

# **5G Multi-Tier Handover with Multi-Access Edge Computing: A Deep Learning Approach**

Percy Kapadia

A thesis submitted to Auckland University of Technology in fulfilment  
of the requirements for the degree **Master of Engineering**.



2021

School of Engineering, Computer and Mathematical Sciences  
Auckland University of Technology, Auckland, New Zealand

# Abstract

---

The research presented in this thesis discusses the potential enhancement of 5G multi-tier handover. This proposal will utilise two of 5G's enabling technologies, multi-access edge computing (MEC) and machine learning (ML). MEC and ML techniques are believed to be the primary enablers for enhanced mobile broadband (eMBB) and ultra-reliable and low latency communication (URLLC). The subset of ML that was chosen for this research is deep learning (DL), as it is great at learning long-term dependencies. A variant of artificial neural networks called a long short-term memory (LSTM) network is used in conjunction with a lookup table (LUT), as part of the proposed solution. Subsequently, edge computing virtualisation methods are utilised to reduce handover latency and increase overall throughput of the network. In addition to the proposed, this thesis analyses the validity of various other potential solutions such as multi-connectivity, cloud centralised radio access networks (Cloud C-RAN) and artificial intelligence (AI).

To implement the proposed algorithm, a software simulation of a multi-tier 5G heterogeneous network is developed, based on the 3<sup>rd</sup> generation partnership project (3GPP) standards for: channel models, schedulers, and handovers. This simulator provided the tools for the author to analyse and evaluate the feasibility of the proposed solution.

The results gained from the research was promising. It showed a 40–60% improvement in overall throughput under high user densities. Although the proposed scheme may increase the number of handovers, it is effective in reducing the handover failure (HOF) and Ping-Ping rates in higher user density scenarios by 30%, and 86% respectively, compared to current state-of-the-art. In conclusion, a detailed analysis was undertaken, and the aims of the research were satisfied.

# Table of Contents

---

<b>ABSTRACT</b>	<b>I</b>
<b>TABLE OF CONTENTS</b>	<b>II</b>
<b>LIST OF FIGURES</b>	<b>V</b>
<b>LIST OF TABLES</b>	<b>VII</b>
<b>LIST OF ABBREVIATIONS</b>	<b>VIII</b>
<b>ASSERTATION OF AUTHORSHIP</b>	<b>XII</b>
<b>ACKNOWLEDGEMENTS</b>	<b>XIII</b>
<b>1 INTRODUCTION</b>	<b>1</b>
1.1 PROBLEM STATEMENT AND MOTIVATIONS	5
1.2 THESIS OBJECTIVES, SCOPE AND CONTRIBUTIONS	6
1.3 THESIS STRUCTURE	7
<b>2 BACKGROUND</b>	<b>8</b>
2.1 CELLULAR TECHNOLOGIES AND CONFIGURATIONS	8
2.1.1 RADIO ACCESS NETWORKS	10
2.1.2 BASE STATION TYPES	14
2.1.3 ANTENNAS AND PATHLOSSES	15
2.1.4 FRAME STRUCTURE	16
2.2 HANDOVER	17
2.2.1 3GPP DEFINED LOGIC AND PROCEDURE	18
2.3 MULTI-ACCESS EDGE COMPUTING	22
2.4 DEEP LEARNING	23
<b>3 LITERATURE REVIEW</b>	<b>25</b>
3.1 SINGLE TIER HANDOVERS	25
3.1.1 MULTI-CONNECTIVITY	25

3.1.2	ARTIFICIAL INTELLIGENCE	26
3.1.3	SELF-ORGANISED NETWORKS	27
3.1.4	SUMMARY	29
<b>3.2</b>	<b>MULTI-TIER HANDOVERS</b>	<b>30</b>
3.2.1	RADIO FREQUENCY MAPPING	30
3.2.2	MULTI-CONNECTIVITY	31
3.2.3	MULTIPLE CRITERIA HANDOVER	33
3.2.3.1	Predefined models	33
3.2.3.2	Other models	36
3.2.4	ARTIFICIAL INTELLIGENCE	37
3.2.5	CLOUD C-RAN	39
3.2.6	SUMMARY	40
<b>4</b>	<b>RESEARCH METHODOLOGY</b>	<b>44</b>
<b>4.1</b>	<b>SIMULATION TOOL</b>	<b>44</b>
4.1.1	SYSTEM LEVEL SIMULATOR	45
4.1.1.1	Base stations and users	45
4.1.1.2	Channel models	46
4.1.1.3	Miscellaneous	47
4.1.2	SCHEDULER	48
4.1.3	USER MOBILITY SIMULATOR	50
4.1.4	HANDOVER	51
4.1.5	DEEP LEARNING SIMULATOR	51
<b>4.2</b>	<b>ASSUMPTIONS</b>	<b>53</b>
<b>4.3</b>	<b>SIMULATION ENVIRONMENT</b>	<b>54</b>
<b>4.4</b>	<b>SIMULATOR OPTIMISATIONS</b>	<b>59</b>
4.4.1	SCHEDULER	59
4.4.2	SYSTEM LEVEL SIMULATOR	59
4.4.3	MISCELLANEOUS	60
<b>4.5</b>	<b>PERFORMANCE METRICS</b>	<b>62</b>
4.5.1	DEEP LEARNING METRICS	62
4.5.2	HANDOVER METRICS	62
<b>5</b>	<b>ALGORITHMS</b>	<b>64</b>
<b>5.1</b>	<b>BENCHMARKING ALGORITHM</b>	<b>64</b>
5.1.1	HANDOVER LOGIC	64

5.1.2	HANDOVER PROCEDURE	64
<b>5.2</b>	<b>PROPOSED ALGORITHM</b>	<b>66</b>
5.2.1	HANDOVER LOGIC	69
5.2.2	HANDOVER PROCEDURE	71
<b>6</b>	<b>RESULTS AND DISCUSSION</b>	<b>72</b>
<b>6.1</b>	<b>DEEP LEARNING EVALUATION</b>	<b>72</b>
<b>6.2</b>	<b>HANDOVER EVALUATION</b>	<b>74</b>
6.2.1	TOTAL HANDOVERS	74
6.2.2	MULTI-TIER HANDOVERS	76
6.2.3	PING-PONG HANDOVERS	78
6.2.4	HANDOVER FAILURES	80
6.2.5	AVERAGE HANDOVER LATENCY	82
<b>6.3</b>	<b>THROUGHPUT EVALUATION</b>	<b>84</b>
6.3.1	TOTAL THROUGHPUT	84
6.3.2	AVERAGE USER DATA RATE	85
6.3.3	USER DATA RATE REQUIREMENTS	87
<b>7</b>	<b>CONCLUSION AND FUTURE RESEARCH</b>	<b>89</b>
7.1	FURTHER RESEARCH DIRECTIONS	89
	<b>REFERENCES</b>	<b>91</b>

# List of Figures

---

FIGURE 1.1-1: GLOBAL DATA TRAFFIC IN THE NEXT 5 YEARS (EXABYTES), IMAGE BY ERICSSON [3]	1
FIGURE 1.1-2: COMPARISON BETWEEN 4G (IMT-ADVANCED) AND 5G (IMT-2020), IMAGE BY ITU [5]	2
FIGURE 1.1-3: NETWORK SLICING AND THE ROLE OF MEC © [2017] [7]	2
FIGURE 1.1-4: THREE PILLARS OF 5G AND APPLICATIONS, IMAGE BY ITU [5]	3
FIGURE 1.1-5: DL COMPARED TO NN AND TRADITIONAL ML	4
FIGURE 2.1-1: BASIC ILLUSTRATION OF A CELLULAR NETWORK	9
FIGURE 2.1-2: BASIC ILLUSTRATION OF A D-RAN CONFIGURATION	10
FIGURE 2.1-3: BASIC ILLUSTRATION OF A C-RAN WITH A VIRTUALISED BBU CONFIGURATION	12
FIGURE 2.1-4: BASIC ILLUSTRATION OF FRONTHAUL, MIDHAUL AND BACKHAUL	13
FIGURE 2.1-5: 5G BASIC FRAME STRUCTURE	16
FIGURE 2.2-1: INTERFACES BETWEEN GNBs, AMFs AND UPFs	18
FIGURE 2.2-2: ILLUSTRATION OF ENTRY AND LEAVING CONDITIONS WITH TTT.	19
FIGURE 2.2-3: 5G INTRA RAT HANDOVER PROCEDURE PREPARATION PHASE [15]	20
FIGURE 2.2-4: 5G INTRA RAT HANDOVER PROCEDURE EXECUTION PHASE [15]	21
FIGURE 2.2-5: 5G INTRA RAT HANDOVER PROCEDURE COMPLETION PHASE [15]	21
FIGURE 2.4-1: SIGMOID ACTIVATION FUNCTION	23
FIGURE 2.4-2: TANH ACTIVATION FUNCTION	23
FIGURE 4.1-1: SIMULATOR PROCESS FLOW	44
FIGURE 6.2-1: TOTAL HANDOVERS (10UEs, 500s)	74
FIGURE 6.2-2: TOTAL HANDOVERS (40UEs, 200s)	74
FIGURE 6.2-3: TOTAL TTT INSTANCES (10UEs, 500s)	75
FIGURE 6.2-4: TOTAL TTT INSTANCES (40UEs, 200s)	75
FIGURE 6.2-5: TOTAL MULTI-TIERS (10UEs, 500s)	76
FIGURE 6.2-6: TOTAL MULTI-TIERS (40UEs, 200s)	76
FIGURE 6.2-7: MULTI-TIER PERCENTAGE (10UEs, 500s)	76
FIGURE 6.2-8: MULTI-TIER PERCENTAGE (40UEs, 200s)	76
FIGURE 6.2-9: TOTAL PING-PONGS (10UEs, 500s)	78
FIGURE 6.2-10: TOTAL PING-PONGS (40UEs, 200s)	78
FIGURE 6.2-11: PING-PONG PERCENTAGE (10UEs, 500s)	78
FIGURE 6.2-12: PING-PONG PERCENTAGE (40UEs, 200s)	78
FIGURE 6.2-13: TOTAL HOFs (10UEs, 500s)	80
FIGURE 6.2-14: TOTAL HOFs (40UEs, 200s)	80
FIGURE 6.2-15: HOF PERCENTAGE (10UEs, 500s)	80
FIGURE 6.2-16: HOF PERCENTAGE (40UEs, 200s)	80
FIGURE 6.2-17: SUCCESSFUL HO ATTEMPTS (10UEs, 500s)	82
FIGURE 6.2-18: SUCCESSFUL HO ATTEMPTS (40UEs, 200s)	82
FIGURE 6.2-19: SUCCESSFUL HO RATIOS (10UEs, 500s)	82
FIGURE 6.2-20: SUCCESSFUL HO RATIOS (40UEs, 200s)	82
FIGURE 6.2-21: AVERAGE LATENCY (10UEs, 500s)	83
FIGURE 6.2-22: AVERAGE LATENCY (40UEs, 200s)	83

---

FIGURE 6.3-1: TOTAL THROUGHPUT (10UEs, 500s)	84
FIGURE 6.3-2: TOTAL THROUGHPUT (40UEs, 200s)	84
FIGURE 6.3-3: AVERAGE USER THROUGHPUTS DL LSTM (10UEs, 500s)	85
FIGURE 6.3-4: AVERAGE USER THROUGHPUTS 3GPP (10UEs, 500s)	85
FIGURE 6.3-5: AVERAGE USER THROUGHPUTS DL LSTM (40UEs, 200s)	86
FIGURE 6.3-6: AVERAGE USER THROUGHPUTS 3GPP (40UEs, 200s)	86
FIGURE 6.3-7: USER THROUGHPUT REQUIREMENTS MET WITH DL LSTM SCHEME (10UEs, 500s)	87
FIGURE 6.3-8: USER THROUGHPUT REQUIREMENTS MET WITH 3GPP SCHEME (10UEs, 500s)	87
FIGURE 6.3-9: USER THROUGHPUT REQUIREMENTS MET WITH DL LSTM SCHEME (40UEs, 200s)	88
FIGURE 6.3-10: USER THROUGHPUT REQUIREMENTS MET WITH 3GPP SCHEME (40UEs, 200s)	88

# List of Tables

---

TABLE 2.1-1: OVERVIEW OF THE EVOLUTION OF CELLULAR GENERATIONS [20] [21] [22] [23] _____	8
TABLE 2.1-2: BASE STATION TYPES AND KEY FEATURES _____	14
TABLE 2.2-1: HANDOVER TRIGGER EVENTS FOR INTRA-RAT HANDOVER [46] _____	19
TABLE 2.2-2: EVENT PARAMETER RANGES [47] _____	19
TABLE 3.1-1: SUMMARY AND REVIEW OF SINGLE TIER HANDOVER LITERATURES. _____	29
TABLE 3.2-1: SUMMARY AND REVIEW OF MULTI-TIER HANDOVER LITERATURES. _____	40
TABLE 4.1-1: KEY BASE STATION PARAMETERS IN THE SIMULATOR, COMPLIANT TO 3GPP [16] [79] _____	45
TABLE 4.1-2: RELEVANT PATHLOSS PARAMETERS FOR THE SIMULATOR [17] [86] _____	47
TABLE 4.1-3: MATLAB NR FDD SCHEDULER OPERATIONS FOR GNB AND UE [18] _____	48
TABLE 4.1-4: 5G NR FDD SCHEDULER OPTIONS [90] _____	49
TABLE 4.1-5: RLCC AND APPLICATION CONFIGURATION PARAMETERS FOR THE SCHEDULER [18] _____	50
TABLE 4.1-6: TYPES OF DL SOLVERS IN MATLAB [93] _____	52
TABLE 4.3-1: SIMULATION ENVIRONMENT IMAGES [94] [95] _____	55
TABLE 4.3-2: BASE STATION SIMULATION PARAMETERS _____	56
TABLE 4.3-3: USER MOBILITY SIMULATION PARAMETERS FOR 10 UES _____	57
TABLE 5.1-1: A1 AND A3 HANDOVER CONDITIONS IN THE SIMULATOR _____	64
TABLE 5.2-1: DL LSTM CLASSIFICATION CATEGORIES _____	68
TABLE 5.2-2: ALL 24 CLASSIFICATION OF THE PROPOSED DL LSTM _____	68
TABLE 5.2-3: HANDOVER LOGIC BASED OFF DL LSTM CLASSIFICATIONS _____	70
TABLE 6.1-1: LSTM SUPERVISED LEARNING PERFORMANCE, 1000 EPOCHS PER SIMULATION _____	72
TABLE 6.1-2: LSTM SUPERVISED LEARNING PERFORMANCE, 20 HIDDEN UNITS PER SIMULATION _____	73
TABLE 6.2-1: HANDOVER LATENCIES FOR VARYING RETRANSMISSIONS. _____	83

## List of Abbreviations

---

1G... 5G	1st generation.... 5th generation
3D-UMa	3D Urban Macro Cell
3D-UMi	3D-Ubran Micro cell
3GPP	3rd Generation Partnership Project
5GC	5th Generation Core
AAU	Active Antenna Unit
AI	Artificial Intelligence
AMF	Access Mobility Management Function
AMPS	Advanced Mobile Phone Systems
AP	Access Point
BBU	Baseband Unit
BS/ BTS	Base Station/ Base Transceiver Station
CDMA	Code Division Multiple Access
CN	Core Network
CNN	Convolutional Neural Networks
CoMP	Coordinated Multi-point
C-RAN	Centralised Radio Access Network
CU	Centralised Unit
D-AMPS	Digital - Advanced Mobile Phone Systems
DL	Deep Learning
DNN	Deep Neural Network
D-RAN	Distributed Radio Access Network
DU	Distributed Unit
eMBB	Enhanced Mobile Broadband
eNB/ gNB	eNodeB/ gNodeB
FDD	Frequency Division Duplexing

FFNN	Feed-forward Neural Network
GRA	Grey Relation Analysis
GRU	Gated Recurrent Unit
GSM	Global System for Mobile Communications
HARQ	Hybrid Automatic Repeat Request
HHO/ VHO	Horizontal Handover/ Vertical HO
HOF	Handover Failure
HSPA	High Speed Packet Access
IMT	International Mobile Telecommunications
IoT	Internet of Things
ITU	International Telecommunications Union
JIT	Just In Time
LCID	Logical Channel ID
Lidar	Light Direction and Ranging
LOS/ NLOS	Line of Sight/ Not LOS
LSTM	Long Short-Term Memory
LTE/ LTE-A	Long Term Evolution/ LTE-Advanced
LUT	Lookup Table
MAC	Media Access Control Layer
MCDM	Multi Criteria Decision Making
MDP	Markov Decision Process
MEC	Multi-Access Edge Computing
MIMO	Multiple Input Multiple Output
ML	Machine Learning
mmAP	Millimetre Access Point
mMTC	Massive Machine Type Communications
MR	Measurement Report
NFV	Network Function Virtualisation

NG	Next Generation
NMT	Nordic Mobile Telephone
NN	Neural Networks
NR	New Radio
OEM	Original Equipment Manufacturer
OFDMA	Orthogonal Frequency Division Multiple Access
O-RAN	Open Radio Access Network
PDSCH	Physical Downlink Shared Channel
PHY	Physical Layer
PRB/ RB	Physical Resource Block
PSCN	Packet Switched Core Network
PSTN	Packet Switched Transport Network
PUSCH	Physical Uplink Shared Channel
QoS/ QoE	Quality of Service/ Quality of Experience
RACH	Random Access Channel
RAM	Random-Access Memory
RAN	Radio Access Network
RAT	Radio Access Technology
RE	Resource Element
REM	Radio Environment Map
RF	Radio Frequency
RL	Reinforcement Learning
RLC	Radio link control Layer
RLCC	Radio Network Temporary Identifier Logical Channel Configuration
RLF	Radio Link Failure
RNN	Recurrent Neural Networks
RNTI	Radio Network Temporary Identifier
RRH/RRU	Remote Radio Head/ Unit

RSRP/ RSRQ	Received Signal Received Power/ RSR Quality
RSS	Received Signal Strength
RSU	Road Side Unit
RTT	Round Trip Time
SC-FDMA	Single Carrier Frequency Division Multiple Access
SDN	Software Defined Network
SDO	Standards Development Organisation
SDU	Service Data unit
SINR	Signal-to-Interference-plus-Noise Ratio
SN	Sequence Number
SON	Self-Organised Network
TACS	Total Access Communications System
TDD/ FDD	Time Division Duplexing / Frequency DD
TN	Transport Network
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
TR	Technical Report
TS	Technical Standard
TTT	Time To Trigger
UE	User Equipment
UM	Unacknowledge Mode
UMTS	Universal Mobile Telecommunications Service
UPF	User Plane Function
URLLC	Ultra-reliable Low Latency Communications
VANET	Vehicular Adhoc Network
VLC	Visible Light Communication
V-RAN	Virtualised Radio Access Network
WiFi	Wireless Fidelity
WiMAX	Worldwide Interoperability for Microwave Access

## Assertion of Authorship

---

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Signature: \_\_\_\_\_



Date: 5<sup>th</sup> March 2021

## Acknowledgements

---

Firstly, I would like to convey my thanks to my supervisor Dr. Boon-Chong Seet, I am extremely grateful and honoured to have had such a wise supervisor. His constant patience, support, and endless guidance helped me persevere through my years of research. Additionally, I would also like to thank Auckland University of Technology (AUT) for providing me with the opportunity to undertake further research in my chosen specialisation.

Next, I want to thank my family and friends for their endless love and support throughout this process. Their encouragement has helped me to achieve new heights and I am privileged to have them in my life.

I would also like to thank my team at Rakon Ltd., as they have provided me with the support and industry knowledge, to offer a unique point of view that has enhanced this thesis.

Lastly, I would like to thank, ITU, 3GPP and Ericsson for granting permission to use specific images in this thesis to help explain certain concepts and trends.

# 1 Introduction

---

The next decade is a major technological transition phase for the world, as the shift from 4G (4<sup>th</sup> Generation) to 5G progresses. A lot of importance has been set into developing a 5G network that can be used to connect all types of devices. Prior to this, the main motivation for developing a new generation of telecom infrastructure networks (such as 4G) was to support the rapid growth of the smart phone industry [1]. To make the most of this major milestone, the last few years have been a vital period for many telecom OEMs (Original Equipment Manufacturers) and operators around the globe, as they strengthen their 4G infrastructure in various locations to prepare for 5G's exponential data growth [1].

Over the last three years the world has seen many OEM giants develop and deploy the initial phases of 5G (led by Huawei and Samsung, followed closely by Nokia, Ericsson and ZTE [2]). The expectation from industry professionals is that many large-scale deployments will occur for at least the next decade, where in the next five years, 5G networks will grow and account for approximately 45% of the world's total data traffic, this is illustrated in Figure 1.1-1 (further details are discussed in [3]).

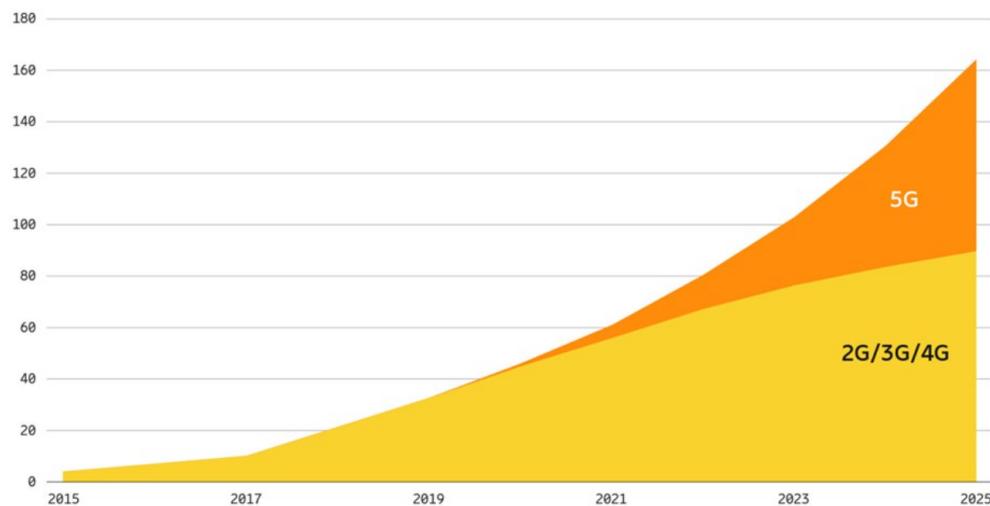


Figure 1.1-1: Global data traffic in the next 5 years (Exabytes), image by Ericsson [3]

There are two major bodies for defining the requirements and standards for 5G. The first is ITU (International Telecommunication Union), this union sets the main visions and goals for the next generation of telecommunication networks. The second is 3GPP (3<sup>rd</sup> Generation Partnership Project), this encompasses seven standards development organisations (SDOs) around the globe that work together to develop the telecommunications standards. 3GPP and ITU collaborate with technical individuals and many industry representatives to standardise what all telecom manufacturers and operators must adhere to everyday [4]. With the new 5G NR standards, it has been made obvious, that the deployment of 5G needed to be a generational

change that is various degrees (10x to 1000x) better than the current 4G LTE and LTE-A ecosystems. Figure 1.1-2 describes a few of the key benefits of 5G.

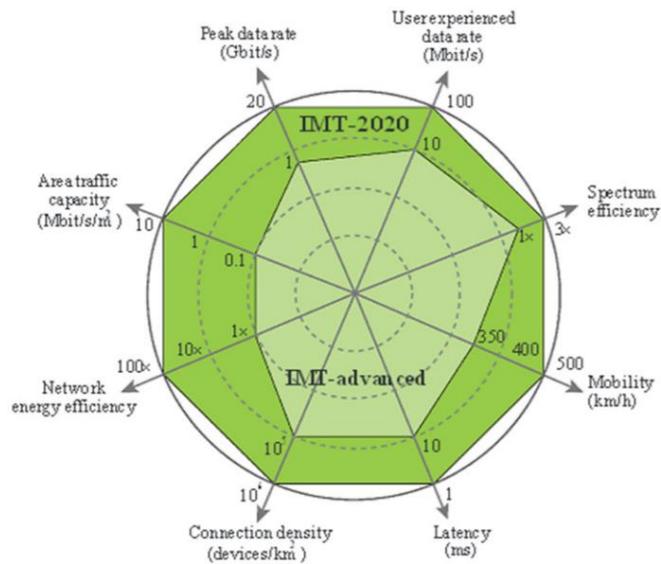


Figure 1.1-2: Comparison between 4G (IMT-advanced) and 5G (IMT-2020), image by ITU [5]

To achieve or get close to the requirements of a more complete 5G network, an important enabling technology that many telecom OEMs are adopting is the introduction of Multi-Access Edge Computing (MEC). MEC provides improvements to various radio access network (RAN) elements, such as:

- Reduction in infrastructure cost through virtualisations,
- Increase in network performance,
- Scalability,
- Network agility [6].

Additionally, the Figure 1.1-3 illustrates the role of MEC in 5G and how a characteristic called network slicing can be used to accommodate the numerous network platforms [7].

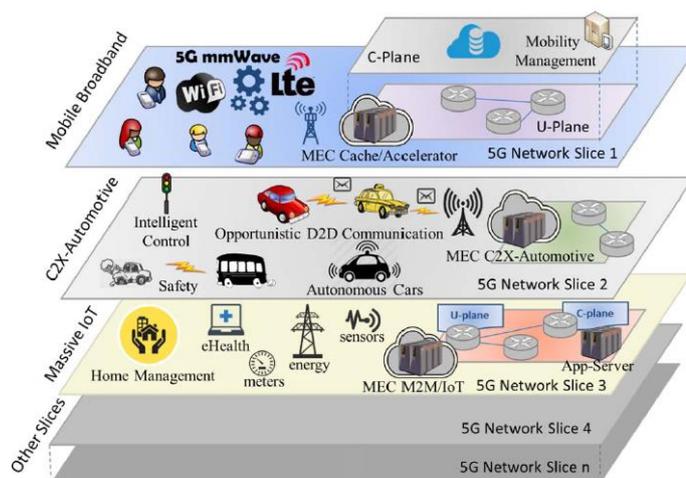


Figure 1.1-3: Network Slicing and the Role of MEC © [2017] [7]

Furthermore, with the introduction of 5G and MEC, a plethora of socio-economic growth opportunities arise as there are endless applications. The points below exhibit a few of the many industries that will begin to see the unveilings of major disruptive technologies:

- Critical high reliability and IoT applications: offsite medical surgeries [8], agricultural farming [9] and smart factories [10].
- Augmented/ Virtual Reality, big data analytics and gaming.
  - o Microsoft’s HoloLens, Amazon web services and Sony’s PlayStation 5, to name a few.
- Vehicular Adhoc networks (VANETs) enhancement for autonomous vehicle communication across various radio access technologies (RATs) such as:
  - o Radio Frequency (RF)
  - o Light direction and ranging (Lidar), these use light in a form of a pulse.
  - o and visible light communication (VLC).

Thus far, MEC have been stated as a key enabler for 5G, although, there is one more key piece required to ensure that all of this performs seamlessly. That piece is Artificial Intelligence (AI), especially the branch of Machine Learning (ML). The diagram in Figure 1.1-4 shows the three major pillars of 5G and the main applications it caters to.

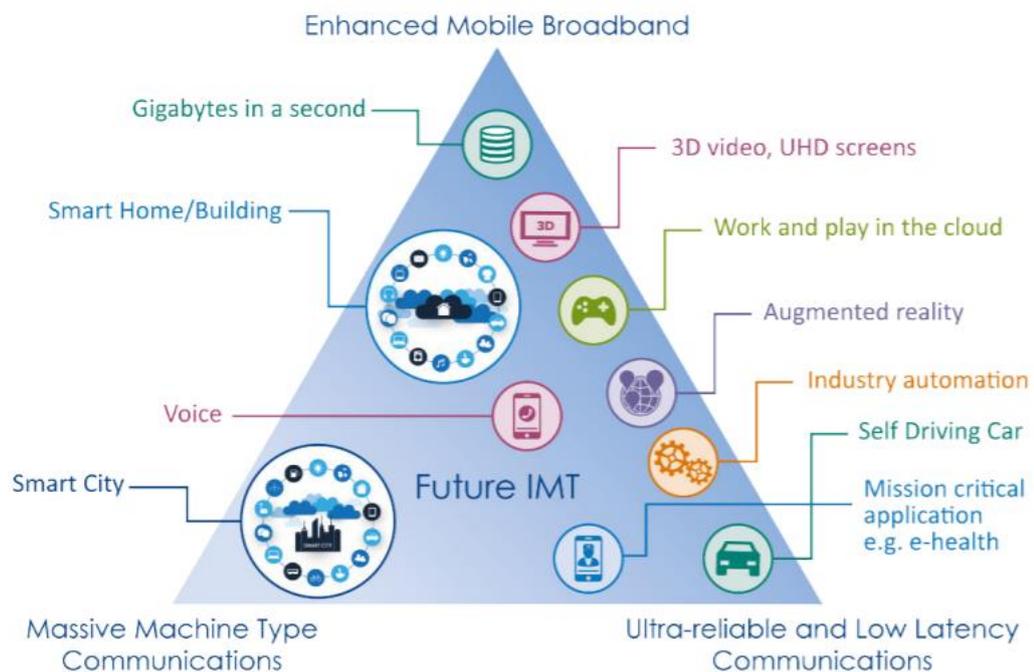


Figure 1.1-4: Three pillars of 5G and applications, image by ITU [5]

The first two Enhanced Mobile Broadband (eMBB) and Massive Machine Type Communications (mMTC) can manage sufficiently with 5G and MEC together, although, to achieve ultra-reliable low latency communications (URLLC) and a more optimised performance for the other two pillars, ML is required.

ML is a subset of AI that is involved with developing computer algorithms for systems to allow it to automatically improve through experience [11]. Traditional optimization techniques are not adaptable enough to handle the complex, real-time analysis required by 5G networks [12]. Over the last decade, ML has proved to be one of, if not the best solution for pattern recognition when computation of big data is required. One of the key contributors to how well a ML algorithm performs is based off the training that is provided. Training helps the algorithm to discover potential relationships between inputs and the desired outputs. There are four different ways a ML algorithm can learn:

1. Supervised learning: Where the algorithm is trained using labelled data. Both input data and output data are known. Commonly used where enough historical data is available.
2. Semi-supervised learning: This can be for the same application as supervised data, but it has labelled and unlabelled data for training. Applications include methods for data classification, regression, and prediction.
3. Reinforcement Learning: Is where data is gained from implementation, the goal is to learn an environment and find the best strategies for a given agent in different environments. Usually used for robotics, gaming, and navigation.
4. Unsupervised Learning: Where tasks are performed without labelled data. The goal is to explore the data and infer a structure from this data to find clusters or anomalies [13].

In addition to the training techniques, there are many types of algorithms used in ML, but for the use in 5G and MEC, the subset of Deep Learning (DL) is the preferred solution by many industry leaders. This is because it can take advantage of its ability to be contextually aware and its dependency on big data. Deep learning has already been applied to various functions within 5G networks, such as traffic classification, routing decisions, and network security [12]. The diagram in Figure 1.1-5 indicates the major performance differences between DL, Neural Networks (NNs) and traditional ML algorithms as the data size increases.

