and reception for downlink. During random access procedure, the gNB communicates to the UE about the timing advance needed in the very first signal and thus is aware of the timing advance used by the UE. Hereafter, UE

sends the required data in alignment with gNB timing advance for uplink (from UE to gNB). This timing alignment is repeated during UE connection lifetime as UE can move closer or away from the gNB.

- Physical Random-Access Channel (PRACH): As the name suggests, it is used for random access control. Their periodicity of occurrence is lower than PUSCH and SRS.

Note: Every transport channel is mapped to a physical channel but the reverse may not be true. For example, PDCCH and PUCCH perform downlink and uplink control information transmission and do not have transport channels mapped to them as shown in figure 2.10.

Reference Signals

Reference signals are predefined signals which occupy certain specific REs in downlink time-frequency resource grid. A drawback of current LTE products is "always-on" cell-specific reference/broadcast signals by eNB. These are used for various purposes like base- station tracking, initial access, system information like channel-state reporting for scheduling, mobility measurements or for downlink channel estimation for coherent demodulation etc. These signals are transmitted in the order of one signal per transmission layer regardless of the fact that downlink data is transmitted or not. The occurrence of these signals take place at an interval of approximately 200 μs which is quite low and eliminates the possibility of reducing power usage by switching the transmitter off. Thus, devices in LTE can expect these cell specific reference signals to be always present. This is because LTE was designed for network traffic with large cells and multiple users per cell, where these always on signals did not cause large impact on overall system performance.

NR is designed for very dense deployments where such "always-on" signals have negative impact as they degrade achievable network energy performance and also data rates by causing interference. The ultra-lean mechanism eliminates such transmissions of cell-specific reference signals. For example, reference signals for demodulation are user-specific in NR and are transmitted only if data gets transmitted and not all the time. This also helps in multi-antenna operations like beamforming along with improving energy performance and reducing interference of the network.

NR reference signals which are used by the UE are as follows:

- Demodulation Reference Signal (DM-RS): It is used for channel estimation to perform coherent demodulation as shown in 2.13. They occur in PDSCH slots and are used by UE for demodulation. They also occur in PUSCH slots and are used gNB. The minimum number of OFDM symbols used for DM-RS is 1 and maximum is 4 [4].The thesis used 1 symbol DM-RS. They are located in the beginning of the slot either occupying the second or third symbol to enable front-loaded design in NR. DM-RS are introduced in multi-antenna precoding section of PDSCH data generation as shown in figure 2.11. In figure 2.12, the slots containing PDSCH and PUSCH also contain DM-RS.
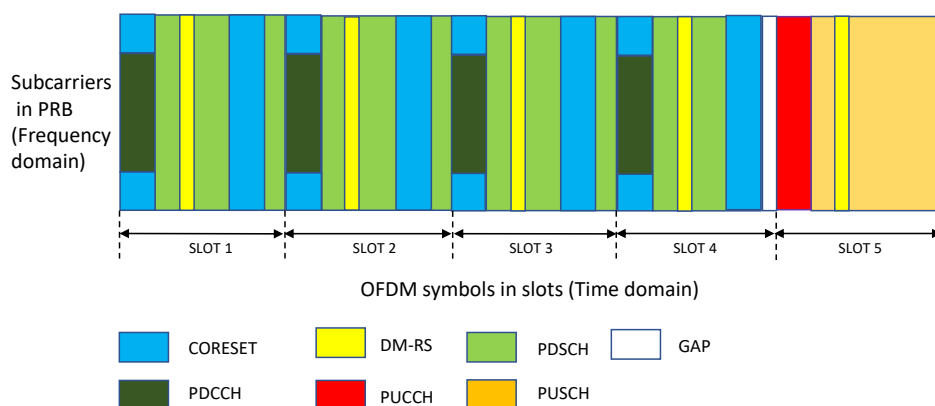
Figure 2.13: Time-frequency resources in TDD 4+1

- SSB: SSBs are used for initial beam establishment and synchronization of P1 process. They are sent in bursts as seen in figure 2.12 and are positioned in slots at the beginning of the TDD pattern. The periodicity of SSB is very low. SSBs consume 4 OFDM symbols of a slot and are positioned in the middle of a slot such that a few guard symbols exists before and after the SSB. These are to prevent interference between SSB and PDSCH symbols. Typically they use slots otherwise dedicated for PDSCH as they are transmitted in downlink. They are covered in a section below.

- Phase-tracking reference signals (PT-RS): It is used for phase-noise reduction and is an extension to DM-RS for PDSCH/PUSCH.

- CSI-RS: It is downlink reference signal used by UE in downlink for CSI. They occur in PDSCH slots and are introduced at the end stage of PDSCH data generation before being mapped to resource grid as can be seen in figure 2.11. It is the main focus of the thesis and covered in detail in the section below.

- Tracking reference signals (TRS): They are formed of a collection of CSI-RS resource sets and used to successfully receive downlink transmissions. Similar to CSI-RS, they also occur in PDSCH slots. The periodicity of TRS slots is very low and they are and are sparse in number. They occur in two slots every TRS period. They are covered in a section below.

- SRS: They are for channel-dependent scheduling in uplink transmission by UE. They perform uplink channel-state estimation and help estimate uplink channel quality at the gNB. They occur in PUSCH slots. They typically consume one OFDM symbol and occupy the last symbol of slot with PUSCH. Their periodicity is less than PUSCH but more than PRACH.

Channel estimation

Detailed knowledge of radio channel characteristics like path loss, amplitude and phase knowledge in all domains and interference level knowledge constitute the base of transmission scheme available today. This knowledge is obtained by measurements on radio link in transmitter or receiver side. UE measurements for downlink channel estimation are reported to gNB for setup of various parameters. Channel reciprocity is often used to assume same characteristics in uplink. Specific signals are used for a UE to perform measurements. This technique is called channel sounding. NR supports channel sounding using CSI-RS for downlink and SRS for uplink [16].

Channel-state-information reference signals and Synchronisation Signal Block

As mentioned earlier, there is a need to reduce "always-on" reference signals for energy conservation. LTE had such signals for channel estimation called cell-specific reference signals (CRS). These were later upgraded to CSI-RS in LTE release 10 [11].

NR also supports SSB reference signal which is "always-on" and is transmitted over a limited bandwidth. The periodicity of SSB is very large. SSB is also used for channel estimation but has low efficiency. This is due to limited bandwidth operation, especially in highly varying channel conditions in time and/or frequency. Thus, CSI-RS from LTE have been further enhanced in NR for CSI acquisition needed or scheduling and LA. These are used in complement with SSB for better channel estimation using RSRP measurements in mobility and beam management scenarios. Unlike SRS for uplink being a general broadcast, NR CSI-RS is configured on per device basis. Keeping this in consideration, SSB and CSI-RS can be transmitted on wide and narrow beams respectively. The benefit of such deployment is increase in robustness to blockage using wide SSB due to their ability to travel in more directions and better Signal to Noise Ratio (SNR) using narrow CSI-RS. It also gives larger range for UE to send and receive data using CSI-RS measurements. UE can detect and decode SSB even in idle mode while it must be in connected state to decode CSI-RS as they are configured in advance by gNB. Figure 2.12 shows different reference signals transmitted in NR TDD structure.

CSI-RS Architecture

NR CSI-RS can relate to up to 32 different antenna ports with each port reflects a single RE of a frequency domain resource block or time domain slot as in figure 2.14. Placing of CSI-RS in resource grid is done such that it does not interfere with Control Resource Set (CORESET), PDSCH demodulation reference signals or SSBs. CSI-RS is also configured in multi-port fashion transmitted orthogonal to each other.
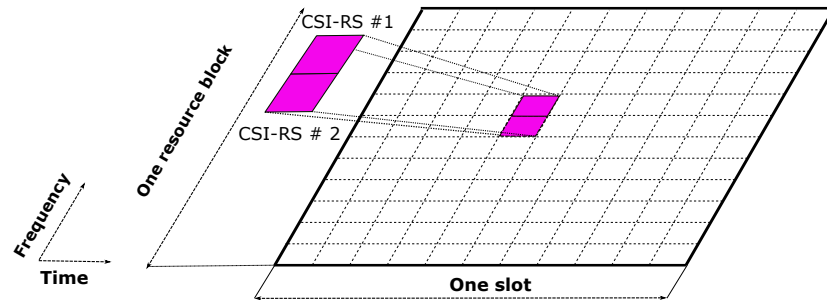
Figure 2.14: Two-port CSI-RS structure based on 2xCDM

Three ways of resource sharing occurs in the event of multi-port CSI-RS transmission [2]:

- Code-domain sharing (CDM): Transmissions occur on same REs but separated by different orthogonal pattern modulation as shown in figure 2.14.

- Frequency-domain sharing (FDM):Transmissions occur on different sub-carriers but within same OFDM symbol.

- Time-domain sharing (TDM): Transmissions occur in different OFDM symbols but within same slot. This is the mode used in the thesis as shown in figure 2.12.

CSI-RS time domain architecture

CSI-RS is configured in groups and not alone called CSI-RS resource set. A CSI-RS resource set can be configured as follows: [2]

- Periodic: Transmissions of configured CSI-RS resource set occur every specified slot. This specified slot can range from 4 to 640.

- Semi-persistent: Transmissions of configured CSI-RS resource set can be activated/ deactivated using MAC CE [7]. After activation, occurrence is similar to periodic case.

- Aperiodic: Transmissions of configured CSI-RS resource set is explicitly triggered by signalling in the DCI. This is the mode used in the thesis.

CSI-RS Types [4]:

- Non-Zero-power CSI-RS (NZP CSI-RS): The discussion so far has been about this type where actual transmission of CSI-RS resource set occurs. UE can perform measurements on this type.

- Zero-power CSI-RS (ZP CSI-RS): In terms of configuration both ZP- CSI-RS and NZP-CSI-RS are same. Using ZP CSI-RS by gNB gurantees that nothing is transmitted on those REs. UE can then assume that any recieved power at those location is due to interference. Thus CSI-RS can also be used for interference measurements, especially for inter-device operations.

Tracking reference signals [2]

CRS in LTE is similar to TRS in NR. However, TRS is better than CRS in reducing overhead with one antenna port and frequency of occurrence in two slots every TRS period. TRS is used to successfully receive downlink transmissions. TRS is a resource set of collection of many periodic NZP CSI-RS with periodicity of 10, 20, 40, or 80 ms respectively as depicted in figure 2.12.

