T.C.

ALTINBAS UNIVERSITY

Institute of Graduate Studies

Electrical and Computer Engineering

IMPLEMENTING BIG DATA TECHNIQUES IN 5G MOBILE NETWORKS

Mahmoud ERRAMAH

Master of Science

Supervisor

Prof. Dr. Oğuz BAYAT

Istanbul, 2020

IMPLEMENTING BIG DATA TECHNIQUES IN 5G MOBILE NETWORKS

by

Mahmoud ERRAMAH

Electrical and Computer Engineering

Submitted to the Institute of Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Science

ALTINBAŞ UNIVERSITY

2020

The thesis titled “IMPLEMENTING BIG DATA TECHNIQUES IN 5G MOBILE NETWORKS” prepared and presented by Mahmoud ERRAMAH was accepted as a Master of Science Thesis in Electrical and Computer Engineering.

____
Prof. Dr. Oğuz BAYAT

Supervisor

Thesis Defense Jury Members:

Prof. Dr. Oğuz BAYAT	School of Engineering and Natural Sciences, Altinbas University	___________________
Asst. Prof. Dr. Abdullahi Abdu Ibrahim	School of Engineering and Natural Sciences, Altinbas University	___________________
Asst. Prof. Dr. Izzet Paruğ DURU	Faculty of Physics to Vocational school, Istanbul Gedik University	___________________

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Approval Date of Institute of Graduate Studies: ____/____/____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Mahmoud ERRAMAH

DEDICATION

I devote and pledge this research work to my supervisor who is salient for guiding me through whole research work as well as my family for always assisting me in my hard time.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Prof. Dr. Oğuz BAYAT for guiding and helping me along the way in writing this dissertation. Discussing my progress, problems, and ideas with my supervisor Prof. Dr. Oğuz BAYAT a couple of times every week helped me tremendously in understanding the logic behind the research. It made me better realize the technical need for this research work. He helped me by looking outside of all the telecommunication-based information, and considering the machine learning based knowledge from a more holistic perspective than just from the productivity point of view. He patiently explained to me a great number of technical details with regards to algorithms, for example how to deal with such an unstable learning method and to provide insight into the reliability of those results. Without his time and help none of this research would have been possible, and I'm grateful to him for giving me the opportunity to work on this interesting subject.

ABSTRACT

IMPLEMENTING BIG DATA TECHNIQUES IN 5G MOBILE NETWORKS

ERRAMAH, Mahmoud,

M.Sc., Electrical and Computer Engineering, Altınbaş University,

Supervisor: Prof. Dr. Oğuz BAYAT

Date: Feb /2021

Pages: 59

With regard to the forthcoming requirements for mobile services in 5G networks, several new technologies have recently become the focus of leading-edge research. One of those is in-band full-duplex communications in processing data with big data based apache spark. The idea is to simply employ the same frequency band to simultaneously transmit and receive information, allowing more spectrally efficient communications when compared to the traditional half-duplex or out-of-band full-duplex counterparts. By breaking a long-held assumption in wireless communications, in-band full-duplex may double the throughput or reduce by half the allocated bandwidth in future transmissions. Nevertheless, the self-inflected interference, that naturally arises, poses the major problem to the use of this technique, enhanced when in presence of system impairments with apache spark. Therefore, this thesis aims to study wireless communication in 5G network with use of big data based apache spark. Particularly, apache spark suppression filters are studied, as well as feedback adaptive filtering is proposed for relay systems with multiple-input multiple-out (MIMO) antennas. Furthermore, the orthogonal properties of large-scale arrays are resorted to, such that an extra level of mitigation is achieved. In this case, the relay system energy efficiency is maximized by finding the optimal transmit powers, while maintaining a certain individual link quality. For this scenario, the effect of massive MIMO is likewise addressed, and an algorithm that maximizes the system total achievable rate is derived. The network of wireless communication requires low latency and high availability of 5G mobile network connections to enable co-operation between the networks and infrastructure. The performance and functional properties of the network slices can be modified optimally with network slicing, to meet these requirements with apache spark. These new 5G technologies, however, have pressing security challenges to secure data integrity and privacy in critical wireless communications requiring extensive research before implementing these technologies into use. This research examines the new possibilities 5G networks offer for wireless network communication with high bandwidth of 90 GHz. The prediction of 5G network bandwidth maximum recoded at 90 GHz on 5 nodes of apache spark at 80% of data trained with total 12 nodes. The main question to answer is will the 5G technology offers sufficient performance for wireless communication.

Keywords: 5G, in-band full-duplex, MIMO, big data, wireless communications, apache spark, spectrum allocation.

TABLE OF CONTENTS

Pages

LIST OF FIGURES

Pages

LIST OF TABELS

Pages

LIST OF ABBREVIATIONS

3GPP	:	The Third Generation Partnership Project
5G	:	Fifth Generation
BER	:	Bit Error Rate
eMMB	:	Enhanced Mobile Network
IoT	:	Internet of Things
LTE	:	Long Term Evolution
Mbps	:	Megabytes Per Second
MIMO	:	Multiple Input Multiple Output
QAM	:	Quadrature amplitude modulation
TDD	:	Time-Division Duplexing
RSU	:	Road Side Unit
WLAN	:	Wireless Local Access Networks

1. INTRODUCTION

1.1 OVERVIEW

The constant demand for higher rates and lower latencies has already driven the scientific community and industry to pursue a new generation of mobile communications. Simply known as the fifth generation (5G), and with expected release date in 2021, telecommunications are about to go through a deep change in their foundations. One of those central changes is the long-held, but now obsolete, assumption that radios can only simultaneously transmit and receive in different frequency bands on the same channel as mentioned in [1]. Thus, the concept of in-band full-duplex communication, i.e., simultaneously employing the same frequency band to transmit and receive data in wireless nodes, will certainly incorporate this new generation, providing a leap forward in terms of spectral usage efficiency. Moreover, when integrated under a physical layer network perspective or when combined with the rise in the number of antennas at base stations, its gains are beyond those expected few years ago.

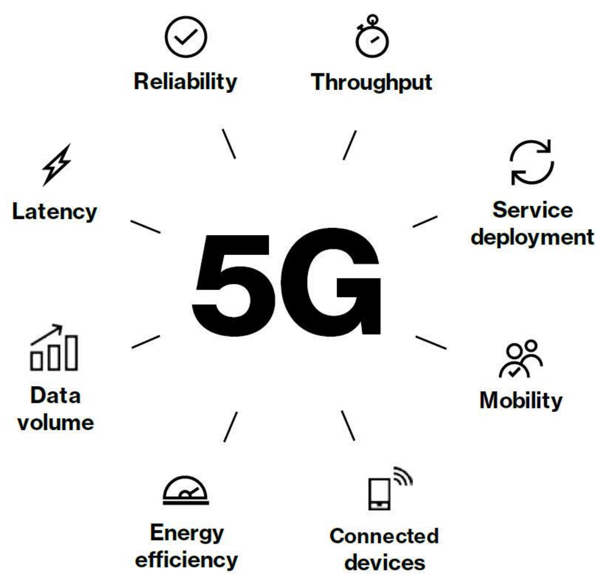


Figure 1.1 : Some of the most common use-cases of 5th generation wireless network [2].

This first chapter presents the main objective of this thesis and the reasons that motivated the work on in-band full-duplex communications. An historical overview of telecommunications, the current cutting-edge research and how this thesis may contribute to the progress of future 5G communications are also assessed.

In recent decades, the topic of mobile communication has been a subject of interest to the research and industry communities. Several different technologies were studied so that a standard system could be implemented, allowing wireless communications to be a reality today. In the last 20 years, different major market requirements have appeared, pushing the telecommunication industry to develop new solutions. Firstly, the need for wireless real-time voice communication boosted the first generation (1G) of analog cellular transmissions as mentioned in [3]. Then, short message service (SMS) was incorporated in a digital system that led to the second mobile generation (2G), the first to use a digital transmission. Due to the popularity of laptops and wireless local access networks (WLAN), wireless data connectivity became a hot topic and with it the third generation (3G) emerged as mentioned in [4]. Currently, the fourth generation (4G) of wireless connectivity has been derived from the need for faster links, aiming to provide higher bandwidth access anywhere.

Currently, the fifth generation (5G) is in a research stage where several methods and techniques are being explored, aiming to be among the wireless connectivity technologies used in 2021 as mentioned in [5]. For that reason, this thesis aims to study the problems rising from in-band full-duplex systems and develop some new considerations to make them a reality in the next wireless generation of communications. The relation between the big data and 5G is very obvious as the both technologies can be used parallel, for instance the big data can be used to store large amount of volume of data with the extra-ordinary network bandwidth of 5G in fraction of time. The 5G protocol also offers new services to support connected 5G wireless networks for big data connected devices such as vehicle platooning and remote driving. These communication nodes might lead to exposure of the end-user's location, thus communicating the privacy feature.

1.2 MOTIVATION

The demand of data exchanged in future wireless networks will certainly pose new challenges to the industry, and particularly to research and development (R&D) departments, requiring a shift of paradigm when compared to that observed today. Exponential growth is expected in cellular data traffic and a thousand-fold increase upon the current mobile data volume, in the next 10 years [6]. This demand for broadband cellular access, mainly for video applications, is driving the conceptualization of future radio mobile networks. The pursuit for the next generation of cellular communications has already started, boosted by European research organizations and projects such as 5GPP [7], 5GNOW and METIS [8], but also by the academia and industry all over the world. The main target is to create a standard communication protocol capable of solving expected application requirements in 2020. Among them are the recently popular topic of the Internet of Things (IoT), involving machine-type communication, the gigabit wireless connectivity (giving the same experience as in a wired connection), the concept of tactile Internet (real-time applications requiring latency of 1 ms) and many others [9]. As defined by the METIS European project [10], the future architecture should support:

- Mobile traffic volume 1000 times higher;
- Support 10 to 100 times more devices connected;
- Increase of up to 100 times throughput capacity;
- 5 times reduced latency.