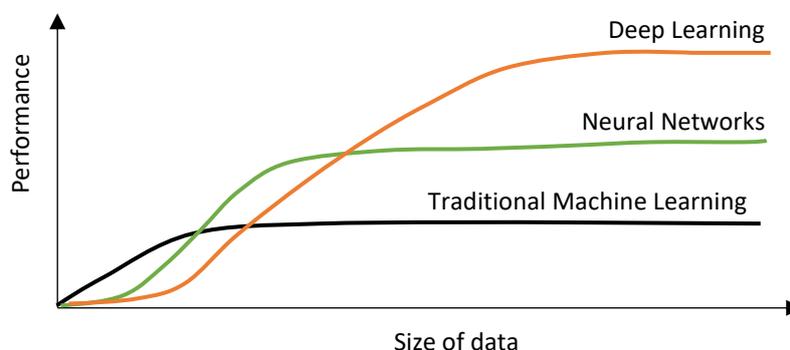


Figure 1.1-5: DL compared to NN and traditional ML

A DL algorithm usually begins with a form of NN design, then several fully connected layers (greater than 3) are added for deeper analytics and superior classifications. A traditional NN has 1 to 3 layers, if layers are more than 3, it is described as a Deep Neural Network (DNN) [14].

## 1.1 Problem Statement and Motivations

With the introduction of 5G, the industry is posed with many diverse challenges. One of the challenges of all the three pillars is, seamless and low latency multi-tier handover. Based off the latest 3GPP standards for 5G, it can be noticed that the event-based triggering for handover ignores various key elements of the user's session that require to be taken into consideration, such as their mobilities and data rate requirements. User requirements are ever-changing; therefore, cellular networks require to be dynamic to react and cater to this demand effectively. There are various channel inefficiencies that occur when a diverse range of requirements are not taken into consideration.

The issues relating to multi-tier handovers have not been effectively resolved to this day. There are quite a few literatures to address the issues of handover, although only a small portion of these adopt a form of AI or cloud computing techniques in their solutions. The use of DL and MEC for optimising handover is still a gap in the industry that has not been explored yet.

With the addition of MEC and DL in 5G, network operators can gather user data and analyse variations in signal strength, mobility patterns and data rate requirements of each user, to achieve optimum user experience. Additionally, with implementing a system that understands the user's requirements, network elements benefit as well, because this helps efficiently manage the base station's resources. Keeping in mind that the network operators should extract/ use all this data without compromising on the compliance of user privacy laws of the specific country.

## 1.2 Thesis Objectives, Scope and Contributions

The objective of this thesis is to develop a DL handover decision algorithm while utilising MEC. This will enable a faster and a more reliable handover system, that would ideally allow the user to switch seamlessly between any cellular network configurations based on key requirements.

This research will be validated by conducting software simulations to compare the proposed method to the handover technique specified by 3GPP in the technical standard (TS) 38.300 [15], Chapter 9.2.3.2. The scope of this research is focused on a 5G heterogeneous environment with various base station tiers (Macro, Micro, and Femtocells), to carry out a comprehensive multi-tier handover evaluation.

There are two key components of the simulator, one is the channel model, and the second is the scheduler. Both the scheduler and the channel models are compliant with 3GPP standards 38.104 [16] and 36.873 [17], respectively. The scheduler supports 5G functionalities [18], and the channel models used are for Macro and Micro base stations for frequencies below 6GHz. Additionally, for frequencies greater than 6GHz, a free space model is employed (further details are in Chapter 4). To conclude this section, the points below highlight the key contributions of this thesis:

1. Propose a new DL Long Short-Term Memory (LSTM) handover decision algorithm that uses a lookup tables (LUT) and is catered to various user requirements (i.e., key Quality of Experience {QoE} and Quality of Service {QoS} requirements).
2. Remove the time to trigger (TTT) and replace it with a dynamic LUT based triggering mechanism.
3. Propose to modify the handover admission control process (when using the DL LSTM logic) to occur at the same timestamp that the base station sends the handover request to the UE. This is assuming that the user plane function (UPF) and access mobility management function (AMF) are located at the MEC aggregated edge.
4. Introduce varying measurement report (MR) instances to accommodate a variety of user mobilities.
5. Propose to reduce the size of measurement reports sent from UEs that subsequently reduce signalling overheads and improve power efficiencies.

## 1.3 Thesis Structure

The remaining parts of this thesis are organised as follows:

### Chapter 2: Background

Presents a detailed view of the specific concepts of cellular technologies, handover, MEC and DL that this thesis will encompass.

### Chapter 3: Literature Review

Delivers a comprehensive review of the various texts encompassing single and multi-tier handover optimisation.

### Chapter 4: Research Methodology

Discusses the various simulation tools, assumptions, and performance metrics required to develop the environment for the researcher to evaluate the findings.

### Chapter 5: Proposed Algorithm

Presents the proposed algorithm and the various other contributors of this thesis. Moreover, a variation of the 3GPP handover algorithm is proposed to be appropriate for comparisons.

### Chapter 6: Results and Discussions

Critically analyses the variations, benefits, and drawbacks of the proposed algorithm and the competitor. Also, the changes in key learning parameters of the DL LSTM are examined.

### Chapter 7: Conclusions, Implications and Further Research

Lastly, this Chapter concludes the thesis and highlights its implications. It will also discuss various prospects for conducting further research.

## 2 Background

A brief overview of, the evolution of the cellular technologies, network configurations, handover, edge computing and deep learning will be discussed in this Chapter. Subsection 2.4 will provide a basic overview of how LSTM networks work and its key features.

### 2.1 Cellular Technologies and Configurations

Cellular technologies and the standardisation of telecommunication networks was first introduced in the 1970s, beginning with 1G, standards such as Nordic Mobile Telephone (NMT) began in 1975 [19] (for Eastern Europe and Russia), followed closely by Advanced Mobile Phone Systems (AMPS) (for North America and Australia), and Total Access Communications System (TACS) (for Western Europe). 1G was the only cellular network that operated primarily on analogue communication systems. Since then, all successive telecommunication standards (2G to 5G and future generations) have been and will be focused on optimising and utilizing digital communication methods. Digital communication provides various benefits compared to its former such as, better signal quality, superior network security and reduced signalling errors.

As cellular networks evolved, each generation had specific goals to be achieved. Table 2.1-1 summarises the few key differentiators between major generations of cellular networks.

*Table 2.1-1: Overview of the Evolution of Cellular Generations [20] [21] [22] [23]*

<b>Technology / Features</b>	<b>1G</b>	<b>2G</b>	<b>3G</b>	<b>4G</b>	<b>5G</b>
<b>Deployment</b>	1970 - 1980	1980 - 2000	2000 - 2010	2010 - 2020	2020 - 2030
<b>Latency</b>	> 1s	1 - 0.3s	500 - 100ms	100 - 30ms	5 to < 1ms
<b>Data Rate</b>	2kbps	50kbps	2 - 8Mbps	0.01 - 3Gbps	1 - 10Gbps
<b>Major/ Key Application(s)</b>	Voice	Secure SMS and Voice	Video calling and GPS navigation	High speed streaming	Low latency for critical work, 1G for machines
<b>Telecom Standard(s)/ RAT(s)</b>	AMPS, NMT, TACS	GSM, D-AMPS, IS-95 A	UMTS, CDMA2000, HSPA+	LTE, LTE-A, LTE-Pro	NR
<b>Radio Access Network(s)</b>	D-RAN	D-RAN	D-RAN	Cloud D-RAN	Cloud C-RAN and V-RAN
<b>Core Network</b>	PSTN	PSTN	PSCN	Internet	Internet
<b>Handover Type(s)</b>	Horizontal	Horizontal	Horizontal	Horizontal and Vertical	Horizontal and Vertical

The explanations below provide a brief background on the functions of cellular network elements and their 5G optimisations. However, before these functions can be described, there are four key terms that require to be defined:

1. User Equipment (UE): Is a device/ body that employs the services provided by the network. The term user and UE will be used interchangeably throughout this thesis.
2. Core Network (CN): Is the main entity in cellular networks, it has many functions, most importantly it provides controls for: user authentications, telephone calls, operators to charge for calls, connections to the internet and the network's handovers (inclusive of both Inter and Intra-RAT handovers) [23]. This function is interlinked with the cloud since 4G, in this thesis it will be termed either the 5GC or Cloud CN.
3. Base station/ Base Transceiver Station (BS/ BTS): Is a transceiver that connects a user to a wider net of users, depending on the desired application [24] (i.e., it connects a user to the telephone network for calling or can connect the user to the internet for streaming/ browsing). Within this thesis a BS can be referred to as an eNodeB (eNB) if it is referring to 4G or prior, otherwise, the term gNodeB (gNB) will be used as the BS provides the UE a connection to the 5GC via next generation (NG) interfaces [15].
4. Transport Network (TN): Is the network that provides transparent transmission of user data between connected entities. This is done by establishing and maintaining a point-to-point or a point-to-multipoint connection between devices [25]. In a simple RAN configuration, TNs can be split into Fronthaul and Backhaul definitions [26].
  - a. Fronthaul: Is the connection between base station components.
  - b. Backhaul: Is the connection between the base station and CN.

An illustration of a basic cellular network is displayed in Figure 2.1-1, combining all these definitions.

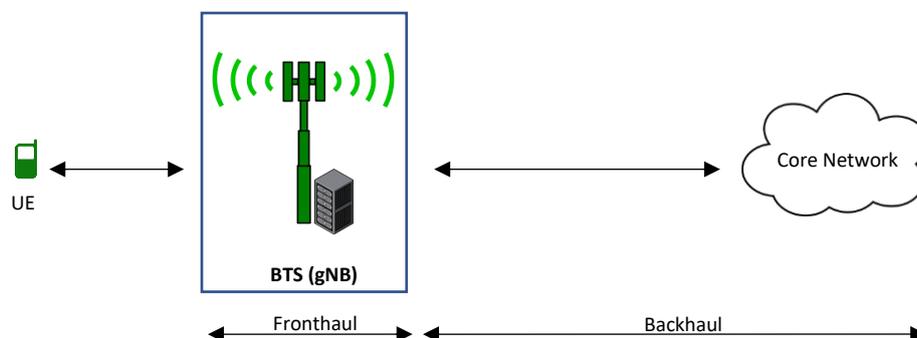


Figure 2.1-1: Basic illustration of a cellular network

## 2.1.1 Radio Access Networks

A RAN provides the user with a connection to the CN to fulfil service requirements [27] [28]. A RAT is the physical underlying communication form that is used for a type of RAN, such as 4G and D-RAN, 4G is the RAT and D-RAN is the RAN used.

Furthermore, there are two types of RAN progressions that will be detailed in this subsection:

1. Distributed RAN (D-RAN) the current RAN used in majority of today's networks.
2. Centralised RAN (C-RAN) a RAN that is being applied to 5G and future networks.

Starting with D-RANs, these networks consist of a set of base stations and organise the BS components into three major segments [28] [23] [29]:

1. Antenna, this is the physical element that emits the electrical signals from the base station into radio waves to be transmitted for the desired UE.
2. Remote Radio Head (RRH/ RRU): Converts digital information into signals with the appropriate encryption that allows it to be securely transmitted through the antenna. The RRH helps to ensure that the transmitted signals are in the right frequency bands and power levels for the UE to correctly receive the data.
3. Baseband Unit (BBU): Provides signal processing functions that makes wireless communication possible between the UE and the CN. This portion of the BS is responsible for helping provide secure connections, mobility management (such as handovers) and radio resource management.

Figure 2.1-2 is an illustration of the Distributed RAN (D-RAN) configuration.

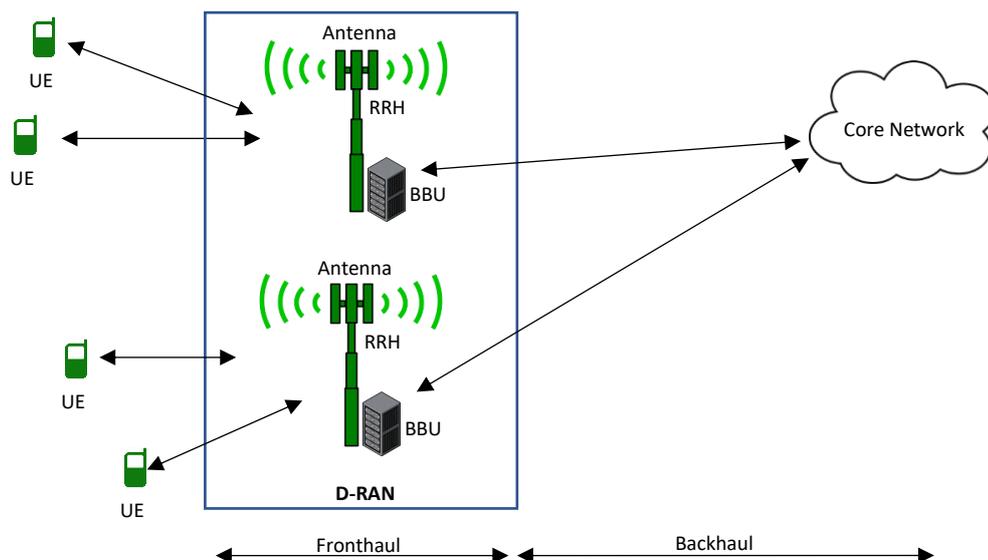


Figure 2.1-2: Basic illustration of a D-RAN configuration

In this design, the BBU is located at each base station to be able to meet high demands when required. Consequently, when there is no/ low demand, BBUs are left underutilized and proves to be a waste of the BBU's capabilities. This BBU also has a limited view of the surrounding environment, therefore, it can only service users that the one base station can reach (shown in Figure 2.1-2). As data requirements grow exponentially, D-RAN systems would have to constantly be upgraded with new hardware to keep up with demand. This proves to be a very expensive effort when all the inefficiencies are taken into consideration [29]. Therefore, with the introduction of 5G a new RAN was introduced into the system, this was C-RAN.

C-RAN is somewhat similar to D-RAN, the major difference is that the BBU is no longer located at the base station this allows the BS to have a smaller footprint. There are three evolutions to C-RAN in relation to the BBUs, these are detailed below [30]:

1. BBU Hotel: This is very much similar to the current D-RAN architecture, although various BBUs are placed in a central location, but these BBUs still only service one BS. Therefore, the same problem of inefficiencies is being run into with this architecture.
2. BBU Pool: This is a more efficient use of resources, as a pool of BBUs can serve many BSs at a time and it has a many-to-many relationship. This allows BBUs to focus on services where it is needed most, this solution is being used for current 5G solutions.
3. BBU Virtualisation (shown in Figure 2.1-3), this is a new technology that is still being deployed, it involves virtualising the whole BBU. Therefore, creating a Virtualised RAN (V-RAN). V-RANs provide benefits in terms of expansion, accessibility, and frequent upgrades. This is currently being implemented by major industry OEMs. These virtualisations are made better by moving parts of the Cloud closer to the user, in a configuration known as Cloud C-RAN (this will be discussed further in Chapter 2.3, when discussing MEC).

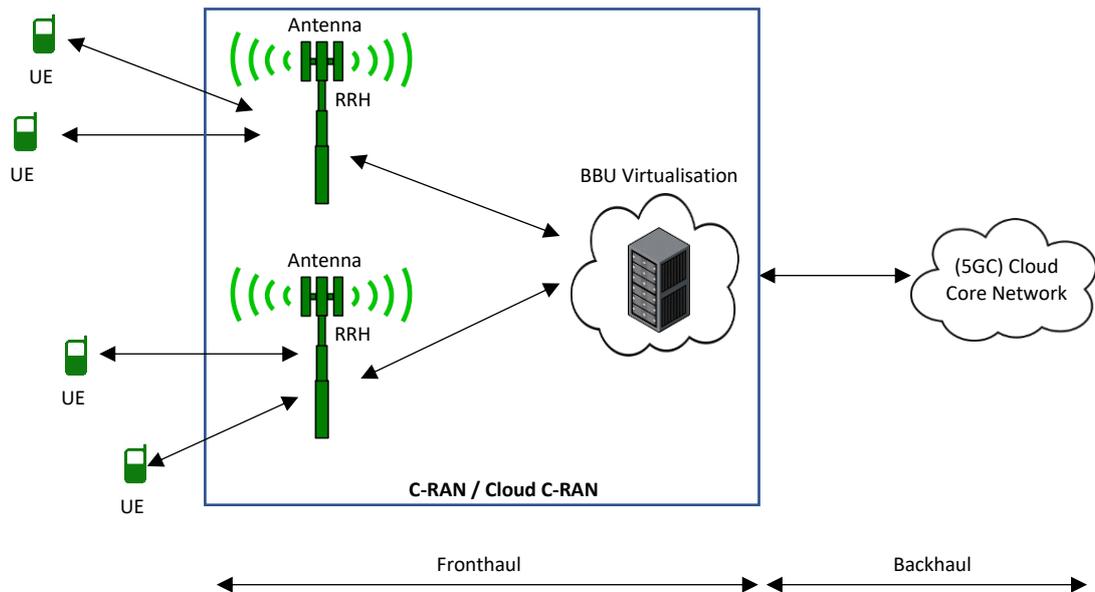


Figure 2.1-3: Basic illustration of a C-RAN with a Virtualised BBU Configuration

Through C-RAN and its virtualisations, 5G has improved in three ways [28] [31].

1. The antenna and radio segments of the RAN are packed closely together with very high computational power to support multiple-input and multiple-output (MIMO) activities. The radio unit used in these 5G base stations allows for complex computations at higher frequencies with improved performance and efficiency.
2. The BBU is coded and virtualised into various different segments. This provides a more cost-effective approach as many physical segments can now be virtualised as a part of the process for implementing a software defined network (SDN) via network function virtualisations (NFVs). As virtualisations occur, latencies of specific functions are reduced due to all major functions of the baseband residing in the same place.
3. Lastly, NFVs can enable easier updates and troubleshooting as majority of the functions are software coded. This becomes more useful as the network grows rapidly. There are various aspects of SDNs that utilise AI or ML, due to their complexity. Ericsson is one of the OEMs that uses ML with their central network management and orchestration in individual base stations, to optimize traffic and load balancing in 5G.

With the introduction of 5G, the definitions of the BBU, RRH and antenna are reimaged to provide a more disaggregated network. The RRH and antenna are integrated into one device called an active antenna unit (AAU). The BBU is split into two components, the distributed unit (DU) and the centralised unit (CU). With this disaggregation, the functionalities will be split. The DU will be responsible for PHY, MAC, and RLC sublayers, whereas the CU will comprise of the remaining sublayers [32].

This redefines the earlier TN definitions.

1. Fronthaul is now the connection between the AAU and the DU
2. Midhaul is the connection between the DU and the CU
3. Backhaul is now the connection between the CU and the 5GC

Figure 2.1-4 provides a visualisation of these three definitions.

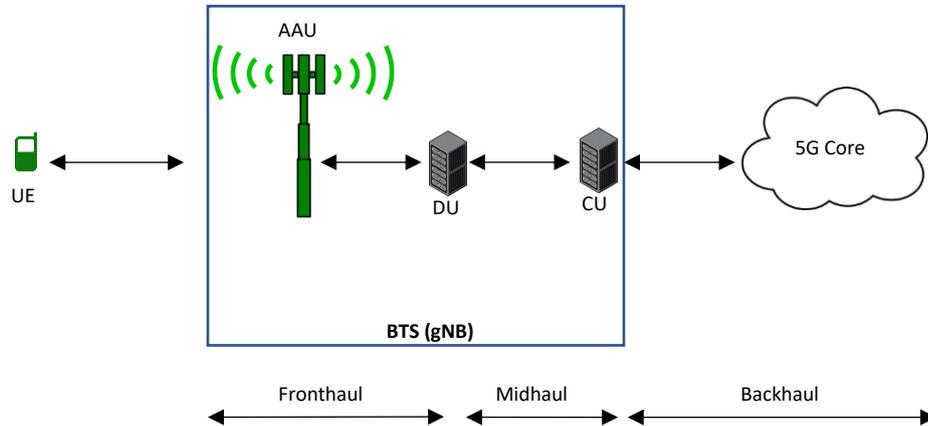


Figure 2.1-4: Basic illustration of Fronthaul, Midhaul and Backhaul

With the introduction of Midhaul, telecommunication operators are pushing for the development of Open RAN (O-RAN) as it assists with standardising mobile networks [33] [23]. O-RAN will not be discussed further as it does not relate to the objectives of this thesis.

The chosen RAN architecture is Cloud C-RANs that have NFVs (with the new BS and RAN architecture including AAU, DU and CU). Further details will be addressed in Section 2.3.

## 2.1.2 Base station types

This subsection expands on the definitions of base stations and highlights their various types and features. There are three types of base stations defined by 3GPP, these are:

1. Wide area BS (Macro cell)
2. Medium range BS (Micro and Pico cell)
3. Local area BS (Femtocell)

Table 2.1-2 summarises their various features [16] [34] [35].

*Table 2.1-2: Base Station types and key features*

<b>Base Station Type</b>	<b>Wide Area (Macro cell)</b>	<b>Medium Range (Micro/ Pico Cell)</b>	<b>Local Area (Femtocell)</b>
<b>Location(s):</b> - <b>Rural</b> - <b>Urban</b>	Rural: Large standalone cell tower  Urban: On top of Buildings	Rural: small standalone cell tower  Urban: Streetlights, stadiums, etc.	Rural: inside homes  Urban: inside shopping malls, Universities, etc.
<b>Typical antenna height (m)</b>	25 to 75	8 to 15	2 to 8
<b>Minimum distance along the ground to UE (m)</b>	32	5	2
<b>Minimum coupling loss (dB)</b>	70	53	45
<b>Carrier Frequency Ranges (GHz)</b>	FR1: 0.45 to 6 FR2: 24.25 to 52.6 Usually in FR1.	FR1: 0.45 to 6 FR2: 24.25 to 52.6 Can be FR1 or FR2.	FR1: 0.45 to 6 FR2: 24.25 to 52.6 Usually in FR2.
<b>Cell Radius, based off 4G (m)</b>	> 1,000	Micro: 250 to 1,000 Pico: 100 to 300	10 to 50
<b>Maximum Carrier Output power (Typical power in 4G) (W)</b>	> 6.31 20 to 160 (40)	$\leq 6.31$ , Micro: 2 to 20 (5), Pico: 0.25 to 2	$\leq 0.25$ 0.01 to 0.25

For this thesis, Macro cells and Micro cells will operate in the FR1 frequency band, whereas Femtocells will operate in the FR2 frequency band. Further details are in Chapter 4.

### 2.1.3 Antennas and Pathlosses

There are various types of antennas used for transmission and reception of signals. Generally, all antennas can fall under the classifications of directional, or Omnidirectional. Directional antennas radiate a signal in a specific direction to provide an improved signal and reduction in interferences. Omnidirectional antennas are antennas that radiate in an isotropic doughnut shaped pattern [36] to provide equal coverage in all directions. This thesis will only consider the use of Omnidirectional antennas, as it is a sufficient method to implement for analysing the proposed handover algorithm.

As the antenna radiates an RF signal out of the BS, there are attenuation/ pathloss factors that require to be taken into consideration. Pathloss is a growing issue in 5G as the frequencies used proceed to get higher into the FR1 and FR2 bands. The relationship between the pathloss and frequency in free space (where there are no obstructions) is defined by the equation below [37].

$$PL_{freeSpace} = 20\log_{10}(4\pi R/\lambda)$$

where:

$PL_{freeSpace}$  is the pathloss in free space in decibels (dB).

R is distance from the transmitter to the receiver in metres (m).

$\lambda$  is the wavelength of the signal in metres (m).

- $\lambda = c/f$ , where  $c$  is the speed of light in metres per second ( $\text{ms}^{-1}$ ), and  $f$  is the frequency in Hertz (Hz).

The equation specified above considers only the pathloss of free space. There are various other factors that require to be understood to develop a comprehensive pathloss model, such as:

- Wall losses, these are the losses due to obstacles that may be in the way.
- Shadow Fading, these are the fluctuations in received signal due to obstructions, such as multipath effects (two or more signals received by the antenna due to reflections from objects in the environment) and weather.
- Small scale fading, these are the rapid changes of the amplitude and phase of a radio signal over a short period of time (on the order of seconds) or a short distance (a few wavelengths) [38].
- User noise powers, interfering and unwanted signals from other users.

For the simulator in this thesis, all four pathloss contributors are considered. Macro cells use the 3D-UMa (Urban Macro cell 3D) model, Micro cells use 3D-UMi (Urban Micro cell 3D) model and Femtocells use a free space model (with the addition of all other pathloss factors). The first two models are defined in 3GPP's TS 36.873, in section 7.2.1 [17] for LOS (line of sight) and NLOS (not line of sight) scenarios.

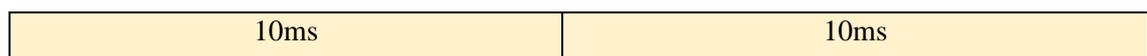
## 2.1.4 Frame structure

For synchronisation between devices, RATs require to have a defined frame structure for the transfer of information between the UE and the network. There are two frame structure modes in cellular networks, time division duplexing (TDD) and frequency division duplexing (FDD). For this research, the chosen frame structure is FDD as it provides various benefits over TDD in terms of channel interferences and better coverage [39]. In 4G and 5G, the FDD technique for transmitting downlink data utilises Orthogonal Frequency-Division Multiple Access (OFDMA), this allows the base station to transmit data in parallel to multiple users. Subsequently, UEs that transmit uplink data back to the BS use a technique called single carrier FDMA (SC-FDMA), as it saves power, is safer for the user, and is cost effective [40].

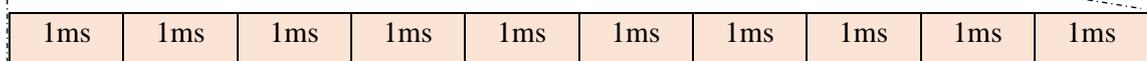
A brief description of the frame structure is described below. A frame in 5G is similar to LTE in the way that each frame is 10ms long and both downlink (from BS to UE) and uplink (from UE to BS) communications are organised into frames of 10 sub-frames of 1ms duration each [41]. A sub-frame consists of slots, the number of slots vary based on the number of subcarriers there are. One subcarrier is known as a resource element (RE). 12 REs complete 1 physical resource block (PRB/ RB), this the minimum resource that can be allocated to a user.

In 5G a subcarrier can vary in sizes from 15kHz (1 slot per sub-frame) to 240kHz (16 slots per sub-frame), in multiples of 15kHz subcarriers. Additionally, there are a number of symbols per slot based off cyclic prefixes, this is used to avoid intersymbol interference (ISI) from its previous symbol. There are 2 cyclic prefixes configurations, normal and extended that occupy 14 or 12 symbols per slot, respectively. The illustration below depicts a simple 5G frame.

### Frame:



### Sub-frame (10 sub-frames per frame):



### Slots (30kHz subcarrier, 2 slots per sub-frame):



### Symbols (14 symbols per slot):



Figure 2.1-5: 5G basic frame structure

## 2.2 Handover

Over the years, as RANs have evolved, there have been various significant enhancements. An enhancement that will be discussed in detail within this thesis is that of handover. There are two types of handovers:

1. Hard handover, consisting of breaking the link between a UE and the current BS before making the new link with the target BS (break-before-make).
2. Soft handover which is when the UE makes the link with the target BS before it breaks the link with the current BS (make-before-break).

In cellular networks (for generations after 3G), hard handover has been the preferred option due to it being less expensive and complex when dealing with OFDMA techniques. There are three main classifications of handover that will be discussed in this subsection, they are horizontal, vertical, and multi-tier handovers.

1. Horizontal handover (HHO): This is a mature concept and has been a part of cellular networks from the start. HHO occurs when a UE hands its context over from a source BS to a target BS as it moves between the two BSs, in all cases, the RAT of these BSs is the same. This type of handover is referred to as Intra-RAT handover [42].
2. Vertical handover (VHO): Is still somewhat in its early stages, it was introduced with the launch of 4G [21]. VHO and HHO have similar definitions, although, the main difference is that VHO occurs when the BSs operate on a different RATs (e.g., handover from 4G to 5G), referred to as Inter-RAT handover [42].
3. Multi-tier handover: like VHO, this is still somewhat in its early stages. This focuses on the handover between various tiers/ types of BSs (Macro, Micro/ Pico, and femtocells). Multi-tier handover can occur for both inter or intra-RAT scenarios. For example, a 4G Macro cell can handover to a 5G Micro cell (this is a multi-tier and an inter-RAT HO).

This thesis will focus on the improvement of 5G multi-tier intra-RAT handover (HHO), although, the same principles could be applied for Inter-RAT (VHO) solutions as well. Intra-RAT handovers occur in the AMF and UPF elements of the 5G architecture.

The AMF (located in the 5GC) manages handovers between different gNBs, this is also referred to as Xn handover, Xn is the communication interface between gNBs [43]. The protocols between these BSs are defined in the 3GPP standard TS 38.423 [44]. Moreover, N2 is the interface between the AMF and the respective gNBs. The UPF (also situated in the 5GC) supports service features for the UE (such as packet routing), it communicates to the BS through the N3 interface.

All these communication interfaces are illustrated in Figure 2.2-1.

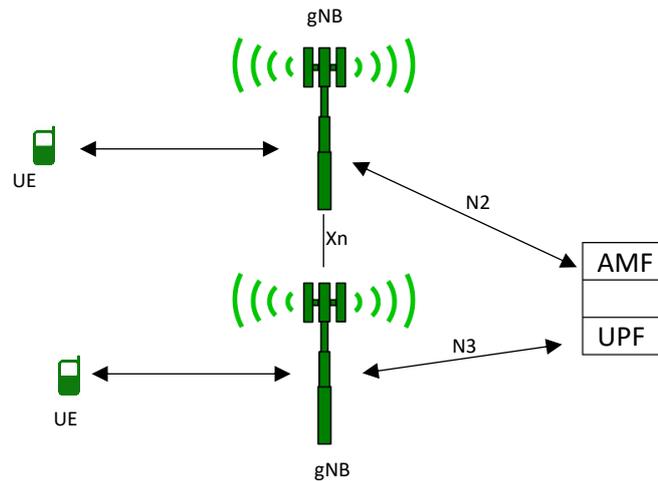


Figure 2.2-1: Interfaces between gNBs, AMFs and UPFs

The full 5G architecture described in the 3GPP standard TS 23.501 [45], this thesis refers to the non-roaming architecture in section 4.2.3 and will only model the handover functions that are performed by the UE, gNB, AMF and UPF.

### 2.2.1 3GPP Defined Logic and Procedure

The handover logic and process below will refer to the following two 3GPP TSs: 38.331 section 5.5.4 [46] for the handover logic, and 38.300 section 9.2.3.2.1 [15] for the handover procedure. Before detailing these steps, an understanding of how a UE switches between two of its states, connected and Idle, are described below (other states are not relevant to this thesis).