Antenna-port mapping

As mentioned earlier in antenna-port sub-section of section 2.4, each antenna port reflects a specific reference signal use by UE for detailed channel-state information. Expanding the concept, multi-antenna ports reflect multi-port CSI-RS used by UE for channel sounding. However, this mapping is done on a linear spatial filter and not directly to a physical antenna. The UE experiences these channels based on CSI-RS ports and not physical antennas. As shown in figure 2.15, CSI-RS port is mapped to filter F implying that the channel being sounded is based on a CSI-RS not the actual physical radio channel. The number of physical antennas (N) to which the CSI-RS is mapped is usually larger than the number of CSI-RS ports (M). UE performing channel sounding using CSI-RS does not see F or N physical antennas. It only experiences M channel corresponding to the M CSI-RS ports. The thesis explored multi antenna-port mapping including for CSI-RS at gNB.
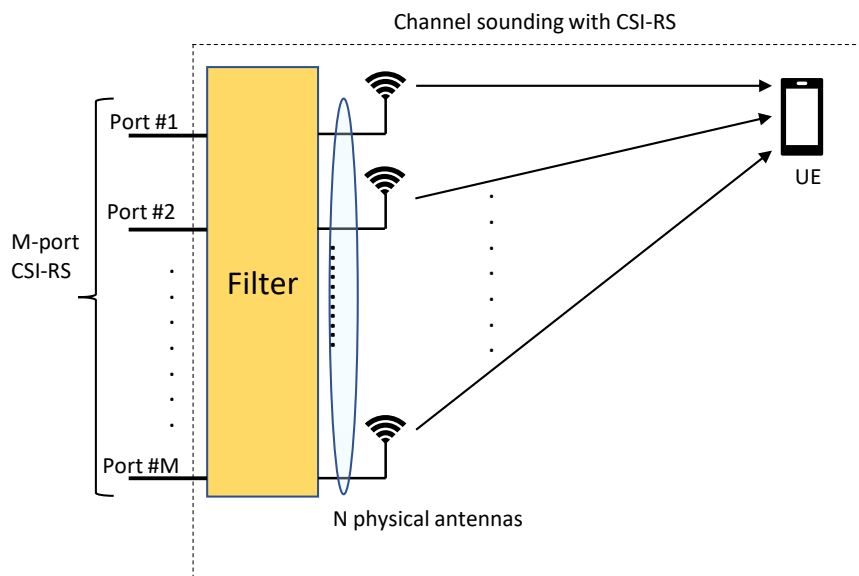

Figure 2.15: Multiple antenna-port CSI-RS mapping

CSI-RS measurements and reports

NR UE does different types of physical-layer measurements and reports to the network using report configuration according to in the 3GPP specifications [16]. A typical report configuration covers set of quantities to be reported, downlink resources for measurements and procedure of actual reporting.

Report content

CSI report consists of Channel Quality Indicator (CQI), CSI-RS Resource Indicator, Layer Indicator (LI), L1-RSRP, Rank Indicator (RI) and Precoder Matrix Indicator (PMI). An important parameter that is stated in the report is RSRP which indicates the received signal strength. L1-RSRP is the key measurement in NR for beam management and is also used for the thesis. In a single report, UE can include measurements for up to four reference signals corresponding to four beams. A typical report includes information about reference signals and corresponding beams, L1-RSRP measurement for the strongest or best beam and L1-RSRP measurement differences between the best beam and three beams below best beam.

Report types

Reporting by UE is as follows:

- Periodic: Reporting occurs every specified period. An example is periodic reporting done on PUCCH.

- Semi-persistent: Reporting can be activated/ deactivated using MAC-CE. After activation, occurrence is similar to periodic case. An example is reporting for semi-persistently allocated PUSCH.

- Aperiodic: Reporting can be explicitly triggered by signalling in the Downlink Control Information (DCI). This method is used for the thesis as the type of CSI-RS triggering used for the thesis was also aperiodic. This is done by reporting for scheduled PUSCH.

Radio Resource Control

Communication between UE and network takes place using radio channel. Various lower layers need to be configured in common for successful exchange of user data. For high end cellular communication such as NR, a generic approach facilitating dynamic change of configurations is required for best results in real time when the actual communication happens. Radio Resource Control (RRC) is a special control mechanism to exchange data on such configurations [16].

RRC control mechanism is prevalent in both UE and gNB. Main functions of RRC cover connection establishment, system information broadcast for connection of device with a cell, radio bearer establishment, RRC connection mobility for connection management and cell re-selection, paging when device is not connected to a cell or system information such as warning messages, beam indication by

assignment of TCI states for PDCCH to each configured CORESET and many more. Signalling functions are used by RRC to configure user and control planes. Radio Resource Management (RRM) strategies are also controlled by RRC.

RRC operates using a state machine. NR device can be in one of three RRC states based on traffic activity, RRC-IDLE, RRC-ACTIVE and RRC-INACTIVE. RRC-INACTIVE is a new state which was not present in LTE while other two were present. These different states correspond to different amounts of radio resources available to UE in it's given state. Thus, the state machine determines the quality of the service (QoS) and energy consumption experienced by UE.

Control Resource Set

A set of physical resources in a specific area on NR Downlink Resource grid is called CORESET [4] where a device can receive PDCCH/DCI. Using CORESET, UE attempts to decode candidate control channels. The size and location of CORE-SET is configured by gNB. It can occur at any position in a slot in time domain and anywhere in the carrier frequency range in frequency domain. It can occupy three OFDM symbols at maximum anywhere in a slot in time domain. In general practice, CORESET OFDM symbols are located in the beginning of the slot followed by DM-RS as scheduling takes place in the very start. This can be seen in figure 2.13. CORESET is smaller than carrier bandwidth in NR due to large bandwidth size. NR has bandwidth upto 400 MHz. A UE may not be capable of receiving such large bandwidth.

The concept of CORESET is novel to NR and was not used in LTE. LTE had control channels called LTE PDCCH spanning over entire carrier bandwidth. It was due to small bandwidth size in LTE. But this proved to be a bottleneck for devices not capable of supporting full carrier bandwidth and also led to frequency domain interference between downlink control channel cells. CORESET defined with respect to the UE indicates where UE can receive PDCCH.

Note: CORESET does not imply that the location is always used for PDCCH. PDCCH can only occur in a CORESET slot but CORESET occurs with a specified periodicity and occurs in slots without PDCCH also as in figure 2.13. Different parameters of CORESET are as follows:

- Resource Element: As defined in earlier sections, it is one sub-carrier in frequency domain and one OFDM symbol in time domain and is the smallest unit of the resource grid as shown in figure 2.3 and 2.4

- Resource Element Group (REG): One REG is one resource block. One resource block is collection of 12 REs in frequency domain and one OFDM symbol in time domain.

- Control Channel Element(CCE): Multiple REGs form a CCE.

- Aggregation Level: Aggregation Level shows number CCEs allocated for one PDCCH as defined in Table 7.3.2.1-1 [4]. Out of every 72 REGs, 18 REs are used for DM-RS and remaining 54 are used for PDCCH. The thesis used aggregation level 8 for CORESET.
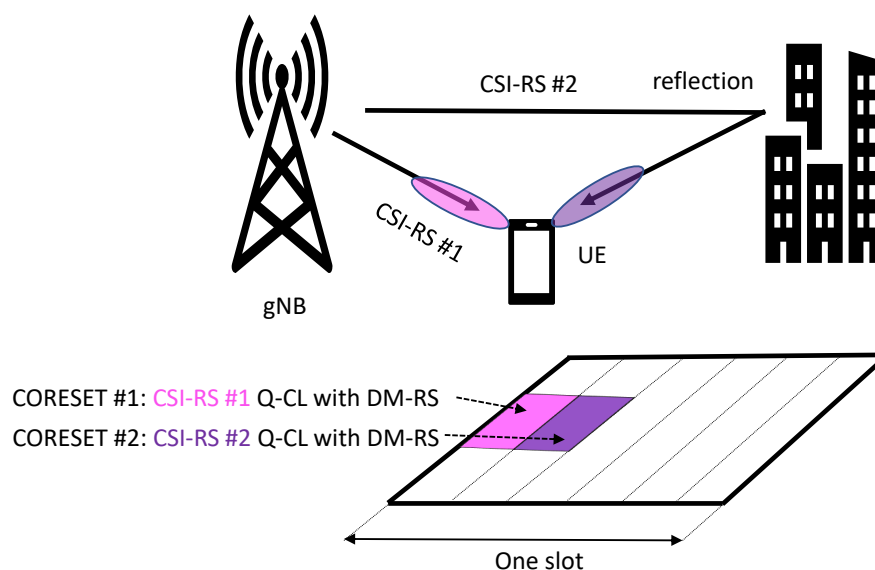
Figure 2.16: CORESET with QCL

Concept of QCL is also applied for reference signals and PDCCH to improved channel sounding by UE. For this, CORESET is configured with a TCI state. If UE is aware that a CORESET and a CSI-RS are spatially QCL, it can decide best receive beam for reception of PDCCH in this CORESET. As illustrated in figure 2.16, two CORESETs are availaible in UE. First CORESET with spatial QCL between DM-RS and CSI-RS 1 and second with spatial QCL between DM-RS and CSI-RS 2. After CSI-RS RSRP measurements, UE knows best receive beam for both CSI-RSes. When UE is supposed to receive PDCCH, it needs to monitor CORESET 1. UE then uses QCL and selects appropriate reception. The same is repeated for CORESET 2. This way, UE can handle multiple reception beams using the blind decoding framework. QCL for CORESET and CSI-RS were explored in the thesis.

Current standards and proposed solution

This chapter explains the present standards and their implementation for beam management and channel sounding related to ABF. The working principles of current products are provided for the reader to get a better understanding of different working ways of beam management to make comparison with the proposed algorithms. Based on the problem description, proposed solutions have been stated for solving the problem. Certain limitations existing with the current products have also been mentioned.

3.1 Baseline solutions

As mentioned in the last chapter, P1, P2 and P3 processes are used for beam management in downlink which is the scope of the thesis. Being more precise, P2 measurement involving selection of best narrow beam among available narrow beams is the area of exploration in this thesis. Channel measurement in the form of CSI-RS measurements are done and reported to the network. Best wide beam and narrow beam is selected based on RSRP values. Each wide beam consists of a fixed number of narrow beams in ABF, 12 in our case. Beam switching takes place from serving wide beam to new best wide beam and from serving narrow beam to new best narrow beam.

Different baseline solutions for beam refinement in NR:

- Inter-narrowbeam selection: This process selects the best narrow beam among the set of all narrow beams available based on the radio product in use. This is done by measuring and comparing RSRP values by UE.

- Inter-widebeam tracking: This is outer loop P2 process covered in Chapter 2.

- Intra-widebeam selection: This is inner loop P2 process covered in Chapter 2. This is carried in combination with outer loop P2 approach and is the focus of the thesis.

- Two narrow beam selection: This process sweeps serving narrow beam and it's two neighbouring narrow beams. In addition, inter wide beam selection tracking mentioned above also repeats aperiodically. This algorithm reduces

the number of narrow beam measurements and overhead greatly but is a sub-standard solution as minimum narrow beam measurements are done.

- No narrow beam tracking: This is an extreme solution that does not perform any measurements for best narrow beam selection and stays on the initial serving narrow beam. In addition, inter wide beam tracking mentioned above also repeats aperiodically. This algorithm reduces the number of narrow beam measurements and overhead completely but is a sub-standard solution as it relies only on wide beam tracking.

- Less narrow beam selection: This process sweeps serving narrow beam and it's neighbouring narrow beams. No inter wide beam selection tracking mentioned above is done. This algorithm reduces the number of narrow beam measurements and overhead but is a sub-standard solution as it does not involve wide beam tracking.

3.2   Proposed algorithms

Each narrow beam RSRP measurement consumes downlink resources for CSI-RS transmission and uplink resources for measurement reports which could otherwise be used for data. CSI-RS overhead grows linearly with the number of connected users. This is a potential problem especially with more users like > 50.

- Increasing measurement period: Increasing the periodicity of triggering CSI-RS measurements to reduce overhead. It method does not require increasing number of trigger states.

- Minimize number of narrow beam measurements: Reducing the number of CSI-RS symbols to measure is another proposed algorithm. This can be tried without sacrificing a loss in beam management procedure of finding the best beam. Based on radio product considered, overhead can be reduced to half by reducing the number of slots used for CSI-RS measurements to half.

Some limitations due to product hardware and specification standard applicable for beam management are stated below:

- Maximum trigger-states: The maximum number of aperiodic trigger states is limited. The selection of the trigger states is done by MAC CE.

- Maximum number of CSI-RS beams to measure in a slot is limited to satisfy QCL.

- Performance: The implementation should be such that beam tracking performance is not sacrificed.