Several different radio technologies are currently being studied and developed in order to achieve the above mentioned specifications. For example, in [10] the five major technologies that are believed to potentially fulfill 5G needs are pointed out:

Device-centric architectures: Changing the interaction between the base-station and the users, exploiting new control and data flows;

Millimeter waves: Moving towards higher frequencies, due to the spectrum scarcity at the current used carrier frequencies;

Massive MIMO: Using various antennas for multiplexing data for users at the same time instant and focusing energy in the user direction;

Smarter devices: Taking advantage of the device capability of processing information;

Machine-to-machine communication: Supporting a massive number of devices connected to the network, and allowing low data rate and very low latency applications.

As mentioned above, one of the possible future technologies is so-called device-centric architectures. The downlink-uplink paradigm between user and a base station in a mobile cell has necessarily to be replaced by a more cooperative network approach. Smarter communication systems are emerging, such as Cooperative multipoint (CoMP) or new more efficient relaying techniques, that will certainly increase the connection throughput of links by exploiting high spectrally efficient transmissions.

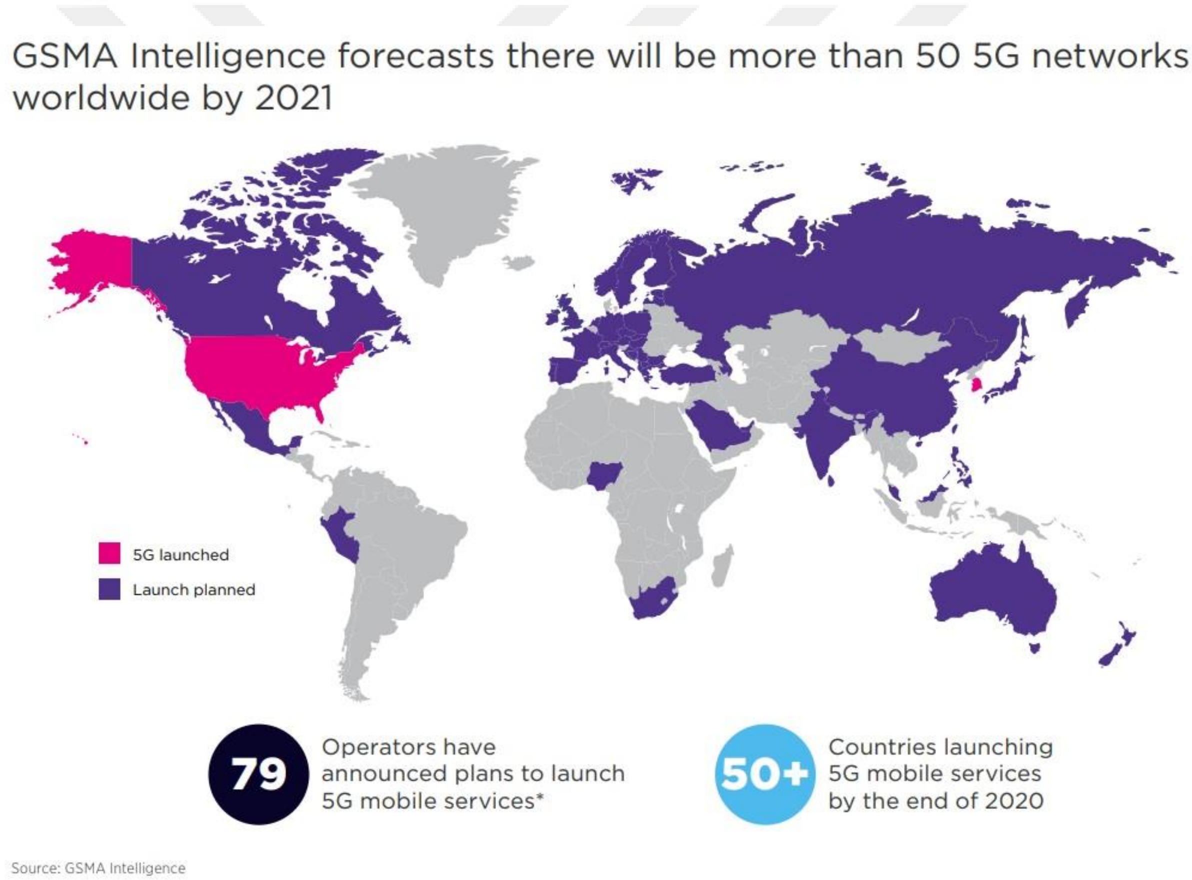


Figure 1.2 : Intelligence forecast and prediction of 5G deployment worldwide [11].

The idea of re-using the same spectral band to transmit and receive information at the same time, therefore operating in in-band full-duplex mode, theoretically doubles the link capacity when compared with the half-duplex mode. As a consequence, the current telecommunication concepts of frequency-division duplexing (FDD) and time-division duplexing (TDD) become obsolete, with the appearance of this new idea of any-division duplexing (ADD) and eliminating the long-held paradigm of any type of digital communication so far used. However, once frequency and time resources are used simultaneously, self-interference cancellation techniques have to be applied in order to isolate the desired signal from the interfering one [12]. Another important aspect for 5G that full-duplex may solve is spectrum management, by means of duplicating the available current spectrum (commonly known as spectrum virtualization). The possibility to enable almost instantaneous retransmissions for network connections, i.e., with no need to wait for a time slot or for the attribution of frequency bands, can also be viewed as an advantage. Combining this new concept of in-band full-duplex with the state-of-the-art ideas coming from physical layer network coding (PLNC) and large/massive multiple-input multiple-output (MIMO), the performance of multi-hop networks may be drastically improved, allowing each relay to serve more and more users with higher requirements. The integration of large-scale data is proposed for the 5G protocol and the performance degradation of the scheme for assessing different numbers of big data system on the network and for channel estimation errors with data limits. Additionally, the achievable sum rate of this system was derived and an optimization algorithm proposed in order to find the optimal transmit powers that maximize it. The scale data is usually in large amount and can only be processed with high speed like 5G and storage volume like big data process in terms of mount binding.

This technology can actually be implemented in current wireless networks. The main target of full-duplex technique is to enhance link capacity for future 5G, although there are some systems that would benefit from it with some slight modifications. Table 1.1 shows some different applications of in-band full-duplex to the industry beyond mobile communications and the process of technology implementation and standard adoption in modern and forthcoming networks.

Table 1.1 : Full-duplex current and future possible applications (adapted from [9]).

NO STANDARDS MODIFICATION	5G ADOPTION	BEYOND 5G CELLULAR
Self-backhaul for small cells	In-band full-duplex for 5G	Ad hoc mesh networks
Full-duplex microwave backhaul	5G spectrum virtualization	Military radio jammers
Adaptive duplexers for mobiles	5G radio spectrum	Multi-channel Wi-Fi AP

1.3 PROBLEM STATEMENT

In this thesis, in-band full-duplex exchange of information under a MIMO and under a physical wireless network perspective is assessed with big data. The main objective is to study and propose some techniques for in-band full-duplex systems for 5G network using the big data techniques, that take tools from current state-of-the-art technologies described in this work. Self-interference mitigation methods in big data, relaying techniques, physical layer networking and the use of massive antennas are further studied and proposed to integrate in-band full-duplex systems in 5G networks. This work focuses mainly on the theoretical side of telecommunications, always assuming transmissions in the digital baseband domain in big data. A simulation framework is then developed to evaluate and confirm the performance of the proposed 5G system. A research combining the in-band full-duplex transmission in a relay station with the use of a massive number of antennas of 5G network. A communication link is established between several pairs through the relay system in big data. This relay incorporates self-interference suppression filters, as well as detector and pre-coder filters, whose performance is evaluated at the relay and at the destinations with BER curves. Further, an optimization problem is solved that ensures a minimum quality link for the pairs, while minimizing the total energy spent.

1.4 RESEARCH CONTRIBUTIONS

With the widespread use of 5G technology, the number of users on communication network have also increased [13]. According to Gartner Research [14], the number of users on communication devices by 2050 will hit 18 billion. The scope of this research is beyond the operations of 5th generation networks as we determined its usage and criteria under the severe weather conditions. This study has prominence because it is state-of-the-art work performed under the research parameters of 5th generation networks using big data technology. Thus, we proposed the big data based on the dataset acquired. Furthermore, some variations in the context of traffic profile are performed, in terms of service throughput, service share percentages and different traffic profile configuration shares in some municipalities as mentioned in [15]. In terms of licensing and legalizing the use of big data frequency for possible because of its designing and development with the open-source material and methodology with high transmission power as well as high transmission rate under uncertain weather conditions using MIMO input/output antennas used. Firstly, this work will fill the gap of research where a 5G technology, which is under field testing and for a part never been tested under different weather conditions using both of big data toolbox and signal processing toolbox. This thesis provides an investigation of 5G technology dimensioning as objective for different weather conditions, while also permitting changes in input parameters for different network deployment configurations from big data. This process can be seen as a high-level approach to 5G network investigation under different conditions, where by the use of propagation models of 5G network and specific big data-based algorithms, it is possible to produce an output solution for communication with an acceptable degree of accuracy of 5G networks as mentioned in [16]. The input and output parameters are generically described in this work for capacity and coverage for communication for coverage planning in 5G network using big data techniques as mentioned in [17]. The most advanced big data techniques and model for 5th generation communication network as well as for coverage and capacity are defined in this work. Besides these essential elements for cellular network planning there is the usage of 5G network in terms of density/volume and share/mix profile throughout a typical load day are also been discussed.