- Idle: A UE is in the idle state when the UEs context is known to the 5GC, but it does not have an established connection to a gNB. In this state, the user ‘listens’ and responds to broadcasted messages from gNBs. It performs measurements and cell reselection methods when it is ready to connect to a gNB. This state is preferred by UEs that do not require to transmit any messages, as it has great power efficiency benefits.
- Connected: A UE is in the connected state when the UEs context is known to the core and the gNB. In this state the user provides periodic measurement reports with channel quality information (CQI). Data is regularly transferred in this phase.

At set timestamps during the connected state, the UE sends measurement reports for the AMF to assess whether a handover necessary. Usually, it is based on the user’s received signal strength (RSS) for a particular BS, although, sometimes other factors such as loading are considered.

The way the AMF decides whether a handover is to occur, is decided based off an event triggered system. Events are triggered by the logic described in [46]. There are various event triggers, and their parameters are specified in Tables 2.2-1 and 2.2-2. Events A1 through A6 are only considered as they relate to intra-RAT handover; other events such as B1 are not relevant to this thesis as they relate to inter-RAT handover.

Table 2.2-1: Handover trigger events for intra-RAT handover [46]

Event Type	Description
Event A1	Serving cell becomes better than a threshold
Event A2	Serving becomes worse than a threshold
Event A3	Neighbour becomes offset better than serving
Event A4	Neighbour becomes better than a threshold
Event A5	Serving cell becomes worse than threshold 1 and neighbouring cell becomes better than threshold 2
Event A6	Neighbour become offset better than secondary cell

Table 2.2-2: Event parameter ranges [47]

Event	Parameter	Minimum	Maximum
A1, A2, A4, A5	RSRP threshold	-156 dBm	-31 dBm
All	Hysteresis	0 dB	15 dB
A3, A6	Offset	-15 dB	+15 dB

Each event has an entry and a leaving condition, if the entry condition is satisfied for longer than a certain period, called the time to trigger (TTT), the BS will initiate the handover procedure to the desired cell. However, if the UE's RSRP drops below the leaving condition, or does not meet the entry condition after the TTT, the UE remains connected to the current BS, as the desired BS no longer meets the criteria. The diagram below in Figure 2.2-2 illustrates these conditions. It assumes that all UEs meet the entry condition at the same time for simplicity, although only UE1 has a signal strength above the leaving condition for an A3 handover.

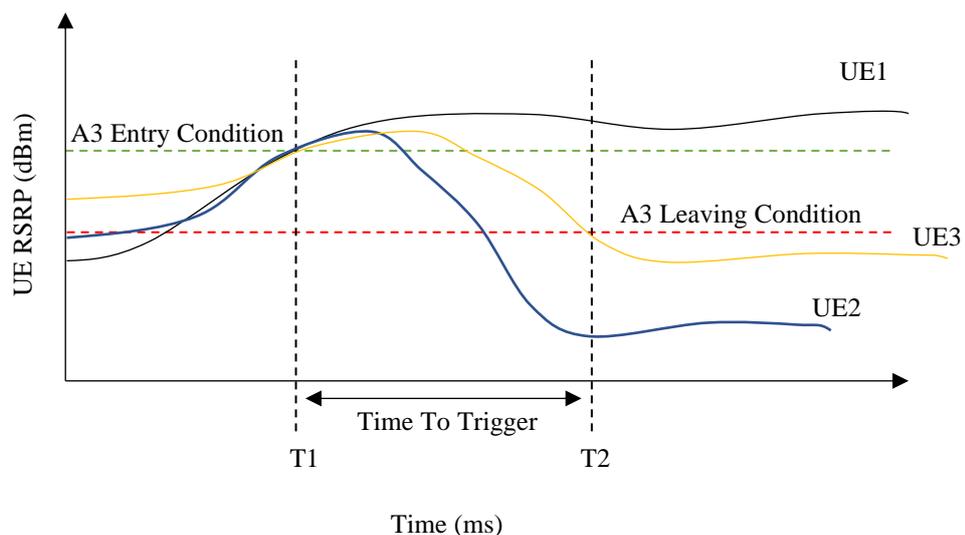


Figure 2.2-2: Illustration of Entry and leaving conditions with TTT.

This thesis focuses on implementing two of the key handover events, A1 and A3 handover. Other scenarios are not considered in this thesis, as there will be unnecessary complexities introduced which will diverge from the thesis objectives. For example, A6 handover would require a form of dual connectivity for the user to correctly perform/ asses this event.

After the UE has met the entry condition for the duration of the TTT interval, a handover is initiated. The flow diagrams below represent the 5G Intra-RAT handover procedure that is described in [15]. There are three phases:

1. Preparation
2. Execution
3. Completion

These steps are described in detail below.

Firstly, the preparation phase (steps 1 to 5, shown in Figure 2.2-3), the discussion above already described steps 0, 1 and 2. Steps 3 to 5 is where the source gNB requests a handover to the target gNB, the target gNB then processes the request and completes admission control, and lastly, a handover acknowledgement is sent back to the source gNB.

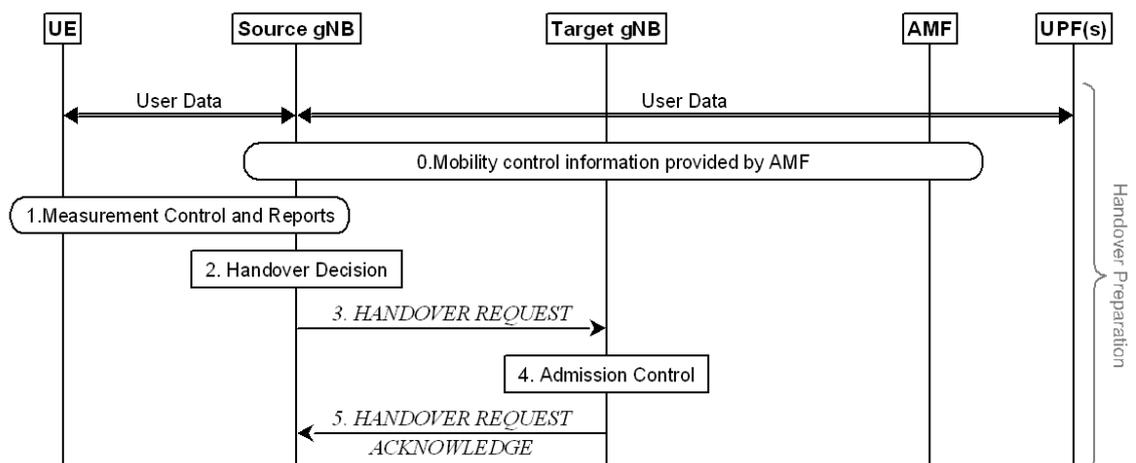


Figure 2.2-3: 5G Intra RAT handover procedure preparation phase [15]

Second, the handover execution process begins (displayed in Figure 2.2-4), where the source gNB initialises the handover by notifying the UE. This prompts the UE to begin to detach from the source gNB and synchronise to the target gNB.

Simultaneously the source gNB executes a sequence number (SN) status transfer for the target gNB to know from which packet that it should send or receive. Then buffer data and new data is delivered from the UPF to the source gNB which is then forwarded to the target gNB. After that the RAN handover is completed.

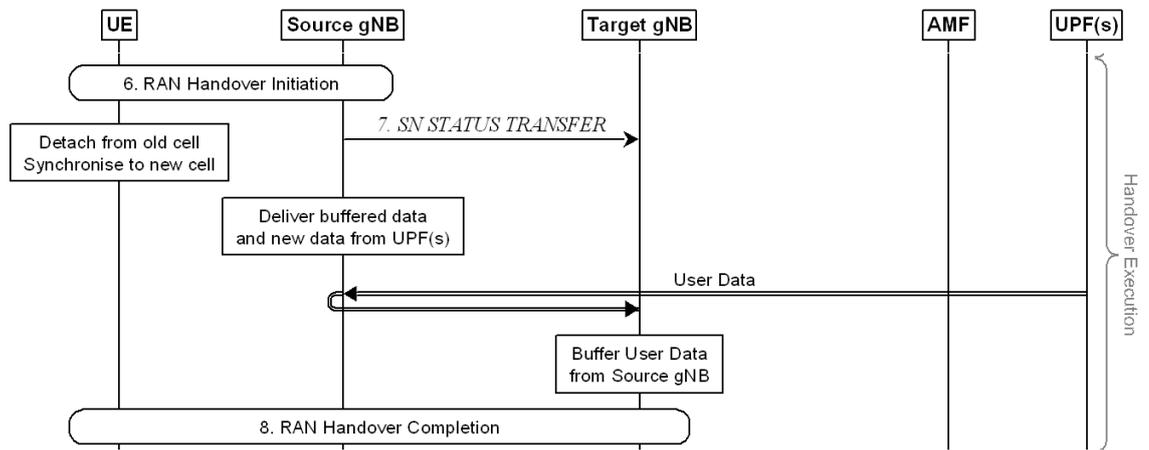


Figure 2.2-4: 5G Intra RAT handover procedure execution phase [15]

Lastly, the completion phase in Figure 2.2-5 begins with a path switch is requested from the target gNB. This triggers the 5GC to switch the path of the UE’s data to the target gNB, via the UPF. Then the UPF sends the end marker for the source gNB via the AMF. The AMF then sends the path switch acknowledgement. The target gNB then sends a message to source gNB to release the context of the UE as the handover was successfully complete.

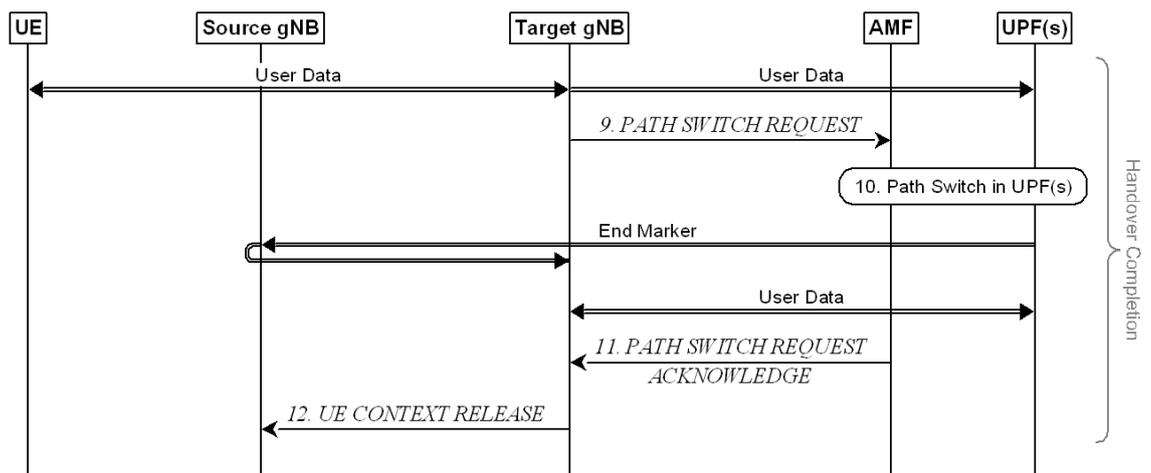


Figure 2.2-5: 5G Intra RAT handover procedure completion phase [15]

The handover process described above is what this thesis will integrate and adjust in Chapter 5.

## 2.3 Multi-Access Edge Computing

As cellular networks are progressing, big data analytics is becoming progressively more vital. With the introduction of 5G, there has been a push to develop more datacentres/ cloud computing services to cope with network growth. Multi-access edge computing (MEC) has become a very key technology in this field. Before discussing the reasons why MEC has become such a compelling solution, it is important to understand exactly what MEC is and how it differentiates from traditional cloud computing platforms.

MEC enables cloud-like services closer to the edge of a network (traditional cloud computing services reside in the core) [48]. The visualisation in [49] by Nokia highlights the key differences between MEC and cloud computing (central datacentres).

This architecture enables a more decentralised cloud service based on user service requirements, allowing for lower latencies and higher throughputs for users. For example, an airport can have deployments of far edge servers to rapidly update location and augmented reality applications [49], although, this optimisation does come with a certain associated cost as the deployment of these edge servers will require to be quite densely packed. A way that OEMs are tackling this is by installing edge servers to highly populated urban areas where large throughputs and low latencies are required, and not deploying them in rural areas, where densification is not required.

This thesis will focus on the two edge data centre architectures highlighted above.

- The far edge services will be more user centric to reduce latency and boost throughput.
- The aggregated edge will be used for lower latency handovers, to approach the user plane eMBB and URLLC latency targets of less than 4ms and 1ms, respectively [50].

## 2.4 Deep Learning

Building on the ideas described in Chapter 1, deep learning (DL) consists of many ML algorithms. The topmost used algorithms are, convolutional neural networks (CNN), long short-term memory (LSTM), and recurrent neural networks (RNN) [51]. This thesis will focus only on LSTM as it is known to be able to learn and memorise long term dependencies [51]. The remainder of this Chapter will give a brief overview of how LSTM networks work, although, a more detailed explanation can be found in [52]. LSTMs consist of one cell state and various gates.

The cell state can be considered as the “memory” part of the LSTM, as it carries all the relevant information throughout the process of the sequence. It progresses straight through the entire network and runs through two linear equations every time stamp. LSTM networks can alter the cell state through various regulated and structured gates. Gates allow optional entry of specific information, comprising of a sigmoid NN layer and a point wise multiplication operation [53].

The diagram in [53] shows an illustration of three repeating modules of a LSTM network. The summary below briefly outlines the key phases, a detailed explanation can be found in [53].

Before proceeding to describe the process for the image above, two key activation functions that will be used in conjunction with the gates, require to be defined.

1. Sigmoid ( $\sigma$ ) activation function: This is a mathematical operation that squeezes the input value between 0 and 1 (defined and displayed in Figure 2.4-1).

$$f(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{e^x + 1}$$

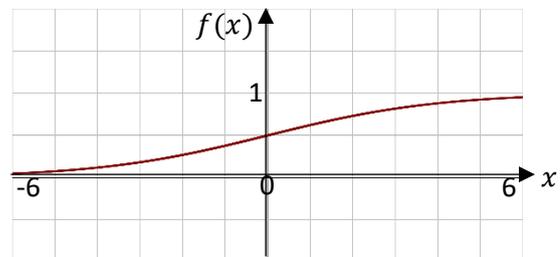


Figure 2.4-1: Sigmoid activation function

2. Tanh activation function: This is a mathematical operation that squeezes the input value between -1 and 1 (defined and displayed in Figure 2.4-2).

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

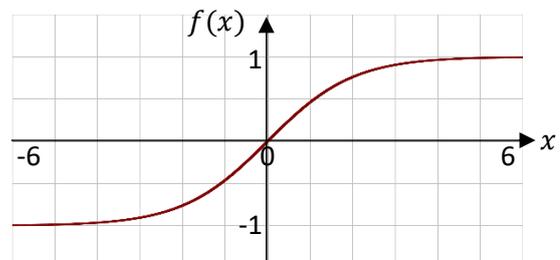


Figure 2.4-2: tanh activation function

These activation functions will be used throughout the LSTM to provide results that will converge to a specific value.

The first step is the ‘forget gate’. This gate helps decide what information is going to be kept from the previous time step’s output  $h_{t-1}$ . The decision for what is kept from both  $h_{t-1}$  and  $x_t$  are decided via a sigmoid activation function, a 0 means that the information is completely discarded, a 1 indicates to completely keep this information. Where  $W_f$  and  $b_f$  are the weight and the bias values of the forget gate, respectively.

The next process step is split into two parts:

1. The ‘input gate’  $i_t$ , this decides the values that the cell gets updated to.
2. The ‘cell candidate gate’  $\tilde{C}_t$  [54], decides what new information can be added to the cell state.

Then progressing on to the next step, the cell state is updated to the most current timestamp and the two values are combined to output the cell’s state. By multiplying the forget gate function with the previous time stamp’s cell state, the cell forgets all the information that is no longer relevant to the cell state.

Then the new candidate values are multiplied with the input gate values to favour how much of each value is going to be added to the cell state.

Lastly, the ‘output gate’ is updated to decide what values are going to be sent out of the LSTM. This step is split into two aspects again:

1. The sigmoid layer decides what is going to be output.
2. Then the cell’s state is squeezed with a tanh function and multiplied out with the output function, allowing the LSTM to decide relevant outputs.

There are many different variations and additions of LSTMs, three variations are:

1. Peephole connections, where each gate layer is able to look at the cell state, this is defined in [55].
2. Gated recurrent unit (GRU), which combines the forget and input gates into one, this is defined in [56].

For this thesis, a traditional LSTM will be used, as these algorithms are ideal for classifications of sequence of data.

## 3 Literature review

---

There has been extensive research conducted on handover and their optimisations, although multi-tier handover still has several gaps that can be addressed. After an exhaustive search, very few works in the area of multi-tier handover with ML/AI and/ or MEC were found. This Chapter classifies and analyses the relevant literature under two main categories: single-tier, and multi-tier handovers, and highlights key areas for improvement.

### 3.1 Single tier handovers

In single tier handovers, there are a few key metrics that require to be controlled to maintain a consistent QoS and QoE for the user. These are:

- The total number of handovers (successful and failures), and a handover latency (the time it takes to prepare, execute, and complete the handover process).
  - o These factors directly influence the user's throughput. As the number of handovers or duration of handovers increase, less time is spent receiving meaningful data, and more time is spent on configuration of the UE.
- The handover failure (HOF) ratio, this is a metric that is expressed as a percentage of the total number of handovers.
- Ping-Pong handovers: This happens when the user is constantly switched between serving and the desired BS in a short amount of time. The defined time can vary from network to network, although, this is typically a value less than 10 seconds.
- Air interface signalling overhead: This is the additional information that is sent in a message for enhancing the performance of the network, although, these overheads can reduce the overall data throughput for the user. In handover, high signalling overheads can occur from factors such as constant measurement reports from the UE.
  - o This impacts the UE's overall power consumption, as well as the cost of radio equipment, to service the demand.

There have been many single tier solutions proposed over the years, this section will review and analyse a select few categories that are relevant to the proposed solution.

#### 3.1.1 Multi-connectivity

Multi-connectivity is when a user is connected to more than one base station to ensure that the connection to the network is not lost. Soft handovers are considered to be a part of multi-connectivity (in this thesis) as when the user is handing over to a new BS, for a small amount of time the user is connected to multiple BSs at once (due to the make-before-break methodology).

Beginning with a pure multi-connectivity solution, the authors in [57] aims to reduce the handover cost due to network densification. The proposed algorithm provides an optimised solution by using an anchor-based multi-connectivity architecture, and it derives compact

expressions of handover probabilities through stochastic geometry analysis. The environment includes three access points (APs) and one UE. The authors simulate the handover probability as the density of the APs increase and speed of the UE increases. The proposed algorithm was compared with a single connectivity algorithm. The results show that the proposed algorithm outperforms the single connectivity algorithm and reduces the handover probability by around 40%. Consequently, due to the multi-connectivity nature of this algorithm, signalling overheads increase as a user has to maintain connection to two base stations. Thus, the improvements in handover performances in [57] could be negated by these increases in signalling overheads.

The proposal, in [58] suffers from the same issues. It attempts to address the matter of seamless low latency handover for mobile UEs in 5G. A random-access channel (RACH)-less handover scheme is proposed to achieve this. The authors combine this scheme with a make-before-break (soft handover) methodology to achieve a lower latency for the UE. The authors compare their algorithm with three different schemes, one soft and two hard handover algorithms. It can be observed that the handover execution time is considerably less for the proposed algorithm by nearly around 33% (from 30ms to 20ms). Additionally, the handover interruption time is on par with the make-before-break algorithms, which are much better than the hard handover schemes.

Moreover, the authors in [57] and [58] do not consider the extra processing time and equipment cost for UEs and OEM vendors, to enable such technologies. The costs could be very high, as each system would have to be able to support a multi-connectivity environment.

### **3.1.2 Artificial Intelligence**

AI based handover is where the system learns user patterns and dependencies and tries to find the most optimal solution through its learning. AI handover can include forms of ML as well as other forms of evolutionary learning. In comparison to the multi-connectivity solutions, AI is less expensive as it is mainly software driven, therefore, it would need little to no extra equipment. Although if implemented incorrectly, it can provide a few drawbacks. These drawbacks will be highlighted throughout this section.

To begin with, the authors in [59] addresses the issue of poor QoE for a UE when the Reference Signal Received Power (RSRP) and Reference Signal Received quality (RSRQ) factors are reduced by the presence of obstacles. The authors propose a ML based algorithm using a feed-forward neural network (FFNN) and compares it to the A3 event-based logic. Two metrics are used, these are, the probability of a successful file download (logical output, 0 or 1), and the duration of the download for it to be successfully completed (single output, in seconds).

The simulated scenario used 3 eNodeBs, 3 UEs and an obstacle partially obstructing coverage between macro cells. With the Machine Learning based handover scheme, the authors achieve a 45% increment in the number of completed downloads (this can correlate into a reduction in

HOFs, and the number of handovers required). Additionally, a decrease of 25% in file download time was achieved, this may insinuate the algorithm as a lower latency solution.

Even Though the authors in [59] highlighted sizeable improvements, there is no consideration for the presence of small cells, as they considerably reduce the impact of obstacles on the UE's QoE. The use of small cells may make this handover algorithm redundant if, the cells are placed strategically where obstruction interferences are minimal.

If obstacle interferences are not considered, the authors in [60] aim to optimize handover in large scale wireless systems that support mobile internet of things (IoT) devices. A two-layer framework is proposed to optimize the handover system; the first layer partitions UEs with different mobility patterns into clusters, where mobility patterns that are similar are in the same cluster. Then, within each cluster, an asynchronous multiuser deep reinforcement learning (RL) scheme is developed, to control the handover processes across the UEs in each cluster.

Supervised learning was used during the initialization phase of the DNN before the execution of RL, to mitigate the random negative effects of exploration while the system is learning.

The simulation environment contained 3 zones of 6 small cell base stations, there are up to 1200 UEs in the given area. The authors compared it to the 3GPP algorithm (from chapter 2.2). The proposed minimizes the number of handover occurrences while ensuring a certain throughput, outperforming the 3GPP protocol by up to 80%, but this is only true when the number of UEs is minimal.

The critical issue highlighted in [60] is that the algorithm suffers from a delayed gradient, which decreases the test accuracy as the number of UEs increase. This poses as an issue in 5G, as the number of densely populated UEs will increase. Additionally, this could insinuate that the DL algorithm in [60] has a higher computational complexity/ strain when compared with [59]. This limits/ decreases the handover performance as the number of users increase.

### **3.1.3 Self-Organised Networks**

Self-organised networks (SONs) are an automation technology that consists of events for configuration, optimization, diagnostics analysis, protection, and healing of cellular networks, this in turn, makes them a simpler and faster solution to incorporate [61].

In [62], the authors propose the use of an energy efficient self-configured handover algorithm. The aim is to reduce the number of unnecessary handovers that occur in vehicles when traveling at certain speeds based on the time of day. The proposed scheme builds on the 3GPP logic already used in today's cellular networks, and it utilises these parameters to calculate a weighted sum. Two variations are proposed, the A3 offset + hysteresis value, this is changed to be based off hourly traffic. Subsequently this value varies a normally distributed TTT interval. The scheme is compared with an A3 handover logic.

The simulation consisted of 4 eNodeBs and a varying set of UEs (from 21 to 210) travelling at different speeds throughout the day (from 50-100kmph). The two performance indicators used to evaluate this are, the HOF ratio and the Ping-Pong rate. The overall improvement to the HOFs and Ping-Pong effects, was around 15% and 17%, respectively. Additionally, the authors developed a performance metric to combine the Ping-Pong ratio, HOF ratio, and the power consumption of the cell. This metric showed an overall improvement of approximately 11%, compared to A3 handover logic.

Although the algorithm in [62] provides good improvements with its minimalistic parameters, the authors do not discuss the weightings of the handover performance equation. It can be assumed that as the speed of the UE increases, the A3 + Hysteresis offset would decrease, causing a TTT interval reduction to support UE connectivity. Consequently, this would cause more frequent measurement reporting, resulting in increases in power consumption and signalling overheads. Furthermore, a lower throughput would be realized.

### 3.1.4 Summary

The table below summarises the literatures reviewed for single tier handovers.

Table 3.1-1: Summary and review of single tier handover literatures.

Type of solution	Ref. Article	Summary of potential solutions	Solution benefits and drawbacks
<b>Multi-connectivity</b>	[57]	Provides an optimised solution by using an anchor-based multi-connectivity architecture. Derives compact expressions of handover probabilities through stochastic geometry analysis.	Benefits: <ul style="list-style-type: none"> <li>- Reduced handovers.</li> <li>- Low latency.</li> <li>- Good throughput.</li> </ul> Drawbacks:
	[58]	A random-access channel (RACH)-less scheme is proposed. Combines with a make-before-break (soft handover) method.	<ul style="list-style-type: none"> <li>- High signalling overhead.</li> <li>- Low channel efficiency.</li> <li>- Expensive.</li> </ul>
<b>Artificial Intelligence</b>	[59]	The authors propose a machine learning (ML) based algorithm using a feed-forward neural network (FFNN) to improve the QoE of the UE in the presence of obstacles.	Benefits: <ul style="list-style-type: none"> <li>- Reduced handovers.</li> <li>- Reduced HOFs.</li> <li>- Low latency.</li> <li>- Good throughput.</li> </ul> Drawbacks: <ul style="list-style-type: none"> <li>- Does not consider adding small cells, makes algorithm redundant.</li> </ul>
	[60]	A two-layer framework is proposed; first layer partitions the UEs with different mobility patterns into clusters. Then, an asynchronous deep RL scheme is developed to control the processes across UEs in each cluster.	Benefits: <ul style="list-style-type: none"> <li>- Reduced handovers.</li> <li>- Good throughput.</li> </ul> Drawbacks: <ul style="list-style-type: none"> <li>- Suffers from delayed gradient (accuracy decreases when UEs increase)</li> <li>- High computational strains.</li> </ul>
<b>Self-Organised Network (SON)</b>	[62]	Proposes the use of an energy efficient self-configured algorithm. Two variations are proposed, the A3 offset + hysteresis value, which subsequently varies a normally distributed TTT interval.	Benefits: <ul style="list-style-type: none"> <li>- Reduced handovers.</li> <li>- Reduced HOFs.</li> <li>- Lesser Ping-Pongs.</li> </ul> Drawbacks: <ul style="list-style-type: none"> <li>- Less power efficient and low throughput for fast moving UEs.</li> </ul>

## 3.2 Multi-tier handovers

In addition to the single tier metric requirements multi-tier handovers are required to put more emphasis on the user mobility and user throughput requirements. The reasons why are explained in the points below:

- User's mobility: If the system can handover between tiers to accommodate varying user mobility scenarios this would help reduce the computational strain and signalling overheads for the system, due to executions of frequent handovers.
- User's throughput requirements: These require to be met to improve the overall QoE that 5G will require in critical applications. The use of the multiple tiers allows the network to adjust to the UE's needs by switching between various throughput mediums.
- Number of multi-tier handovers: These should be limited, as this can cause larger signalling overheads compared to single tier handovers, due to increased coordinating messages between tiers, including vertical handover (between different RANs/ RATs).
  - o This could also be linked to Ping-Pongs due to the power variations between different types of tiers.

Multi-tier handovers are usually more complex to perform and require a more comprehensive approach to achieve significant improvements in handover logic and/ or process. The articles below discuss the complexities of multi-tier environments and attempt to improve handover performances. The next sub-sections are divided into various classifications, to collectively assess the benefits and drawbacks that they pose.

### 3.2.1 Radio frequency mapping

The proposal in [63] aims to reduce abrupt behaviours such as Ping-Pong effects in multi-tier networks. It is based on predicted incomplete channel states from a Radio Environment Map (REM). In particular, the work is trying to improve mobility performance by reducing the overall number of handovers, without sacrificing network capacity. A REM is a heat map for the RSSs in a specific band. The environment consisted of 19 macro cells with 8 small cells and 5 users per macro cell site. The user uses 2 metrics to highlight improved performance, the Ping-Pong ratio, and the radio link failures (RLF) per user.

The performance was compared to the traditional LTE handover logic. The algorithm postpones handovers for as long as possible until a strong cell which will stay strong for at least a certain number of seconds shows up (in this case it was 2.5s). The number of Ping-Pong handovers was reduced by around 90% at higher user mobilities (30-120kmph), but the improvements reduce dramatically down between 5 to 10% when the user mobility is low (3kmph). It was observed that the number of RLFs was not improved, it actually degraded by a 5-10%, this can be a consequence of the long delays before a handover is executed.

From the research above, RLF can become a problem when such long delays of 2.5s is used, as this corresponds to a 40m travelling distance for fast moving users. Additionally, one of the parameters highlights an additional 120ms for recovery of the RLFs. Plus, the handover preparation and execution time of 90ms. This adds up to around a 2.7s delay from when the user detects that the target cell is a candidate. This is far too long to evaluate a handover, as it leads to RLFs (especially for fast moving users) and reduced throughputs (by at least 5 to 10%).

Subsequently, the authors in [64] uses a very similar method when they propose to address the issue of multi-RAT integration and handover management that 5G will cause. The goal was to dynamically adjust the handover triggering instance so that a better throughput can be achieved. The novel approach exploits the RF coverage maps to determine the optimum instance for handover triggering. The authors use geometric means to map out the exit regions of the base stations and UE round-trip time (RTT) estimations, to predict what the handover latency is compared to the packet latency.

The environment consists of a UE that has WiFi and WiMAX network access options. The authors have set a fixed velocity of the UE over the duration of a 550 second telephone call (which is low in data size). The algorithm builds on current 3GPP VHO schemes and uses an optimised algorithm using predicted RTT values. Based off the set RTT values, the author achieves a 20% optimisation of the overall throughput, and a 15% increase in the time spent in the preferred network.

Though the authors achieved these improvements, the key metric of handover Ping-Pongs is not addressed, as mentioned earlier, if vertical handovers are occurring frequently to maintain connection to the preferred network, it would result in undesired signalling overheads. Additionally, the environment is very limited and only considers one UE, thus, it is not able to capture user interferences correctly.

### **3.2.2 Multi-connectivity**

As reviewed in section 3.1.1, the same theories of multi-connectivity are employed in this section, however, these algorithms are adapted for more complex scenarios.

In [65], the issue of high handover latency and signalling overhead due to unnecessary frequent handovers are addressed. The handover scheme proposes that the UE is always connected to two 5G milli-meter access points (mmAPs) and one LTE BS. The prediction scheme compares the RSS of the active set of 5G-mmAPs and the one LTE-BS to another candidate LTE-BS or 5G-mmAP. If the candidate device has a better RSS, the network will initiate a handover request to the core network.

The proposed algorithm is compared to two systems, A3 event-based logic, and the second is the current scheme, but without prediction. The simulation environment consists of one LTE-BS (that covers the whole simulation area of 1 square kilometre), 100 5G-mmAPs, and 50 UEs. The

proposed algorithms average throughput is shown to be better than the other methods in higher mobility scenarios (30-50kmph) by 12%. Additionally, the number of handovers is also reduced by 5%.