Theoretical calculations

This chapter contains details about various parameters used for the thesis and results from theoretical calculations. The results are provided for the reader to get better understanding of expected or ideal throughput values obtained on varying periodicity for channel estimation measurements and to make comparison with the results from simulations using the proposed algorithms.

4.1 Transport Block Size (TBS)

Transport block is the payload for physical layer. It is mostly calculated for downlink, uplink or special sub-frame in TDD pattern. We have calculated TBS for shared channels PDSCH and PUSCH. The size of the transport block is calculated using several parameters. Successful reception of a downlink transmission depends on knowledge about various parameters. These are resource blocks, modulation scheme, transport block size and redundancy version. Information about some of these parameters is provided by 5-bit MCS field I_{MCS} and the redundancy version field (rv) in the DCI to the UE. I_{MCS} for NR is decided by UE based on three look-up tables 5.1.3.1-1, 5.1.3.1-2 and 5.1.3.1-3 provided by 3GPP [2]. The table used depends on whether 256 Quadrature Amplitude Multiplexing (QAM) is configured or 64QAM is configured. LTE also uses these look-up tables to determine MCS. It also tabulates transport block sizes with respect to MCS field. This helps in resource-block allocation. However, NR supports large range of transport block sizes which are dynamic in nature. This is because of large bandwidths used in NR. NR also supports different transmission duration and variations of the reference overhead signals. Hence, together they would result in many tables required to cover all this which would also require modifications with changing parameters. Thus, a formula-based approach combined with a table for the smallest transport-block sizes is used in NR for flexibility. TBS is determined based on number of layers and PRB. It is determined separately for PDSCH and PUSCH. This formula is complex and is based on certain derived parameters that in turn are further based on certain derived parameters as shown in figure 4.1. These have been covered precisely in the sections below. For more detailed reading, readers are advised to refer to the specification document covering these calculations by 3GPP [2].
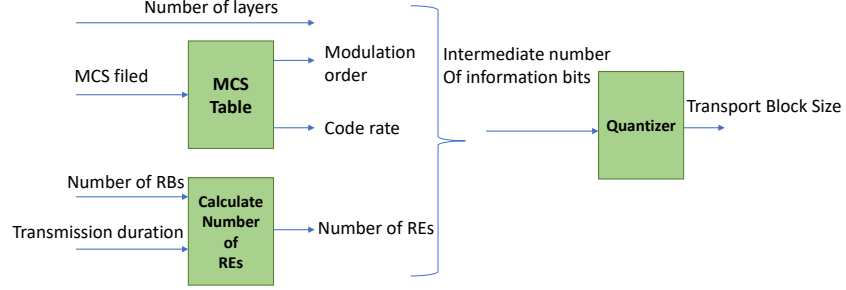
Figure 4.1: NR TBS calculation

TBS formula for PDSCH

Calculation of information about PDSCH

$$N_{\text{info}} = N_{\text{Re}} \cdot R \cdot Q_m \cdot v \tag{4.1}$$

Different values in the equation 4.1 are as follows:

- N_{info}: Intermediate number of information bits, used for calculation of TBS.
- R: Target Code rate, which can extracted from table using MCS index.
- Q_m: Modulation order, which can extracted from table using MCS index.
- v: Number of transmission layers
- N_{Re}: Number of Resource Elements determined by UE

N_{Re} is further defined as the following:

$$N_{\text{Re}} = \min(156, N'_{\text{Re}}) \cdot n_{\text{PRB}} \tag{4.2}$$

Different values of equation 4.2 are as follows:

- n_{PRB}: Total number of allocated PRBs for the UE.
- N'_{Re}: Number of REs allocated for PDSCH within a PRB.

N'_{Re} is further defined as the following:

$$N'_{\text{Re}} = N_{\text{SC}}^{\text{RB}} \cdot N_{\text{sh}}^{\text{symb}} \cdot N_{\text{DMRS}}^{\text{PRB}} \cdot N_{\text{oh}}^{\text{PRB}} \tag{4.3}$$

Different values of equation 4.3 are as follows:

- $N_{\text{SC}}^{\text{RB}}$: 12 and is number of sub-carriers in a PRB
- $N_{\text{sh}}^{\text{symb}}$: This is the number of scheduled OFDM symbols of the PDSCH allocation within the slot (maximum value as 12 or 14).
- $N_{\text{DRMS}}^{\text{PRB}}$: Number of REs for DM-RS per PRB in the scheduled duration including the overhead of the DM-RS CDM groups without data.

- N_{oh}^{PRB}: The overhead configured by a higher layer parameter in PDSCH Serving Cell Configuration.

Upon calculation of N_{info}, value of TBS is determined using different cases. Overall structure has been mentioned below. However, covering all types of cases in detail is beyond the scope of the report.

- TBS for $N_{info} \leq 3824$

$$N'_{info} = \max \left[24, 2^n \cdot \left\lfloor \frac{N_{info}}{2^n} \right\rfloor \right] \tag{4.4}$$

Different values of equation 4.4 are as follows:

- $N'info$ = Quantized intermediate number of information bits

$$n = \max(3, \lfloor \log_2(N_info) \rfloor - 6) \tag{4.5}$$

Table 5.1.3.2-1 by 3GPP [2] is used to find the closest TBS that is not less than $N'info$.

- TBS for $N_{info} > 3824$

$$N'_{info} = \max \left[3840, 2^n \times \text{round} \left[\frac{N_{info} - 24}{2^n} \right] \right] \tag{4.6}$$

Different values of equation 4.6 are as follows:

- N'_{info}: Quantized intermediate number of information bits

$$n = \lfloor \log_2(N_{info} - 24) \rfloor \tag{4.7}$$

- round: The ties in the round function are broken towards the next largest integer according to a very complex chain of algorithms mentioned in 3GPP [2].

TBS for PUSCH is calculated using series of more complex algorithms. The thesis is focused on downlink throughput increment. TBS for uplink is beyond the scope of the thesis and we will limit our discussions to downlink.

4.2   Maximum throughput calculation

The approximate downlink and uplink data rates (Mbps) for a given number of aggregated carriers in a band or band combination is computed in NR as follows [3]:

$$DR = 10^{-6} \cdot \sum_{j=1}^{J} \left(v_{Layers}^{(j)} \cdot Q_m^{(j)} \cdot f^{(j)} R_{max} \cdot \frac{N_{PRB}^{BW(j),\mu} \cdot 12}{T_s^\mu} \cdot \left(1 - OH^{(j)}\right) \right) \tag{4.8}$$

Different values of equation 4.8 are determined by different high-layer parameters. Readers are advised to go through specification sheet by 3GPP [2] for details. Functions of different values are as follows:

- J = Component carrier (CC) index

- $v_{Layers}^{(j)}$ = Maximum number of supported transmission layers for uplink and downlink

- $Q_m^{(j)}$ = Maximum supported modulation order for uplink and downlink

- $f^{(j)}$ = Scaling factor which can take up values 1, 0.8, 0.75 and 0.4

- R_{max} = 948/1024

- μ = Numerology [4]

- T_s^{μ} = Average OFDM symbol duration in a subframe for numerology. It is calculated as $\frac{10^{-3}}{14.2^{\mu}}$ assuming normal cyclic prefix

- $N_{PRB}^{BW(j),\mu}$ = Maximum RB allocation in maximum supported UE bandwidth BW(j) and numerology μ [5]. The term $\frac{N_{PRB}^{BW(j),\mu} \cdot 12}{T_s^{\mu}}$ can also be written as product of bandwidth BW(j) and spectral efficiency S_u^j

- OH = Overhead and takes values of 0.14 for frequency range FR1 for downlink, 0.18 for frequency range FR2 for downlink, 0.08 for frequency range FR1 for uplink and 0.10 for frequency range FR2 for uplink.

The above can also be summarized and calculated using TBS for EUTRA in case of MR-DC. Our implementation in NR high-band NSA is such a case which uses MR-DC. The approximate data rate for a given number of aggregated carriers in a band or band combination is computed as follows [3]:

$$\text{Data rate (Mbps)} = 10^{-6} \cdot \sum_{j=1}^{J} \text{TBS}_j \qquad (4.9)$$

Different values of equation 4.9 are

- J: EUTRA CC index in MR-DC band combination

- TBS_j: Total maximum number of downlink-SCH transport block bits received within a 1 ms Transmission time interval (TTI) for j-th CC [8]. This is based on the supported maximum MIMO layers for the j-th carrier by UE. It is also based on the modulation order and number of PRBs based on the bandwidth of the j-th carrier.

4.3   Algorithm 1: Increasing periodicity of CSI-RS

Following section describes the calculation of throughput with varying P2 periodicity:

4.3.1   Various parameters used:

- TDD: Pattern 4+1

- SSB periodicity: Fixed for current products

- Number of connected UEs: Varied with values of 1, 5, 10, 20, 50 and 100

- P2 CSI-RS period per active UE: Varied with values of 20 ms, 40ms, 80ms, 160ms, 320ms and 640ms

- Slots: PUSCH and PDSCH slots available in the TDD 4+1 pattern are counted containing symbols with CSI-RS for P2 and LA, PUCCH and gap

- Bandwidth of operation: 100 MHz

- CC: Throughput is calculated per CC (J = 1 in (4.9))

- DM-RS: TBS is calculated for 1 DM-RS

- TBS: Different TBS pertaining to PDSCH and PUSCH slots are calculated using equations (4.1)-(4.7)

- Block Error Rate (BLER): It defined as ratio of the number of erroneous transmissions to the total number of transmissions. The outputs are multiplied with a factor of 2×10^{-2} is used to account for re-transmissions.

4.3.2   Theoretical results

Equation (4.9) is used to calculate the maximum data rate for downlink. The obtained data rates are tabulated below. When number of users is 100, the value for throughput for baseline periodicity of 40 ms is negative. This means that all TDD slots will be consumed for channel sounding and no resources will be left for data. Decreasing periodicity below baseline to 20 ms has a negative impact on data rates and the effect is more profound when the number of users become 50 and more.

The results show that increasing periodicity of CSI-RS for a fixed number of users increases overall system throughput. The results also show decrease in throughput with increase in users. This is intuitive as more resources are used for channel measurements with increase in number of users. But even with more users, increasing periodicity has as positive effect on throughput indicating that this algorithm is effective for P2 process. In fact, the increase is 0.88% for 1 user while it goes up to 95.81% for 50 users when compared to baseline. We will verify the same in simulations and make comparisons.

Users = 1

Periodicity	Throughput
20	526.8
40	531.8
80	534.8
160	535.5
320	536.2
640	536.5

Users = 5

Periodicity	Throughput
20	483.3
40	508.4
80	521.0
160	527.2
320	530.4
640	532.0

Users = 10

Periodicity	Throughput
20	429.0
40	479.2
80	504.3
160	516.9
320	523.1
640	526.3

Users = 20

Periodicity	Throughput
20	320.4
40	420.8
80	471.0
160	496.1
320	508.7
640	514.9

Users = 50

Periodicity	Throughput
20	-5.4
40	245.6
80	371.1
160	433.8
320	465.2
640	480.9

Users = 75

Periodicity	Throughput
20	-276.9
40	99.6
80	287.8
160	381.9
320	429.0
640	452.5

Users = 100

Periodicity	Throughput
20	-584.4
40	-46.4
80	204.6
160	330.1
320	392.8
640	424.2

Table 4.1: Throughput (Mbps) by varying P2 periodicity (ms) for different number of users*

*Negative data rates indicate all TDD slots consumed for channel sounding and no resources left for data.