1.5 THESIS ORGANIZATION

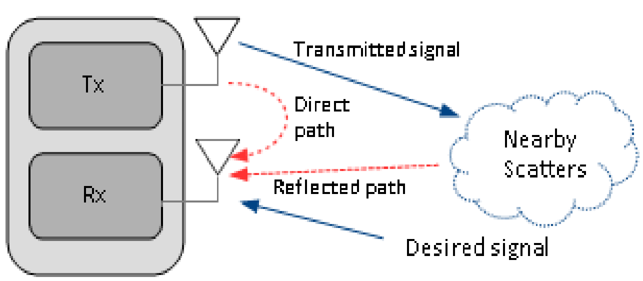
In the first chapter, background and motivation, research question, methodology and outline are presented. The 5G network based on high spectrum allocation faces challenges which are presented with a specific emphasis in introduction of big data with dependencies of problem statement and research contribution. The chapter ends with a description of the characteristics of an organization of thesis. The second chapter gives a basic description of the literature review on 5G network using big data, beginning with a small historical note, a definition of the commercially available types of available solutions and ending with a description of the major drawbacks of not classifying the big data needs for 5G network and the definition of high spectrum allocative system for real time communication usage monitoring in large scale environment. The third chapter presents the 5G network system with the methodology background; the concept of big data techniques with networking concepts is presented, followed by the methodology for intelligence based 5G network with spectrum allocation monitoring techniques, a characterization of the 5G network, and the method for energy consumption classification and reducing cost of 5G network energy monitoring system with big data by providing the cost analysis, the purpose of this chapter, go through a sensitivity analysis and are presented and discussed. The forth chapter in terms of structure is very similar to the third chapter but for the result of dissertation with aspect of this work, that is the 5G network with communication system using time related features on 5G network with high spectrum allocation using big data. A typical framework outline is followed: first the concept of 5G network based smart phones results is described, then the objective and scope of the analysis are being found, followed by data collection for 5G network using the big data analysis. The fifth chapter provides the discussion of the results with big data analysis and their respective discussion based on the limitations and comparison with other existing systems on 5G network with communication with high spectrum allocation monitoring for communication. The sixth chapter presents the main conclusions of this dissertation with a summary of the main findings, the consequences these results have in policy making, followed by the main limitations of this study and ends with leads for future work.

2. RELATED WORK

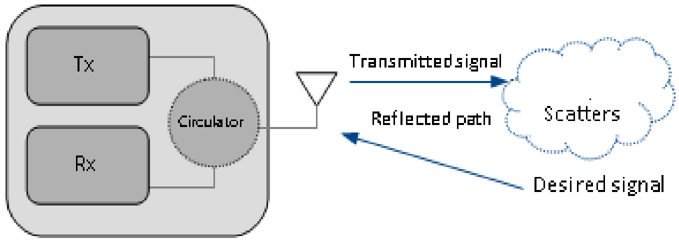
This thesis work is motivated by the goal of overcoming the challenges and difficulties that energy consumption units in 5G network faces with the help big data techniques in real time with their structure and the environmental hostility surrounds them, that is by creating a better environment solution, a solution that will avoid the shortcomings of the traditional networks for communication with data traffic handling and controlling operations. For the sake of explaining the motivation of this thesis work, a detailed review of all these difficulties and constraints is discussed in depth in the sections below; furthermore, a comprehensive review of 5G wireless network and their advantages with spectrum range allocation in this chapter.

2.1 LITERATURE REVIEW

The first approach to mitigate self-interference was based solely on propagation properties of electro-magnetic waves in 5G networks. The need to connect terminals that are too far to establish reliable communication brought about the idea of using relay stations as mentioned in [18]. In order to increase the coverage of networks, relays can operate within the same bandwidth and act as passive amplifiers by receiving and retransmitting an amplified version of the desired signal, known in literature as amplify-and-forward (AF). Therefore, wireless propagation insulation techniques were developed, aiming for separation of the receiving side from the transmitting side of the terminal as mentioned in [19]. This procedure consists of the first stage of self-interference suppression, and it is essential to reduce the signal power so that its range fits in the downstream receiver hardware RF band. There are two different schemes of propagation techniques, based on the design of antennas.



(a) Separated-antenna design.



(b) Shared-antenna design.

Figure 2.1 : Different possibilities to mitigate self-interference in 5G networks using wave propagation properties [19].

The first, separated-antenna design, consists of using different antennas to transmit and receive, as shown in Fig. 2.1(a). Using this procedure, it is then possible to suppress interference by path loss attenuation, increasing the antennas distance to further increase the attenuation or by placing absorptive material between them [20]. However, this approach is limited by the terminal design that today plays a serious and important part in commercial products. Cross-polarization solutions are also interesting possibilities to insulate antennas and enhance the interference suppression [21], for example, by transmitting in vertical polarization and by receiving in horizontal polarization. Another possibility is to exploit the antenna radiation pattern, carefully placing the receive antennas in radiation null points of the transmit antenna array [22]. The second one, shared-antenna design, in Fig. 2.1(b), consists of using the same antenna to transmit and receive. This is possible by means of a circulator, a circuit that routes the signals from the transmit chain to the antenna and from the antenna to the receive chain [23]. In the case of a relay station, it is common to place antennas back-to-back, i.e., arraying antennas together that connect each relay end. Furthermore, one can apply techniques such as wave traps [24], which extend the relay isolation by connecting short circuit resonant transmission lines, such as band-gap structures [25]. These materials have high surface impedance and for that reason attenuate the wave propagation. Also, one may use slots on the ground plane [26], which consist of internal dual band diversity antennas with dual frequency notched resonators to enhance electromagnetic attenuation.

The problem of both of the designs described above is sensitivity to environment conditions. For a rich scattering channel, the self-interference is composed of multiple copies of the transmitted signal, which may vary rapidly in time. It is possible that the desired signal is also deteriorated by applying these propagation methods, since they are blind to the carried information, i.e., the desired signal may also be attenuated by the propagation methods along with the self-interference component.

2.2 ANALOG RADIO CIRCUITS IN 5G

Analog circuit mitigation techniques are performed to increase attenuation on the self-interference signal and used as the second suppression stage of in-band full-duplex architectures. These circuits are placed before the analog-to-digital converter (ADC) and act either over the baseband or broadband analog received signal as mentioned in [27]. The main idea is to track the propagation effect of the interfering signal and by means of electronic processing in the analog domain subtract it from the received signal. Moreover, it is also possible to tap the transmitted signal in the digital domain, apply the interference channel effect by means of gain and phase adjustment, and convert it back to the analog domain so that one can subtract it from the received signal [28].

As in the propagation domain, some analog circuits are not able to track channel variations in multiple scatter conditions. These so called channel-unaware circuits are focused only on suppressing the interference direct path [29]. However, analog processing development has brought channel- aware circuits, that are capable of tracking channel variations and, therefore, more effectively cancel self-interference [30]. Both consist of a delay tapped line with amplitude and phase correction. For the first case, the circuit is calibrated once and before its operation stage. For the second case, it is possible to continuously adjust those values in order to track channel variations.

As is known from the literature, cellular channels are commonly frequency-selective, due to multipath received copies of the desired signal at the receiver. For most current systems, it is essential to have this channel property in order to exploit spatial diversity, for example. Thus, analog circuits, even if channel- aware, cannot completely cancel out the self-interference multipath components as mentioned in [31]. For that reason, a stage of digital signal processing is used so that the remaining self-interference, usually referred to as residual component, is further suppressed, thus, allowing reliable in-band full-duplex communication.

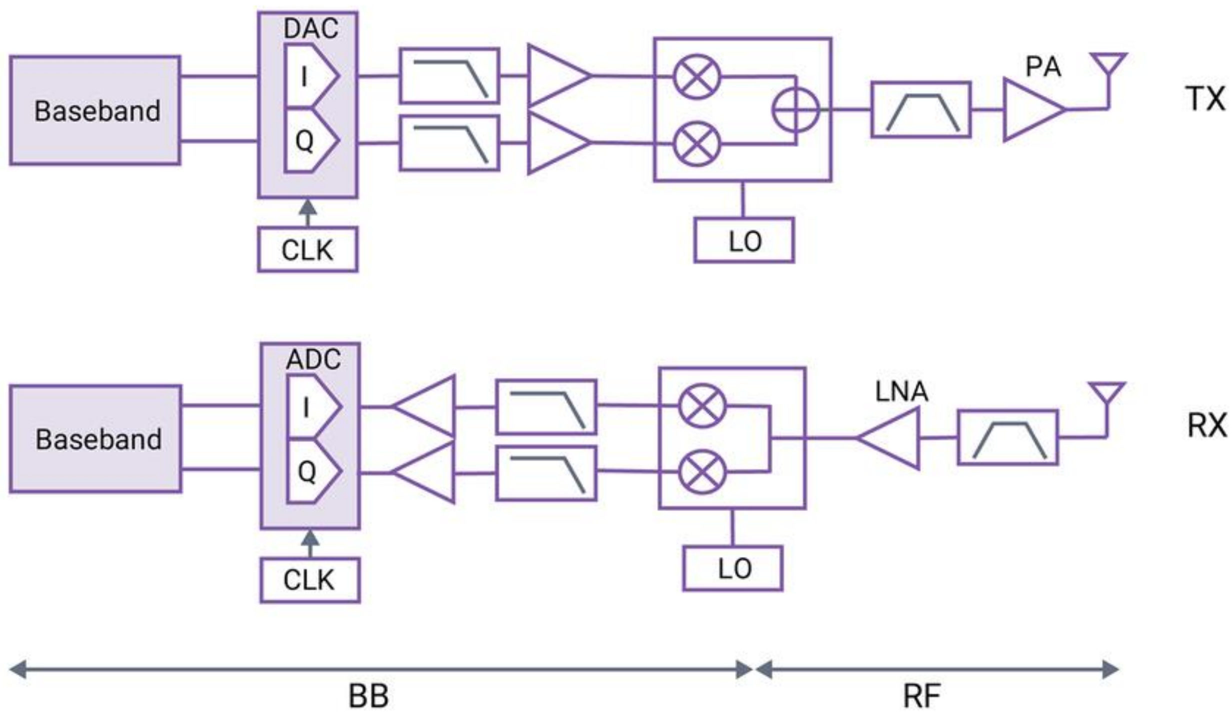


Figure 2.2 : Very high speed data converter for 5th Generation network with the help of baseband and radio frequency [31].