An observation that was made with this proposal, is that it lacked a comprehensive study on multi-tier handover, the focus was only on the 5G-mmAPs. There was no analysis done for the handover between LTE-BSs. Furthermore, in addition to the signalling overhead issues discussed in the section prior, the channel efficiency is also reduced, since the UE uses three forms of connectivity to achieve minor improvements. This issue can pose as a major constraint when this is deployed in a populated environment.

The same channel efficiency and signalling overhead issues are accentuated in [66], where the authors aim to address the issue of mmWaves being highly susceptible to blockages and degradation in channel quality. The proposed solution consists of a heterogeneous dual-connectivity solution that is connected to the 4G AP as well as to the 5G AP. This provides rapid switching from 5G to 4G for failures on any one link. Additionally, the authors add the complexity of including static and dynamic TTT delays. The simulation environment consisted of 4 eNBs, 1 LTE eNB, 3 mmWave eNBs and 1 UE.

The authors compare the proposed algorithm with the conventional 3GPP algorithm being used for vertical handover. When comparing the number of handovers executed, the 3GPP algorithm proved to be more efficient by 5-10%, as the dual-connectivity approach requires handovers to occur more frequently as user requirements change. The overall packet loss ratio was reduced, due to the dual connectivity attribute of the algorithm, providing a more reliable connection. Additionally, the latency of the handover reduced by 70% and the overall throughput improved by 2-5%.

A combination of soft and hard handover solution has also been proposed for VANETs). In [67] the authors provide a decision algorithm for horizontal and vertical handover. The algorithm that is proposed uses circular geometric calculations to model the cell's coverage, relying on the vehicle's GPS coordinates to accurately trace the path. The algorithm uses soft handovers between roadside units (RSUs) and hard handovers between an RSU and BSs. The algorithm considers a log normal pathloss and shadow fading effects. A method that was used to counteract these effects was by implementing a sliding window technique. Additionally, the handover latency is considered when the algorithm executes its decision, resulting in a combination of the cell with the lowest latency and the best QoS. Software simulations of various scenarios were performed to understand how the algorithm adapts to the changes in vehicular velocities. The proposed solution was compared with the 3GPP threshold and signal hysteresis methods, to see if the number of handovers and HOFs are reduced. The algorithm provided a reduction of 30% in unnecessary handovers, and a 25% reduction in HOFs at speeds of 100kmph.

It was noticed that this improvement is reduced dramatically as user mobility reduced. In addition to this, two other major drawbacks related to the radio equipment are listed below:

- Assuming that the GPS signals are transmitted within 100ms every time. The signal overhead for the RSU and the BS will increase dramatically as the density of vehicles grow. Therefore, the cost of the equipment (on the network side) to support this will increase.
- The system present in the vehicle would require at least three separate systems to be able to execute handovers, which further would add to the cost of the system.

### **3.2.3 Multiple Criteria Handover**

In this section, authors aim to solve multi-tier handover issues with different multi-criteria algorithms. As this section is quite dense, it will be broken up into two sections. The first will consist of an article that discusses a novel multi-criteria decision making (MCDM) model, and three predefined models. Then, the next sub-section will discuss the other remaining non-conventional methods of MCDM improvements.

#### **3.2.3.1 Predefined models**

Highlighted in [68], the authors aim to improve the vertical handover decision making. The MCDM algorithm selects the network that best meets the UE's connectivity requirements. The authors state that this algorithm can be used for other highly dynamic mobile networks. An algorithm called NAIRHA (Neighbourhood-aware vertical handover algorithm) is developed, it has 3 different modules; a GPS module that calculates the navigational route and geolocation, secondly, a neighbourhood database that stores information in the on board unit regarding the current and soon-to-be reached neighbourhoods. The last module, called the useful coverage time, calculates the time that the UE spends in the cell to achieve the peak data rate from that cell. When selecting the most suitable network, there are 4 metrics that the algorithm considers: throughput, latency per packet, packet loss ratio and price per MB. The authors use vehicles moving at a constant speed of 32km/h over a 5.5km distance between two universities. The environment includes one UMTS, eight WiFi and three WiMAX APs with distinct data rates.

This algorithm is compared with 3 other algorithms called, MACHU, Geo-MACHU, and technology aware (the former two are prior proposals done by the group). The results show that the amount of vertical handover events with the proposed algorithm has dramatically reduced by around 50%. When comparing the other parameters such as, throughput, and packet loss, the proposed performed better than the rest by, 35% and 90% respectively.

The algorithm above provides a robust solution for vehicular VHO systems. However, the authors do not show any comparisons to the current algorithms used today. Furthermore, there are some other concerns described in the points below:

- Only static velocities were analysed, since vehicular networks will have UEs with dynamic velocities, it would be interesting to see what these results look like when the speed is varied.
- Memory of the neighbourhood database, as there was no mention on how the database was being updated when the memory was full. Additionally, based on how often the data refresh occurs, there were no calculations done on how fast the database would reach maximum capacity.

Shifting to more established MCDM methods, [69] attempts to solve the issue of seamless connectivity in heterogeneous wireless networks. To address the issue the authors propose a novel algorithm. Consists of a VHO management scheme, that is triggered based off the data rates required by the UE. The authors use the Grey Relation Analysis (GRA) decision scheme based on delay, jitter, bandwidth communication cost, and network load. A feature of this algorithm that would leads to an improved performance is the multi-threshold mechanism for controlling the scanning window time. The authors simulate the proposed scheme on multiple mobility scenarios (ranging from 1 to 9 m/s) with three different networks, WiFi, WiMAX, and LTE.

The algorithm was compared to the IEEE 802.21 scheme (Media Independent Handover standard). It shows that as the number of mobile nodes increases to 40, the handover delay is reduced by approximately 66% (from 400ms to 100ms), subsequently, the throughput over a 40-minute period of the simulated time, increases by around 5%. Lastly, the frequency of handovers is reduced by around 55%.

It is noticed that the battery usage on the mobile node for this algorithm is higher than the other compared schemes. Over a 11-hour simulated time it is 5% more power hungry. This could suggest that the measurement reporting frequency is higher than the compared algorithms. The enhancements become minimal when the user's battery is drained as the duration of a session increases.

Secondly, the authors in [70] talks about the issues related to the high number of handovers, signal interference, and signalling overhead, due to the densification of small cells during 5G deployment. The authors propose a novel multiple attribute decision making method, TOPSIS (Technique for Order Preference by Similarity to Ideal Solution). For this proposal, the authors propose 2 modified TOPSIS methods for handover management; the first method incorporates entropy weighting for handover metrics weighting, and the second proposed method uses a standard deviation weighting technique to score the importance of each handover metric. The simulation environment is a two-tier heterogeneous network consisting of 1 macro-cell and 50 small cells within the range of the macro-cell. There are multiple UEs placed within the simulation software and set to move uniformly in a randomly selected direction.

The authors compare the algorithm to three conventional methods, the 3GPP method, the network controlled HO method called NCH, and another TOPSIS method which uses a predefined weighting vectors with fixed values. The algorithm has a reduction in the number of handovers from small cell to small cell by 40%, for low mobility scenarios. Moreover, the number of handovers from small cell to macro cell for higher mobility scenarios is lower too, by 50% at 100kmph. For the second set of results, the RLF rate was measured, in this, the proposed method had a significantly lower RLF rate by 85% when in high mobility scenarios (80-100kmph). Lastly, the user mean throughput was 5% better when travelling at 100kmph.

It can be noticed that, although the RLF reduced dramatically as mobility scenarios increased, the throughput only showed a 2-5% boost. This can indicate that the frequency of measurement reporting is increased to maintain the connection, consequently, signalling overheads are increased.

Lastly, the authors in [71] attempt to address the issues of user mobility and its handover management in future 5G networks. A novel approach is taken by using a Markov Decision Process (MDP) to optimise the QoS that the UE experiences in mmWave heterogeneous networks. The authors propose to add an elimination method to the MDP by exploiting unique handover properties to improve the computational efficiency. The elimination method removes the suboptimal actions from each state to reduce the computational complexity. Additionally, three states are proposed for this MDP, these are: the UE's network connection, location, and velocity. The simulation environment consisted of 2 macro base stations each having a radius of 500m, and a varying density of small cells, from sparse to dense (20 small cells as per the 3GPP technical standards) scenarios.

This is compared with four other algorithms, these are: conventional signal-to-interference-plus-noise ratio (SINR)-based, a simple additive weighting based on analytic hierarchy process (AHP), Q-learning based and lastly, a state-action-reward-state-action (SARSA) scheme. The authors compare performances of the algorithms as the density of the small cells increases. The proposed algorithm's mobility reward is better than the competitors by 5%, although, in high mobility scenarios margin improves to around 10%. Furthermore, the algorithm had a considerably better overall reward for the UE by at least 50% (this considered a weighted sum of the downlink throughput and handover cost {control signal overhead and service delay}). This can translate to approximately 50% less computational strains due to lower complexities.

The algorithm in [71] provided some key improvements in the scope of throughput and user latencies. However, impacts on key performance indicators such as, the number of handovers and Ping-Pong effects were neglected.

### 3.2.3.2 Other models

The work in [72] addresses the issue of the long delays that occur during multi-tier handovers for delay sensitive UEs. The authors focus on macro-cell to small-cell (M2S) handover, as it is proposed that handover from S2M can be analysed in similar ways. The solution optimises the handover delay performance by combining three different approaches: stochastic geometry, resource reservation schemes and convex optimization theory. The authors propose three improvements, optimising the probability factor for initiating handover to the targeted small cell. Then an effective capacity model that captures the maximum arrival rate that can be supported. Lastly, the resource reservation is improved by applying a scheme that adopts a fixed reserved channel, this effectively reduces the blocking probability of handovers.

The BS deployments are in terms of density, for macro cells and small cells, there is one every  $1\text{km}^2$  and  $0.1\text{km}^2$ , respectively. A traditional 3GPP algorithm is used for comparisons, this algorithm selects the best BS based on the average RSRP. The handover probability factor reduces from approximately 95% to 55% as the ratio of delay sensitive UEs increase by a factor of 10. The second portion of the results depict the effective resource allocation, for executing handovers. When compared to the reference algorithm, it can be noticed that the blocking probability is reduced by around 17% when the ratio of delay sensitive UEs increases by a factor of 10.

It can be noticed that in [72], a large portion of the bandwidth is used up as the ratio of delay-sensitive users increases, which reduces the blocking probability of a handover occurring. This causes a trade-off of the effective capacity, which is consequently reduced. Additionally, the proposed scheme that quantitatively shows that their algorithm can be used to serve delay-sensitive users, but this is at a cost of capacity, which will become a lot more critical in future networks when the densification of users and data rate requirements increase.

Lastly, in [73] the authors attempt to reduce the number of frequent handovers caused by fast moving UEs between small cell coverages. The authors propose an algorithm that will handover users to the macro cell from a small cell based off their mobility behaviour. An additional aim is to reduce the number of Ping-Pongs that occur during handover. The authors classify the users into three categories: fast-moving, Ping-Pong and the remaining users. Fast-moving users are handed over to the macro cell layer, where the authors develop an estimation algorithm for the time that the UE spends in each small cell. Ping-Pong users are identified by an algorithm that finds patterns in the serving cell history where dwell times are low, these are considered Ping-Pongs. For all the other users, a remaining service time is introduced to determine how long the UE is going to remain the current cell.

They simulated 3 macro cells, and 20 small cells. The macro cells were deployed in a three-sectored hexagon grid, and the small cells were deployed in a Manhattan style layout inside the macro coverage. Two performance metrics were used: the number of handovers, and the

prediction accuracy of the algorithm. The authors compare this algorithm with three other algorithms, a Ping-Pong reduction algorithm, a mobility behaviour-based user classification algorithm and a sojourn time estimation-based small cell-selection algorithm. Each of these algorithms is compared for Ping-Pongs, fast moving users, and the other types of users. The scenario showed that the algorithm was around 70% lower in the number of handovers, when both Ping-Pong and fast-moving users were present. Moreover, there is an improvement in throughput by approximately 10%.

Nevertheless, it can be noticed that the bandwidth is the same for both types of cells in this simulation, thus, it could simply be that the throughput is a measurement of the UE having a better channel quality indicator (CQI) instead of an algorithm that runs faster. Other missing considerations include the presence of obstacles on the macro cells which could degrade the UEs QoS and/or QoE, and how frequent measurement reporting occurs.

### **3.2.4 Artificial intelligence**

In addition to the articles analysed in section 3.1.2, this section divulges into details relating to the multi-tier improvements that AI based schemes provide. The three reviews below consist of a fuzzy logic algorithm, a ML algorithm and a DL algorithm, respectively.

An AI based contribution that is highlighted in [74], proposes to use fuzzy logic to solve issues relating to redundant handovers and HOF ratios in dense small cell networks in LTE. The novel self-optimising system proposed, analyses the UEs velocity and radio channel quality, to adapt hysteresis values for handover decisions. The inputs for the system were, the UEs velocity, the RSRP and the RSRQ. The proposed algorithm was compared to four different algorithms, these were: Best Connection (BC), Conventional LTE handover, Fuzzy Multiple-Criteria Cell Selection (FMCCS) integrated with TOPSIS, and Self-Tuning Handover Algorithm (STHA). The proposed algorithm reduced the average number of handovers by 20%, the overall HOF ratio by 25% and the Ping-Pongs events to less than 2.5% (which results in a 50% decrease in Ping-Pong events compared to the other algorithms, assuming a minimum stay time of 10s).

However, the impact on latency and throughput were not analysed, which could insinuate a possible increase in computational strains, resulting in a reduction of UE QoE.

In [75] the authors address the inefficiencies of handover for in-building systems. The proposal is to optimize these inefficiencies by using machine learning and data mining techniques by developing a clustering algorithm based off shapelets and wavelet decompositions at the cell's edge. The environment consisted of two in-building systems, one where the university food court is located (Building A), and the other in the health and sport building (Building B), and a 3 sector macro-cell. This experiment captured measurement reports (MRs) at a rate of approximately 300ms. The feature of this algorithm that led to an improved performance was the clustering algorithm, in just a few seconds of MRs to achieve accurate predictions.

The authors also talk about the three scenarios that have been simulated, all scenarios relate to the loading of the macro-cell and in-building systems, assuming the UE is exiting the building. The user develops objective functions for all three scenarios to achieve the optimal operating point (OP), which is a combination of the A2 and A3 handover thresholds, and a TTT period. The optimal OPs were based off, HOF rate, Ping-Pong rate and the average achievable data rate. The algorithm also compares data rate gain as the A3 threshold is altered. The data rate gain that this algorithm provided varies between 25 to 65% compared to the static A3 algorithm.

The authors discuss about Ping-Pong, HOF rates etc., but does not provide evidence of improvements in these areas. Only throughput gains are provided. There are a pair of other drawbacks/ gaps that can be noticed:

1. There are no discussions on mass UE movements, such as leaving the building when a class finishes (which is very likely). This could cause vast disruptions to the proposal.
2. Secondly, the simulation is only focused on users exiting the building, which may suggest the algorithm does not perform as well when the scenario is reversed.

Lastly, a DL approach is analysed, where in [76] the authors propose to significantly reduce service traffic that is transmitted through communication channels in 5G and optimise handovers. By using a GRU RNN, the algorithm will provide a rapid response to changes in the environment. The GRU was used to predict how many users would be in a particular cell for a given time of day. The prediction scheme varies the size of the cell coverages by a factor  $K$ , based off the time of day. For example, in overloaded scenarios, the cell that is overloaded will have its coverage reduced by a factor of  $K$ , and the surrounding less loaded cells would have their coverage areas increased by a factor of  $K$ . Subsequently, this allows underloaded cells to be easily handed over to, and overloaded cells to become harder to connect into.

The authors use supervised learning for these predictions and compares these predictions to a DL LSTM over 300 epochs. Both systems provided very accurate and similar results for prediction. It shows that the GRU is able to achieve a better result than the LSTM in a short time frame, although, as the number of epochs increase, the LSTM becomes more accurate than the GRU system. It is shown that the GRU is able to accurately model daily user traffic at an accuracy of 90%, this can be improved if an increased sample size is taken.

A major gap that is realised when analysing this literature is that the authors do not state when all cells are overloaded, this would cause the coverages to become so small that dead zones in various areas within the network appear, causing mass RLF failures for mobile users.

Furthermore, the authors mention that the LSTM is a better option if the sample sizes or the number of epochs increase, which questions whether a DL LSTM could have been used instead, to provide more accurate results for big data analytics.

### 3.2.5 Cloud C-RAN

A more modern way of implementing handover that is beginning to become more popular, is by utilising the cloud C-RAN, due to the low latency and seamless handover requirements of 5G. This contributing factor is why many industrial companies are making a push for cloud-based solutions. Cloud based handover can provide a few significant benefits in terms of lower latency and larger throughput solutions, when compared to systems that do not utilise the cloud.

This can be observed in [77], where the authors address the issues due to with cooperative interference mitigation and handover management in, heterogeneous cloud small cell networks (HCSNet). Firstly, the author employs a network architecture that combines the cloud RAN with small cells. The authors specifically target UEs moving between macro cells and small cells. Then a low complexity handover management scheme is proposed, and its signalling procedure is analysed in a HCSNet. The authors develop this algorithm based off UE speed estimations (using an autocorrelation function) and UE latency requirements. Additionally, to avoid user interference at the cell's edge, the authors propose a cooperative multipoint (CoMP) joint transmission clustering scheme, specifically the affinity propagation methods. The results show that the signalling overhead reduces significantly by at least 40% and 90% related to the call holding time and the proportion of high mobility users, respectively (this would result into an overall throughput increase).

Although, the solution provides a very good reduction in signalling overheads, there is a lack of a few key performance indicators (highlighted at the beginning of Chapter 3.1 and 3.2), such as the number of handovers, Ping-Pongs, handover latency.

In comparison to the previous effort, the authors in [78] focuses on the latency benefits of using a Cloud C-RAN architecture. As this is an important enabler for URLLC and ultra-high reliability services for high mobility IoT applications. The authors analyse and compare the performances of different Cloud RAN architectures, then develop a new concept for reducing the handover preparation time, called early admission control (EAC). This algorithm is done with respect to synchronous handovers without random access. This is then compared to today's D-RAN configurations. A diverse range of interface and processing latency assumptions were used. The results show a clear reduction in handover preparation time of up to 30% when compared to cloud C-RAN architectures that did not use their EAC algorithm, although, this improvement proves to be larger when compared to D-RAN architectures, offering improvements greater than 60% (resulting in better throughput and lower signalling overheads).

The research performed in [78], can be considered for a multi-tier environment for current 5G and future looking cloud C-RAN enabled technologies. Furthermore, the authors could consider moving the EAC preparation closer to the edge of the RAN to achieve an even lower latency. This will be discussed further in Chapter 5 where the handover algorithm is proposed.

### 3.2.6 Summary

The table below summarises all the literatures that have been reviewed. Some literature benefits and drawbacks are grouped as they have the very similar benefits and drawbacks.

Table 3.2-1: Summary and review of multi-tier handover literatures.

Type of solution	Ref. Article	Summary of potential solutions	Solution benefits and drawbacks
<b>Radio frequency mapping</b>	[63]	Proposes a Radio Environment Map (REM) to improve mobility performance by reducing the overall number of handovers, without sacrificing network capacity.	Benefits: <ul style="list-style-type: none"> <li>- Reduced handovers.</li> <li>- Reduced RLFs.</li> <li>- Less Ping-Pongs.</li> </ul> Drawbacks: <ul style="list-style-type: none"> <li>- Lower throughput.</li> <li>- Slow response rate to fast moving UEs.</li> </ul>
	[64]	The goal was to dynamically adjust the handover triggering instance so that a better throughput can be achieved. The novel approach exploits the radio frequency coverage maps to determine the optimum instance for handover triggering.	Benefits: <ul style="list-style-type: none"> <li>- Reduced handovers.</li> <li>- Increased time spent in desired cell.</li> <li>- better throughput.</li> </ul> Drawbacks: <ul style="list-style-type: none"> <li>- Ping-Pongs not mentioned.</li> <li>- Limited situation, one UE only moving between cells.</li> </ul>
<b>Multi-connectivity</b>	[65]	The scheme proposes that the UE is always connected to two 5G milli-meter (mm) access points (APs) and one LTE BS, to improve high handover latency and signalling overhead due to unnecessary frequent handovers.	Benefits: <ul style="list-style-type: none"> <li>- Reduced handovers.</li> <li>- Low latency.</li> <li>- Good throughput.</li> </ul> Drawbacks: <ul style="list-style-type: none"> <li>- High signalling overhead.</li> <li>- Low channel efficiency.</li> <li>- Expensive.</li> <li>- No comprehensive study of multi-tier handovers. Mainly focuses on a particular tier and uses the other tier as a fall-back system.</li> </ul>
	[66]	The solution consists of a connection to a 4G AP as well as a 5G AP, this is due to mmWaves being highly susceptible to blockages and degradation in channel quality	
	[67]	The algorithm uses circular geometric calculations (to model the cell's coverage), relying on the vehicle's GPS coordinates to	

		accurately trace the path. The algorithm uses soft handovers between RSUs and hard handovers between an RSU and BSs	
<b>Multi-criteria</b>	[68]	The MCDM algorithm that selects the network that best meets the UE's connectivity. Proposed a novel algorithm called NAIRHA. It consists of three different modules: a GPS module, a neighbourhood database, and a useful coverage time module. This is at the UEs end in an OBU.	<p>Benefits:</p> <ul style="list-style-type: none"> <li>- Reduced multi-tier handovers.</li> <li>- Low packet loss.</li> <li>- Low latency.</li> <li>- Good throughput.</li> </ul> <p>Drawbacks:</p> <ul style="list-style-type: none"> <li>- Majority of the HO calculations at UE.</li> <li>- Expensive.</li> <li>- No study on varying user movements.</li> </ul>
	[69]	The authors use the Grey Relation Analysis (GRA) scheme based on delay, jitter, bandwidth communication cost, and network load. To improve seamless connectivity and meet the data rates required by the UE.	<p>Benefits:</p> <ul style="list-style-type: none"> <li>- Reduced handovers.</li> <li>- Low latency.</li> <li>- Good throughput.</li> </ul> <p>Drawbacks:</p> <ul style="list-style-type: none"> <li>- Power hungry, drains UE batteries due to increased measurement reporting.</li> <li>- HOFs not mentioned.</li> </ul>
	[70]	Proposes a novel multiple attribute decision making method, TOPSIS. For this proposal, the authors propose 2 modified TOPSIS methods, the first method incorporates entropy weighting for handover metrics, and the second uses a standard deviation weighting to score the importance of each metric. To improve, the number of handovers, signal interference, and signalling overhead, in 5G small cells deployment.	<p>Benefits:</p> <ul style="list-style-type: none"> <li>- Reduced multi-tier and single tier handovers.</li> <li>- Reduced RLFs.</li> <li>- Low latency.</li> <li>- Good throughput.</li> </ul> <p>Drawbacks:</p> <ul style="list-style-type: none"> <li>- Power hungry, drains UE batteries due to increased measurement reporting.</li> <li>- Higher signalling overheads.</li> </ul>

	[71]	A novel approach is taken by using a MDP to optimise the QoS that the UE experiences in mmWave heterogenous networks. The authors propose to add an elimination method to the MDP by exploiting unique handover properties to improve the computational efficiency. Three states are proposed, the UEs link, location, and velocity.	<p>Benefits:</p> <ul style="list-style-type: none"> <li>- Low latency.</li> <li>- Good throughput.</li> <li>- Low computational complexity</li> </ul> <p>Drawbacks:</p> <ul style="list-style-type: none"> <li>- Impact on number of handovers, signalling overheads and Ping-Pong effects not mentioned.</li> </ul>
	[72]	Addresses the issue of long delays that occur during cross cell tier handover for delay sensitive UEs. The solution optimises the handover delay performance by combining three different approaches: stochastic geometry, resource reservation schemes and convex optimization theory.	<p>Benefits:</p> <ul style="list-style-type: none"> <li>- Handover probability reduced.</li> <li>- Blocking probability reduced.</li> </ul> <p>Drawbacks:</p> <ul style="list-style-type: none"> <li>- The authors focus on M2S handover, assumes it acts the same for S2M.</li> <li>- Reduction in overall throughput.</li> </ul>
	[73]	The authors are trying to reduce the number of Ping-Pongs and frequent handovers caused by fast moving UEs between small cells. Proposes an algorithm that will handover users to the macro cell from a small cell based off their mobility behaviour. The authors classify the users into three categories: fast-moving, Ping-Pong and the remaining users.	<p>Benefits:</p> <ul style="list-style-type: none"> <li>- Reduced handovers</li> <li>- Reduced Ping-Pongs.</li> <li>- Good throughput.</li> </ul> <p>Drawbacks:</p> <ul style="list-style-type: none"> <li>- No analysis with the presence of obstacles.</li> <li>- How often measurement reporting occurs, may result in high signalling overheads.</li> </ul>
<b>Artificial Intelligence</b>	[74]	Proposes to use fuzzy logic to solve issues relating to redundant handovers and handover failure (HOF) ratios in dense small cell networks (in LTE). The novel self-optimising system proposed, analyses the UEs velocity and radio channel quality, to adapt	<p>Benefits:</p> <ul style="list-style-type: none"> <li>- Reduced handovers</li> <li>- Reduced Ping-Pongs.</li> <li>- Reduced HOF.</li> </ul> <p>Drawbacks:</p> <ul style="list-style-type: none"> <li>- No data on latency or throughput.</li> </ul>

		hysteresis margins for handover decisions.	- Possible high computational strains.
	[75]	the authors address the inefficiencies of handover for in-building systems. The proposal is to optimize these by using ML and data mining techniques (clustering algorithm based off shapelets and wavelet decomposition) at the cell's edge.	Benefits: - Good throughput - Low latency Drawbacks: - No data on HOF, number of handovers or Ping-Pong impacts. - Only focused on a one-way scenario (exiting UEs).
	[76]	The authors propose to significantly reduce service traffic that is transmitted through communication channels in 5G. By using a GRU RNN, the algorithm will provide a rapid response to changes in the environment. The prediction scheme varies the size of the cell coverages by a K factor based off the time of day.	Benefits: - Accurately models daily user traffic, up to 90%. - Can vary cell coverages for balanced loading. Drawbacks: - When all cells are overloaded RLFs increase. - Requires lot of data to learn. - Lower accuracy compared to LSTMs.
<b>Cloud RAN</b>	[77]	where the authors address the issues due to with cooperative interference mitigation and handover management in, HCSNets. To avoid user interference at the cell's edge, the authors propose a CoMP joint transmission clustering scheme, specifically the affinity propagation methods.	Benefits: - Reductions in signalling overhead. - Improved throughput. Drawbacks: - Only targets M2S environments. - No data for, number of handovers, Ping-Pongs, or latency.
	[78]	Focuses on the latency benefits of using a Cloud RAN architecture. As this is an important enabler for URLLC and ultra-high reliability services for high mobility IoT applications. This algorithm is done with respect to synchronous handovers without random access.	Benefits: - Reductions in signalling overhead. - Improved throughput. - Improved latency Drawbacks: - Only uses cloud, could improve latency by utilising MEC.

## 4 Research Methodology

To analyse and implement a simulation for the proposed theory in Chapter 5, a diverse range of software tools are required. This chapter will highlight these tools, implementations and performance metrics following the structure below:

Section 4.1 highlights the various simulation tools and parameters required to simulate:

- User mobilities (for cars, pedestrians, and cyclists)
- Buildings, roads, and footpaths.
- Carrier frequencies (FR1 and FR2 ranges), multiple bandwidths.
- Pathlosses, shadow fading, wall losses, LOS and NLOS losses,
- Base stations (Macro cells, Micro cells, and Femto cells)
- 5G gNB and UE schedulers, including full uplink and downlink transmissions.
- Deep learning and LSTM networks

Section 4.2 lists the assumptions made for the simulation of the systems.

Section 4.3 discusses the simulation environment parameters in detail.

Section 4.4 details the key optimisations performed on the simulator.

Section 4.5 emphasizes the key performance metrics that will be used to evaluate the results.

### 4.1 Simulation Tool

The chosen simulation tool was MATLAB. This software is readily available to students at the university, and it provides an extensive range of simulation libraries and applications to cater to the diverse requirements of this research. Additionally, MATLAB is a common tool used for academic simulations and it is very familiar to the author. The sub-sections below describe the various libraries and simulators that were required to develop the final simulator, to evaluate the 5G with MEC handover performance. The simulator segments described in the following sub-sections help model the process flow shown in Figure 4.1-1.

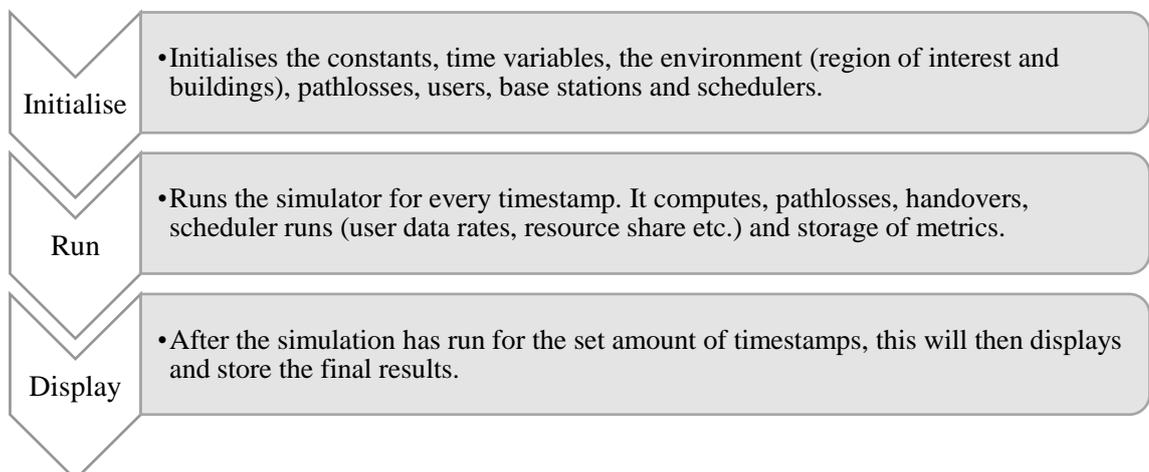


Figure 4.1-1: Simulator process flow

### 4.1.1 System level simulator

The system level simulator used is the Vienna 5G system level simulator [79] [80]. This simulator allows the user to simulate a complete 5G environment and it evaluates the average performance of large-scale networks by means of Monte Carlo simulations [81]. For this simulator there are two MATLAB toolboxes required:

- Statistics and Machine Learning toolbox [82]
- Parallel Computing toolbox [83]

The simulator provided a multitude of features that can accurately simulate real world scenarios with references to the 3GPP standards. For the simulations carried out in this thesis, the base station and user schedulers provided by this simulator is removed and replaced with the scheduler in Section 4.1.2. All other parameters will be highlighted in this section. The section will begin with all base station and user parameters. Then the various channel models, and lastly, the other miscellaneous items (such as building creations etc.). All the relevant pathloss models used in this simulator are compliant to LTE 3GPP standards (applicable for frequency ranges between 2 and 6 GHz).