4.4   Algorithm 2: Decreasing CSI-RS measurements

The following section describes the calculation of expected gain when reducing 12 beams to 6 beams:

4.4.1   Various parameters used:

- TDD: Pattern 4+1

- Periodicity for measurements: 20 ms or 160 slots

- maximum number of P2 occasions in 20 ms: Defined as 32 as mentioned in CSI-RS architecture sub-section but is limited currently to 27. We assume this upper bound is hit due to lot of inactive UE

- Slots: PUSCH and PDSCH slots available in the TDD 4+1 pattern are counted containing symbols with CSI-RS for P2 and LA, PUCCH and gap

- Bandwidth of operation: 100 MHz

- CC: Throughput is calculated per CC (J = 1 in (4.9))

- DM-RS: TBS is calculated for 1 DM-RS

4.4.2   Theoretical results

First, calculation of total number of PDSCH symbols was done. With 12-beam CSI-RS, 12*27 PDSCH symbols are consumed for channel sounding and remaining are left for data. Calculating the same with proposed algorithm of 6 beams decreases to 6*27. Thus, overall increase in downlink resources with new algorithm is 13.73%. In general, every OFDM symbol saved leads to gain in 2.29% in downlink resources. We will verify the same with simulations.

Methodology

This chapter contains details of simulation environment, different parameters used, various use cases and the process followed to obtain results in an Ericsson simulator.

5.1 Proposed algorithms

Two different solutions are proposed to mitigate the problem of overhead in P2 beam management procedure. These have been discussed in an earlier chapter in detail. These are as follows:

- Increasing measurement period of P2 process

- Minimizing number of narrow beam measurements of P2 process

5.2 Simulator overview

This section gives an overall picture of the simulation environment used for testing the effectiveness of the proposed algorithms. A very powerful simulator developed internally by Ericsson has been used for the thesis. It provides ideal state of the art 5G NR environment and supports multiple 5G features. Especially, beam-forming and beam management procedures mimicking the current products are provided in the simulator which were very helpful in the thesis. Iterating a certain parameter of interest over many values was also facilitated by the simulator. A simulation seed encompasses all possible random configurations of parameters related to the simulation. Randomness of the seed influence the surroundings in which the solutions are tested. Thus, simulations which are run with multiple seeds add credibility to the results as in the case of the thesis. In addition to iterating the simulations over multiple seeds, they were also iterated over multiple simulations. Thus, results obtained from multiple seeds as well as multiple simulations strengthened their trustworthiness. The simulations were run with existing baseline algorithm of beam management and also with proposed solutions for comparison. These were run on powerful servers provided by Ericsson minimizing the processing time. The simulator has several interfaces to run simulation on. Two such user interfaces were used to run simulations as follows:

Main user interface

It is the central user interface available to run simulations. It gives the functionality of choosing the simulator file along with associated log file defined with parameters for which results are to be stored. Thus, multiple jobs can be submitted simultaneously using this interface. One job refers to running a simulation with one specific set of parameters. Hence, instead of changing the parameters every time after a successful simulation, this interface helps a user define all iterations in the simulator file and then submit the file for a run. Numerous output log files were generated by the simulator storing various results from different use cases. It also generates multiple MATLAB files associated with various seeds and simulations storing results which are post-processed later. This was the primary interface for running simulations and storing results.

Graphical user interface

As the name suggests, it is the visual graphical user interface available to run simulations. It also gives the functionality of choosing the simulator file. However, unlike main interface, multiple jobs can not be submitted together in this interface. One specific set of parameters are run each time and are changed every time after a successful simulation. The interface has several user window planes to configure parameters and view results. It has a simulator setup plane window showing all possible set of parameters to change. It also has a logging window to choose different log values such as Cell downlink (DL) throughput, File transfer Protocol (FTP) throughput, RSRP values and so on. When a simulation run begins, different logs can be viewed in the result window console. Also, these results can be viewed as plots in charts console. It also gives the functionality of saving the changes made here in new simulator files. This was also heavily used along with main interface to visualize different logs and see the results in plots.

5.3   Traffic models

Three different traffic models have been explored in the simulator. These models have been curtailed to match real time data traffic. For achieving better results, real time data collected from three operators from different parts of the world at different times has been used. Distribution of downloaded user data, uploaded user data and session length assuming an RRC inactivity time of 10s was used to fine-tune the models.

Full buffer traffic model

This model is close to the ideal situation where the traffic has full buffer. The cell always serves certain number of UE and the packets are always available with gNB to send to connected UE. So at every TTI, there is data to transmit. On successful download of a package, a new package is always generated. The deployment is interference limited. UEs always keep downloading files and the gNB is always transmitting. This model can be assumed to be like User Datagram

Protocol (UDP) which is suitable for cases where error checking and correction is not done and overhead from such processing is avoided. In this model, we have a fixed number of UE always connected to the system. In case a UE dies out due to inactivity, a new UE is instantly created thus maintaining the overall head-count. This model was to have a complete understanding of the system behaviour change on implementing the proposed algorithms.

FTP traffic model

This bursty model involves small, medium and large files of upto 100 MB size being downloaded by all UE. The download is requested by the UE. The gNB buffer gets exhausted after successful download and UE need to wait for a certain time before next arrival. The packets for download also arrive at a fixed pre-defined arrival rate. This model can be assumed to be like Transmission Control Protocol (TCP) where active open, re-transmission and error-detection is done. This adds to reliability to network. In this model, we have a fixed number of UEs always connected to the system as for full buffer model. This deployment is also interference limited. The simulations also include incomplete sessions such as those for cell-edge UEs. Excluding such sessions could skew statistics. This type of model is used dominantly for web applications.

Realistic FTP traffic model

This traffic model is a variation of FTP model and more close to the real life scenario. Unlike the other two models, we do not have fixed number of UEs always connected to the system. Number of active UE depend on a factor called intensity. For example, if intensity $= 1$, we have 1 user generated per second. If the simulation runs for 10 simulation seconds, we will have a total of 10 UE. However, not all of them will be active or connected at all times. Hence, number of generated users is a function of time. The final results are a combination of the two FTP models.

5.4   Key Performance Index

To investigate reliability of obtained results from proposed algorithms and comparing them with existing baseline algorithm, Key Performance Index (KPI) was selected. Various parameters such as Bit Error rate (BER) and RSRP values were considered. Finally, best KPI found for the thesis was overall system throughput. However, other parameters not considered as KPI were also studied to increase the understanding of the simulation environment. Two types of throughput used for running simulations are described as following:

Cell throughput

This kind of throughput in the simulator gives the instantaneous value of throughput generated every specified time interval. Typically, the time interval is very simulation second but can be customized to occur more or less frequently. The

values of throughput used in results is mean of several instantaneous throughput values taken over several simulation seconds. This is further classified into Cell downlink (DL) throughput and Cell uplink (UL) throughput for downlink and uplink respectively. This throughput is used for full buffer traffic model.

Note: The measurements taken are in simulation seconds. This is not same as real time seconds. Simulation seconds takes all processes for the simulation into account and is usually larger than real time. Henceforth, any mention of time in seconds reflects simulation seconds and not real time seconds.

FTP throughput

This kind of throughput in the simulator gives the value of throughput generated per UE basis at time intervals which can not be customized. It is calculated based on the time taken for the entire file to get downloaded or uploaded. This throughput is used for FTP traffic model. It is more reliable for FTP UE as compared to cell throughput. Cell throughput is generated every simulation second. Thus, if a UE completes a download at 0.2 seconds and another one does so at 0.5 seconds, the throughput value displayed for both of them will be same. It may also be possible that since a packet is arriving from an upper layer, the instantaneous value of Cell throughput may be zero at an instant. FTP throughput is further classified into FTP DL throughput and FTP UL throughput for downlink and uplink respectively.
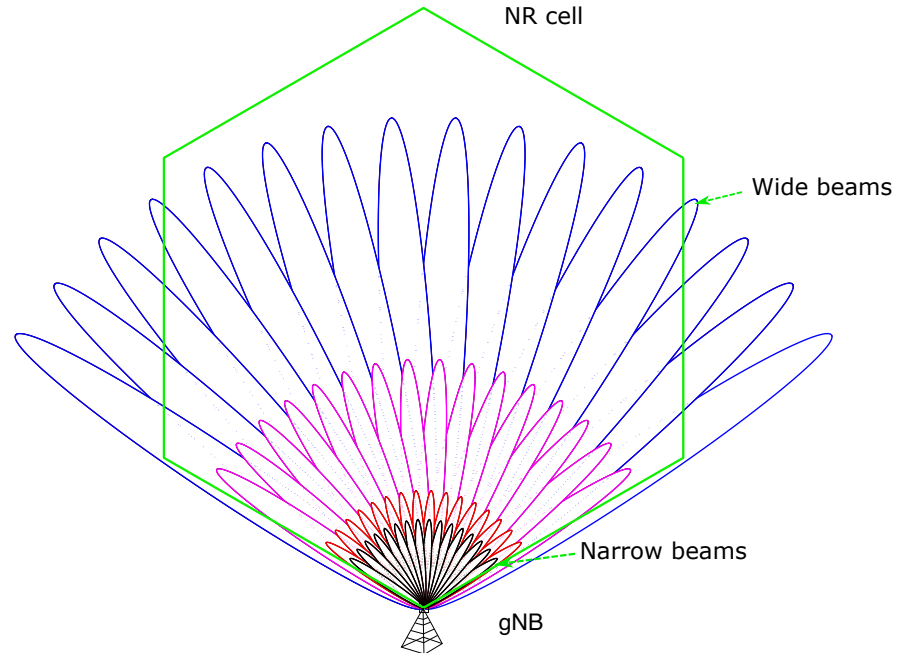

Figure 5.1: Simulation layout with one cell and one gNB

5.5   Simulator parameters

Various parameters used for running simulations are as follows:

- Simulation time: Variations of 5s, 10s and 20s

- Seeds: Combinations of 10, 25, 50 and 100

- Carrier bandwidth: 100 MHz with

- Resource block: 66 PRBs and 8 CCE CORESET Aggregation 8 [4]

- Carrier frequency: 39 GHz high-band

- Deployment scenario: 1 gNB with 1 cell as shown in figure 5.1

- Cell radius: 100 m as shown in figure 5.2

- Numerology: μ=3 (120 KHz OFDM sub-carrier spacing)

- Processing speed: 8000 slots/second corresponding to μ=3.

- TDD pattern: 4+1

- Radio link: Multi-path slow fading multi-user propagation model with Additive White Gaussian noise (AWGN) was used. This design is based on Extended Pedestrian A model (EPA) [10]

- Periodicity of CSI-RS: Different periodicity values were used to see effect on throughput values. However, the simulator is not designed take variations of periodicity. Instead, it counted variations in term of TDD slots used for CSI-RS measurements. Therefore, variations of slots amounting to variations of periodicity was used. Variation used were 160, 320, 640, 1280, 2560 and 5120 slots reflecting 20 ms, 40 ms, 80 ms, 160 ms, 320 ms and 640 ms in periodicity respectively. Henceforth, we will refer to periodicity in ms for ease of representation and discussion.

- Re-transmissions: Maximum number of H-ARQ DL and UL re-transmissions was set to 7

- Antenna deployment:

  - Type: MACRO MU-MIMO using muti-port antenna mapping

  - Dimensions: Length 0.5 m and width 0.5 m [9].

  - Height: 23 m in accordance with urban macro cell (3D-UMa) with high UE density [9].

  - Antenna tilt: Measured by zenith angle and set to 14°. Antenna zenith angle in the simulator is defined as antenna tilt towards the ground and applies to both narrow and wide beams. Thus, an angle tilt of zero degrees reflects horizontal antenna.