2.3 5TH GENERATION FREQUENCY BAND AND SPECTRUM ALLOCATION

There are essential elements that must be clear concerning the evaluation of cell capacity. The first is the signal-and-noise to interference ratio (SINR), essentially dependent on the MT distance to the BS, and the channel spectrum efficiency (SE), dependent on the level of quantization [32]. Both these elements will define the appropriate modulation coding scheme (MCS) to be assigned to each user, which in 5G is defined as from QPSK, 16-QAM, 64-QAM and 256-QAM. The implications of the MCS order can be defined as:

• **Lower 5G orders (m < 32):** more robust and tolerant to higher values of interference at the cost of lower transmission throughput, thus it performs better on low path-loss.

• **Higher 5G orders (m > 64):** much higher transmission throughput at the cost of lower tolerance to noise and interference, and increased transmission and channel estimation errors.

Generically speaking, since radio spectrum resource management and optimization is a fundamental concept in wireless communications, it is important to consider improvements in the field of spectrum efficiency. States that advances in 5G architecture, domain processing and network architecture are important steps towards a higher spectral efficiency, as well as improvements in physical layer technologies, such as multiple access scheme, channel coding scheme and waveform [33]. This presents a key advantage of 5G towards 4G, since with an equivalent amount of spectrum. It can transmit a greater number of bits, leading to faster 5G experienced bit rates.

In regard to carrier aggregation techniques, they are to provide the operator with the opportunity of combining and operating separate, often non-contiguous, blocks of spectrum as one block with the goal of maximizing available and future spectrum, in order to provide a higher level of performance consistency [34]. There are three configurations for carrier aggregation:

• **Contiguous frequency band:** carriers are contiguous and located within the same frequency band. In this case the MT can handle signals using a single trans-receiver.

• **Non-contiguous frequency band | intra-band:** carriers are located within the same frequency bands, but they are not adjacent to each other. In this case the MT may require a separate transceiver for each carrier.

• **Non-contiguous frequency band | inter-band:** carriers are in different, distinct parts of the radio spectrum. It requires a transceiver for each carrier and the MT must operate in multiple bands at the same time. It is very useful for operators with limited or fragmented spectrum, though at the cost of increased system design complexity, cost and power consumption.

Due to a commercial competition for spectrum at lower bands, operators have often to use carrier aggregation techniques to make use of their non-contiguous frequency band. Although there have been successful tests in LTE with advanced carrier aggregation producing data rates up to 600 Mbit/s per user in controlled environments [35], they were based on higher MIMO order (8x8) and on limited/research spectrum, thus it does not and cannot still apply to most of real-life deployments as mentioned in [36].

Regarding bandwidth in 5G, the main difference with LTE (4G) is that 5G will consider a single bandwidth of up to 700 MHz, whereas in LTE the maximum is 26 GHz, for single channels. Hence, this is a clear indicator of the potential of 5G to achieve even higher data rates since with carrier aggregation [37], in the future, the maximum achievable bandwidth will be even considerably higher than those in LTE with the same techniques.

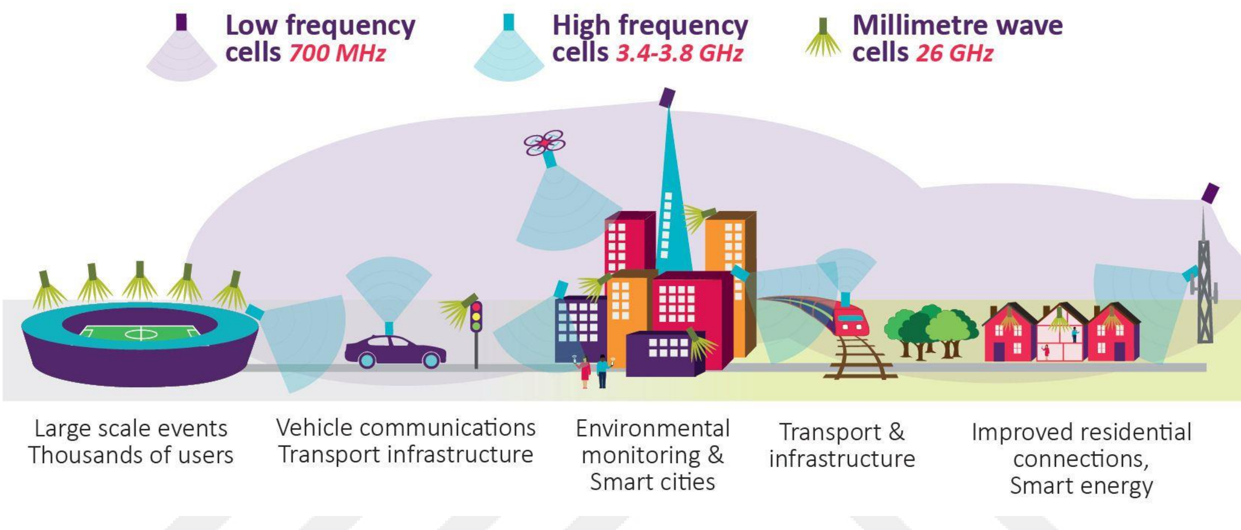


Figure 2.3 : The 5th Generation network continuous frequency bands with spectrum allocation [37].

2.4 5G NETWORK INFRASTRUCTURE SPENDING

As reported by [38], due to the wide range of capabilities required to be developed for 5G, its network interface is expected to be comprised of more than one RAT. The established common ground is to have each RAT optimized for certain use cases, or for certain spectrum lot. However, one must pay attention not to consider unnecessary development of RATs to attend niche optimizations, as doing so would hinder RATs economy of scale development [39]. The underlying approach is to have different RATs being defined as modes of operation of a single, unified RAT. At last, there have been proposed three possible technical migration paths from previous generations of mobile communications to 5G, according to [40]. Assuming the best choice, the global 5G network infrastructure spending.

According to [41], the frequency spectrum under consideration for 5G networks spans from the 50 to 100 GHz frequency bands, with expected different deployment dates. This wide spectrum range has to do with different entities being interested in a wide range of devices to be served by 5G [42]. However, in this thesis, one considers the initial spectrum bands that are considered for the first 5G field deployments, the cm-spectrum, specifically the 50 and 100 GHz frequency bands. Therefore, this thesis is focused on the propagation and coverage effects related to these frequency bands. It is possible to provide the best coverage with more power although at the cost of increased interference with adjacent areas using the same frequency, while on the contrary there is a strong possibility of blind spots and thus lack of coverage at antenna edges [43]. There is then the need to establish a compromise between path loss models and interference when designing a network for a frequency of 100 GHz.

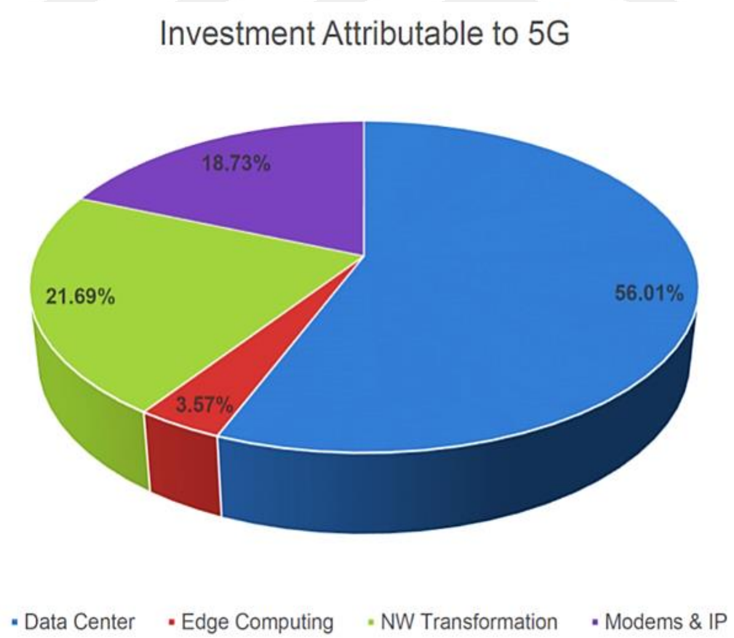


Figure 2.4 :The 5th generation wireless network infrastructure spending in terms of attributes [43].