#### 4.1.1.1 Base stations and users

Base stations:

For each base station, the simulator provided the following antenna types:

- Omnidirectional
- 3-sector
- 6-sector

With configurable number of antennas, azimuth angles (in degrees), gain values (in dBi), and transmit powers (in W). In addition to this, the remaining parameters in Table 4.1-1 discuss the applicability ranges.

*Table 4.1-1: Key base station parameters in the simulator, compliant to 3GPP [16] [79]*

	<b>Macro cell</b>	<b>Micro/ Pico cell</b>	<b>Femto cell</b>
<b>Coupling loss</b>	70dB	53dB	45dB
<b>Minimum distance along the ground (2D distance)</b>	35m	5m	2m
<b>Carrier Frequencies</b>	3D-UMa, 2 to 6 GHz	3D-UMi, 2 to 6 GHz	Free space, Any frequency.

The simulator also supported the placement of macro cells on top of buildings. Lastly, it included indoor to outdoor, and outdoor to indoor support for all types of base stations.

### Users:

The types of users were configurable, the author could simulate, pedestrians, low-speed vehicles (such as cars in urban or suburban areas) and high-speed vehicles (for highway scenarios or high-speed trains). Additionally, the simulator supported down to 1ms granularity for updates on user positions. Other relevant configurable parameters include:

- Number of transmitting and receiving antennas for heterogeneous environments.
- Channel model types (stated in Section 4.1.1.2).
- User heights (default is 1.5m)
- User velocities, these can be following a distribution or can be predefined.
- Transmit power and frequency (in W and Hz, respectively).

The user mobilities are mapped with a simulator defined in Section 4.1.3 and input into this simulator as predefined positions, further details are stated in Section 4.3.

#### **4.1.1.2 Channel models**

##### Small-scale fading models:

The small-scale fading models used for the simulator in this thesis were:

- Typical Urban (TU) model in 3GPP technical report (TR) 25.943 [84], for Pedestrians.
- Vehicle A (VehA) model in TR 25.890 [85], for vehicles and cyclists.

##### All other pathloss parameters:

The equations in Table 4.1-2 are from Table 7.2-1 in [17], these depict the pathloss models for 3D urban macro and micro cell environments in 4G (LTE and LTE-A). The frequency applicability ranges for both LOS and NLOS range from 2 to 6 GHz. The terminologies below describe characters used in the table that is subsequent.

##### Terminology:

$d_{3D}$  is the distance in 3D from the BS and UE antennas in metres.

$f_c$  is the carrier frequency in GHz (2 to 6GHz).

$d'_{BP}$  is the breakpoint distance (defined below).

$h_{BS}$  is the height of the base station antenna in metres (10 to 150m).

$h_{UT}$  is the height of the UE antenna in metres (1.5 to 22.5m).

$W$  is the street width in metres (5 to 50m).

$h$  is the average building height in metres (between 5 to 50m).

$PL$  is the pathloss (any subscripted text will be relevant to the parameters discussed in the specific row).

Table 4.1-2: Relevant pathloss parameters for the simulator [17] [86]

	3D-UMa	3D-UMi
<b>LOS pathloss (dB)</b>	Below breakpoint distance: $PL_{LOS} = 22.0 \log_{10}(d_{3D}) + 28.0 + 20 \log_{10}(f_c)$ Above breakpoint distance: $PL_{LOS} = 40 \log_{10}(d_{3D}) + 28.0 + 20 \log_{10}(f_c) - 9 \log_{10}((d'_{BP})^2 + (h_{BS} - h_{UT})^2)$ Breakpoint distance: $d'_{BP} = 4 h'_{BS} h'_{UT} f_c / c$	
<b>NLOS pathloss (dB)</b>	$PL_{3D-UMa-NLOS} = 161.04 - 7.1 \log_{10}(W) + 7.5 \log_{10}(h) - (24.37 - 3.7(h/h_{BS})^2) \log_{10}(h_{BS}) + (43.42 - 3.1 \log_{10}(h_{BS})) (\log_{10}(d_{3D}) - 3) + 20 \log_{10}(f_c) - (3.2 (\log_{10}(17.625))^2 - 4.97) - 0.6(h_{UT} - 1.5)$	$PL_{3D-UMi-NLOS} = 36.7 \log_{10}(d_{3D}) + 22.7 + 26 \log_{10}(f_c) - 0.3(h_{UT} - 1.5)$
<b>Total pathloss (dB)</b>	$PL = \max(PL_{3D-UMa-NLOS}, PL_{3D-UMa-LOS})$	$PL = \max(PL_{3D-UMi-NLOS}, PL_{3D-UMi-LOS})$
<b>Shadow fading Standard deviation (dB)</b>	LOS: 4 NLOS: 6	LOS: 3 NLOS: 4
<b>Wall loss (dB)</b>	$PL_{concrete} = 5 + 4(f_c)$ (concrete pathloss)	$PL_{concrete} = 5 + 4(f_c)$ (concrete pathloss)

Additionally, the simulator included, thermal noise powers (in dB) per sub carrier frequency for each network element, and the UE receiver noise figure (default value of 9 dB).

The simulator sampled all these pathlosses at 10ms intervals, although, it allowed for faster updates of channel models if speeds increased faster than the channel model was applicable for. The refresh rates can be updated if required.

#### 4.1.1.3 Miscellaneous

The simulator helped prove a default time interval of 1ms for the scheduler. Additionally, every 10ms the simulator had a cell association phase that was configured to either RSS or SINR. This method of handing over was removed, as this is not the correct way of implementing handovers. However, this cell association strategy was used to initialise base stations and users for the first timestamp. Section 4.1.4 will discuss the author's own developed handover process that was incorporated in this simulator.

For the simulation environment, the simulator provided support to create a region of interest where it applies relevant pathlosses and other parameters.

- This also included a region of interference (with a default factor of 1).

Within that region of interest, buildings and other obstructions were able to be defined. Buildings could be created with random locations (such as a Manhattan layout) or they can be predefined. Streets were subsequently created after the buildings.

Furthermore, this simulator allowed visual displays of the user movements and base station associations.

#### 4.1.2 Scheduler

The base station and UE scheduler used for this simulator is a 5G NR scheduler that is developed by MATLAB, it can be found in [18]. This scheduler is compliant with the 3GPP standards, these standards will be listed throughout this section, where applicable. For this simulator, one MATLAB toolbox is required, the 5G Toolbox [87].

Table 4.1-3 lists a few of the important operations that this simulator performs.

Table 4.1-3: MATLAB NR FDD scheduler operations for gNB and UE [18]

Equipment	Operation
<b>gNB</b>	<ul style="list-style-type: none"> <li>- Assigns uplink and downlink resources.</li> <li>- Sends uplink and downlink assignments to UEs.</li> <li>- Receives the Physical Uplink Shared Channel (PUSCH) transmissions from the UEs.</li> <li>- Adheres to downlink assignments for the Physical Downlink Shared Channel (PDSCH) transmission.</li> <li>- Receives feedback of PDSCHs from the UEs.</li> </ul>
<b>UE</b>	<ul style="list-style-type: none"> <li>- Sends pending buffer status reports to the gNB.</li> <li>- Receives the uplink and downlink assignments from the gNB.</li> <li>- Adheres to the received uplink assignments from the gNB for the PUSCH transmission.</li> <li>- Receives PDSCH transmissions from the gNB.</li> <li>- Sends feedback for the received PDSCHs.</li> </ul>

In addition to these operations, the scheduler allowed many customisations, a few key customisations are listed below.

#### Frame structures and Bandwidths:

This scheduler has support for both TDD and FDD. It allowed for slot based or symbol-based scheduling depending on the mode selected.

- TDD uses symbol based (full preconfigured scheduler example can be found in [88]).
- FDD uses slot based.

Additionally, it supported full configurations of bandwidths and sub carrier spacings in accordance with the 3GPP standards [16].

- Inclusive of 15kHz, 30kHz and 60kHz subcarriers.
- Able to support up to 100MHz bandwidths (for 30kHz and 60kHz, 50MHz for 15kHz).
- Can support up to 256 QAM (from 3GPP TS 38.214 Table 5.2.2.1-3 [89]).

This is configurable for both uplink and downlink bandwidths, plus the ability to change the limit of the maximum number of resource blocks that can be assigned to a UE.

Additionally, CQIs can be improved or deteriorated at a certain periodicity. Lastly, hybrid automatic repeat request (HARQ) can be toggled, and the number of processes are able to be varied. All wireless communications between UE and gNB have a set success probability, this will be detailed in Section 4.2 (where the assumptions are discussed).

Scheduler strategy:

Table 4.1-4 describes the simulator’s three scheduler strategies:

*Table 4.1-4: 5G NR FDD Scheduler options [90]*

<b>Scheduler Type</b>	<b>Description</b>
<b>Proportionally fair (PF)</b>	Tries to maximise total throughput of the network while maintaining a minimum level of service. <ul style="list-style-type: none"> <li>- This scheduler provides the user with a moving average data rate weight parameter that can be configured.</li> </ul>
<b>Round Robin (RR)</b>	Ensures that each user gets an equal share of the resources. Offers the best fairness for all users.
<b>Best CQI</b>	Assigns the most resource blocks for the user with the best CQI, although this is not the best at being fair.

More in-depth explanations can be found in [90]. Lastly, the PUSCH can be configured, the default value of 200µs is used to ensure that the assignments are received before the transmission time. The scheduler then relies on the correct information to schedule provided from two tables:

- the radio network temporary identifier (RNTI) logical channel configuration (LCC) shortened to RLCC table.
- the application configuration tables.

The parameters of both these tables are described in Table 4.1-5.

Table 4.1-5: RLCC and Application configuration parameters for the scheduler [18]

RLCC configuration parameters	Application configuration parameters
<ul style="list-style-type: none"> <li>- RNTI of the UE.</li> <li>- The UEs Logical channel identifier (LCID) and LCID group.</li> <li>- Sequence field length, either 6 or 12.</li> <li>- Maximum buffer for service data units (SDUs) (in number of packets).</li> <li>- Reassembly timer (in ms).</li> <li>- Entity type, for the RLC unacknowledge mode (UM) entity: <ul style="list-style-type: none"> <li>o unidirectional downlink</li> <li>o unidirectional uplink, or</li> <li>o bidirectional UM.</li> </ul> </li> <li>- Priority of the logical channel</li> <li>- Prioritized bit rate (in kbps).</li> <li>- Bucket size duration (in ms).</li> </ul>	<ul style="list-style-type: none"> <li>- RNTI of the UE</li> <li>- LCID</li> <li>- Packet Interval between two consecutive packet generations (in ms).</li> <li>- Size of the packet (in bytes), which can also be translated from required data rate.</li> <li>- Host Device, this is the device (UE or gNB) on which the application is installed with the specified configuration. The values indicate the application is configured on either: <ul style="list-style-type: none"> <li>o gNB side,</li> <li>o UE side, or</li> <li>o both UE and gNB.</li> </ul> </li> </ul>

Visualisations:

Lastly, this scheduler also allowed visualisations of:

- CQIs for each RB.
- RB assignments to a particular UE, for each RB in the base station.
- Every attached UEs data rates by their LCIDs.

Furthermore, as discussed earlier in Section 4.1.1, this scheduler from MATLAB was combined with the Vienna 5G system level simulator to develop a fully functioning uplink and downlink 5G system level simulator.

**4.1.3 User mobility simulator**

The predefined user mobilities are modelled using an application called the Driving Scenario Designer in MATLAB. This requires the use of the automated driving toolbox [91].

The key features that attracted the author to using this for developing user positions include:

- Modelling roads, vehicles, pedestrians, and cyclists.
  - o Speeds can be varied to simulate turning vehicles and cyclists.
  - o Delays can be added to simulate signals, parked cars etc.
  - o Sensors allow the author to measure localisations of all actors in a set environment.

- It allowed the author to simulate timestamps as low as 1ms (this is a requirement for the Vienna simulator for accurate user mobility calculations).

This toolbox provides an easy-to-use user interface that allowed the user to model all these parameters without the need for coding.

#### **4.1.4 Handover**

The handover logic and process that was added to this simulator follows key 3GPP standards, as well as some documents from industrial leaders for latencies and various other timing values.

- This is described in Chapter 5, as this is the benchmark algorithm used for comparisons.

In this simulator the A3 and A1 handover logic is implemented. Additionally, the handover process and the appropriate delays are implemented to simulate UE, BS, AMF and UPF communications during a handover.

Furthermore, the MEC far edge and aggregated layers are implemented to provide a lower latency implementation that can occur in the 5G environments today.

#### **4.1.5 Deep learning simulator**

For the deep learning section of this thesis, it was simulated in MATLAB using the Deep Network Designer Application. The appropriate library required was the Deep Learning Toolbox [92].

Key features that attracted the author to this were:

- All the coding for LSTMs NNs and other forms of ML are embedded into the toolbox, the user only requires designing the layers of the architecture.
- Allowed modifications of various layers.
- The sequence training and testing data is simple to input into the system.
- Visualisations of training shows the learning rates and losses.
- A final trained LSTM is outputted and is ready for classifying live data.
- Easy to use with completely configurable training options, such as:
  - o The Number of epochs
  - o Initial learning rates
  - o Gradient thresholds
  - o Shuffling of the training data
  - o Various solvers

The various solvers are defined in Table 4.1-6.

Table 4.1-6: Types of DL solvers in MATLAB [93]

<b>root mean square propagation (rmsprop)</b>	<b>stochastic gradient descent model (sgdm)</b>	<b>adam</b>
Learns per-parameter and is adapted based on the average of latest magnitudes of the gradients for the weight. This allows the system to perform well in fluctuating problems.	This maintains a single weight for all updates, which consequently does not change the learning rate during training.	This is a combination of the rmsprop algorithm as well another one called adaptive gradient algorithm (AdaGrad), which maintains the same per-parameter learning rate, but it is more for sparse gradients.

Lastly, for the deep learning simulator to learn, Microsoft Excel power query editor was used to easily model the simulated data in discrete dimensions. This data is then used by the simulator to easily comprehend and output a DL LSTM that can be used for classifications.

## 4.2 Assumptions

This section will list the assumptions that were made to simulate the environment.

1. The pathloss model used for 3D-UMa and 3D-UMi scenarios can be used to model this 5G scenario, as the frequency ranges are applicable for the types of base stations used.
  - a. Additionally, subcarrier frequencies of 15kHz are also being used, so this should generally be a good model for the simulation of these pathlosses.
2. Shadow fading standard deviation remains constant at a value of 4 for ease of calculations.
3. Wall losses are constant, and they use the carrier frequency of 2GHz (based off the macro cell's carrier frequency).
4. CQIs of the signal remains constant throughout a frame, there are no degradations or improvements within one frame.
5. There are no users travelling above 80kmph, therefore, correlation distance is sufficient to not require channel updates more than once every 10ms.
6. There are no trees or streetlights in the simulation environment for simplicity.
7. MEC far and aggregated edge deployments improve throughput and latencies by bringing virtualisations of the AMF and UPF to the aggregated edge. Other optimisations are not considered for this research.
8. All types of users have compatible hardware to connect to any of the stated carrier frequencies in the simulation.
9. All users transmit powers are the same.
10. The packet success probability of the UE or gNB receiving the packet of data through the air interface, was described by the equation below:

$$\text{Packet Success Probability} = 0.7 + (UE\ CQI) / 20$$

Additionally, this probability is also applied for messages between UE and gNB for handover execution.

### 4.3 Simulation Environment

For the handover simulation, there were two sets of simulations:

- 10 UEs simulated for 510s.
- 40 UEs simulated for 210s to help provide an insight into what happens in higher density environments.
  - o Section 4.4 highlights the reasons why 210s was used instead of using 510s.

The simulation environment and all its relevant parameters are stated below:

#### Simulator timings and scheduler:

The chosen scheduler mode was round robin, as it provided an overall fair throughput for all users regardless of their priorities and CQIs. Additionally, this scheduler mode can help provide fair results for handover evaluations.

It can be noticed that the simulations were performed for an additional 10s. This is because the first 10s was used as a buffer, ensuring that all parameters are in a steady state scenario when the actual simulation begins. The remaining time is when the useful results are recorded.

The frame mode and subcarrier spacing for all three types of base stations is FDD and 15kHz, respectively. Therefore, there is 1 slot per sub-frame and each frame is 10ms long. In addition:

- User positions, schedulers and handover parameters are updated every 1ms.
- Pathloss parameters are updated every 10ms.
- User MRs occur every 160ms in accordance with one of the values in [46]. This value was chosen as it provides a balance between being faster than normal LTE measurement reports, but not being so frequent that it has high power inefficiencies.
- TTT value of 160ms was used, this matched the occurrences of MRs, additionally, it is a value compliant with [46].
- Retransmission timings after a failed handover communication is 10ms, this is to match the timings of each frame and it provided sufficient time for message acknowledgements.

The handover procedure timings will be discussed in Chapter 5, as modifications are made to the 3GPP solution to incorporate MEC and NFVs.

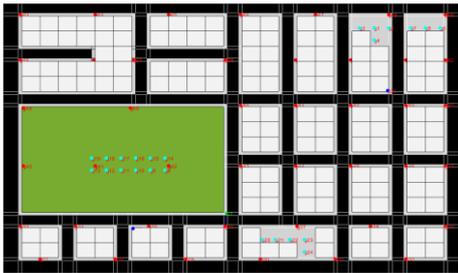
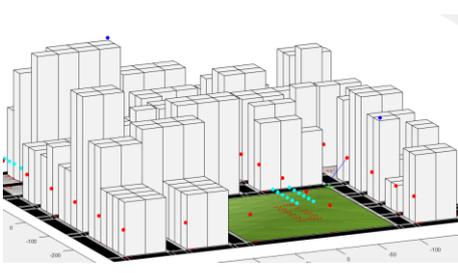
#### Region of interest:

The region of interest is rectangular, this spans 600m-by-700m (0.42km<sup>2</sup>), with varying building heights between 10 and 45m. Building widths and lengths are mapped in blocks of 25m-by-25m. If a building has a larger width or length, then a 25m-by-25m block is placed adjacent to it to develop a wider or longer building. Street widths were designed to be 25m wide, this is because it can accommodate all types of users and be split easily into the following segments:

- Pedestrian walkways: 2.5m on either side of the road.
- Cycle ways: 2.5m on either side of the road.
- Road lanes: 7.5m for each lane.
  - o Moving vehicles lane: 4.0m lane width.
  - o Carpark lane: 3.5m lane width.

The simulation environment is based off New York University (NYU), with Manhattan style building configurations. The images in Table 4.3-1 show two perspectives of the environment. Both satellite images were obtained using Google Maps.

Table 4.3-1: Simulation environment Images [94] [95]

	Aerial view	Elevated and angled view
NYU		
Simulation		

The coloured dots in the simulation image depict base stations. There is a total of 69 BSs:

- 2 Macro cells (dark blue dots).
- 43 Micro cells (red dots).
- 24 Femto cells (teal dots).

#### Base stations and MEC deployments:

The points below give a brief explanation for why certain carrier frequencies and locations were used for the base stations.

- Macro base stations are placed on top of buildings (5m above the building's height). They have been strategically placed to provide coverage and to minimise dead zones within the simulation.  
Carrier Frequency: 2 GHz was chosen as this is at the lowest end of the spectrum in 3D-UMa model and will help provide the widest coverage for UEs within the area.
- Micro cells have been placed on streetlights approximately 10m above the ground.

Carrier Frequency: 3.5 GHz was used as this one of the key FR1 frequency bands for New Zealand's 5G networks [96].

- Femto cells are placed 6.5m above the ground on top of lampposts.

Carrier Frequency: 26 GHz was selected as this one of the key FR2 frequency bands for New Zealand's 5G networks [96].

All the other relevant base station parameters are summarised in Table 4.3-2.

Table 4.3-2: Base station simulation parameters

Parameters	Macro Cell (Wide area BS)	Micro/ Pico Cell (Medium area BS)	Femtocell (Local area BS)
Number of BSs	2	43	24
BS coverage range	500-1000m	50-100m	10-20m
BS height	50m	10m	6.5m
Min 2D distance to UE	35m	5m	2m
Frequency Band	FR1		FR2
Carrier Frequency	2.0 GHz	3.5 GHz	26 GHz
Bandwidth	20MHz	40MHz	
Duplex Mode	FDD		
Transmit Power	40 W	6.31 W	0.25 W
Antenna Gain	0 dBi		
Antenna type	Omnidirectional		
Number of Antennas	1 Transmit antenna, 1 receiving antenna		
Pathloss model (for both LOS and NLOS)	3D-UMa	3D-UMi	Free space with all added pathlosses*.
Coupling loss	70dB	53dB	45dB
Shadow fading	4dB		
Wall losses	13dB		

\* Wall losses, shadow fading and user interferences.

In addition to the base stations, there will 44 MEC deployments, 1 aggregated edge server (at the CU) and 43 far edge servers (one for each Micro/ Pico cell).

- The far edge servers provide multiple benefits (highlighted in Chapter 2.3), although, in this thesis the far edge MEC will only help boost throughputs. Furthermore, the Micro/ Pico and femto cells have an increased bandwidth of 40MHz compared to 20MHz for Macro cells.
- The aggregated edge will be used for providing lower latency handovers by bringing AMF and UPF handover functions to the edge of the network.

### Users:

The author has simulated 40 UE movements:

- 26 higher mobility users, consisting of 22 cars 4 cyclists.
- 14 low mobility pedestrians.

These are simulated for at least 510s. The pedestrians were simulated for 100s longer for validation purposes, to ensure base station were located where coverages were sufficient and dead zones minimised.

For the handover simulation, a subset of 10 users is simulated for 510s. Additionally, the whole set of 40 UEs were simulated for 210s to help provide an insight into what happens in higher density environments. All user movements were varied in a random pattern over the simulated time frames. User speed variations are detailed below:

- Cars slowed down around turns, came to stops at certain intersections.
- Cyclists slowed down around turns and slowed down around pedestrian areas.
- Pedestrians had varying speeds between for walking, running, and resting. UEs only stopped for a few seconds (a maximum of 10s).

All users remained in the region of interest. The remaining UE parameters are listed in Table 4.3-3.

*Table 4.3-3: User mobility simulation parameters for 10 UEs*

<b>Parameters</b>	<b>Vehicles</b>	<b>Cyclists</b>	<b>Pedestrians</b>
<b>Number of UEs (510s)</b>	4	2	4
<b>Number of UEs (210s)</b>	22	4	14
<b>Speeds</b>	0 to 80kmph	0 to 20kmph	0 to 5kmph
<b>Channel model</b>	Vehicle A	Typical Urban	
<b>Number of Antennas</b>	2 transmit and receive antennas. One supports FR1 range, the other for FR2.		
<b>Transmit power</b>	1W		

User RNTIs were uniquely numbered 1 to 40 and each user had independent LCID and LCID groups. Each user had a randomised data rate ranging from 0.1MBps to 20MBps. This accommodated for, calls, browsing and high-quality video streaming.

User priorities are ignored, as this simulation will be using the round robin scheduler. All other user parameters relating to the RLCC and application configurations were identical.

- Sequence field length was 12.
- Maximum buffer for SDUs was 25kB.
- Reassembly timer was 5ms.

- Entity type was bidirectional UM.
- Packet interval was one every 1ms.
- Host devices were both UE and gNB.

Majority of these values were chosen for lower complexity of calculations. The packet interval was chosen to be every 1ms because of the entity type being in bidirectional UM mode. Therefore, providing an overall consistent data rate where retransmissions are not required.

#### Deep Learning:

Lastly, this portion will highlight all the justifications for the parameters used for training the DL system. Beginning with the solver, the adam solver was used, because of the various benefits that were highlighted in Section 4.1.5. Other key configurable parameters are described below:

- Initial learning rate was set to 0.001, A value of 0.1 or 0.01 creates a less accurate model. A value of 0.0001 or less takes very long for the system to learn with little to no improvements over 0.001.
- The gradient threshold was set to 2, to prevent the gradients from diverging from the desired learnings.

The LSTM was taught to learn a sequence of 4 dimensions and 24 classifications, all these classifications and their logical reasonings are described in Chapter 5.

- The number of epochs (1000) and hidden units (20) were varied and analysed to find the optimal values, this analysis is detailed in Chapter 6.

For the DL LSTM, a training and testing data set of approximately 25,000 data points were taken based off actual UE movements.

- 10 users were simulated for 100s (4 cars, 4 pedestrians, 2 cyclists).
  - o 8 were used for training (20,000 points).
  - o 2 were used for testing and evaluating the prediction accuracy (5,000 points).
- These 10 users were independent to the 10 users in the first 510s simulation, although they were part of the 40 UEs that were simulated for 210s.
- Base stations were only considered if their CQI  $\geq 1$ , the remainder were not included.

User data point sizes varied from 2,200 to 2,700 based off the number of BS coverages that could be quantified as a potential base station.

## 4.4 Simulator Optimisations

Several memory saving optimisations were achieved to improve the overall performance of the simulation. The impact of the five stated below provided the largest impact. The impact is measured against the simulation of 10 UEs over 510s.

All memory saving optimisations were implemented to reduce the size of the virtual random-access memory (RAM), on the device that the author was using to perform the simulations.

### 4.4.1 Scheduler

#### Memory allocation:

A large memory saving optimisation in the scheduler was achieved. The 5G NR scheduler initialised a simulation logger consisting of a cell array of 7,140,000 by 13 cells when a BS scheduler was initialised (one for every symbol in a 1ms simulation, this is not required for FDD and is unnecessary). This was a massive strain on the simulator and was very unnecessary.

- This would take around 57 GB of memory per base station to initialise. Therefore, for even 5 of the 69 base stations it would account for 285 GB at least. For this obvious reason, the computer constantly ran into many out of memory errors.

This issue was solved by removing this pre-allocation of memory and replacing it with a less memory intensive solution. This solution added an instance (one row) of the cell array when required, therefore, removing the need to initialise such a large array.

- This reduced the initial size of each array to less than 100 bytes.
- Additionally, at the end of the simulation, the array size for all used base stations increased to around 45GB of the memory.

This optimisation was put in place because it was noticed that all 69 base stations were not required to initialise so many instances, when majority of the BSs would only be used for small periods of the simulation time.

Moreover, this optimisation reduced the initialisation time for each base station. From approximately 120-300s down to 5-10s, depending on the number of users attached to that respective base station.

### 4.4.2 System level simulator

#### Simulation progress reporting:

It was noticed that the authors device was crashing due to the large array of messages that gets printed to the command window, at set intervals after simulating a 10ms frame. The author found that the constant writing of progress to the command window resulted in critical failures. This feature was removed, resulting in no more computer or simulation crashes.

### Memory allocation:

After the scheduler optimisation was enabled, another memory intensive flaw was noticed from the system level simulator. It was noticed that to simulate 10 users for 510s, the simulator would initialise 51,000 individual structures (1 for every 10ms) to capture changes in user mobilities at the beginning of every 10ms frame. The volume of samples posed to be a very big strain on the computer (in terms of memory as well as computational capabilities), which in turn reduced the speed of the simulation. It was observed that the:

- Simulation took approximately 8 hours to initialise all 51,000 instances.
  - o Memory allocated for the 51,000 instances was approximately 40GB.
- Time for one 510s simulation took 5 days (approximately 120 hours) to complete.
- Memory allocated for the whole simulation was 75GB (excluding initialisation).
- Time taken for one 10ms timestamp was approximately 8.5s.

This was very undesirable for the author. It had caused a major issue (out of memory) when attempting to save the results, resulting in an unsuccessful save and all the data was lost.

An examination was done to understand what could be done to improve the efficiency and time of the simulation. The author found that the simulator could initialise one instance of the simulation, as everything is identical for subsequent instances. The only parameter that required to be varied was the user positions. Additionally, it was noticed that the volume of samples and the duplication of results could be made redundant for memory efficiency. These changes were implemented to radically reduced the:

- Simulation initialisation time down to 10 seconds.
  - o Memory allocation for initialisation down to 4GB.
- Duration for one 510s simulation to 2.5 days (approximately 60 hours).
- Memory allocated for the whole simulation 50GB (excluding initialisation).
- Time taken for 1 timestamp reduced to 4.0s.

This resulted in enough of a reduction to save the results. Although, a further optimisation put in place to provide nearly an 8x, discussed below.

### **4.4.3 Miscellaneous**

#### Reduction in the volume of samples:

It was noticed that the simulator was recording samples of user data every 1ms. This causes the simulator's speed to decrease dramatically over time. This optimisation reduced this volume by a factor of 10 and record values over the 10ms frame. This did slightly impact the accuracy of the simulator, although, the impact that was noticed was minimal (a variation of 2-5% for the data rates) and the timing benefits outweighed the accuracy drawbacks. The optimisations reduced the:

- Duration for one 510s simulation to 12 hours.
- Memory allocated for the whole simulation 25GB (including initialisation).
- Time taken for 1 timestamp reduced to 0.85s.

Clear memory and Just in Time (JIT) compiler utilisations:

Additionally, in combination with the optimisation stated above, the author was able to notice an even further reduction by implementing the following steps:

- Clearing larger functions. Even though MATLAB does clear variables when the simulator exits the function, it was noticed that clearing larger functions before exiting optimised the speed of the simulator.
  - o Duration for one 510s simulation to 10 hours.
  - o Memory allocated for the whole simulation 20GB (including initialisation).
  - o Time taken for 1 timestamp reduced to 0.72s.
- Lastly, the feature of MATLAB's JIT compiler was exploited. It was noticed that if a small sample of the simulator was run, the speed was increased. This is because of the JIT compiler's property to save the initialisation of the users functions and classes.

Therefore, further reducing the time, resulting in the final timings of:

- o Duration for one 510s simulation to 8 hours.
- o Memory allocated for the whole simulation 20GB (including initialisation).
- o Time taken for 1 timestamp reduced to 0.55s.

All these optimisations allowed for a more manageable and fast simulation of the scenarios. Subsequently, it was noticed that for the 40 UEs simulation, these optimisations did not provide enough of a reduction to practically simulate these mobilities over 510s, as it would take over 7 to 10 days for completion. Therefore, a simulation time of 210s was used as this required approximately 36hrs to complete.

## 4.5 Performance Metrics

This thesis consists of various performance metrics to evaluate the handover and the DL LSTM. All performance metrics that are used in Chapter 6 for evaluations are described below.

### 4.5.1 Deep learning metrics

For the DL LSTM, one metric is considered vital for the handover optimisation in this thesis. This is the Accuracy (%) of the LSTM. This metric evaluates how accurate the DL LSTM algorithm is. An individual can test this by varying two key parameters:

1. The number of hidden units: This value can correlate directly to the computational latency of the network. The key reason behind this parameter being varied, is to find a balance between the number of hidden units and computational speed.
2. The number of epochs: This parameter is varied to ensure the most optimal time required for training the system is found.

For both these metrics, a chance of over or under fitting of the solution is possible. Therefore, a range of values is considered, to ensure over and under fitting are effectively captured. These concepts will be discussed in detail in Chapter 6 where the analysis is performed.