  - Antenna elements: 128 antenna elements divided in 8 rows were used. Antenna elements are individual antennas used to beamform the signal.

- Antenna ports: 2 transceiver ports per antenna which also function as CSI-RS ports
- Sector: Geographical sector of width 120 m and height 30 m
- Wide beams:
    * Horizontal wide beams: 3
    * Vertical wide beams: 4
- Narrow beams:
    * Horizontal narrow beams: 34
    * Vertical narrow beams: 4

- UE:
    - Number of UE:
        * Fixed UE numbers: 1, 5, 10, 20, 30, 40, 50, 75, 100
        * Intensity: 0.5, 1, 5, 10, 20
    - Height: 1.5 m
    - UE antenna elements: 2 elements per UE with 2 transceiver ports
    - UE speed: Different UE speeds were used to test the effect of the algorithms proposed. The simulator is designed to take measurements in meter per second (m/s). The values used were 0, 1.38, 2.78, 13.89 and 27.78 m/s reflecting 0, 5, 10, 50 and 100 kilometer per hour (km/hr). Henceforth, we will refer to speeds in km/hr for ease of representation and discussion.
    - UE power class: Class 3 of maximum cell downlink power 23 dBm (with tolerance of ± 2) for non-CA configuration and maximum uplink power of 0.25 dBm was used [5]
    - Inactivity timing: UE was supposed to be inactive if it did not transmit or receive anything for 10s or more.
    - UE types:
        · DL FTP UE: single file transfers for download
        · UL UE: continuous file transfer for upload
        · Short UE: single transmission
        · Chatty UE: chatty traffic
          Overall number of UE is divided in these 4 types based on data collected from 3 different operators. Chatty and Short UE are used for impact of background light traffic, which represent the majority of RRC sessions.
    - Movement direction:
        · Straight mover UE
        · Origination of UE at any position in the cell is determined by the randomness of the simulation seed
        · Continues to move until the end of the simulation

· Moves straight in any direction until they reach a circle that
  circumscribes the hexagonal simulation area
· Hereafter, bounce back and continue straight movement in a
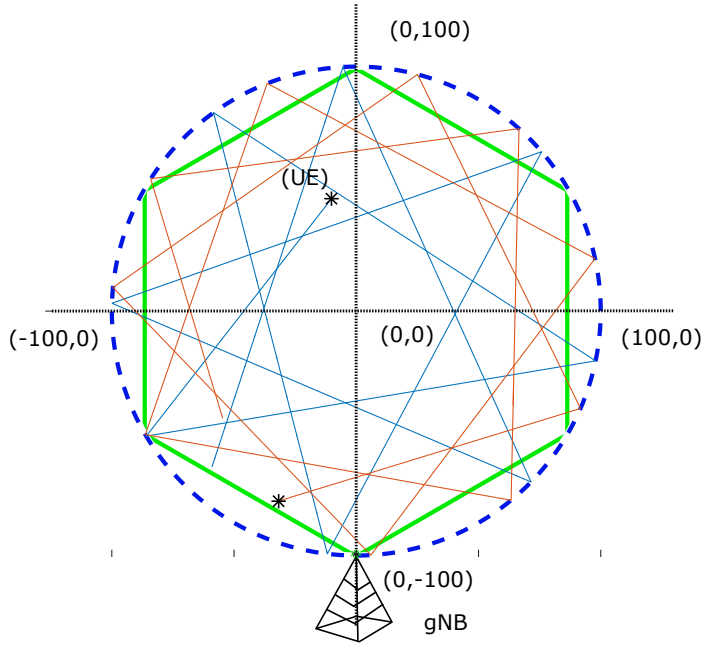 new direction as shown in figure 5.2.

Figure 5.2: Straight mover

5.6   Simulation procedure

In the simulation procedure, UEs are generated according to environment pattern
in a hexagonal cell belonging to a gNB. The distance between hexagon's centre to
its six corners is 100 m and is called radius of the cell. The BS along with wide
beams containing narrow beams has been shown in figure 5.2. Beamforming is
done in MU-MIMO layers and beams overlap each other as shown in the figure
5.1. UE tries to establish a connection with the BS through initial access by beam
sweeping over SSBs transmitted using P1 process. This results in connection of
UE to best wide beam by measuring RSRP values. After P1, UE establishes
itself with the best narrow beam of wide beam by P2 process. The beam tracking
algorithm ensures stationary or moving UE is always connected to the best narrow
beam by measuring and comparing RSRP values through CSI-RS measurements.
Different RSRP values for serving beam measurements were recorded to look at
the effectiveness of the algorithms. The periodicity of L1 SSB RSRP measurement
and CSI-RS RSRP measurement was aligned to be same at 20 ms. To reduce this
overhead, two proposals are mentioned. These were simulated as follows:

5.6.1   Algorithm 1: Increasing periodicity of CSI-RS

CSI-RS measurement periodicity at baseline was simulated and then varied for all combinations mentioned in section 5.5. Further, the speed of UE was also varied for all combinations mentioned in section 5.5. These simulations were further classified in the three traffic models:

- Full buffer: For idealistic full buffer model, the number of UE was varied as 1, 5, 10, 20, 50 and 100. Cell DL and UL throughput was recorded in log files.

- FTP model: For FTP model, the number of UE was varied as 1, 5, 10, 20, 50, 75 and 100. FTP DL and UL throughput was recorded in log files.

- Realistic FTP model: For realistic FTP model, the number of UE was varied using intensity parameter as mentioned in section 5.5. FTP DL and UL throughput was recorded in log files.

5.6.2   Algorithm 2: Decreasing CSI-RS measurements

12 narrow beams exists in a wide beam. CSI-RS measurements take place for all 12 narrow beams in the serving wide beam. This was already simulated in algorithm 1 and then changed to 6 beams in this algorithm. The periodicity for CSI-RS was set to baseline. Further, the speed of the UEs was also varied for all combinations mentioned in section 5.5. Single periodicity was used for these simulations. These simulations were further classified in two traffic models:

- Full buffer: For idealistic full buffer model, the number of UEs was varied as 5, 10, 20, 30, 40, 50, 75 and 100. No simulation was run for a single UE as it would not be enough to gain reliability of this solution. The number of combinations of UE tested was also increased to understand system behaviour in a better way. Cell DL throughput and cell UL throughput were recorded in log files.

- FTP model: For FTP model, the number of UE was varied similar to full buffer model. FTP DL throughput and FTP UL throughput were recorded in log files. No simulations were run for realistic FTP model as that model does not have fixed number of users and would not suit this algorithm.

Results

This chapter presents output of the algorithm behaviour in different scenarios. The results are also compared to theoretical outputs and baseline solutions.

6.1 Theoretical result for periodicity increase

Figure 6.1 shows the theoretical results for different number of stationary users. This is also represented with absolute values in table 4.1. The results show increasing periodicity has as positive effect on DL throughput at 0 km/hr. Also, there is decrease in throughput with increase in users. However decreasing periodicity below baseline to 20 ms case has a negative impact on data rates especially with increase in number of users beyond 50. Similarly, values less than 20 ms were also tried and the results were found to be even worse. Those results have been excluded from the report.
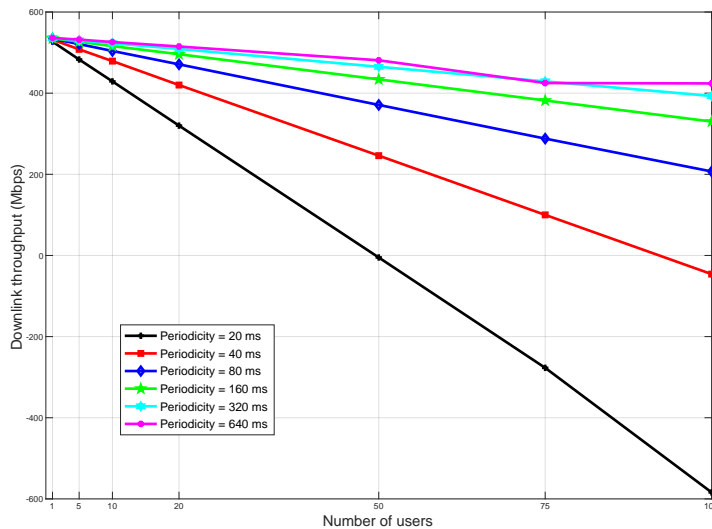
Figure 6.1: Theoretical DL throughput varying P2 periodicity. Negative values show all resources consumed for channel sounding.

6.2   Algorithm 1: Increasing periodicity of CSI-RS

6.2.1   Traffic Model: Full buffer
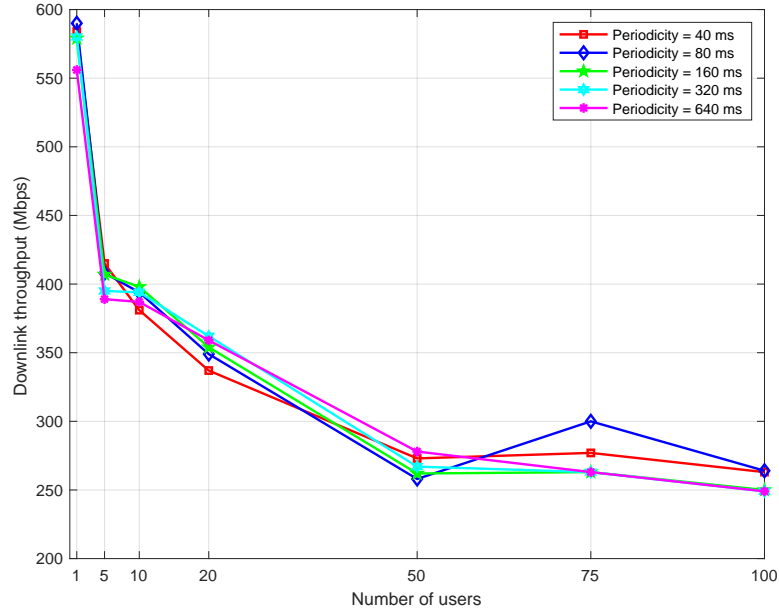
Figure 6.2: Average Cell DL throughput with varying P2 periodicity
(5km/hr)

Figure 6.2 shows the DL throughput for different number of pedestrian users at 5km/hr. When compared to figure 6.1, we find that the behaviour is similar with better data rates with increase in periodicity. However, comparing absolute values shows that the result is full buffer is mostly lower than theoretical case but sometimes higher, especially with less users. This can be due to the fact that interference comes into play in simulator which is absent while making theoretical calculations. Also, higher throughput may be present for full buffer than theoretical due to effect of multi-path adding constructively thus giving better data rates.

Figure 6.3 shows behaviour of periodicity change at speed of 50km/hr. Here, we again find that effect of periodicity change is better for all number of users.