2.5 RADIO SPECTRUM RESOURCE MANAGEMENT IN 5G

Generically speaking, since radio spectrum resource management and optimization is a fundamental concept in wireless communications in 5G coverage, it is important to consider improvements in the field of spectrum efficiency. The standard value considered for throughput evaluation is with the numerology of 50 GHz, from [44]. There are relevant aspects concerning band guards on higher orders of numerology in 5G which can change the multiplying throughput factor from the 50 GHz throughput curve, but for simplicity purposes one assumes a multiplying factor of 2 every time the frequency bandwidth doubles. It should be noted that this approximation may not be suitable for use with much higher orders of numerology (e.g., frequency of 50 or 100 GHz), however since those are not within the scope of this thesis due to it being defined for mm-waves (100 GHz), the factor of 2 is considered in the 25, 50 and 75 GHz throughput curves.

As previously discussed, the model can take any numerology configuration although only the one with highest distance is considered for coverage distance purpose [45]. That is to say, if numerologies 0, 1 and 2 are selected, coverage estimation will always be based on numerology and consequently the same applies for computing the maximum achievable throughput at the cell edge [46]. Although there is a liberty to pick any numerology configuration, there are certain restrictions imposed by 3GPP. Essentially there are numerologies that are not supported by some pairs of frequency/bandwidth.

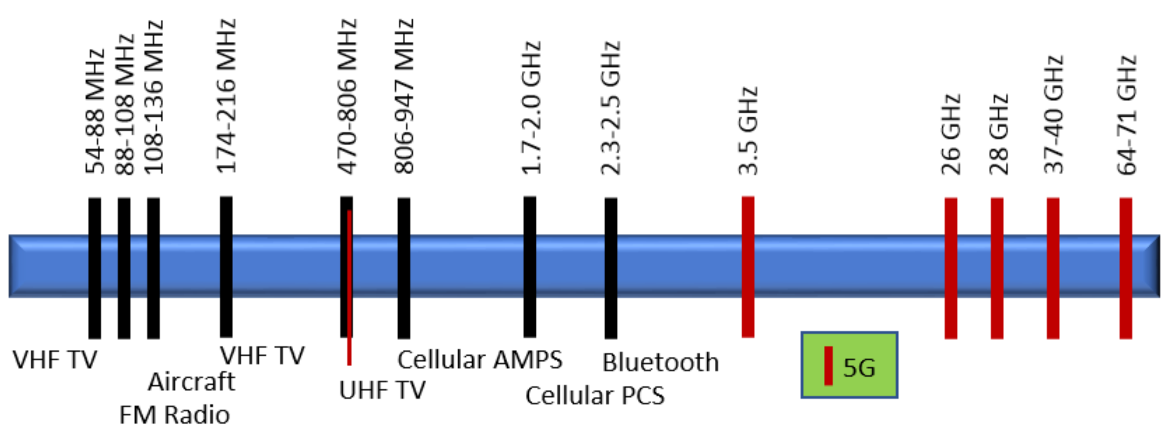


Figure 2.5 : The radio spectrum in 5th generation wireless networks [46].

In this sense, considering the user population randomly distributed in the cell and the capacity possibilities in terms of throughput from weather regions and numerologies, the model picks the most efficient 5G in use ratio for every service [47]. Due to a commercial competition for spectrum at lower bands, operators have often to use carrier aggregation techniques to make use of their non-contiguous frequency band. Although there have been successful tests in LTE with advanced carrier aggregation producing data rates up to 600 Mbit/s per user in controlled environments, they were based on higher MIMO order (8x8) and on limited/research spectrum, thus it does not and cannot still apply to most of real-life deployments.

Regarding bandwidth in 5G, the main difference with LTE (4G) is that 5G will consider a single bandwidth of up to 100 GHz, whereas in LTE the maximum is 20 MHz, for single channels [48]. Hence, this is a clear indicator of the potential of 5G to achieve even higher data rates since with carrier aggregation, in the future, the maximum achievable bandwidth will be even considerably higher than those in LTE with the same techniques [49]. The initial, and maximum, number of users is given by the coverage estimation, from which the capacity estimation indicates if such users can be served with the required traffic profile. If there is cell capacity overload, the number of users is reduced through cell radius reduction, until or if there is possibility of the cell withstanding the required number of users on 5G network. After this estimation, the final number of users in the cell and in the network can be computed.

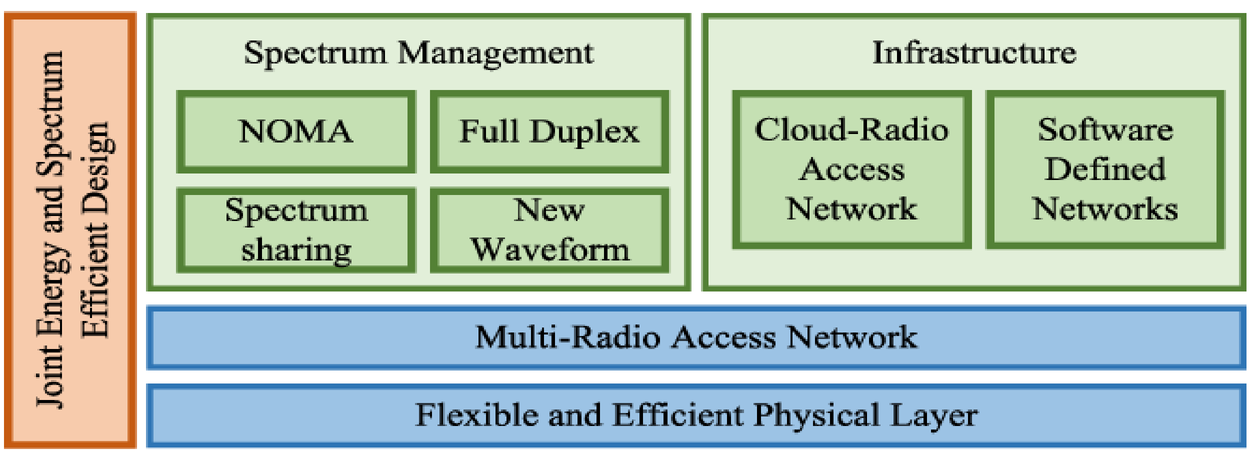


Figure 2.6 : The orthogonal radio spectrum and resource management in 5th generation networks [49].

2.6 SPECTRUM REGIONS IN 5G NETWORK

A considerable number of spectrum regions for 5G are currently being considered, at both the mm-wave (e.g., 50, 100-GHz) and the mm-wave spectra (e.g., 100 GHz), in terms of technical and economic feasibility as mentioned in [50]. Due to the very different and complex nature of mm-wave propagation and economic implications in terms of operator network configuration, such as with the need for much denser base station scheme, mm-waves are unlikely to be the first choice in early 5G deployment phases as mentioned in [51]. The main objective is to study the 5G network coverage and capacity implications with big data in different scenarios for the mm-wave spectrum, the most likely spectrum to be considered in first network deployments, more specifically in the 50 and 100 GHz bands as mentioned in [52]. The former will provide MTs general access to the network, while the latter will be able to provide MTs with higher data rates from larger bandwidth allocation or higher MIMO orders as mentioned in [53]. Regarding the possibility of using MIMO at these frequencies, the paper [54] shows equipment with different dimensions with the highest achievable MIMO order for different MT devices from (2.1) where the number of antennae is the rounding to the closest multiple of 2.

$$N_{Antenna} = [\frac{d_{[m]}}{\frac{c_{m/s}}{f_{[Hz]}}}] \quad (2.1)$$

- d: device distance (length or width);
- c: speed of light;

It is possible to formulate the observations based on the most common devices to be supported in 5G networks. For devices such as smartphones, tablets or laptops, in most cases even at the 100 GHz band it will not be possible to assure a MIMO configuration higher than 2x2, posing a serious limitation for the very high throughput values normally referred for 5G deployments as mentioned in [55]. In the case of very high throughput-demanding devices, such as in the context of virtual reality (e.g., Samsung Gear V.R glasses) or autonomous vehicles' devices (e.g., vehicle rooftop serving as antenna for infotainment), in the first the requirement for high throughput may not be sufficiently met with the maximum MIMO order of 2x2 at 100 GHz, while for the second there is the possibility of achieving interesting MIMO orders, especially at the 100 GHz band, for cases such as V2V communications or infotainment [56]. Even in difficult propagation scenarios, at the 50 GHz band achieving up to 4x4 MIMO in a bus/train scenario is feasible.

In what concerns the present commercial, military or other spectrum usage at the two bands mentioned, the 50 GHz band has been issue of a frequency reframing initiative from the European Union in 2018 [57] to guarantee an available band to 5G mobile communications by 2021 while the 100 GHz band is currently on the process of being opened from previous satellite and non-licensed services, according to [58]. In some countries, such as the Turkey, spectrum auctions are currently being held for the 50 and 100 GHz bands. Regarding the frequency bands and bandwidths for early 5G, they have been defined for the 50 GHz band, although for the 100 GHz the specific and differential values are still to be defined officially for frequency ranges designation.