### 4.5.2 Handover metrics

For the evaluation of handovers, there are 8 metrics that are thought to be key metrics for multi-tier handover evaluations.

These are split into 2 categories, one category evaluates handover performance, the other evaluates overall throughput and user throughput requirements. All throughput metrics are measured for uplink and downlink communications.

#### Handover performance metrics (QoS metrics):

1. Total handovers:
  - This is the total number of handovers in the whole network, inclusive of HOFs.
2. Total number of multi-tier handovers:
  - This is the total number of multi-tier handovers; it will also be presented as a percentage of the overall handovers.
3. Total number of Ping-Pong handovers:
  - This is the total number of ping-ping handovers; it will also be presented as a percentage of the overall handovers.
4. Total number of handover failures:
  - This is the total number of HOFs; it will also be presented as a percentage of the overall handovers.

- These can be related to either desired base station CQI dropping to a value lower than desired, or due to a gross handover failure (described in Chapter 5).
5. Average handover latency:
- This includes the time it takes for a handover failure to become successful after retransmissions.
  - This will not include gross failures (described in 5.1.2), as is a rare occurrence and it can skew the latency to something that does not accurately reflect the handover latency.

Overall throughput metrics (QoE metrics):

6. Total throughput:
- This is the total overall throughput of the whole network that is being simulated.
7. Average UE throughputs:
- This is the average throughput of each user, then averaged again to provide an overall average throughput.
8. Percentage of the time UE data rate needs are met:
- This will consider the percentage of the time that the user's data rate requirements have been met.

# 5 Algorithms

This Chapter details the benchmarking and the novel DL LSTM algorithms.

## 5.1 Benchmarking Algorithm

The benchmark scheme will be a 3GPP defined handover logic and procedure, these were described in Chapter 2.2. The procedure will be adapted with aggregated edge components for enabling lower latencies.

### 5.1.1 Handover Logic

For the UE to trigger a handover, its RSS has to be greater than the A3 entry condition (stated in Section 2.2 Figure 2.2-2), which is:

$$\text{Target gNB RSS} > \text{source gNB RSS} + \text{A3 offset (3dB)}$$

Table 5.1-1 describes the equations and offset values used. This table is used for the decision that is made after the TTT period, whether it is an A1 or an A3 handover.

Table 5.1-1: A1 and A3 handover conditions in the simulator

	<b>A3 Event</b>	<b>A1 Event</b>
<b>Exit point decision</b>	<i>Target gNB RSS &gt; source gNB RSS + A3 offset (3dB)</i>	<i>source gNB RSS &gt; A1 threshold (minimum RSS for a CQI of 1)</i>

If it is an A1 handover, it will revert back and remain connected to the source gNB. If it is an A3 handover, the UE will move onto handover initiation phase. Otherwise, if the UE's CQI for the serving BS drops below a value of 1 during this process, the UE becomes idle and begins to reconnect to the base station that meets the A1 handover conditions. For all cases, the handover trigger instance will be recorded.

### 5.1.2 Handover Procedure

Key handover procedure communication delay metrics are stated below, all latency values are obtained from [97] and [78]:

- Handover request from source gNB to target gNB: 2ms between both DUs [97].
- Admission control: 1ms for admission control at the target gNB [78].
- UE handover initiation message: 1ms for data transmission over air interface [78].
- UE Handover configurations [78]:
  - o HO request processing: 5ms
  - o HO reconfiguration: 10ms
- Status transfer from source gNB to target gNB: 1ms [97].
- Target gNB and UE synchronisation messages: 2ms [78].

All wireless communications have a success probability and a retransmission delay, these values were discussed in section 4.2. Additionally, two HOF types are identified:

1. If at any point during the handover procedure, the desired BS's CQI is less than 1, the handover is stopped, and the UE is moved to the connected state.
2. If more than 16 communication failures occur in a set handover period, these are considered as gross handover failures [98], then the UE will be disconnected from the BS and become idle. This is considered as another type of HOF.

## 5.2 Proposed Algorithm

To develop a DL LSTM, an understanding what the inputs and outputs must be realised. First, the desired outputs are decided. These are based off what is desirable and what challenges that this proposal is trying to address. The metrics are:

1. User's CQI: This is chosen to be an output to ensure that data connections are never lost and a good QoS is maintained.
2. User's data rate requirements: This is required to ensure that the user's data rate requirements are met for as long as possible.
3. User's velocity: This ensures that the algorithm is dependent on user mobility when connection requirements become more important.

The parameters above were chosen because it provides the algorithm the best opportunity to meet the UEs QoS and QoE requirements. Therefore, from these desired output metrics, the key input dimensions were chosen. This LSTM consisted of 4 dimensions. The user's velocity is split into direction and speed for a smoother and faster learning process for the DL LSTM.

### 1<sup>st</sup> Dimension: User's CQI

This is the user's CQI rating for the potential BS, only BSs that have a  $CQI \geq 1$  are considered. This eliminates any base stations which are not within the range.

- This method if implemented correctly, can reduce signalling overheads and UE costs. This is because there is no longer a need for the capability to monitor and report on a minimum of 8 (4 Inter-RAT and 4 Intra-RAT) BSs, as stated in section 9.3.3.1 and 9.3.3.2 in [99].

### 2<sup>nd</sup> Dimension: User's data rate ratio

The user's data rate ratio is derived from the equation below. For the compactness of the equations data rate has been abbreviated to DR.

$$DR \text{ Ratio} = \max \text{ user DR requirements} / \min \text{ potential DR the BS support.}$$

The max data rate is the maximum of the uplink and downlink requirements. A simplified version of the minimum potential data rate that the BS can support is given by:

If attached users > 0:

$$\text{Min Potential DR} = \min \text{ BS DR for a CQI of 3} / \text{number of UEs attached}$$

If attached users = 0:

$$\text{Min Potential DR} = \min \text{ DR the BS can support for a CQI of 3}$$

A CQI of 3 was chosen as this is the average CQI that a user will have when connecting to a base station, at a distance equivalent to approximately 70% ( $\pm 10$  to 20%) of the BS's coverage.

- This distance was chosen, as in most cases CQIs of potential BSs that are can be handed over to in most cases will not be higher than 50 to 70% of the BS coverage. Therefore, a value of 20% of the maximum CQI value (of 15) was taken.
- The variation takes into consideration of small scale fading and shadow fading effects, which can at the least cause a  $\pm 10$  to 20% variation in the channel quality. The variations in the presence of cars and NLOS obstructions can be seen in [100], where the authors simulate shadow fading models for VANETs.

3<sup>rd</sup> and 4<sup>th</sup> Dimensions: User's direction and speed:

The values for both these dimensions are using an RSS values that are measured in dBm. The variations can also be measured in watts, although, classification parameters may need to be adjusted appropriately to accommodate for this change. This was a measure that was chosen by the author for ease of comprehension.

Firstly, the user's direction is measured from the variation in the RSS between two successive MRs of the potential BS. A negative value denotes a user is moving away, a positive value denotes the user is moving closer, both expressions are with respect to the potential BS being assessed.

Additionally, the variation in speed is then calculated as an absolute value of RSS variation. All values below are with respect to RSS variations (for completeness a velocity equation has been included).

$$UE\ velocity = -RSS_{t-1} + RSS_t$$

$$UE\ direction = \begin{cases} -UE\ velocity, & UE\ velocity < 0 \\ \dots & \\ UE\ velocity, & UE\ velocity \geq 0 \end{cases}$$

$$UE\ speed = |UE\ velocity|$$

A variation of 5dBm or more was chosen to be the value of a fast-moving user, as a 1m variation in 100ms (equivalent to a vehicle traveling at approximately 36kmph), accounts for an RSS change of 10-15% in free space. When assuming a dBm value of 20dBm, a value of 2-3dBm would be the free space RSS variation.

Additionally, due to the additional pathloss factors discussed in Chapter 2 and 4, a 3dBm offset is added to avoid majority of the misrepresentations. From these definitions, each of the input dimensions were classified and concatenated into one output. These are specified in Table 5.2-1.

Table 5.2-1: DL LSTM classification categories

Dimension	Classifications	Value ranges
user CQIs	Good	$CQI > 5$
	Ok	$3 < CQI \leq 5$
	Poor	$CQI \leq 2$
User Data rate ratio	Meets	$DR \geq 1$
	Not Met	$DR < 1$
UE Direction	Closer	$UE\ Direction \geq 0\ dBm$
	Away	$UE\ Direction < 0\ dBm$
UE Speed	Fast	$UE\ speed \geq 5\ dBm$
	Low	$UE\ speed < 5\ dBm$

With these chosen output types and classifications, there are 24 possible combinations.

These combinations are shown in Table 5.2-2.

Table 5.2-2: All 24 classification of the proposed DL LSTM

Index	Code	Classification			
		CQI	Data rate	Direction	Speed
1	GMCF	Good	Meets	Closer	Fast
2	GMCL				Low
3	GMAF			Away	Fast
4	GMAL				Low
5	GNCF		Not met	Closer	Fast
6	GNCL				Low
7	GNAF			Away	Fast
8	GNAL				Low
9	OMCF	Ok	Meets	Closer	Fast
10	OMCL				Low
11	OMAF			Away	Fast
12	OMAL				Low
13	ONCF		Not met	Closer	Fast
14	ONCL				Low
15	ONAF			Away	Fast
16	ONAL				Low
17	PMCF	Poor	Meets	Closer	Fast
18	PMCL				Low
19	PMAF			Away	Fast
20	PMAL				Low
21	PNCF		Not met	Closer	Fast
22	PNCL				Low
23	PNAF			Away	Fast
24	PNAL				Low

Now that the classifications and their reasonings are clarified, the adjustment to the handover logic and procedures are discussed below.

## 5.2.1 Handover Logic

This algorithm relies on previous MRs to predict the best base station to handover to. This decision happens in the current MR time stamp. Each user's last 7 MRs for potential BSs are stored in the aggregated edge of the MEC.

- In addition to this, the connected BS CQIs are also stored for the past 7 timestamps. These 7 CQIs are averaged to ensure shadow fading and small-scale fading effects are minimised. Furthermore, the same is done for the actual data rate, to ensure fluctuations are filtered out.

If the base station has a CQI that is higher than 1 for longer than 7 consecutive time stamps, MR variations are considered, to efficiently save power for low mobility users. The pseudocode below describes these logical steps.

---

### Algorithm 1: Classification logic

---

1. **procedure:** Classify BSs based off MRs
2.   **for** each potential detected BS
3.     **if** CQI  $\geq 1$  **then**
4.       Calculate all remaining parameters to input into the LSTM
5.       Predict potential BS classification based off the inputs
6.       Store the classification for the user at the aggregated edge
7.     **end if**
8.   **end for**
9. **end procedure**

---

### Algorithm 2: Handover logic

---

1. **procedure:** vary measurement reporting
  2.   **if** consecutive MRs for potential BS = 7 **then**
  3.     UE state = potential BS handover
  - 4.
  5.   **if** UE speed is fast for  $\geq 5$  MR instances, **then**
  6.     Decrease MR interval by 40ms
  7.     **if** MR interval is  $\leq 80$ ms **then**
  8.       MR interval = 80ms
  9.     **end if**
  10.   **end if**
  - 11.
  12.   **if** UE speed is low for  $\geq 5$  MR instances, **then**
  13.     Increase MR interval by 40ms
  14.     **if** MR interval is  $\geq 400$ ms **then**
  15.       MR interval = 400ms
  16.     **end if**
  17.   **end if**
  18. **end if**
  19. **end procedure**
-

The reasons why these MR occurrence limits were chosen are highlighted below:

- For high mobility users, the MRs will not go below 80ms, as it will drain the UE's battery at a high rate.
- For low mobility users, the MRs will not go above 400ms, as this will impact the response of a handover decision if it is required for sudden changes in movements.

For this algorithm there is no longer a TTT value, this is replaced with the third and final logical process called, 'potential BS handover'. This process is described via the LUT below for ease of comprehension. LUTs provide a very fast and simple approach to solving repetitive problems. Additionally, outcomes can be easily modified to achieve the desired outcomes.

All handover decisions require 5 (~70%) or more instances of each predicted classification. For example, if a UE is a fast-moving UE, the classification 'fast' within the last 7 predictions must occur at a minimum of 5 times, otherwise, it will be considered as a 'low' moving UE.

Table 5.2-3: Handover Logic based off DL LSTM Classifications

		Current BS Parameters				Key	
		Meets DR		Not Met DR		Letter/ Number	Classification
		CQI < 3	CQI ≥ 3	CQI < 3	CQI ≥ 3		
Potential BS Classifications	GMCF	1	0	1	1	G	Good CQI
	GMCL	1	0	1	1	O	Ok CQI
	GMAF	1	0	1	1	P	Poor CQI
	GMAL	1	0	1	1	M	Meets DR
	GNCF	1	0	1	0	N	Not met DR
	GNCL	1	0	1	0	C	Moving Closer
	GNAF	1	0	1	0	A	Moving Away
	GNAL	1	0	1	0	F	Fast speed
	OMCF	1	0	1	1	L	Low speed
	OMCL	1	0	1	1	1	Handover
	OMAF	1	0	1	0	1*	Exception Handover
	OMAL	1	0	1	1	0	No Handover
	ONCF	1	0	1	0		
	ONCL	1	0	1	0		
	ONAF	1	0	1	0		
	ONAL	1	0	1	0		
	PMCF	1*	0	1*	0		
	PMCL	1*	0	1*	0		
	PMAF	0	0	0	0		
	PMAL	0	0	0	0		
PNCF	1*	0	1*	0			
PNCL	1*	0	1*	0			
PNAF	0	0	0	0			
PNAL	0	0	0	0			

In Table 5.2-3:

- 1 denotes a handover is to be performed.
- 0 implies no handover required and revert to the connected state.
- 1\* is an exception handover, that should only occur if the CQI is better than the current BS CQI, to avoid the risk of RLF. After a handover decision is made, all past UE predictions for all BSs are cleared.

If a handover is requested, the user moves to the handover procedure phase, where the current BS initiates the handover to the desired BS.

### **5.2.2 Handover procedure**

The handover procedure being proposed in this section implements a faster variation to the current being used today.

The proposal is to send the admission control for handover request at the same time instant that the UE begins to process the handover request. This can be made possible because of the aggregated edge architecture, due its centralised nature, the MEC can orchestrate both events to execute simultaneously. Hence, a further reduction of 3ms in latency could be made in addition to the latency optimisations discussed previously in Section 5.1.2.

All other remaining parts of the handover process remain the same as described in Section 2.2.

## 6 Results and Discussion

This Chapter will discuss the various results of each simulation. Beginning with the DL evaluation, then progressing to analysing and comparing the handover performance metrics.

### 6.1 Deep Learning Evaluation

The training for this evaluation was using a technique called supervised learning, the LSTM was made aware to all 24 of the classification variations. The parameters that were varied to evaluate accuracy impacts were, the number of epochs and hidden layers. Three simulation runs were carried out for each variation, to estimate average values of the results. For both results, percentages are rounded to the nearest 0.01%, simulation time is rounded to the nearest second.

To start, the table below highlights the impact on LSTM performance as the number of hidden units increase. Each of the simulations were 1000 epochs long. This was experimented upon and chosen to be the right value based on the results of Table 6.1-2.

Table 6.1-1: LSTM Supervised Learning Performance, 1000 epochs per simulation

Number of Hidden Units	Duration of simulation (s)				Prediction Accuracy (%)			
	Run 1	Run 2	Run 3	Mean	Run 1	Run 2	Run 3	Mean
5	241	239	254	<b>245</b>	81.35	76.94	73.35	<b>77.21</b>
10	273	284	294	<b>284</b>	88.44	71.08	86.32	<b>81.95</b>
20	352	353	360	<b>355</b>	99.45	99.86	99.47	<b>99.59</b>
40	540	550	533	<b>541</b>	94.12	98.82	99.27	<b>97.40</b>

The test data above, it is a perfect example of all three types of capacity fittings [101]:

1. Underfitting: Where the solution is not complex enough to understand the data, it causes a bias underfitting issue. This can be seen where the hidden units values were 5 and 10.
2. Overfitting: Where the solution learns the training data but fails to aptly generalise the training set for new unseen testing data. Can be slightly observed with 40 hidden units.
3. Appropriate fit: where the solution can appropriately generalise as well as learn the trend to accurately predict new data. This is the optimal solution of 20 hidden units.

The same methods can be applied to the number of epochs. The number of epochs should be stopped when the error of the learning is at its minimum. Further efforts to optimising the learning leads to overfitting and compromises the performance of the LSTM.

This is noticed in Table 6.1-2, again all three types of fit are visible. For finding the correct number of epochs, the optimal value of 20 hidden units was used, as this was the best performing LSTM from Table 6.1-1.

Table 6.1-2: LSTM supervised Learning Performance, 20 hidden units per simulation

Number of Epochs	Duration of simulation (s)				Prediction Accuracy (%)			
	Run 1	Run 2	Run 3	Mean	Run 1	Run 2	Run 3	Mean
<b>10</b>	4	4	4	<b>4</b>	22.12	8.17	19.86	<b>16.72</b>
<b>100</b>	37	34	34	<b>35</b>	57.03	67.89	54.17	<b>59.70</b>
<b>1000</b>	352	353	360	<b>355</b>	99.45	99.86	99.47	<b>99.59</b>
<b>2000</b>	710	819	739	<b>756</b>	90.05	90.68	99.35	<b>93.36</b>

For the final simulation, the most accurate hidden units and epochs combination was used. The accuracy of predicting the test data was at 99.86%.

## 6.2 Handover Evaluation

From the DL LSTM evaluations, simulations comparing the benchmark and the proposed algorithm was run. The results analysed the impact on handover performance when the densities of the users are varied.

- It has to be noted, that the density of users increasing is also including a higher number of varying user movements. A major impact will be due to the larger the volume of fast moving users increasing (from 4UEs to 22UEs), as user interferences will impact these handovers the most.

### 6.2.1 Total handovers

The Figures 6.2-1 and 6.2-2 highlight the total handovers for both scenarios.

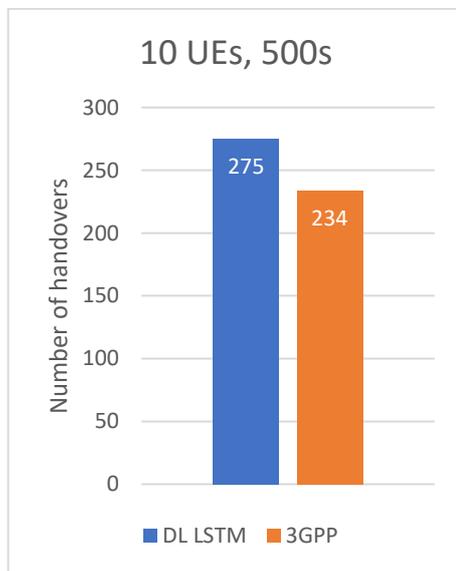


Figure 6.2-1: Total handovers (10UEs, 500s)

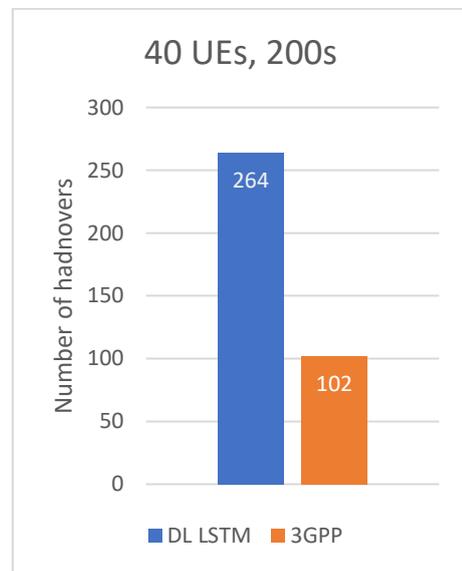


Figure 6.2-2: Total handovers (40UEs, 200s)

From this, it can be noticed that the total number of handovers for the proposed solution is higher, although, the 3GPP scheme had a large number of entry condition triggering events. This is due the scheme only taking the current time instance to compare the RSSs against, introducing a higher chance of the decision being impacted greatly by random channel effects. The results showed that the 3GPP scheme had 556 and 286 triggering events for the 10UEs and 40UEs, respectively. From this data it can be noticed that as the density of the users increase, the number of times that the network has to monitor the triggering conditions per user also increases (by 15% in this case).

Furthermore, the proposed scheme has a lower triggering instance when compared to the 3GPP benchmark. It would only trigger once (when it required to continue onto the procedure phase). Therefore, if we compare the triggering instances, the proposed scheme provides a lower number of triggering instances, shown in Figure 6.2-3 and 6.2-4.

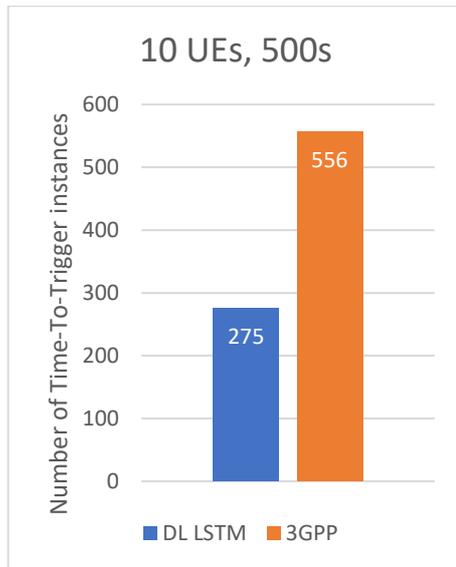


Figure 6.2-3: Total TTT instances (10UEs, 500s)

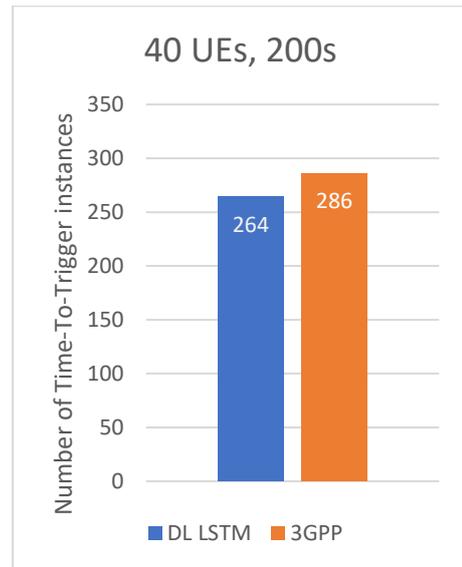


Figure 6.2-4: Total TTT instances (40UEs, 200s)

The reasons why the gap decreases when user densities increase can be explained by the following two reasons:

1. The emphasis on the LUT parameters. Mainly the proposed scheme focuses on improving data rates or providing the best CQI possible. This is observed especially for the higher density scenarios, as a lot more handovers occur to accommodate user requirements.
2. In the 3GPP scheme the RSS of the desired base station was lower than the current BS, due to the number of user interference factors that come into play.

## 6.2.2 Multi-tier handovers

Beginning with the volume of multi-tier handovers, the figures below highlight the total number of multi-tier handovers (Figures 6.2-5 and 6.2-6), as well as their respective percentages of the total handovers (Figures 6.2-7 and 6.2-8).

- There was no macro cell to macro cell handovers, as the distance and the obstacles between the two did not provide many avenues for overlapping coverages.
- There was also very rarely a micro to micro cell handover in the scenarios, due to the fact that the RSS of the macro cell was quite high compared to the other small cells.

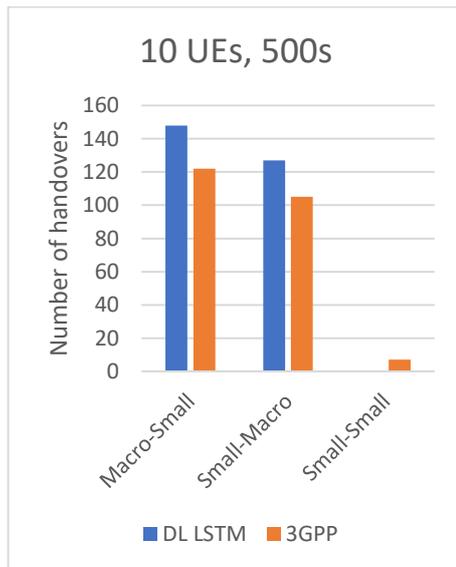


Figure 6.2-5: Total multi-tiers (10UEs, 500s)

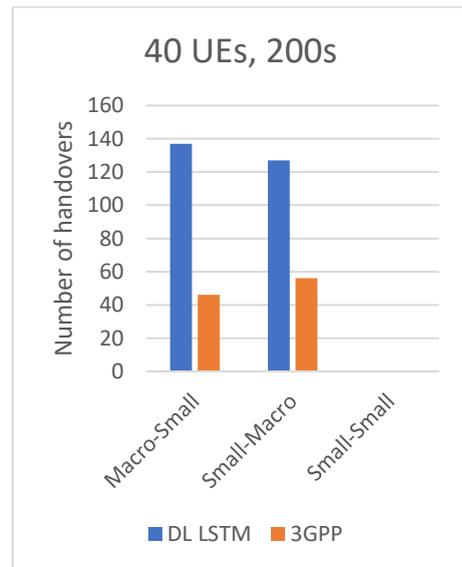


Figure 6.2-6: Total multi-tiers (40UEs, 200s)

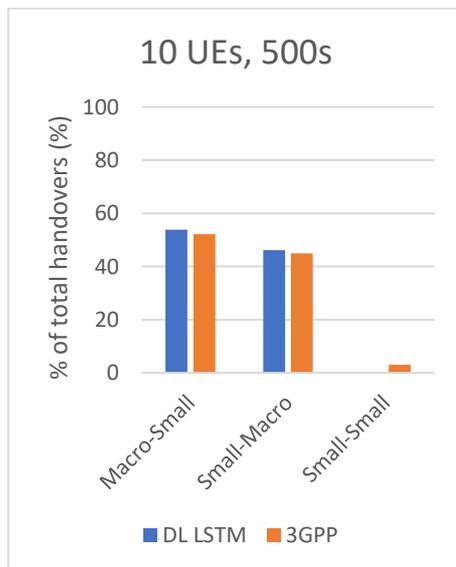


Figure 6.2-7: Multi-tier percentage (10UEs, 500s)

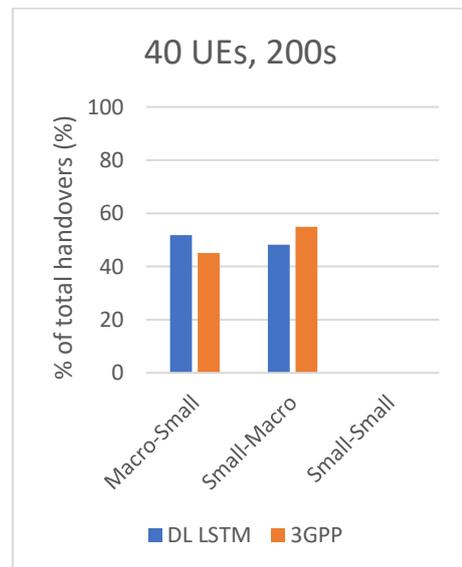


Figure 6.2-8: Multi-tier percentage (40UEs, 200s)

Both schemes mainly did multi-tier handovers, as it primarily required to handover due to the high variations in RSRP. For the 3GPP scenarios it can be noticed that as the user densities

increase, there is a higher micro to macro handover rate, from around 45% to 55%. This is due to the UEs connectivity being the only important parameter in this scheme.

To elaborate further, when the algorithm was in a lower density scenario, the connectivity was not a large obstacle, as user interferences were not significant. Consequently, as the densities increased, connectivity issues became more apparent.

For the proposed scheme these high number of multi-tier handovers was made a bit more apparent, as the connectivity requirement outweighed the data rate requirement, or vice versa. It can be seen it is approximately 60% to 40% in favour of the data rate importance in the lower density, as the CQI was quite good. Subsequently, this reduced to around 50% each way when channel qualities were reduced. This provided the UE with the opportunity to connect to the cell that best met its data rate requirements.

For example, it was noticed that the macro cell would provide the best CQI in an area where a single tier handover could have occurred. Therefore, the macro cell provided the best connection and the least chance of HOF, whereas for the single tier handovers the CQIs were undesirable. That is the reasons why the proposed scheme did not attempt to do any single tier handovers, as the CQI of the macro cell was always higher than the CQI of the desired micro cell in question (at the beginning of the handover decision).

Then after the handover had been completed, the UE was handed over to the desired small cell when the CQI was no longer an obstacle, and a data rate boost was required. Section 6.3.3 shows the benefits of this.

### 6.2.3 Ping-Pong handovers

The improvements on the Ping-Pong handovers are displayed in Figures 6.2-9 through to 6.2-12. It provides an insight into how both schemes perform in these two scenarios.

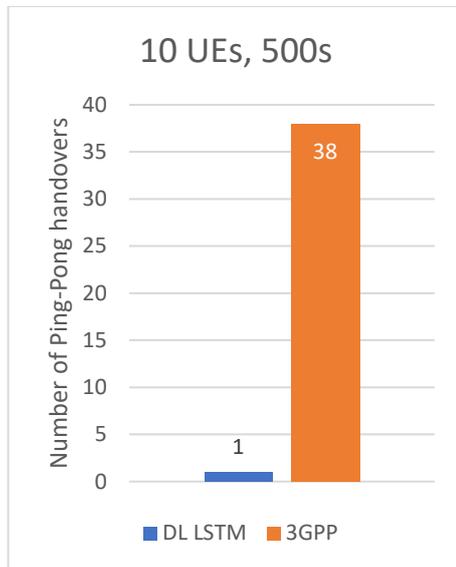


Figure 6.2-9: Total Ping-Pongs (10UEs, 500s)

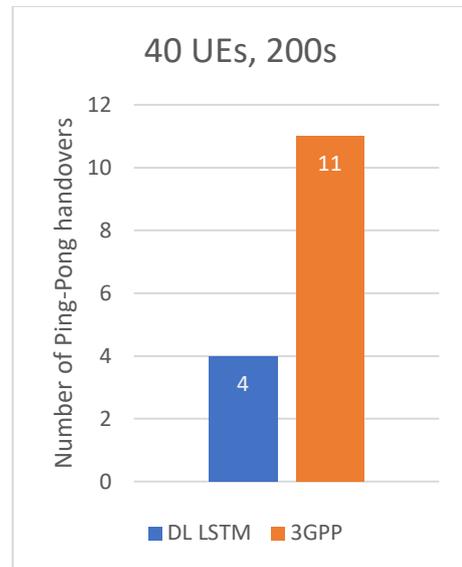


Figure 6.2-10: Total Ping-Pongs (40UEs, 200s)

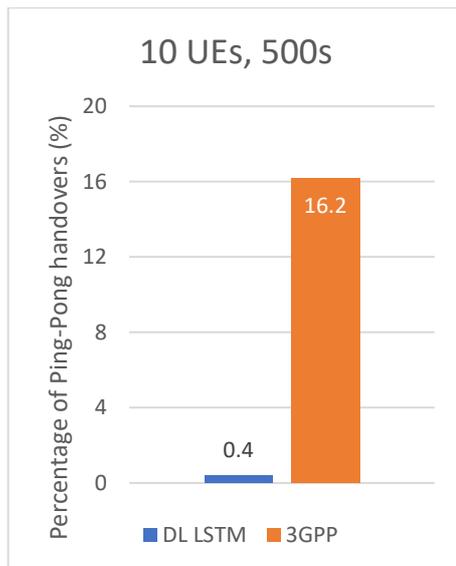


Figure 6.2-11: Ping-Pong percentage (10UEs, 500s)

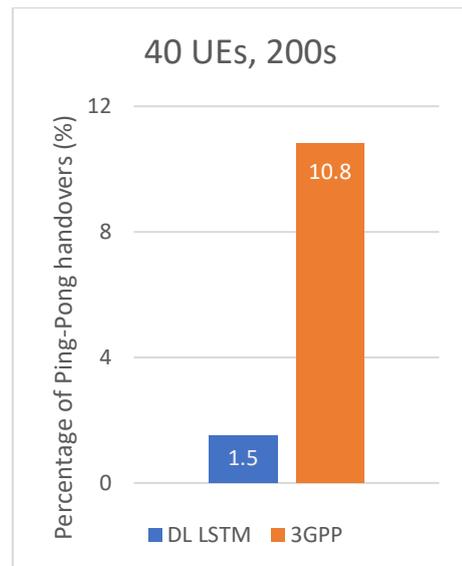


Figure 6.2-12: Ping-Pong percentage (40UEs, 200s)

For the proposed scheme in both low and high densities, the feature that helped the most was the averaging of the past 7 time stamp CQIs. Therefore, mitigating Ping-Pongs due to random small-scale and shadow fading. This rate remained at a very low Ping-Pong rate for the higher density of users as well. Providing a more stable transition to the desired base station.