Similarly, results for all other use cases have been included in the Addendum section for readers to have a complete idea of system behaviour. In full buffer model, cell DL throughput increases with the increase in periodicity. Overall data rates decrease as the number of users increase. However, the gain is less when number of UE increases beyond 50. Also, the effect of increasing periodicity is more profound when the speed of UE increases.
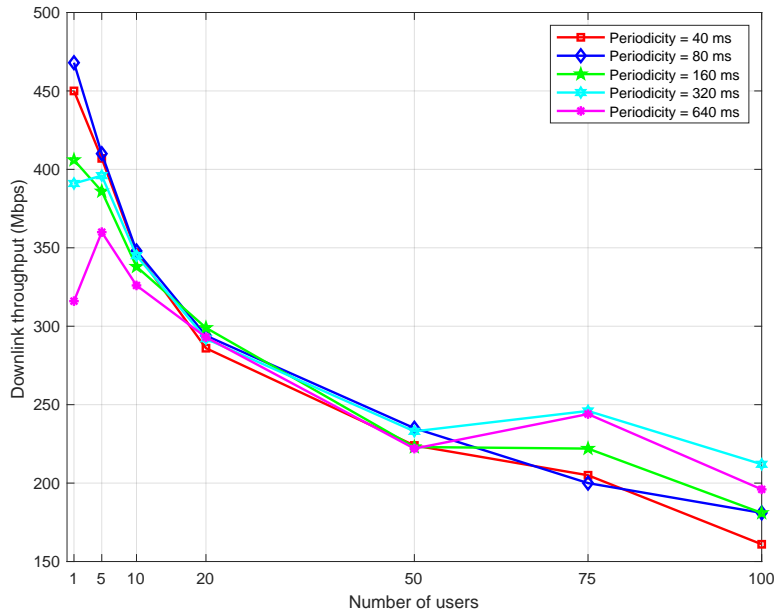
Figure 6.3: Average Cell DL throughput with varying P2 periodicity
(50km/hr)

6.2.2 Traffic Model: FTP

Figure 6.4 shows the results for different number of pedestrian users at 5km/hr.
It can be seen that increasing periodicity has positive effect on rise in FTP DL
throughput. Decreasing periodicity to 20 ms is found to be a bad choice as the
throughput values fall drastically as was seen in theoretical results too.

Figure 6.5 shows results for different number of high-speed users at 100km/hr.
An example of such a case in reality can be when travelling on a train where many
users may be connected to a single cell from a single gNB. It can be seen that
increasing periodicity has positive effect on rise in FTP DL throughput for all
users.

Similarly, results for all other use cases have been included in the Addendum
section for readers to have a complete idea of system behaviour. In FTP buffer
model, FTP DL throughput increases with the increase in periodicity. Overall
data rates decrease as the number of users increase. Unlike full buffer model, the
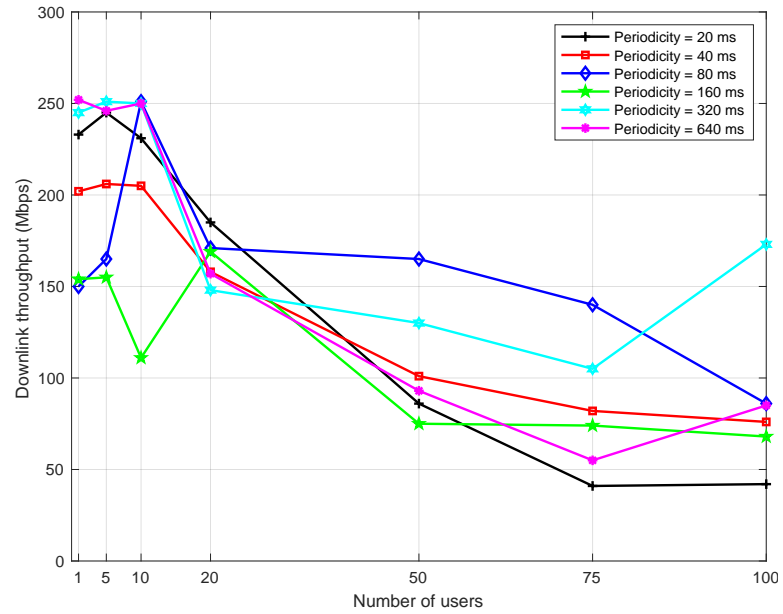gain is good even when number of UE increases beyond 50.

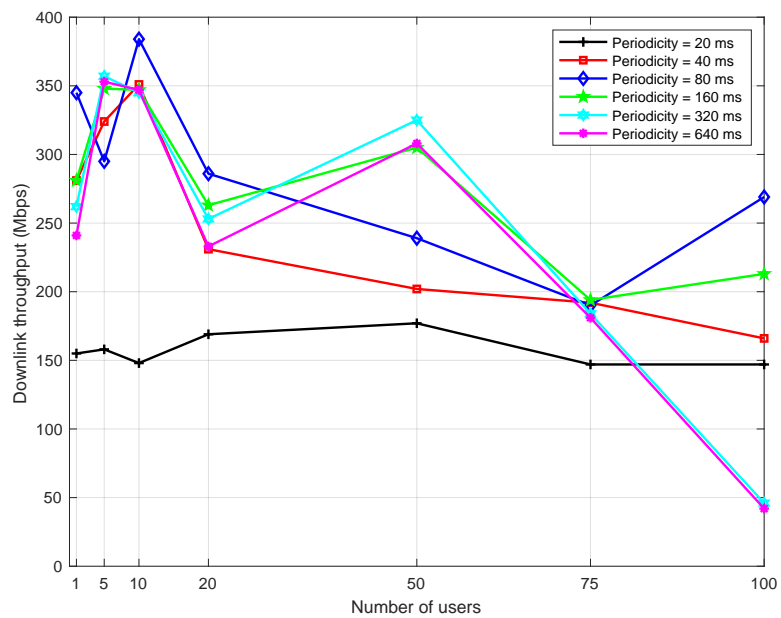**Figure 6.4:** FTP DL throughput with varying P2 periodicity (5km/hr)

Figure 6.5: FTP DL throughput with varying P2 periodicity (100km/hr)

6.2.3   Conclusion

Using results from FTP and full buffer model, it can be clearly seen that increasing periodicity has a positive effect on data rates. Decreasing periodicity below baseline was not found to be a good approach. For full buffer model, that gain in throughput is more for users less than 50 while it is good for all user cases in FTP model. This may be because of the fact that FTP model is more realistic than full buffer. Also, the effect of increasing periodicity is more profound when the speed of UE increases. Sometimes, the system throughput behaves abruptly as can be seen in the Addendum section. This may be due to the fact that FTP throughput accounts for file transfers which may be effected by some higher-layer effects. Relying on absolute results may be harmful and thus focus is based on the larger picture of overall system behaviour. Since real time traffic is more close to FTP, we can conclude that this algorithm is good for decreasing CSI-RS overhead.

The simulator shows good results in most use cases for even very high periodicity values like 640 ms. This may not translate into reality and UE may start dropping out due to less channel sounding. The simulator takes interference as well as fading into consideration. However, since the simulation is limited to one cell, most of the communication between UE and gNB is via Line of Sight (LOS). For Non-Line of Sight (NLOS) propagation in fading dips, increasing periodicity to a very large value may be harmful. This hypothesis can only be verified when the algorithm is put to practice in real-time measurements which could not be done for the thesis. Overall, with more than 50 UE, 80 ms and 160 ms are found to be good choices for periodicity.

6.3   Algorithm 2: Decreasing CSI-RS measurements
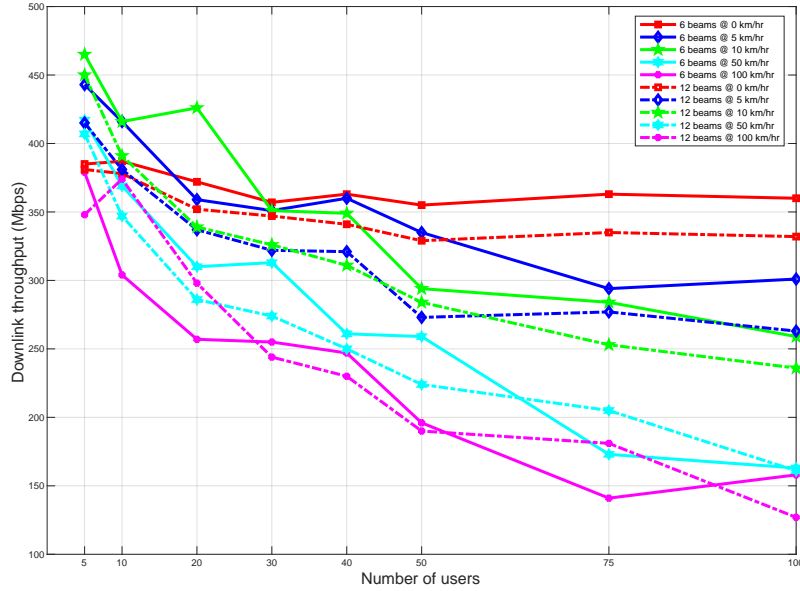
6.3.1   Traffic Model: Full buffer

Figure 6.6: Cell DL throughput comparison between 6 beams and
12 beams with varying speed

Figure 6.6 shows comparison between cell DL throughput values for 6 beams
and 12 beams CSI-RS measurements for different UE speed in full buffer model.
The periodicity has been fixed to baseline. As seen, the downlink throughput gain
is visible for 6 beams when compared to 12 beams. There is an exception of speed
= 100km/hr where degradation occurs for some number of users. This may be
due to the high speed of users which might be unfavourable for migration from
12 to 6 beams. Also, the gain % achieved is approx 9-25 % which is in range as
theoretical calculation if 13.73 % (Chapter 7, section 4.5) and even more. The
rise in downlink gain can be explained from the fact that theoretical gain was
calculated based on an assumption that maximum UE is inactive leading to upper
bound on maximum CSI-RS measurements of 27 being achieved. This leads to
more data PDSCH slots being used for CSI-RS instead. This upper bound may
not always be true which can result in higher downlink gain as in these cases. The
same can be seen in figure 6.7. Based on full buffer model, this algorithm seems
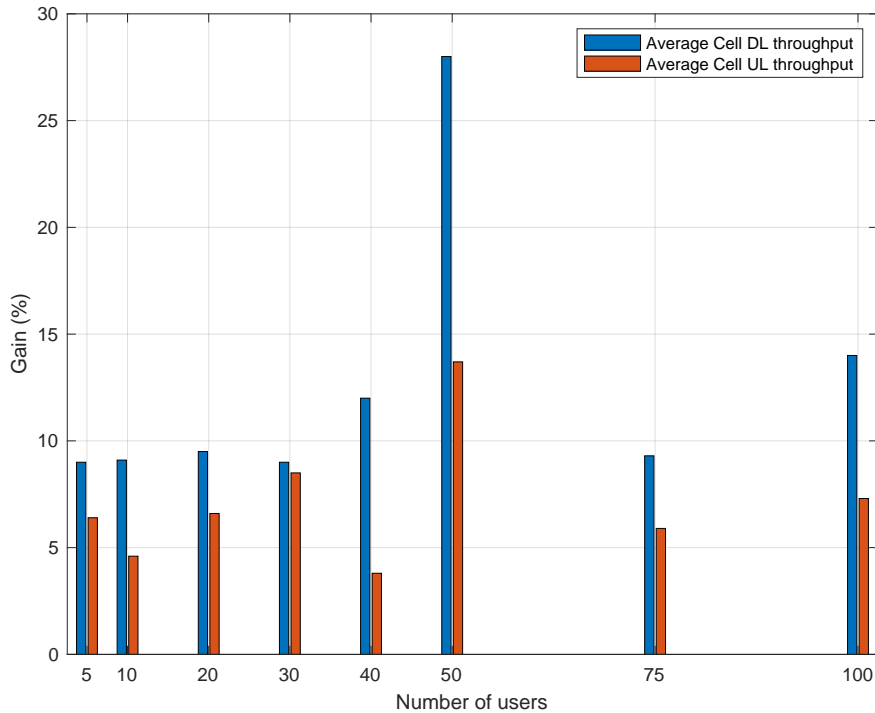to produce good results.