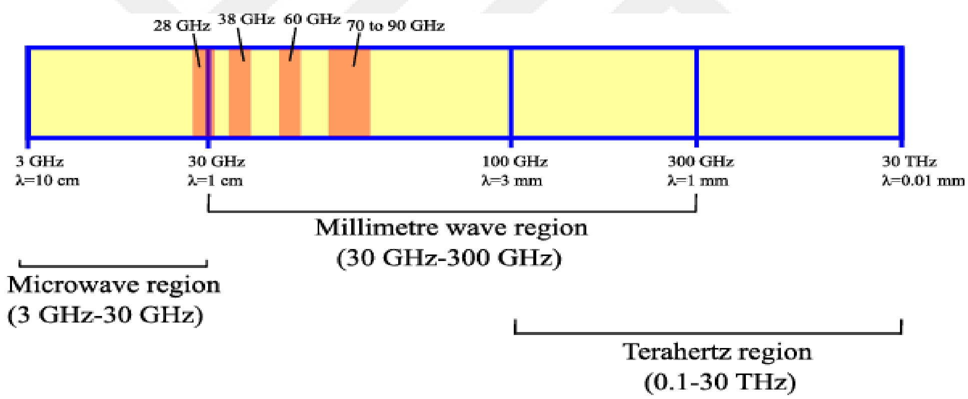


Figure 2.7 : The mm-wave regions of electromagnetic spectrum in 5th generation network with specific range [58].

In most countries the main band for the 50 GHz spectrum usage will be the NR-band [59], whose maximum available bandwidths have already been defined by 3GPP (45 MHz per channel). For the 100 GHz band, the maximum available bandwidth value per channel has not yet been fully defined by 3GPP although a total available bandwidth of around 2 GHz is expected to be the norm. Although in the context of this thesis the distinction is of no relevant effect, it may be important to note the 50 GHz band will be FDD based while the 100 GHz band will be TDD based.

The main propagation scenarios to be examined are urban, suburban and rural for the cm-wave spectrum (50 and 100 GHz), whose main differences consist of the presence and distribution of obstacles, the density of buildings and the density of users as mentioned in [60]. Besides the geospatial distribution, the model will consider an empirical measure of traffic density fluctuation throughput the day with the ROM configuration metric (Residential, Office, Mixed).

2.7 COVERAGE AND CAPACITY PLANNING IN 5TH GENERATION

The main goal of this section is to analyses and define the dimensioning of 5th generation network requirements and characteristics with coverage and capacity for the investigation purpose in the context of expected 5G use-cases. In detail, the goal is to analyses the requirements in terms of throughput, latency and other relevant KPIs of services and applications, such as video streaming, augmented reality, Internet of Things (IoT) and self-driving cars [61].

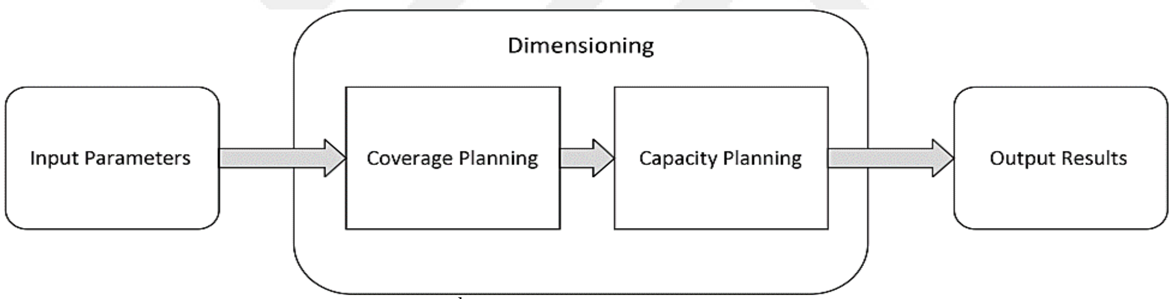


Figure 2.8 : The dimensioning of 5th generation network which includes both coverage and capacity.

The network development behind 5G networks has the objective of making possible the creation of a wide range of new and unique use cases to be deployed by operators, which were not yet available or efficient to implement in previous generations of mobile communications, such as with UMTS or LTE. Therefore, the current panorama for use cases can be split into three main groups [62]:

- **Enhanced Mobile Broadband (eMBB):** diverse services (download file, streaming), usually larger packets. It includes services such as virtual reality and infotainment for vehicles.

- **Massive Machine Type Communications (MMTC):** periodic communication, small packet size. It includes IoT services such as Smart Cities and Demotics.
- **Ultra-reliable and Low-Latency Communications (URLLC):** usually small packets, latency critical. It includes services such as communication between vehicles (V2V) and industry automation.

Nevertheless, service categories defined for previous generations of mobile communications are anticipated to still be supported by 5G networks. The model dimensioning should be as objective and detailed as possible, while also permitting changes in input parameters for different network deployment configurations as mentioned in [63]. This process can be seen as a high-level approach to network dimensioning, where by the use of propagation models and specific algorithms it is possible to produce an output solution with an acceptable degree of accuracy.

The input and output parameters are generically described in the next section, as well as the propagation models used for coverage planning. The most relevant models for coverage and capacity are defined in the next section. Besides these essential elements for cellular network planning there is the traffic density/volume and share/mix profile throughout a typical load day.

From the input parameters the first step is to compute the maximum cell distance through link budget and propagation models, given the reference throughput, then to define the maximum network load, from the traffic profile, which produces an estimation cell radius for the different propagation scenarios from which the antenna density configurations are retrieved for network and mobile terminal.

Table 2.1 : Description for the coverage and capacity with Network and Mobile Terminals.

	NETWORK	**MOBILE TERMINALS**
CAPACITY	Device/connection density, throughput, cell size, latency	Throughput, latency, traffic volume profile
COVERAGE	Propagation scenario, penetration losses, cell size, obstructions	Signal strength, mobility

While there may exist the same parameter in the capacity and coverage field, its meaning is not necessarily the same. For example, a macro-cell could be a suitable solution for coverage purpose while cells could be the most adequate solution to support or increase network capacity in the same coverage area. On the other hand, it is the received power at the MTs that will define their access to the network, but it is through network capacity that MTs can achieve certain parameters.

Besides these, there are others specifically relevant to 5G network deployments, such as the numerology configuration and the maximum supported MIMO order. Although not considered in the early model development in this thesis, latency requirements could be added at a later stage. In this case, after having computed the cell radii and density for different propagation scenarios, it could very well be useful to consider the work from [64] for estimating backhaul infrastructure to support certain latency requirements. Feature Reduction Based on Hybrid Efficient Weighted Gene Genetic Algorithms with Artificial Neural Network for Machine Learning Problems in the Big Data [65]. Hybrid Efficient Genetic Algorithm for Big Data Feature Selection Problems [66].

3. METHODOLOGY

Current big data with 5G-connectivity have been able to transmit selected data, such as location, speed, and turning angle to the network to be used for automated driving. However, big data are producing an enormous amount of data through the multiple sensors and cameras that the 5G-network throughput is not able to cope with. On top of the sensor data, big data are using data for various applications such as infotainment and safety functions. Due to the nature of applications' data usage, the requirements for each connection are different. Infotainment such as streaming multiple video inputs simultaneously, requires high data throughput without the need for low latency or high reliability. Safety functions require high availability, ultra-low latency, and high throughput if uncompressed data streams are used. Automated driving will increasingly demand more and more reliable network-based structures, requiring redundant, real-time architectures.

An increasing number of sensors, inside and outside the car, are producing an enormous amount of data every second. Based on some estimates, each automated car uses up to 4 terabytes per day, requiring data rates far beyond 100 GBPs. Currently, cars are processing the data mostly with onboard computers. To improve the communication experience 5G withdraw the data to servers and platforms offering maps or crowd-sourced data such as Waze.

A crucial prerequisite for safe automated driving is the real-time transmission of data with ultra-high availability and reliability. Current standards such as LTE based mobile communications can deliver communication with a latency of 30-40 milliseconds, which is not sufficient enough to fill the requirements for real-time transmission. Developers of automated driving functions are focusing on implementing the technology on the 5G standard. The 5G is capable of offering communication with very low latency, less than 1 ms, and with high availability up to 99.9999% with big data thus filling the requirements of wireless communication.

Network slicing included in 5G enables applications with different network requirements to be served by a single base station. With the use of the apache spark in big data can stream infotainment, communicate with roadside infrastructure, and maintain automated driving specific functions such as radar and video stream analyzing with the help of big data based apache spark.

3.1 DATA ANALYSIS IN 5G WIRELESS COMMUNICATION

Regarding works based not only but mostly on the mm-wave spectrum in 5G, focused on coverage, capacity and protocol design layers for mm-waves, basing the studies on features such as data-rate, latency, energy and cost requirements, as well the implications in terms of propagation characteristics, channel modelling, link outage scenarios and deployment configurations and scenarios using the big data based apache spark. Investigated characteristics of indoor coverage at high frequency ranges, but also focused on outdoor to indoor coverage modelling by adapting already existing models into a single one. Established methods for computing capacity and coverage solutions for 5G mobile broadband in rural and remote areas at the mm-wave spectrum. Presented the case for propagation mechanisms at mm-waves and characterization of wireless channels in site-specific and stochastic channel approaches, both for narrow and wideband configurations, proposing a method to estimate capacity provisioning.

In broader terms, regarding general 5G network schemes is focused on the study of mill-metric waves in the access radio, front haul and backhaul contexts, in a variety of scenarios. The group has performed analysis on the use of mm-waves in micro-cells for dense radio-access areas in indoors and outdoors, to infer the technical feasibility of network increase capacity. The Fantastic 5G project has researched the primary key indicators needed to be achieved to improve capacity, coverage, reliability and latency for networks operating below 6 GHz, based on LTE and ANN for model simulations. We developed a model that computes capacity and coverage for mill-metric waves in the 100 GHz band, in an urban scenario. The implementation of mm-waves for future 5G networks, having as underlying modelling interface an open-source MATLAB software, capable of simulating 5G-network parameters with a 100 GHz bandwidth.