When analysing the 3GPP scheme, it was noticed that the handover failure ratio reduced as the density of the UEs increased. This can be due to the 3dB threshold offset in addition to the current BS's RSRP. Therefore, mitigating as many false readings as possible.

Subsequently, it can be noticed, for the lower densities, Ping-Pongs for the proposed scheme were reduced by 97.5%, this can also show that the algorithm performs well in lower densities where user interference are small. Successively, it also performs very well in high density environments, where it provided a reduction of approximately 86%.

## 6.2.4 Handover failures

Figures 6.2-13 through to 6.2-16 display the handover failures, these handover failure types have been explained in Chapter 5.1.2, although, the summary is provided below:

1. If at any point during the handover procedure, the desired BS's CQI is less than 1, the handover is stopped, and the UE is moved to the connected state.
2. If more than 16 communication failures occur in a set handover period, these are considered as gross handover failures [98], then the UE will be disconnected from the BS and become idle. This is considered as another type of HOF.

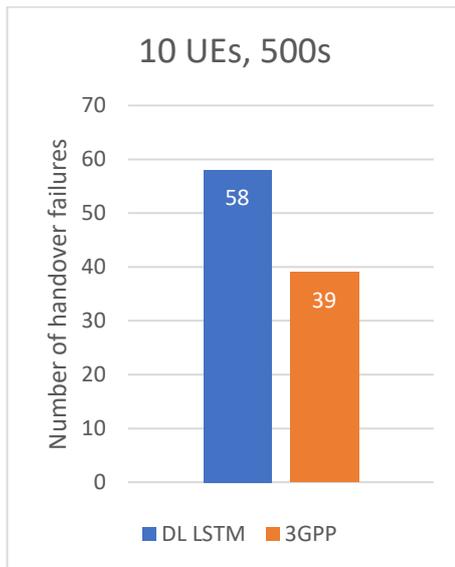


Figure 6.2-13: Total HOFs (10UEs, 500s)

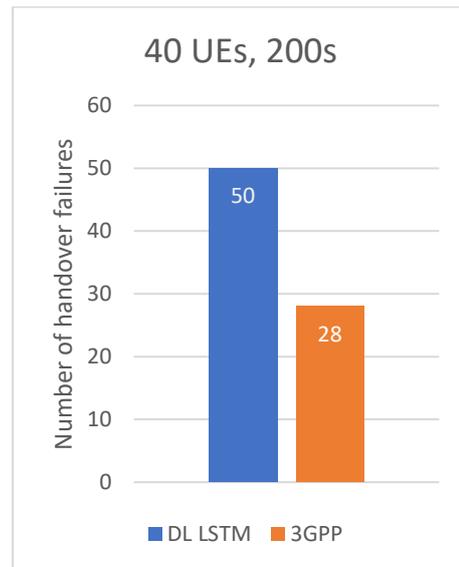


Figure 6.2-14: Total HOFs (40UEs, 200s)

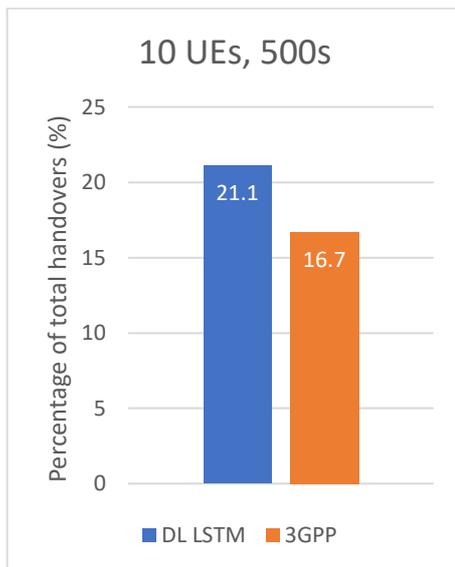


Figure 6.2-15: HOF percentage (10UEs, 500s)

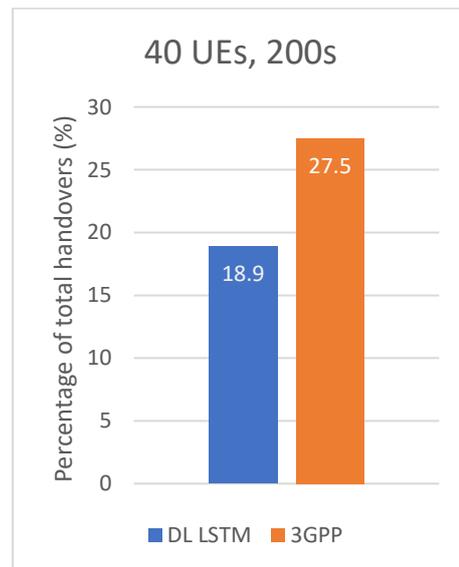


Figure 6.2-16: HOF percentage (40UEs, 200s)

The handover failures of the proposed algorithm are 25% higher in lower density scenarios. This is attributed to the emphasis of the desired CQI in the LUT being quite strict. Resulting in handovers occurring in the CQI exception phase, where CQI values for both serving and desired

BSs are at a minimum, giving it less of a chance for a handover to occur successfully for lower user mobilities. This was reduced when in the higher density scenario, as user interferences helped the proposed scheme make a stronger decision, that failed 30% less compared to the 3GPP (when focusing on the percentage of total handovers).

Furthermore, the 3GPP scheme performed poorly when considering the percentage of total HOFs. This can be attributed to the higher user interferences. It had to check if the A3 handover condition was met for only two instances, which could vary a lot due to the interfering factors. Therefore, as the handover was progressing, the chances of failure were elevated. Whereas the proposed scheme was more confident as it checked the LUT conditions for the average value of the past 7 instances to compare.

Additionally, the variation in the frequency of the MRs can be attributed to this, as the higher the user mobilities are, the more rapid the response is of the proposed scheme is. The proposed is faster by one TTT instance, when compared to the 3GPP scheme (simply because of the instant triggering functionality).

## 6.2.5 Average handover latency

Lastly, the average handover latency is measured and compared. Handover latency is only calculated if a handover successfully completes. These values are compared in Figures 6.2-17 through to 6.2-20.

The first graphs show the number of successful handovers after a set number of retransmissions (abbreviated to reTx in the graph). These contribute directly to the average handover latency calculation.

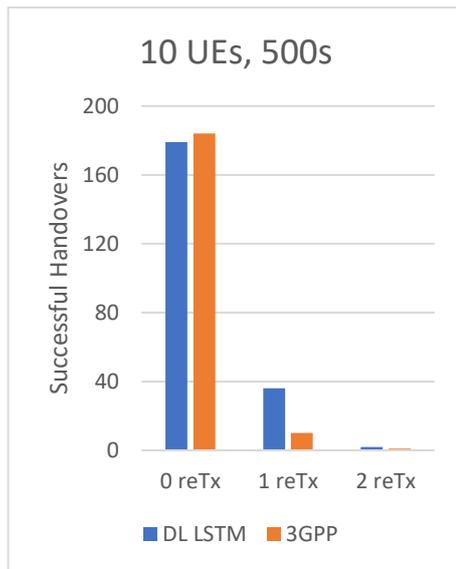


Figure 6.2-17: Successful HO attempts (10UEs, 500s)

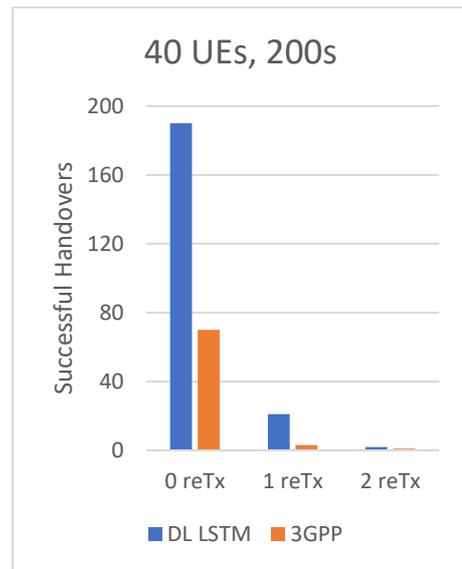


Figure 6.2-18: Successful HO attempts (40UEs, 200s)

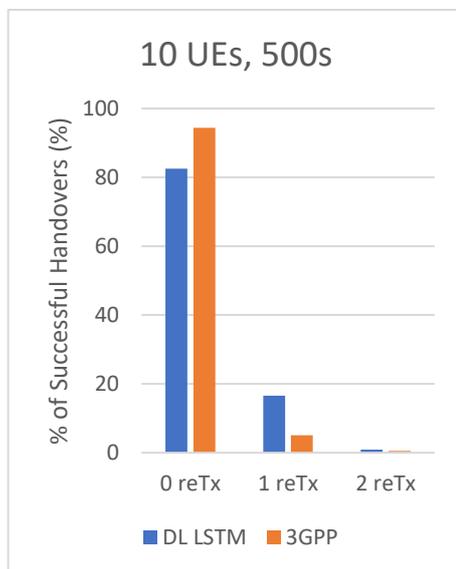


Figure 6.2-19: Successful HO ratios (10UEs, 500s)

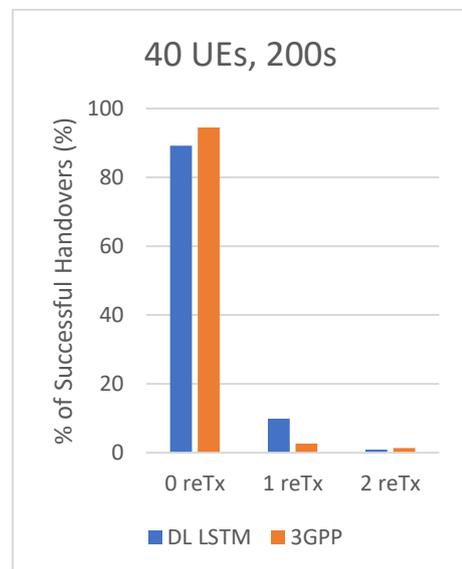


Figure 6.2-20: Successful HO ratios (40UEs, 200s)

From the results above, it can be noticed, that the 3GPP provides quite a good performance in successfully executing handovers in all cases. Although, this gap narrows as the density of users increase and the proposed scheme's characteristics work towards enhancing the handover performance.

From these values an average handover latency calculation is performed. The handover latency metrics used for the calculations are described in the table below.

Table 6.2-1: Handover latencies for varying retransmissions.

Algorithm	0 reTx	1 reTx	2 reTx
<b>DL LSTM with MEC and LUT</b>	17ms	27ms	37ms
<b>3GPP with MEC</b>	20ms	30ms	40ms

From the assumption in Chapter 4.2, the retransmission delay is 10ms. With these values, a simple weighted sum is calculated and averaged, to provide the average handover latency of the UEs. The results can be viewed in Figures 6.2-21 and 6.2-22.

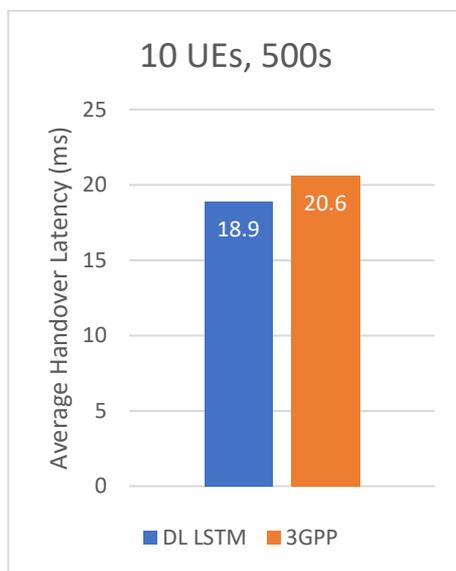


Figure 6.2-21: Average latency (10UEs, 500s)

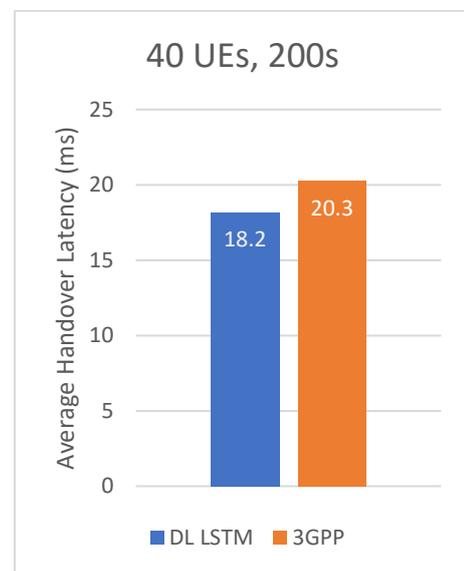


Figure 6.2-22: Average latency (40UEs, 200s)

The average handover latency for the proposed solution for both user densities is lower than the 3GPP solution, even though there were more retransmissions in the proposed. This is due to the optimisation of simultaneous activations of admission control, and UE handover process steps (described in Section 5.2.2). Therefore, reducing the impact of these handover retransmissions.

## 6.3 Throughput Evaluation

Continuing on from the handover evaluations, this section investigates the impact that these handovers had on the throughput.

### 6.3.1 Total throughput

Figures 6.3-1 and 6.3-2 reveal the total throughput for both scenarios and schemes. This is measured in megabytes per second (MBps).

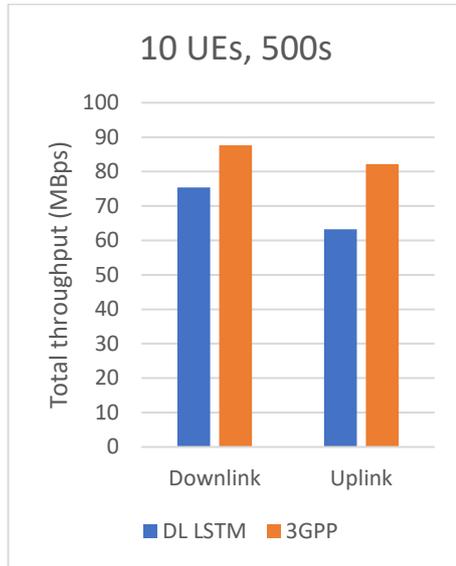


Figure 6.3-1: Total Throughput (10UEs, 500s)

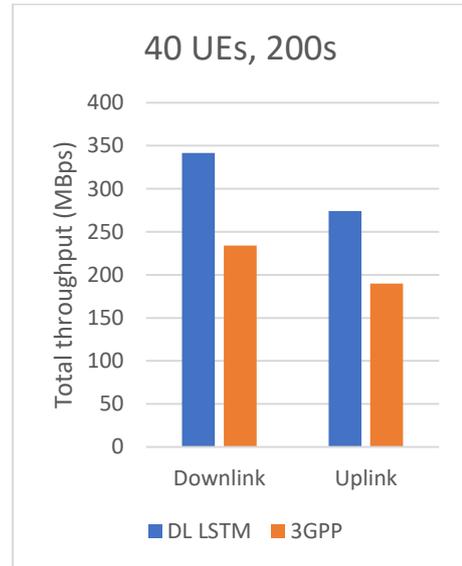


Figure 6.3-2: Total Throughput (40UEs, 200s)

The results above show that the 3GPP scheme does perform better in lower density scenarios by 15 and 30% respectively for downlink and uplink. This is due to the CQI of the connected users to the macro cell not being dramatically impacted from interference factors. Therefore, from the LUT requirements, a handover does not require to be done as the macro cell's CQI is a lot better than the desired micro cell.

When noticing the performance in the high user density scenario, the proposed solution performs a lot better, as user interference factors become very apparent and impact the users. Then the CQI of the macro cell is very similar to that of the micro cell and handovers can occur to optimise the user's data rate. This proves to boost the throughputs in the proposed solution to beat the 3GPP solution by approximately 45% for both downlink and uplink.

### 6.3.2 Average user data rate

Figures 6.3-3 and 6.3-4 highlight all user data rates, for both uplink and downlink. Starting with the lower density of users.

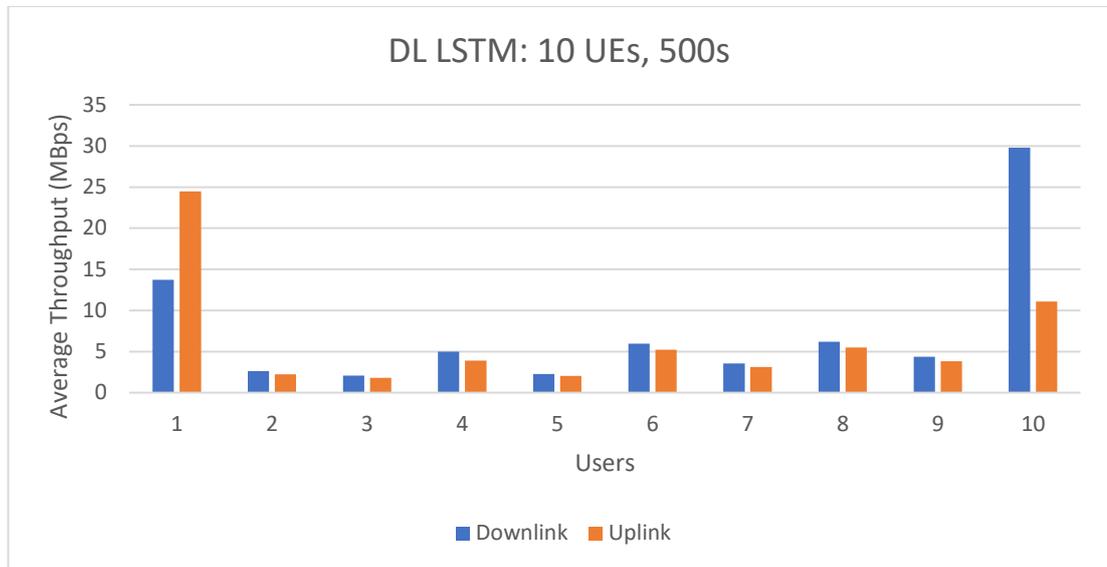


Figure 6.3-3: Average user throughputs DL LSTM (10UEs, 500s)

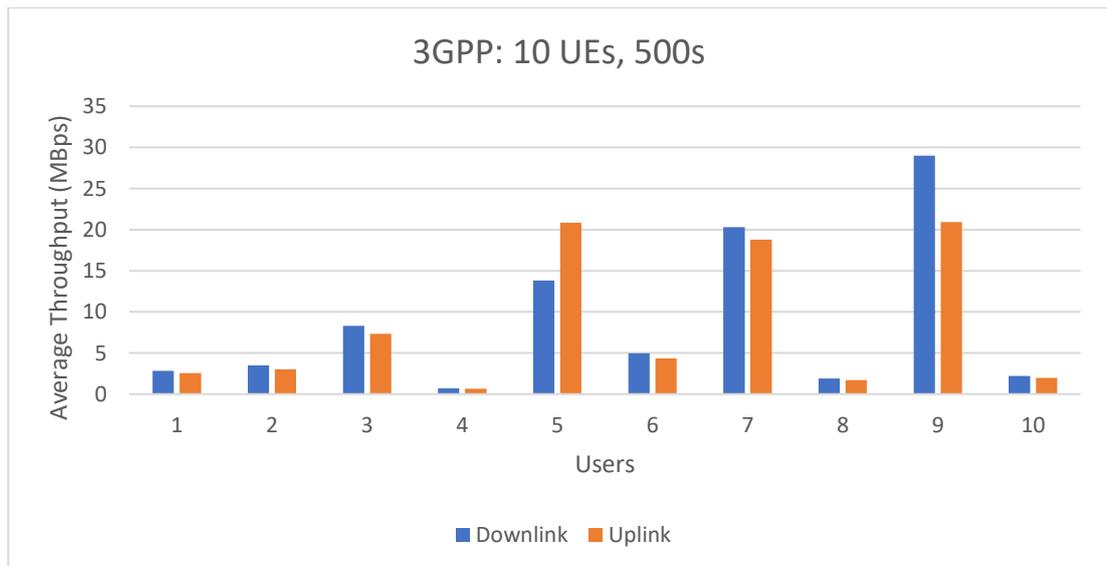


Figure 6.3-4: Average user throughputs 3GPP (10UEs, 500s)

From the data above, it can be noticed that the 3GPP provided a better overall throughput for the users, although, for the users connected to the macro cell, they suffered from a lower data rate than the DL LSTM solution (all users below 4MBps are macro cell users).

This is due to one of two reasons:

1. The proposed solution has an importance on the CQI if it is very good, therefore, if the UE is connected to the macro cell, the CQI is quite good. Furthermore, if the UE is connected to the macro cell, it enhances the data throughput where possible, but in some cases this may not be enough.

- The UE's data rate is being met by the macro cell, therefore, not requiring the higher data rate from the small cell.

These throughput metrics are also compared in higher mobility scenarios. Those values are considerably better due to the impacts discussed in Section 6.3.1. These results are displayed in Figures 6.3-5 and 6.3-6.

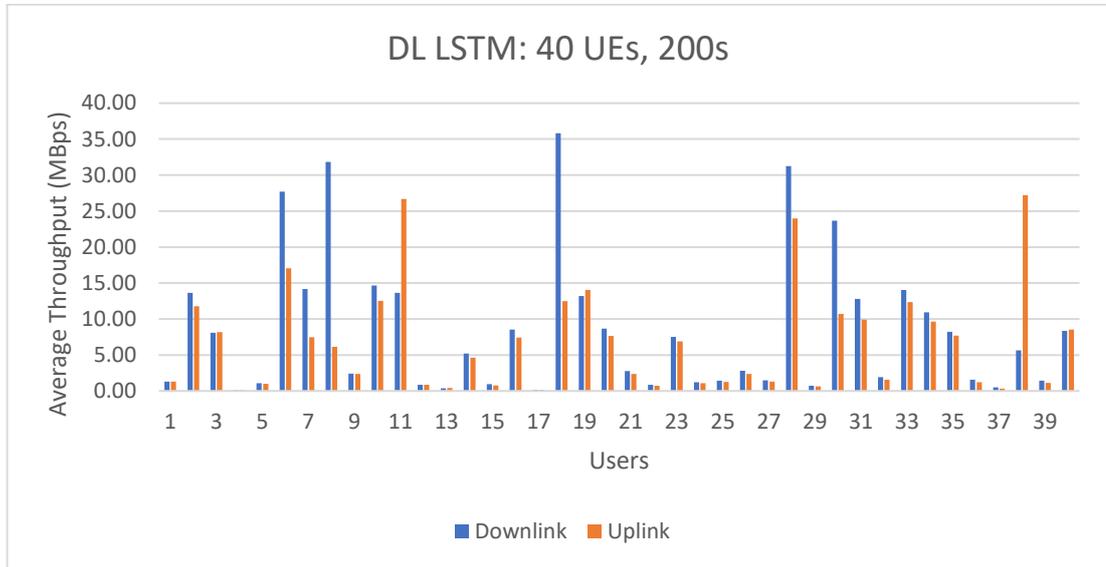


Figure 6.3-5: Average user throughputs DL LSTM (40UEs, 200s)

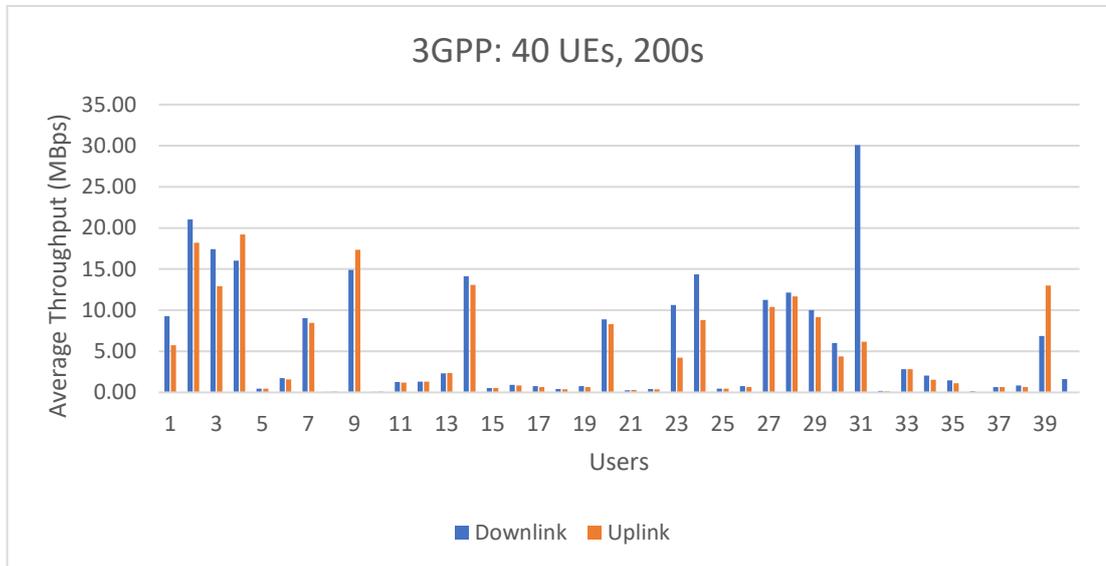


Figure 6.3-6: Average user throughputs 3GPP (40UEs, 200s)

The data shows some users have a data rate < 1MBps, this is because there are many users connected to the macro cell. Unlike the 3GPP scheme, it can be noticed that the proposed scheme was able to connect as many users to the micro cell as possible. Therefore, this subsequently enhanced the average macro cell throughputs higher than the 3GPP solution. Resulting in, 15 users having a data rate < 1MBps for the 3GPP scheme, compared to only 7 for the proposed scheme (a 53% reduction in these types of events).

### 6.3.3 User data rate requirements

Lastly, the percentage of the time the UEs data rate needs were met was measured. The results for the lower density of users are shown first (in Figures 6.3-7 and 6.3-8). One definition requires to be stated is that of good accommodating service. In this research it is defined to be, where the user's data rates are met for at least 50% of the time.

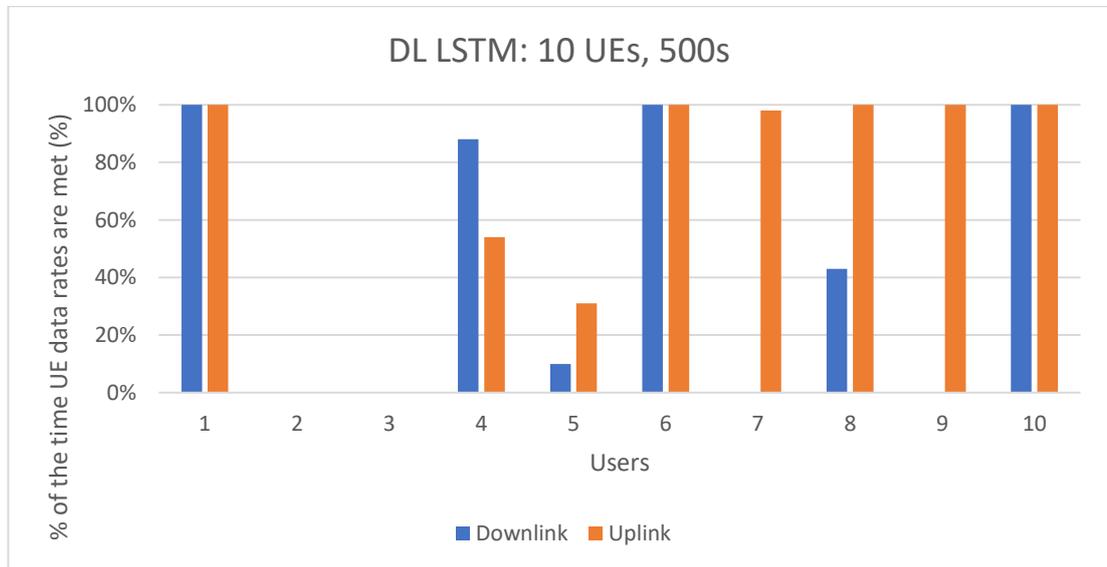


Figure 6.3-7: User throughput requirements met with DL LSTM scheme (10UEs, 500s)

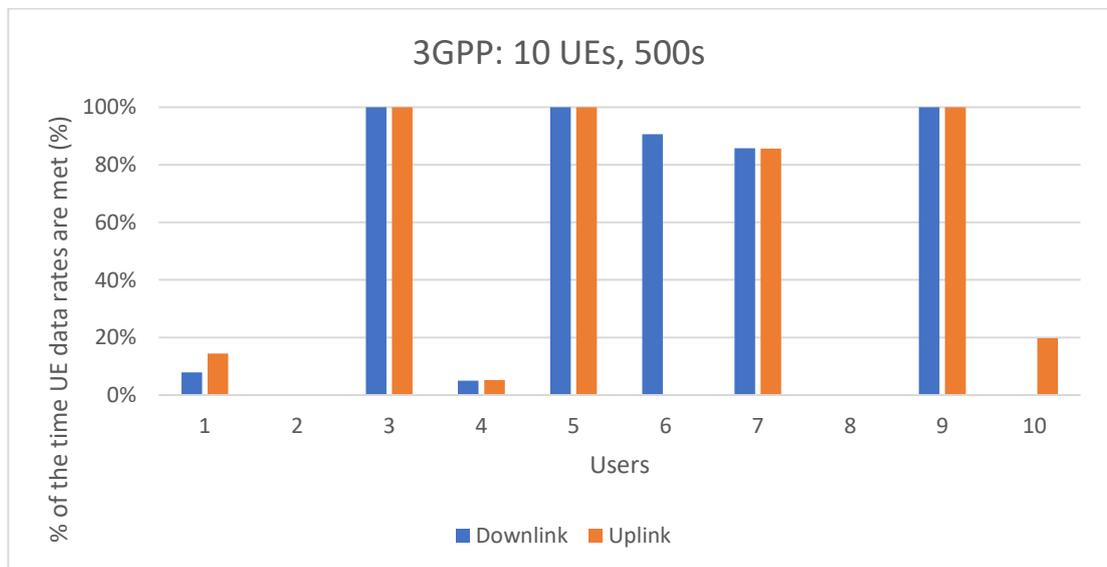


Figure 6.3-8: User throughput requirements met with 3GPP scheme (10UEs, 500s)

In majority of the cases, the proposed scheme provided a better accommodation to user data rate requirements. From the results the proposed scheme provided a good accommodating service 55% of the time, compared to 45% of the time for the 3GPP solution.