Figure 6.7: Cell DL and UL throughput gain (5km/hr)

6.3.2   Traffic Model: FTP

Figure 6.8 shows comparison between FTP DL throughput values for 6 beams and 12 beams CSI-RS measurements for different UE speed in FTP model. As seen, the throughput gain is visible for some cases of 6 beams when compared to 12 beams. However, for some cases, decrease in throughput is observed. For few high speed cases, no FTP DL throughput values were obtained. Even in case of the gain, only 2.3 % approx is achieved which is way less than expected theoretical calculation of 13.37% as in section 4.5 and as obtained for full buffer model. The same can be seen in figure 6.10. To understand this difference, uplink throughput was investigated.

Figure 6.9 shows comparison between FTP UL throughput values for 6 beams and 12 beams CSI-RS measurements for different UE speed in FTP model. As seen, the throughput gain is visible for 6 beams when compared to 12 beams for some cases. However, for most cases, decrease in throughput is observed especially for high speed UE. Gain % achieved can be seen in figure 6.10.
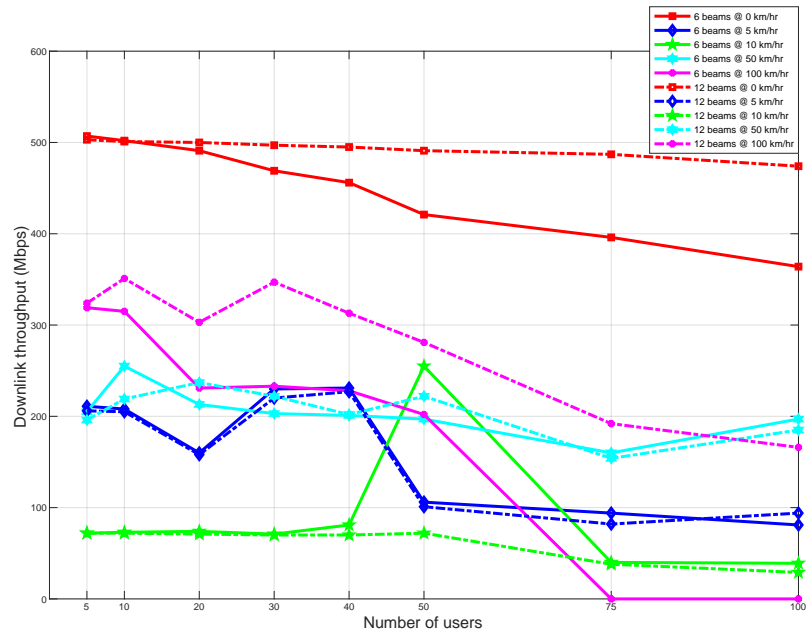
Figure 6.8: FTP DL throughput comparison between 6 beams and 12 beams with varying speed
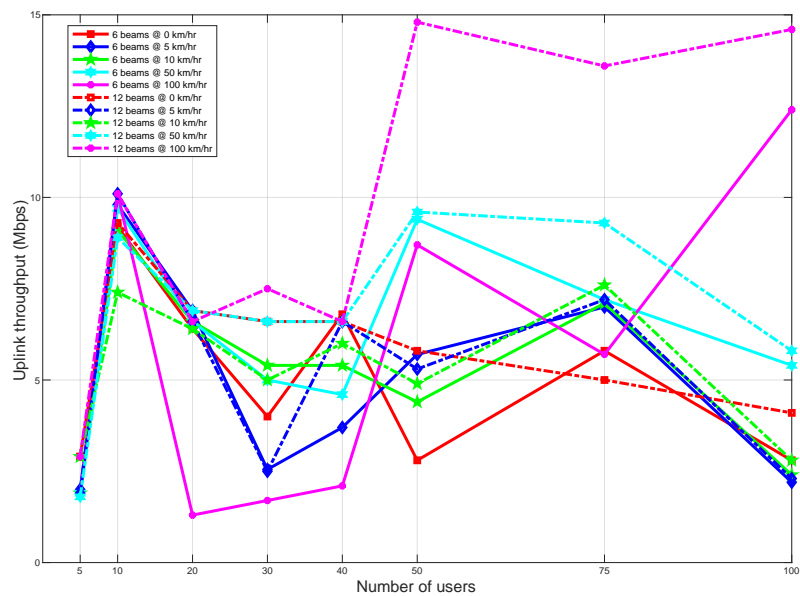

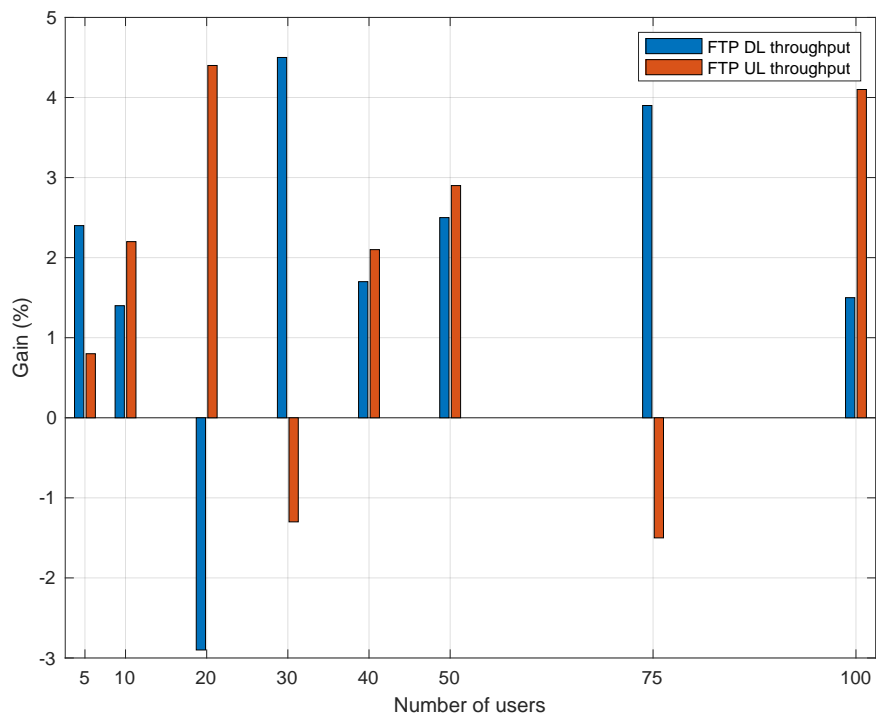**Figure 6.9:** FTP UL throughput comparison between 6 beams and 12 beams with varying speed

Figure 6.10: FTP DL and UL throughput gain (5km/hr)

6.3.3   Conclusion

Using results from full buffer model, it can be seen that decreasing number of CSI-RS measurements has a positive effect on data rates. However, FTP traffic model shows that this algorithm has little gain of DL data rates. The reason for this is degradation UL throughput. As UL data rates go down, TCP acknowledgements that are needed for DL transmissions become very slow. Congestion control is another feature used by TCP to avoid congestive collapse. This is performed using congestion window which limits total number of unacknowledged packets that might be in transit completely. This causes DL incoming rate from higher layers to go down. This is the reason for some cases where no FTP DL throughput is obtained in simulations. In these cases, only FTP UL throughput is obtained. The effect is more profound when speed of UE increases. As mentioned in previous chapter, FTP throughput is generated on per UE basis. Since UE is typecast in DL and UL before simulation run, no FTP DL throughput being generated indicates that DL UE is unable to download data and becomes inactive due to RRC session time-out and only UL UE is active. On an overall, there is decrease in the number of FTP DL throughput occasions for 6 beams when compared to same run with 12 beams. For the present simulated environment, UL is throttled and becomes a bottleneck which then effects DL data rates. Little overhead saving for DL does not lead to large expected gain while system degrades when shifting from 12 beams to 6 beams. Although algorithm 2 has been explored for P2 inner-loop narrow-beam refinement in the simulator, the same can also be extended to P2 outer-loop wide-beam refinement. Using 6 beams instead of 12 for both loops in P2 process is unhealthy.

Although gain exists for full buffer model, real time traffic is more like TCP or QUIC than UDP. Therefore, we conclude that this algorithm is not suitable for decreasing overhead of CSI-RS measurements. Some measures can be taken to amend this behaviour which has been covered in the next chapter.

The results obtained and the conclusion are under specific parameters of the simulated environment. They may or may not translate similarly in real-time runs or other simulated environments.

Future Scope

This chapter contemplates obtained results and provides thoughts on extension of beam-management in future.

7.1   Summary

Reliability of simulated throughput is difficult to testify. Data rates vary a lot depending on number of users, type of users, traffic model, multi-path environment etc. Care was taken to have an array of use cases and models in close relation to real-time data to provide reliability to obtained results.

Looking in retrospection, algorithm 1 of increasing P2 periodicity of CSI-RS measurements was found to be a good idea for saving resources. This algorithm is easy to implement in baseline without making lot of changes in existing product code. Also, the number of aperiodic trigger states for P2 does not increase in this process. Considerable gain in data rates was observed which can be verified from various plots in Chapter 6 and the Addendum section.

Algorithm 2 of decreasing P2 CSI-RS beam measurements on the other hand was not found to be a good idea for saving resources in the simulations. This algorithm was relatively more complex in implementation. The gain in throughput for downlink is negligible if any. Uplink was found to be the prime tailback.

But as George E.P.Box quoted "All models are wrong, but some are useful", key takeaway from algorithm 2 is that more focus in today's communication standards is made on downlink. It is the biggest demand of the market and user downlink speed always takes the center stage. However, this should not lead to overshadowing of uplink. In order for downlink to work effectively, uplink needs to be managed properly. Algorithm 2 could work well if some new balanced duplex pattern is used in future with more focus on uplink and increase in uplink slots like TDD 4+2. Another way to amend this problem could be to implement Machine Learning (ML) algorithm which might record/predict correlated measurements and then decrease number of beams for measurements. Having said that, radio propagation environment is highly volatile and the reliability of such algorithm may be less.

7.2   Future work

As mentioned in Chapter 4, various processes of beam management exists and exploring all was beyond the scope of the thesis due to limitations of time. Such cases may be potentially exploited of the future have been stated in this section.

P1 process of initial beam establishment and P3 process of UE side beam sweeping constitute very important part of beam management which could have similar aspects to explore.

The thesis was implemented in Ericsson's internal simulator. Analysing and predicting real-time output and behaviour of the algorithms still remains unexplored and could present challenges that may lead to further studies.

Also, the simulations are run for one cell with one BS. The scenario of handovers between multiple BSs can be investigated.

One beam management feature is LA measurement which happens on the serving narrow beam twice as often as P2 measurement. A lot of research is on-going in this process and could be possibly explored in future too.

A PDSCH slot containing CSI-RS resources for P2 measurement cannot be used for anything else in current product for simplicity. However, there is on-going work to remove this limitation. This could be a very good area to work and explore.

Extended Kalman Filter (EKF) is an area of gaining popularity, especially for mmWave beam tracking. It has also been applied in new trends of vehicular communication beam tracking for more frequent channel estimations due to use of narrow beams [43]. Vehicular communication is more challenging, particularly because of fast-changing environments. The same could be applied to beam management process leading to better results.

Finally, Machine Learning (ML) is an area of fast growth. It could also be explored for beam management

References

[1] 3rd Generation Partnership Project (3GPP), "Study on New Radio (NR) access technology (Release 15)", 3GPP TR 38.912 V15.0.0, Technical Specification, 2018.

[2] 3rd Generation Partnership Project (3GPP), "Study on New Radio (NR) access technology - Physical layer procedures for data (Release 15)", 3GPP TS 38.214 version 15.3.0, Technical Specification, 10-2018.

[3] 3rd Generation Partnership Project (3GPP), "Study on New Radio (NR) access technology - User Equipment (UE) radio access capabilities (Release 15)", 3GPP TS 38.306 version 15.5.0, Technical Specification, 05-2019.