At last, the works that relate more to the frequency spectrum in the scope of this thesis and network characteristics, such as path-loss or MIMO, presents a simplified but comprehensive overview on 5G deployment considerations, from use-cases to frequency bands' implementation roadway, as well as interoperability architectures with LTE core network. We proceed to study the requirements for new channel models, from frequency bands, bandwidth, types of antenna arrays, mobility and handover issues, as well as presenting an experimentation and validation method for typical outdoors propagation scenarios in micro and macro-cells and outdoor-to-indoor penetration loss. For this method, it considers elements such as signal blockage, path loss models and shadow fading The flowchart diagram that is applied in this study as we show in Figure 3.1.

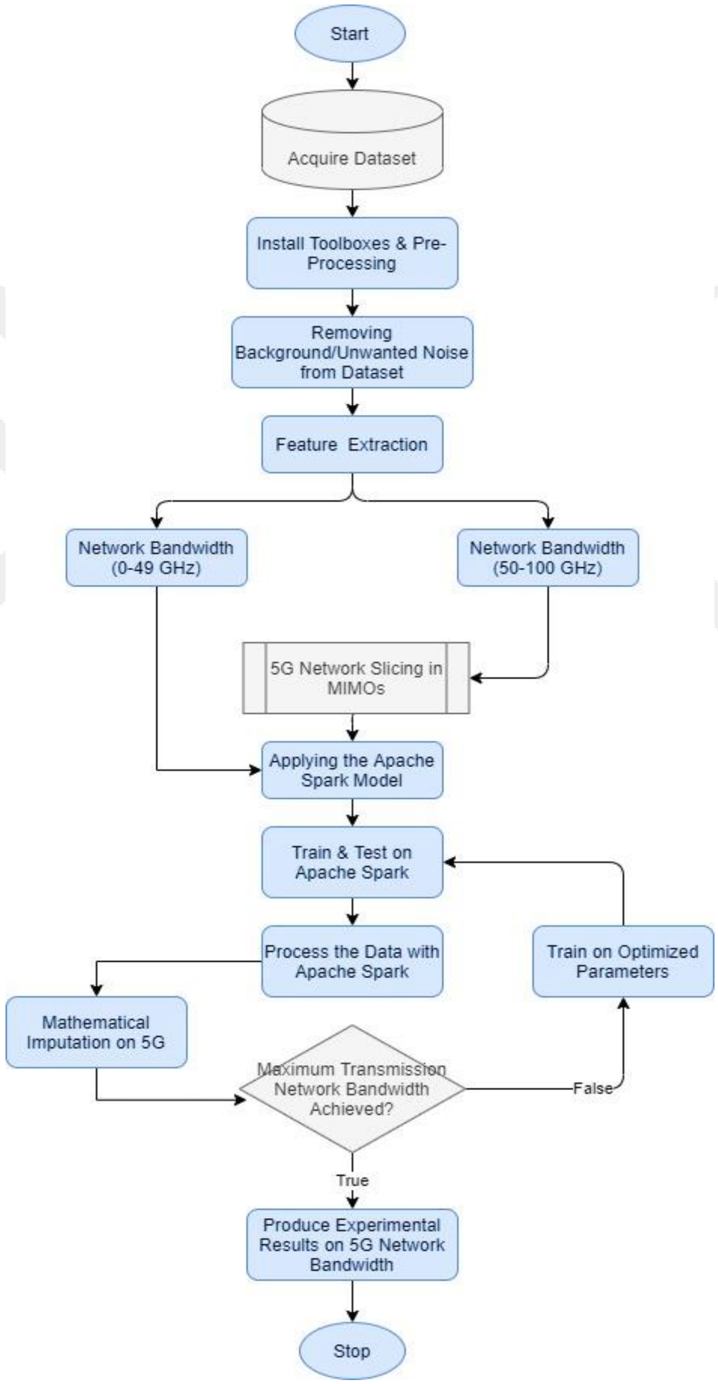


Figure 3.1 : The flowchart diagram that is applied in this study.

3.2 DATASET DESCRIPTION

The dataset has been acquired from an open-source repository crawdad: http://crawdad.org/keyword-5G.html. A dataset is composed of rows and columns, where, usually, each row represents an observation and each column a different variable. The 5G wireless transmissions have provided the tools to design faster and more reliable systems for wireless communications networks. The idea is to exploit the space diversity introduced by using multiple antennas, in order to increase user capacity. Using a 5G MIMO channel in this dataset for communication provides several data streams sent simultaneously, called spatial multiplexing, where information is processed and stored by apache spark to serve different users. The key concept is to take advantage of the multiple scattered uncorrelated signals that arrive via different paths at the receiver to extract the information under transmission.

3.3 BIG DATA BASED APACHE SPARK

Apache Spark is an open source tool for processing high volume of data by Apache Software Foundation falls in the domain of big data. The 5G mobile communication technology that is about to be harnessed into use during 2021 has been an anticipated evolution of the currently used 4G technology. The extensive performance and multiple new services are enabling innovations in multiple different verticals. The 5G wireless mobile communication with apache spark is one of the verticals that benefits from this technological leap. The research evaluates if the 5G technology offers sufficient performance for the fully automated communication using the big data based technology for processing. Modern automated communication is requiring reliable and low-latency data transfer from the communication channel to exchange information with other entities. With the help of network slicing, the 5G network can be split into different data profiles serving customers with various needs simultaneously. The big data based apache spark will serve the end-users with ultra-low latency and highly reliable communication that is required by autonomous devices to communicate with smart city infrastructure. The apache spark offers high data throughput that can be used to transfer data-intensive sensor input to the cloud for further analysis. The spark architecture is divided into 4 steps:

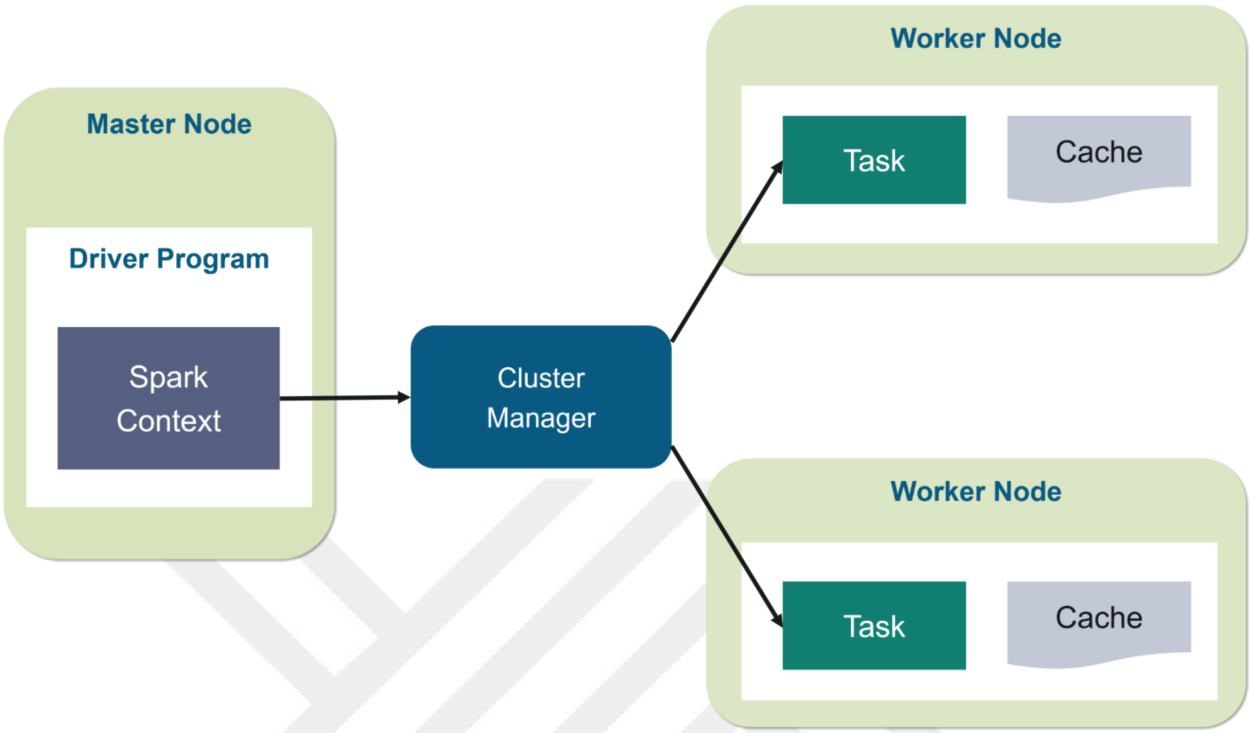


Figure 3.2 : Architecture of apache spark that is being used in this research for the implementation and experiments on the 5G network.

- Driver
- Worker
- Cluster Manager
- Executor

The improvements in the performance and changes in the services and require changes in the network architecture compared to the 4G technology. These changes, however, are adding challenges to the network design, especially to the security of the network. The pressure to deliver high performance might lead to compromises for the security and privacy of the network. For example, extensive authentication adds latency and limits the throughput of the communication. The architectural changes might also open up unexpected attack surfaces in services such as network slicing. The research focus was on the communication features of the proposed 5G protocol. The proposed communication and parallel processing protocol of big data for 5G was apache spark found to be prone to replay communication.

3.4 TRAINING AND TESTING ON APACHE SPARK

The first tool applied in this step is a network traffic flow crawler, a tool which gather information, content and the whole structure of a determined set of network traffic flow pages. We are going to talk briefly about the whole process later in this chapter and also deeply in a more specific section.