The only time the proposed scheme was not effective was when the CQI values of the connected users to the macro cell was very high. From the LUT, the only time the user would have handed over was if the CQI for the potential base station was above a value of 3, which was not the case

for UE 2 and 3. Therefore, they remained connected to the macro base station the whole simulation, and UE data rate requirements remained unmet.

Then when analysing the higher density scenarios, the gap between the two schemes is increased dramatically. This can be noticed in Figures 6.3-9 and 6.3-10.

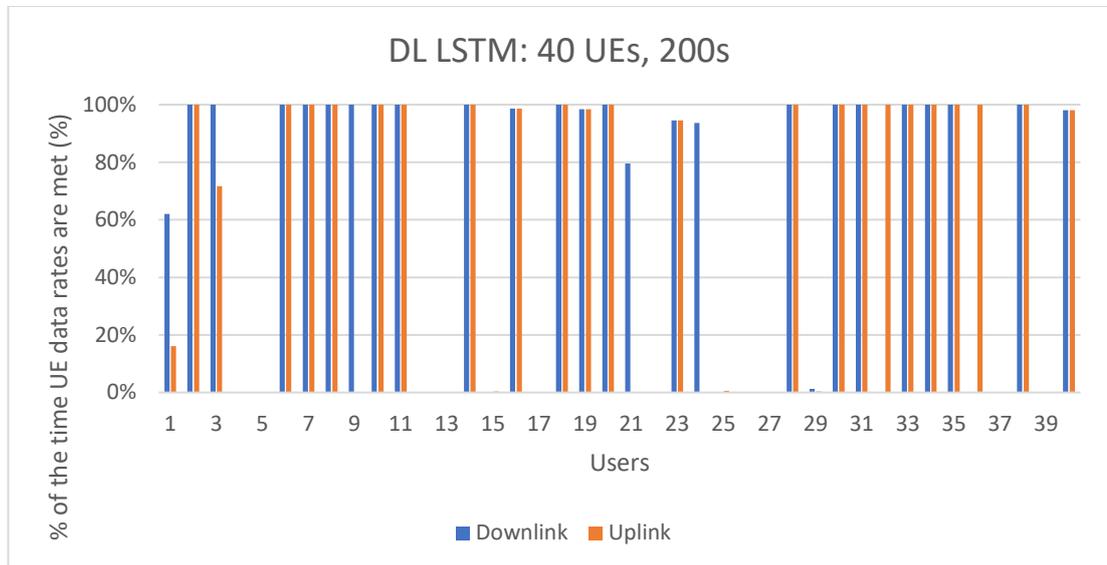


Figure 6.3-9: User throughput requirements met with DL LSTM scheme (40UEs, 200s)

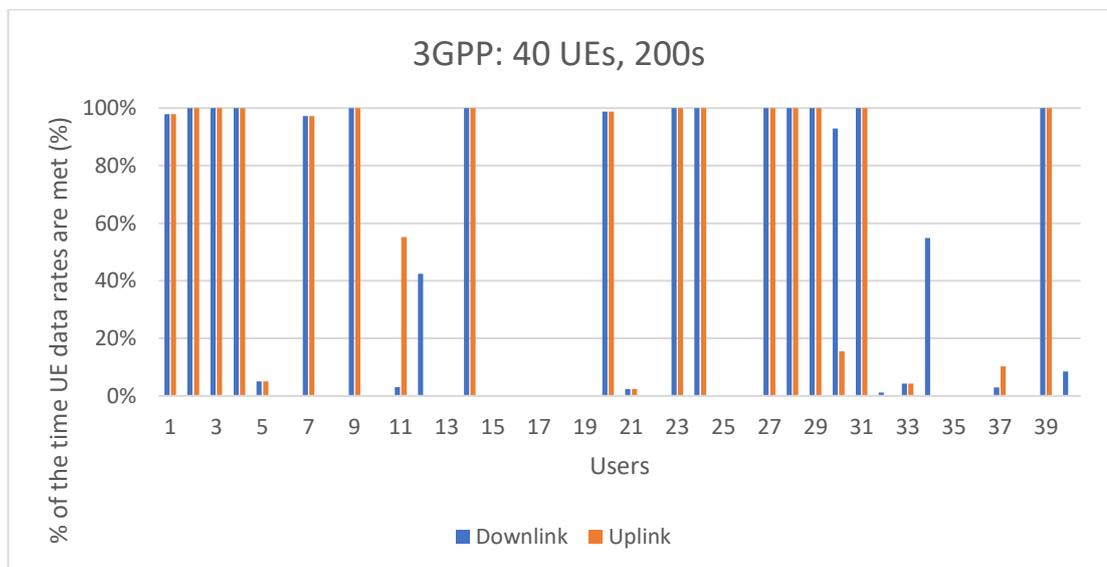


Figure 6.3-10: User throughput requirements met with 3GPP scheme (40UEs, 200s)

It can be observed that the user’s requirements are met 60% of the time, compared to the 3GPP which could only meet the requirements 41.25% of the time.

The reduction for the 3GPP solution as the density increased can be attributed to the points discussed in 6.2.2. When it was made apparent that the UE was performing small cell to macro cell handovers, therefore, compromising data rates to maintain connectivity.

The proposed scheme was able to maintain both, a good connectivity, and data rate.

## 7 Conclusion and Future Research

---

The project undertaken was interesting and it provided the author with a great deal of insight into various strains of 5G, edge computing and deep learning. Additionally, the learnings developed and sharpened various other skillsets while progressing throughout this research. These include acknowledging diverse perspectives, ethical reasoning, and applied learnings. The objective and contribution targets of this thesis have been achieved.

A DL LSTM handover decision algorithm while utilising LUTs, 5G and MEC was proposed, to investigate the impact that it would have on UEs and BSs when compared to the benchmark 3GPP algorithm. It was noticed that from the results that QoE targets were achieved and provided an increasing gap of up to 45% higher, as the user densities increased. Subsequently, this helped improve the time that the UE requirements are met by 20%.

When removing the TTT and replacing it with a novel and dynamic triggering function, the proposed scheme provided a very fast response to UE mobilities, when the LUT requirements were met. This allowed majority of the QoS targets to be achieved, providing a lower HOF and Ping-Ping rate in higher user density scenarios by 30% and 86% respectively. The main drawback that was observed was a higher occurrence of handovers in high density scenarios. This is due to the algorithm attempting to accommodate user data rate requirements and/ or user CQI expectations.

The proposed modification to the handover admission control process reduced the handover latency down to approximately 4.2ms. This allowed the algorithm to nearly meet the user plane eMBB latency targets of less than 4ms (stated in Section 2.3). Lastly, the proposal to reduce the size of MRs sent from UEs, subsequently reduced the size of signalling overheads, and improved power efficiencies. In this simulation, this factor has been reduced by 2, as only Intra-RAT environments are considered (where a minimum of 4 BS readings per MR are needed). However, these improvements can be negated when the increase in the number of handovers are taken into consideration.

In conclusion, a detailed analysis was undertaken, and the aims of the research were satisfied. Given the constraints of time and the degree of knowledge, the research delivered useful results and learnings, that are deemed to be sufficient. These learnings provide a good foundation for broader research prospects.

### 7.1 Further Research Directions

The author has attempted to touch on one of the most important aspects of 5G multi-tier handover. There are various gaps in this research that have yet to be explored to provide a more comprehensive solution for all generations of telecommunications networks.

The avenues for further research are described below, although, these should not be interpreted as the only or the most important opportunities.

- Tuning of the LUT table parameters. Due to time constraints, the author was not able to tune and test out various LUT parameters. This will help investigate the optimal solution.
- An analysis on the practical aspects of implementing this research.
- Inter-RAT handovers for the older generation (4G) and the impact of not being able to utilise MEC to reduce latencies and computational time.
- How retraining of LSTMs based on specific environmental conditions impacts handover decisions. Whether it provides a better result in comparison to a pretrained model. This may help understand if an adaptive solution is required to ensure the best service is implemented for each localised area.

## References

---

- [1] Nokia, Nokia Bell Labs, "<https://www.nokia.com/>," October 2019. [Online]. Available: [https://onestore.nokia.com/asset/205407?\\_ga=2.237682304.1617205958.1603514543-1708625130.1586847988](https://onestore.nokia.com/asset/205407?_ga=2.237682304.1617205958.1603514543-1708625130.1586847988). [Accessed October 2020].
- [2] European 5G Observatory, "<https://5gobservatory.eu/>," European 5G Observatory, September 2020. [Online]. Available: <https://5gobservatory.eu/market-developments/5g-products/#:~:text=While%20first%20integrated%205G%20SoC,SD765%2F765G%20with%20its%20updated>. [Accessed October 2020].
- [3] Ericsson, "Ericsson Mobility Report," Fredrik Jejdling, Executive Vice President and Head of Business Area Networks, 2019.
- [4] Samsung Electronics Co., Ltd., "<https://www.samsung.com/global/business/networks/>," Samsung Electronics Co., Ltd., 2018. [Online]. Available: [https://images.samsung.com/is/content/samsung/p5/global/business/networks/insights/white-paper/who-and-how\\_making-5g-nr-standards/who-and-how\\_making-5g-nr-standards.pdf](https://images.samsung.com/is/content/samsung/p5/global/business/networks/insights/white-paper/who-and-how_making-5g-nr-standards/who-and-how_making-5g-nr-standards.pdf). [Accessed 2020].
- [5] ITU (International Telecommunication Union), "Setting the Scene for 5G: Opportunities and Challenges," ITU (International Telecommunication Union), Geneva, Switzerland, 2018.
- [6] Nokia, "<https://www.nokia.com/>," Nokia, August 2019. [Online]. Available: <https://www.nokia.com/blog/giant-leap-5g-cloud-ran/>. [Accessed October 2020].
- [7] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta and D. Sabella, "On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration," *IEEE COMMUNICATIONS SURVEYS & TUTORIALS*, vol. 19, no. 3, pp. 1657-1681, 2017.
- [8] Huawei Technologies Co., Ltd., "<https://www.huawei.com/>," Huawei Technologies Co., Ltd., January 2019. [Online]. Available: <https://www.huawei.com/nz/industry-insights/outlook/mobile-broadband/wireless-for-sustainability/cases/worlds-first-remote-operation-using-5g-surgery>. [Accessed October 2020].
- [9] AgriTech Tomorrow, "<https://www.agritechtomorrow.com/>," AgriTech Tomorrow, July 2020. [Online]. Available: <https://www.agritechtomorrow.com/article/2020/07/5g-is-coming-to-agriculture/12275>. [Accessed October 2020].
- [10] S. Galea-Pace, "<https://www.manufacturingglobal.com/>," September 2020. [Online]. Available: <https://www.manufacturingglobal.com/technology/transformation-5g-manufacturing>. [Accessed October 2020].
- [11] T. Mitchell, *Machine Learning*, McGraw-Hill Science/ Engineering/ Math, 1997.

- [12] M. McClellan, C. Cervello-Pastor and S. Sallent, "Deep Learning at the Mobile Edge: Opportunities for 5G Networks," *Applied Sciences*, vol. 10, no. 4735, 2020.
- [13] M. Chen, U. Challita, W. Saad, C. Yin and M. Debbah, "Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3039-3071, 2019.
- [14] E. Kavlakoglu, "https://www.ibm.com," IBM, May 2020. [Online]. Available: <https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks>. [Accessed October 2020].
- [15] 3GPP, *TS 38.300 - NR; NR and G-RAN Overall Description*, 3GPP, 2020.
- [16] 3GPP, *TS 38.104 - NR; Base Station (BS) radio transmission and reception*, 3GPP, 2020.
- [17] 3GPP, *TS 36.873 - Study on 3D channel model for LTE*, 3GPP, 2018.
- [18] Matlab, "https://mathworks.com," MathWorks, 2020. [Online]. Available: <https://au.mathworks.com/help/5g/ug/nr-fdd-scheduling-performance-evaluation.html>. [Accessed March 2020].
- [19] S. Karlsson and A. Lugn, "https://www.ericsson.com/en," Telefonaktiebolaget LM Ericsson, [Online]. Available: <https://www.ericsson.com/en/about-us/history/changing-the-world/the-nordics-take-charge/nmt-approved-as-standard>. [Accessed October 2020].
- [20] I. Grigorik, "https://hpbn.co/," 2013. [Online]. Available: <https://hpbn.co/mobile-networks/>. [Accessed 2020].
- [21] S. Mukhopadhyay, V. Agarwal, S. Sharma and V. Gupta, "A Study On Wireless Communication Networks Based On Different Generations," *International Journal of Current Trends in Engineering & Research (IJCTER)*, vol. 2, no. 5, pp. 300-304, 2016.
- [22] A. Ghayas, "https://commsbrief.com/," Commsbrief Limited, 3 March 2020. [Online]. Available: <https://commsbrief.com/what-do-the-terms-1g-2g-3g-4g-and-5g-really-mean/>. [Accessed November 2020].
- [23] Nokia, "https://www.nokia.com," Nokia, 16 October 2020. [Online]. Available: <https://www.nokia.com/about-us/newsroom/articles/open-ran-explained/>. [Accessed December 2020].
- [24] Gartner, "https://www.gartner.com," Gartner, [Online]. Available: <https://www.gartner.com/en/information-technology/glossary/bts-base-transceiver-station>. [Accessed December 2020].
- [25] M. Murakami, H. Li and J.-d. Ryou, "https://www.itu.int," 26 August 2014. [Online]. Available: [https://www.itu.int/en/ITU-T/C-I/interop/Documents/20140826/CI-2\\_INP-09\\_Packet\\_Transport\\_Networks\\_Overview\\_and\\_Future\\_Direction.pdf](https://www.itu.int/en/ITU-T/C-I/interop/Documents/20140826/CI-2_INP-09_Packet_Transport_Networks_Overview_and_Future_Direction.pdf). [Accessed December 2020].
- [26] Viavi Solutions, "https://www.viavisolutions.com," Viavi Solutions, [Online]. Available: <https://www.viavisolutions.com/en-us/fronthaul>. [Accessed December 2020].

- [27] S. Hsu, "https://www.ufispace.com," UfiSpace, 13 November 2020. [Online]. Available: <https://www.ufispace.com/company/blog/what-is-cran-the-evolution-from-dran-to-cran>. [Accessed December 2020].
- [28] Ericsson, "https://www.ericsson.com," 2020. [Online]. Available: <https://www.ericsson.com/495922/assets/local/policy-makers-and-regulators/5-key-facts-about-5g-radio-access-networks.pdf>. [Accessed December 2020].
- [29] Lambda Gain, "https://www.lambdagain.com," Lambda Gain, June 2018. [Online]. Available: <https://www.lambdagain.com/learning-center/chapter-2-d-ran/>. [Accessed December 2020].
- [30] Lambda Gain, "https://www.lambdagain.com," Lambda Gain, October 2019. [Online]. Available: <https://www.lambdagain.com/learning-center/chapter-3-c-ran/>. [Accessed December 2020].
- [31] Search Networking, "https://searchnetworking.techtarget.com," Search Networking, October 2018. [Online]. Available: [https://searchnetworking.techtarget.com/definition/radio-access-network-RAN?\\_ga=2.203974288.1847071535.1599037084-940098927.1599037084](https://searchnetworking.techtarget.com/definition/radio-access-network-RAN?_ga=2.203974288.1847071535.1599037084-940098927.1599037084). [Accessed December 2020].
- [32] B. Lavallee, "https://www.ciena.com," Ciena, June 2020. [Online]. Available: <https://www.ciena.com/insights/articles/spotlight-on-5g-midhaul-networks.html>. [Accessed December 2020].
- [33] O-RAN ALLIANCE, "https://www.o-ran.org," O-RAN ALLIANCE, [Online]. Available: <https://www.o-ran.org/about>. [Accessed December 2020].
- [34] 3GPP, *TS 38.101 - NR; User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone*, 3GPP, 2020.
- [35] Fujitsu, "https://www.fujitsu.com," March 2013. [Online]. Available: <https://www.fujitsu.com/us/Images/High-Capacity-Indoor-Wireless.pdf>. [Accessed November 2020].
- [36] C. A. Balanis and P. Ioannides, "3.2 Antenna Classification," in *Introduction to Smart Antennas*, Arizona, Morgan and Claypool, 2007, pp. 22-23.
- [37] Matlab, "https://au.mathworks.com," MathWorks, 2011. [Online]. Available: <https://au.mathworks.com/help/phased/ref/fspl.html>. [Accessed December 2020].
- [38] A. Grami, "Chapter 12 - Wireless Communications," in *Introduction to Digital Communications*, Academic Press, 2016, pp. 493-527.
- [39] Qualcomm, "https://www.qualcomm.com," [Online]. Available: <https://www.qualcomm.com/media/documents/files/fdd-tdd-comparison.pdf>. [Accessed 2020].
- [40] M. Ergen, "Chapter 7: SC-FDMA," in *Mobile Broadband Including WiMAX and LTE*, Boston, Springer, 2008, pp. 261-267.

- [41] 3GPP, *TS 38.211 - NR; Physical channels and modulation*, 3GPP, 2021.
- [42] E. Shakshuki, X. Xinyu and M. Haroon, "An Introduction to Wireless Multimedia Sensor Networks," in *Handbook of Research on Mobile Multimedia, Second Edition*, Hershey, IGI Global, 2009, p. 15.
- [43] Metaswitch, "<https://www.metaswitch.com/>," Metaswitch, [Online]. Available: <https://www.metaswitch.com/knowledge-center/reference/what-is-the-5g-access-and-mobility-management-function-amf#:~:text=Assuming%20the%20responsible%20for%20mobility,is%20termed%20an%20Xn%20handover..> [Accessed December 2020].
- [44] 3GPP, *TS 38.423 - NG-RAN; Xn Application Protocol (XnAP)*, 3GPP, 2021.
- [45] 3GPP, *TS 23.501 - System architecture for the 5G System (5GS)*, 3GPP, 2020.
- [46] 3GPP, *TS 38.331 - NR; Radio Resource Control (RRC); Protocol specification*, 3GPP, 2021.
- [47] Techplayon, "<http://www.techplayon.com/>," Techplayon, 26 February 2020. [Online]. Available: <http://www.techplayon.com/5g-nr-measurement-events/>. [Accessed December 2020].
- [48] Juniper Networks, "<https://www.juniper.net/>," Juniper Networks, June 2017. [Online]. Available: [https://www.juniper.net/documentation/en\\_US/design-and-architecture/mobile-cloud/information-products/topic-collections/mca-data-centers-wp.pdf](https://www.juniper.net/documentation/en_US/design-and-architecture/mobile-cloud/information-products/topic-collections/mca-data-centers-wp.pdf). [Accessed December 2020].
- [49] Nokia, "<https://onestore.nokia.com/>," 2018. [Online]. Available: [https://onestore.nokia.com/asset/202184?\\_ga=2.17617593.1474324557.1610241338-1708625130.1586847988](https://onestore.nokia.com/asset/202184?_ga=2.17617593.1474324557.1610241338-1708625130.1586847988). [Accessed December 2020].
- [50] ITU, "Minimum requirements related to technical performance for IMT-2020 radio interface(s)," ITU, Geneva, 2017.
- [51] A. Biswal, "<https://www.simplilearn.com/>," Simplilearn, October 2020. [Online]. Available: <https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-algorithm>. [Accessed December 2020].
- [52] S. Hochreiter and J. Schmidhuber, "LONG SHORT-TERM MEMORY," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.
- [53] F. Gers, F. Cummins, S. Fernandez, J. Bayer, D. Wierstra, J. Togelius, F. Gomez, M. Gagliolo and A. Graves, "<http://colah.github.io/>," Github, 27 August 2015. [Online]. Available: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>. [Accessed June 2019].
- [54] Matlab, "<https://au.mathworks.com/>," Mathworks, [Online]. Available: <https://au.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html#:~:text=LSTM%20networks%20support%20input%20data,batch%20have%20the%20specified%20length..> [Accessed December 2020].

- [55] F. A. Gers and J. Schmidhuber, "Recurrent nets that time and count," in *International Joint Conference on Neural Networks (IJCNN)*, Como, 2000.
- [56] K. Cho, D. Bahdanau, F. Bougares, H. Schwenk and Y. Bengio, "Learning Phrase Representations using RNN Encoder–Decoder for statistical machine translation," in *Conference on Empirical Methods in Natural Language Processing*, 2014.
- [57] H. Zhang, W. Huang and Y. Liu, "Handover Probability Analysis of Anchor-Based Multi-Connectivity in 5G User-Centric Network," *IEEE Wireless Communications Letters*, vol. 8, no. 2, pp. 396-399, 2019.
- [58] J.-H. Choi and D.-J. Shin, "Generalized RACH-less Handover for Seamless Mobility in 5G and Beyond Mobile Networks," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1264-1267, 2019.
- [59] Z. Ali, N. Baldo, J. Mangues-Bafalluy and L. Giupponi, "Machine learning based handover management for improved QoE in LTE," in *NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium*, Istanbul, 2016.
- [60] Z. Wang, L. Li, Y. Xu, H. Tian and C. Cui, "Handover Control in Wireless Systems via Asynchronous Multiuser Deep Reinforcement Learning," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4296-4307, 2018.
- [61] N. Agoulmine, "Chapter 7 - Autonomics in Radio Access Networks, From Concepts to Applications," in *Autonomic Network Management Principles*, Academic Press, 2010, pp. 141-166.
- [62] A. B. Ozyurt, M. Basaran and L. Durak-Ata, "Impact of Self-Configuration on Handover Performance in Green Cellular Networks," in *2018 Advances in Wireless and Optical Communications (RTUWO)*, Riga, 2018.
- [63] C. Suarez-Rodriguez, B. A. Jayawickrama, F. Bader, E. Dutkiewicz and M. Hiemlich, "REM-Based Handover Algorithm for next-generation multi-tier cellular networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Barcelona, 2018.
- [64] K. Ahmed, I. Bilal, I. L. Khan, A. Shakeel and R. Hussain, "Improved Handover Triggering Estimation for end-to-end Vertical Handover Schemes," in *International Conference on Information Communication and Management (ICICM)*, Hatfield, 2016.
- [65] F. Zhao, H. Tian, G. Nie and H. Wu, "Received Signal Strength Prediction Based Multi-Connectivity Handover Scheme for Ultra-Dense Networks," in *24th Asia-Pacific Conference on Communications (APCC)*, Ningbo, 2018.
- [66] M. Polese, M. Giordani, M. Mezzavilla, S. Rangan and M. Zorzi, "Improved Handover Through Dual Connectivity in 5G mmWave Mobile Networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 2069-2084, 2017.
- [67] S. C. Evangelina and V. B. Kumaravelu, "Decision Process for Vertical Handover in Vehicular Adhoc Networks," in *International conference on Microelectronic Devices, Circuits and Systems (ICMDCS)*, Vellore, 2017.
- [68] J. M. Marquez-Barja, H. Ahmadi, S. M. Tornell, C. T. Calafate, J.-C. Cano, P. Manzoni and L. A. DaSilva, "Breaking the Vehicular Wireless Communications Barriers: Vertical

- Handover Techniques for Heterogeneous Networks,” *IEEE Transactions on Vehicular Technology*, vol. 64, no. 12, pp. 5878-5890, 2014.
- [69] M. Khan, C. Jung, P. C. Uzoh, C. Zhenbo, J. Kim, Y. Yoon, A. Nadeem and K. Han, “Enabling vertical handover management based on decision making in heterogeneous wireless networks,” in *International Wireless Communications and Mobile Computing Conference, IWCMC*, Dubrovnik, 2015.
- [70] M. Alhabo and L. Zhang, “Multi-Criteria Handover Using Modified Weighted TOPSIS Methods for Heterogeneous Networks,” *IEEE Access*, vol. 6, pp. 40547-40558, 2018.
- [71] S. Zang, W. Bao, P. L. Yeoh, B. Vucetic and Y. Li, “Managing Vertical Handovers in Millimetre Wave Heterogeneous Networks,” *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1629-1644, 2018.
- [72] X. Xu, X. Tang, Z. Sun, X. Tao and P. Zhang, “Delay-Oriented Cross-Tier Handover Optimization in Ultra-Dense Heterogeneous Networks,” *IEEE Access*, vol. 7, pp. 21769-21776, 2019.
- [73] M. M. Hasan, S. Kwon and S. Oh, “Frequent-Handover Mitigation in Ultra-Dense Heterogeneous Networks,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 1035-1040, 2019.
- [74] K. Da Costa Silva, Z. Becvar and C. Renato Lisboa Frances, “Adaptive Hysteresis Margin Based on Fuzzy Logic for Handover in Mobile Networks with Dense Small Cells,” *IEEE Access*, vol. 6, pp. 17178 - 17189, 2018.
- [75] D. Castro-Hernandez and R. Paranjape, “Optimization of Handover Parameters for LTE/LTE-A in-Building Systems,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, pp. 5260-5273, 2018.
- [76] B. Shubyn, N. Lutsiv, O. Syrotynskyi and R. Kolodii, “Deep Learning based Adaptive Handover Optimization for Ultra-Dense 5G Mobile Networks,” in *IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*, Lviv-Slavske, 2020.
- [77] H. Zhang, C. Jiang, J. Cheng and V. C. M. Leung, “Cooperative Interference Mitigation and Handover Management for Heterogeneous Cloud Small Cell Networks,” *IEEE Wireless Communications*, vol. 22, no. 3, pp. 92-99, 2015.
- [78] T. Kolding, L. C. Gimenez and K. I. Pedersen, “Optimizing Synchronous Handover in Cloud RAN,” in *IEEE Conference on Vehicular Technology (VTC)*, Toronto, 2017.
- [79] M. K. Muller, F. Ademaj, T. Dittrich, A. Fastenbauer, B. R. Elbal, A. Nabavi, L. Nagel, S. Schwarz and M. Rupp, “Flexible multi-node simulation of cellular mobile communications: the Vienna 5G System Level Simulator,” *EURASIP Journal on Wireless Communications and Networking*, vol. 1, pp. 1-17, 2018.
- [80] Tuwien, “<https://www.nt.tuwien.ac.at/>,” Tuwien, 2018. [Online]. Available: <https://www.nt.tuwien.ac.at/research/mobile-communications/vccs/vienna-5g-simulators/>. [Accessed September 2019].

- [81] Tuwien, "<https://www.nt.tuwien.ac.at/>," 2018. [Online]. Available: [https://www.nt.tuwien.ac.at/wp-content/uploads/2020/10/listOfFeatures\\_v1\\_1.pdf](https://www.nt.tuwien.ac.at/wp-content/uploads/2020/10/listOfFeatures_v1_1.pdf). [Accessed September 2019].
- [82] Matlab, "<https://au.mathworks.com/>," Mathworks, [Online]. Available: <https://au.mathworks.com/products/statistics.html>. [Accessed September 2019].
- [83] Matlab, "<https://au.mathworks.com/>," Mathworks, [Online]. Available: <https://au.mathworks.com/products/parallel-computing.html>. [Accessed September 2019].
- [84] 3GPP, *TR 25.943 - Deployment aspects*, 3GPP, 2020.
- [85] 3GPP, *TR 25.890 - High Speed Downlink Packet Access (HSDPA); User Equipment (UE) radio transmission and reception (FDD)*, 3GPP, 2002.
- [86] 3GPP, *TR 38.901 - Study on channel model for frequencies from 0.5 to 100 GHz*, 3GPP, 2020.
- [87] Matlab, "<https://au.mathworks.com/>," Mathworks, 2020. [Online]. Available: <https://au.mathworks.com/products/5g.html>. [Accessed 2020].
- [88] Matlab, "<https://au.mathworks.com/>," Mathworks, 2020. [Online]. Available: <https://au.mathworks.com/help/5g/ug/nr-tdd-symbol-based-scheduling-performance-evaluation.html>. [Accessed 2020].
- [89] 3GPP, *TS 38.214 - NR; Physical layer procedures for data*, 3GPP, 2021.
- [90] M. H. Habaebi, J. Chebil, A. G. Al-Sakkaf and T. H. Dahawi, "Comparison between scheduling techniques in long term evolution," *IJUM Engineering Journal*, vol. 14, no. 1, pp. 66-75, 2013.
- [91] Matlab, "<https://au.mathworks.com/>," Mathworks, 2020. [Online]. Available: <https://au.mathworks.com/products/automated-driving.html>. [Accessed 2020].
- [92] Matlab, "<https://au.mathworks.com/>," Mathworks, 2020. [Online]. Available: <https://au.mathworks.com/products/deep-learning.html>. [Accessed 2020].
- [93] J. Brownlee, "<https://machinelearningmastery.com/>," Machine Learning Mastery, 13 January 2021. [Online]. Available: <https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/>. [Accessed January 2021].
- [94] "Satellite image of New York University campus," Google, 2020. [Online]. Available: <https://www.google.com/maps/@40.7312586,-73.9976468,1052a,35y,32.79h,1.21t/data=!3m1!1e3>. [Accessed 2020].
- [95] "Satellite image of New York University campus," Google, 2020. [Online]. Available: <https://www.google.com/maps/@40.7297774,-73.9960968,665a,35y,32.46h,5.97t/data=!3m1!1e3>. [Accessed 2020].
- [96] Ministry Of Business, Innovation and Employment, New Zealand, "<https://www.rsm.govt.nz/>," 2018. [Online]. Available: <https://www.rsm.govt.nz/assets/Uploads/documents/consultations/2018-preparing->

for-5g/5a7e7b8f73/preparing-for-5g-in-new-zealand-discussion-document.pdf.  
[Accessed 2020].

- [97] G. Brown, "New Transport Network Architectures for 5G RAN," Heavy Reading.
- [98] M. L. Chávez, "Chapter 6 - Communication Network Architecture," in *Fieldbus Systems and Their Applications 2005*, Elsevier, 2006, pp. 107-114.
- [99] 3GPP, *TS 38.133 - NR; Requirements for support of radio resource management*, 3GPP, 2021.
- [100] T. Abbas, K. Sjöberg, J. Karedal and F. Tufvesson, "A Measurement Based Shadow Fading Model for Vehicle-to-Vehicle Network Simulations," *International Journal of Antennas and Propagation*, vol. 2015, 2015.
- [101] I. Goodfellow, Y. Bengio and A. Courville, "Chapter 5 - Machine Learning Basics," in *Deep Learning*, Massachusetts, MIT Press, 2016, p. 110.