[4] 3rd Generation Partnership Project (3GPP), "Study on New Radio (NR) access technology - Physical channels and modulation" (Release 15)", 3GPP TS 38.211 version 15.2.0, Technical Specification, 07-2018.

[5] 3rd Generation Partnership Project (3GPP), "Study on New Radio (NR) access technology - User Equipment (UE) radio transmission and reception" (Release 15)", 3GPP TS 38.101 version 15.3.0, Technical Specification, 2018.

[6] 3rd Generation Partnership Project (3GPP), "Study on New Radio (NR) Access Technology - Physical Layer Aspects (Release 14)", 3GPP TR 38.802, 2017.

[7] 3rd Generation Partnership Project (3GPP), "Study on New Radio (NR) Access Technology - Medium Access Control (MAC) protocol specification (Release 15)", 3GPP TR 38.321, 2018.

[8] 3rd Generation Partnership Project (3GPP), "Study on New Radio (NR) access technology - Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures " (Release 13)", 3GPP TS 36.213 version 13.2.0, Technical Specification, 08-2016.

[9] 3rd Generation Partnership Project (3GPP), "Study on 3D channel model for LTE" (Release 12)", 3GPP TR 36.873 version 1.0.0, Technical Specification, 09-2013.

[10] 3rd Generation Partnership Project (3GPP), "Study on LTE - Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) conformance testing" (Release 10),TS 36.141 version 10.8.0 , Technical Specification, 09-2013.

[11] 3rd Generation Partnership Project (3GPP), "Study on LTE - Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures" (Release 10), TS 36.213 version 10.4.0 , Technical Specification, 01-2012.

[12] I. Da Silva et al., "Tight Integration of New 5G Air Interface and LTE to Fulfill 5G Requirements," 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), Glasgow, 2015

[13] 3GPP TS 37.340 v15.0.0, 3rd Generation Partnership Project; NR; Technical Specification Group Radio Access Network; Multi-connectivity;Overall description; Stage-2 (Release 15), January 2018

[14] ITU-R, IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond Recommendation M.2083-0, September 2015. URL `http://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2020/Pages/default.aspx`

[15] 3rd Generation Partnership Project (3GPP), "Beam Management study for New Radio (NR) Access Technology", 3GPP TR 38.802, 6.1.6.1, 2017.

[16] 3rd Generation Partnership Project (3GPP), "Radio Resource Control (RRC) Protocol specification for New Radio (NR) Access Technology", 3GPP TR 38.331, 2018.

[17] 3rd Generation Partnership Project (3GPP), "Study on test methods for New Radio (NR) Access Technology", 3GPP TR 38.810, 09-2018

[18] 3rd Generation Partnership Project (3GPP); "Group Radio Access Network; Study on Scenarios and Requirements for Next Generation Access Technologies"; 3GPP TR 38.913 v14.3.0, (Release 14),August 2017

[19] 3rd Generation Partnership Project (3GPP), "Study on Non-Orthogonal Multiple Access (NOMA) for NR", 3GPP TR 38.812, 2017

[20] Wonil Roh ; Ji-Yun Seol ; Jeongho Park ; Byunghwan Lee ; Jaekon Lee ; Yungsoo Kim ; Jaeweon Cho ; Kyungwhoon Cheun ; Farshid Aryanfar, "Millimeter-wave beam forming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results", 2014.

[21] Robert W. Heath ; Nuria González-Prelcic ; Sundeep Rangan ; Wonil Roh ; Akbar M. Sayeed,"An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems", 2016.

[22] Z. Yang, Z. Ding, P. Fan, and G. K. Karagiannidis, "On the performance of non-orthogonal multiple access systems with partial channel information," IEEE Trans. Commun., vol. 64, no. 2, pp. 654–667, Feb. 2016.

[23] S. Shi, L. Yang, and H. Zhu, "Outage balancing in downlink non-orthogonal multiple access with statistical channel state information," IEEE Trans. Wireless Commun., vol. 15, no. 7, pp. 4718–4731, Jul. 2016.

[24] J. Cui, Z. Ding, and P. Fan, "A novel power allocation scheme under outage constraints in NOMA systems," IEEE Signal Process. Lett., vol. 23, no. 9, pp. 1226–1230, Sep. 2016.

[25] P. Xu, Y. Yuan, Z. Ding, X. Dai, and R. Schober, "On the outage performance of non-orthogonal multiple access with one-bit feedback," IEEE Trans. Wireless Commun., vol. 15, no. 10, pp. 6716–6730, Oct. 2016.

[26] V. F. Monteiro, I. L. da Silva and F. R. P. Cavalcanti,"5G Measurement Adaptation Based on Channel Hardening Occurrence," in IEEE Communications Letters, vol. 23, no. 9, pp. 1598-1602, Sept. 2019, doi: 10.1109/LCOMM.2019.2926268.

[27] Ericsson. "World's first 5G NR radio.", URL https://www.ericsson.com/en/networks/offerings/5g/5g-nr-radio

[28] E. Dahlman, S. Parkvall, J. Sköld, "5G NR: The next generation Wireless Access Technology", Academic press, 2018.

[29] M. Giordani, M. Polese, A. Roy, D. Castor and M. Zorzi, "A Tutorial on Beam Management for 3GPP NR at mmWave Frequencies", IEEE Communications Surveys and Tutorials, vol. 21, no. 1, pp. 173-196, Firstquarter 2019. doi: 10.1109/COMST.2018.2869411

[30] H. Elgendi, M. Mäenpää, T. Levanen, T. Ihalainen, S. Nielsen and M. Valkama, "Interference Measurement Methods in 5G NR: Principles and Performance," 2019 16th International Symposium on Wireless Communication Systems (ISWCS), Oulu, Finland, 2019, pp. 233-238, doi: 10.1109/ISWCS.2019.8877215.

[31] Ericsson "Beamforming, from cell-centric to user-centric.", URL https://www.ericsson.com/en/networks/trending/hot-topics/5g-radio-access/beamforming

[32] S. Han, C. I, Z. Xu and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," in IEEE Communications Magazine, vol. 53, no. 1, pp. 186-194, January 2015, doi: 10.1109/MCOM.2015.7010533.

[33] Alejos, Ana Vázquez, Manuel Garcĺa Sanchez, and Iñigo Cuinas. "Measurement and analysis of propagation mechanisms at 40 GHz: Viability of site shielding forced by obstacles." IEEE Transactions on Vehicular Technology 57.6 (2008): 3369-3380

[34] Ericsson. "Ericsson Mobility Report", URL https://www.ericsson.com/en/mobility-report/reports/november-2019

[35] Ericsson. "Ericsson Mobility Report", URL https://www.ericsson.com/en/mobility-report/reports/june-2020

[36] W. Roh et al., "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results", IEEE Comm. Magazine, vol. 52, no. 2, pp. 106–113, Feb. 2014.

[37] J. Palacios, D. De Donno and J. Widmer, "Tracking mm-Wave channel dynamics: Fast beam training strategies under mobility," IEEE INFOCOM 2017 - IEEE Conference on Computer Communications, Atlanta, GA, 2017, pp. 1-9. doi: 10.1109/INFOCOM.2017.8056991

[38] F. Sohrabi and W. Yu, "Hybrid Digital and Analog Beamforming Design for Large-Scale Antenna Arrays," in IEEE Journal of Selected Topics in Signal Processing, vol. 10, no. 3, pp. 501-513, April 2016. doi: 10.1109/JSTSP.2016.2520912

[39] Ericsson. "Beamforming, from cell-centric to user-centric", URL https://www.ericsson.com/en/networks/trending/hot-topics/ 5g-radio-access/beamforming

[40] E. Peralta et al., "Reference Signal Design for Remote Interference Management in 5G New Radio," 2019 European Conference on Networks and Communications (EuCNC), Valencia, Spain, 2019, pp. 559-564, doi: 10.1109/EuCNC.2019.8802014.

[41] Q. Gao et al., "A Hybrid Channel State Information Feedback Mechanism for Massive MIMO System," 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring), Nanjing, 2016, pp. 1-5, doi: 10.1109/VTCSpring.2016.7504073.

[42] T. Zeng, Y. Chang, M. Hu and Y. Zhang, "CSI-RS Based Joint Grouping and Scheduling Scheme with Limited SRS Resources," 2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Bologna, 2018, pp. 1-6, doi: 10.1109/PIMRC.2018.8580702.

[43] Sina Shaham and Matthew Kokshoorn and Ming Ding and Zihuai Lin and Mahyar Shirvanimoghaddam,"Extended Kalman Filter Beam Tracking for Millimeter Wave Vehicular Communications", 2019, URL https://arxiv. org/abs/1911.01638v1
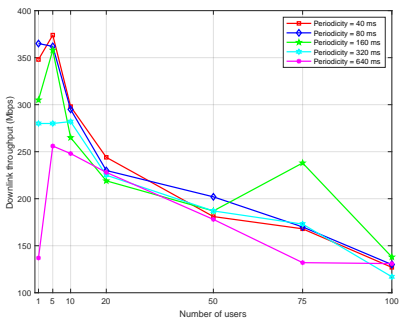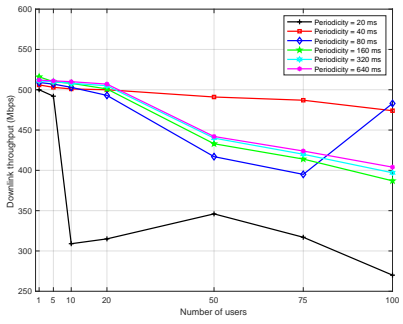
Addendum
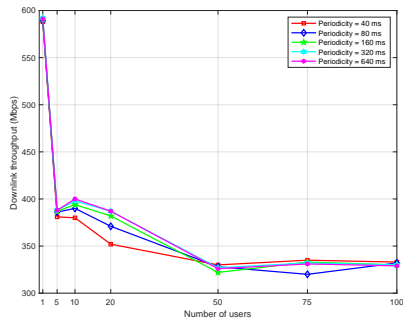
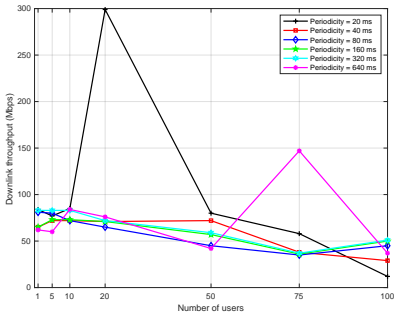Cell DL throughput (0 km/hr)
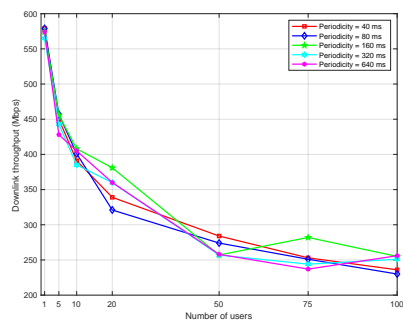
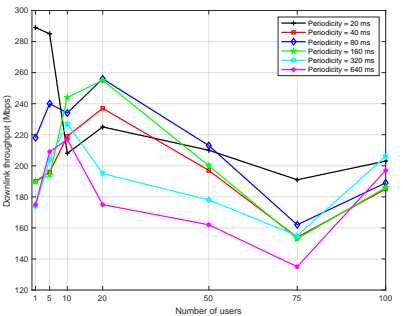

Cell DL throughput (10 km/hr)

Cell DL throughput (100 km/hr)

FTP DL throughput (0 km/hr)

FTP DL throughput (10 km/hr)

FTP DL throughput (50 km/hr)