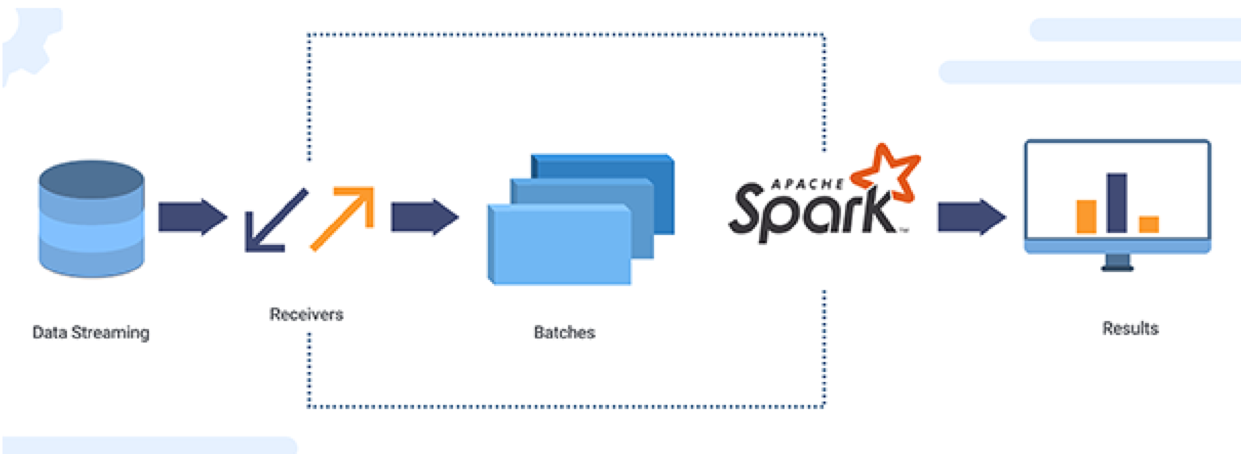


Figure 3.3 : The architecture of big data based apache spark.

Besides extracting the features from a dataset, which is going to be well explained in this work, we have also written some code to extract information from those files, the features to analyses. To do so, we have used the Apache Spark environment in big data, a very concise and powerful framework for Big Data problems. However, we only used a small portion of data in the beginning, but increased the amount over time. The research is dedicated to cover the Apache Spark framework.

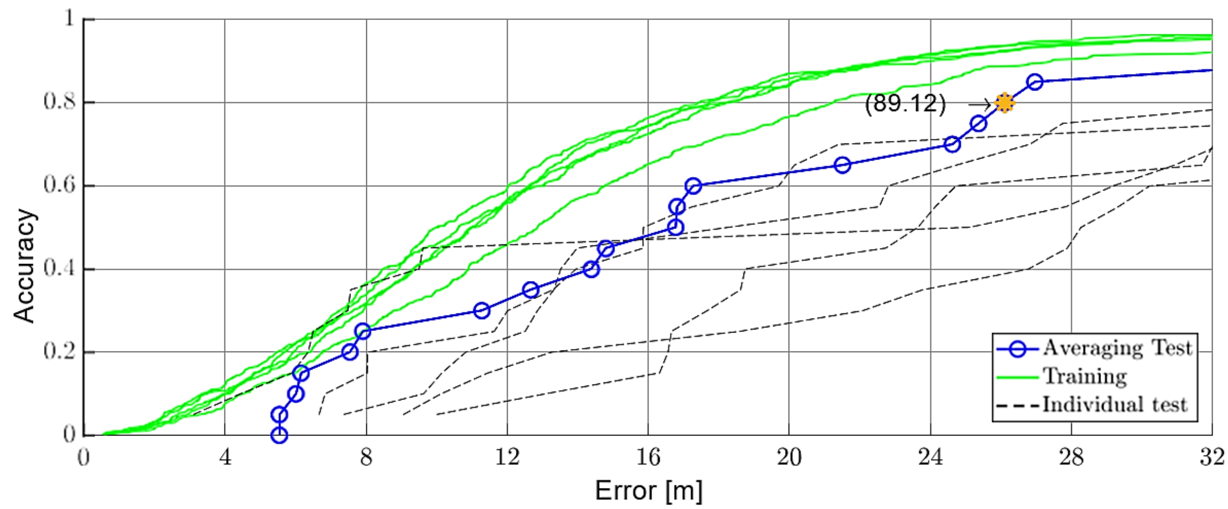


Figure 3.4 : The accuracy and error rate in meters for training and testing of 5G wireless communication using the apache spark.

Finally, the last process was application of Big Data techniques, still in Spark environment. The framework is built on top of libraries, including a Big Data on parallel processing toolbox, and is therefore specially designed and developed to deal with Big Data problems. Techniques broadly known and used, e.g. cross-validation for model selection, classification algorithms and evaluations were applied in practice and we introduce it in the end for 5G wireless communication.

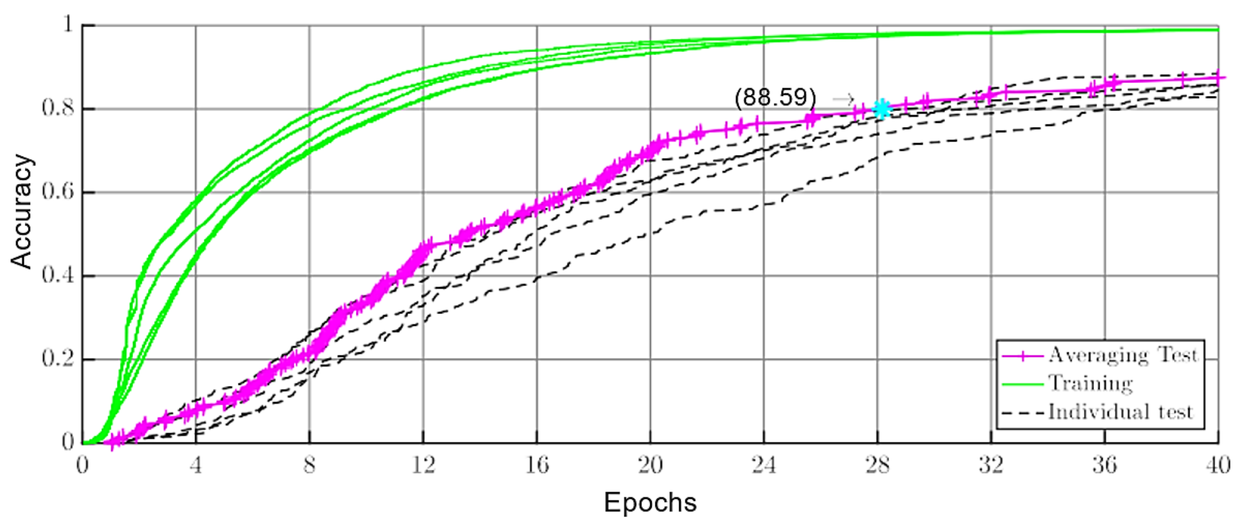


Figure 3.5 :The accuracy in terms of number of epochs the apache spark mode is trained and testing for 5G wireless communication.

3.5 MATHEMATICAL IMPUTATION & CORRELATION IN 5G NETWORK

The purpose of 5G wireless network and communication analysis is to infer the requirement in terms of percentage of antenna range allocated to the cell edge to ensure that there is always a nominal throughput available at this region using big data based apache spark. For this, it is necessary to establish a maximum percentage of resource blocks associated to the cell edge region, which will define the maximum achievable throughput and then define the cell radius. In order to understand the impact of numerology on the available throughput at cell edge, one must be aware of yet another cell-radius reduction process in the model. In the dimensioning process, the cell radius can not only be reduced to make it possible to avoid capacity overload, but it can also be reduced to meet the throughput requirement at cell edge. Since as the radius decreases the SNR increases, and thus the available throughput per mmWave increases from 50 GHz, the total available throughput also increases to 100 GHz. This being said, a proper analysis on this process should consider a user density of zero, which sets the threshold for the maximum available throughput, that is to say, it provides a theoretical mathematical modeling below:

$$y(true) = x_f \tag{3.1}$$

$$\frac{dy(t)}{dt} = x_2 = x_1 \tag{3.2}$$

$$\frac{d^2y(t)}{dt^2} = x_3 = x_2 \tag{3.3}$$

$$\frac{d^{n-1}y(t)}{dt^{n-1}} = x_n = \dot{x}_{n-1} \tag{3.4}$$

$$\frac{d^n y(t)}{dt^n} = \dot{x}_n \tag{3.5}$$

$$x(p) = x_f \tag{3.6}$$

$$\frac{dx(p)}{dt} = x_{p=}\dot{x}_f \tag{3.7}$$

$$\frac{d^2x(p)}{dp^2} = x_3 = \dot{x}_2 \tag{3.8}$$

$$\frac{d^{n-1}x(p)}{dx^{n-1}} = x_n = \dot{x}_{n-1} \tag{3.9}$$

$$\frac{d^n x(p)}{dx^n} = \dot{x}_n \tag{3.10}$$

$$\frac{d^n x(p)}{dx^n} + y_{n-1}\frac{d^{n-1}x(p)}{dx^{n-1}} + .. + y_1\frac{dx(p)}{dx} + y_0 x(p) = x(p) \tag{3.11}$$

Where the estimate of the variable *x*, indication function of mm-wave by *y*, and *dy* as the first-order derivative, Any user density different from zero yields a lower throughput at the cell edge. In order to infer the highest throughput value for each configuration, the algorithm computes every hypothesis while the radius is higher than a minimum pre-defined radius value, which for this simulation is 100 m, otherwise in some cases the algorithm would keep reducing the cell radius until a zero distance with antenna range as calculated using the mathematical model.

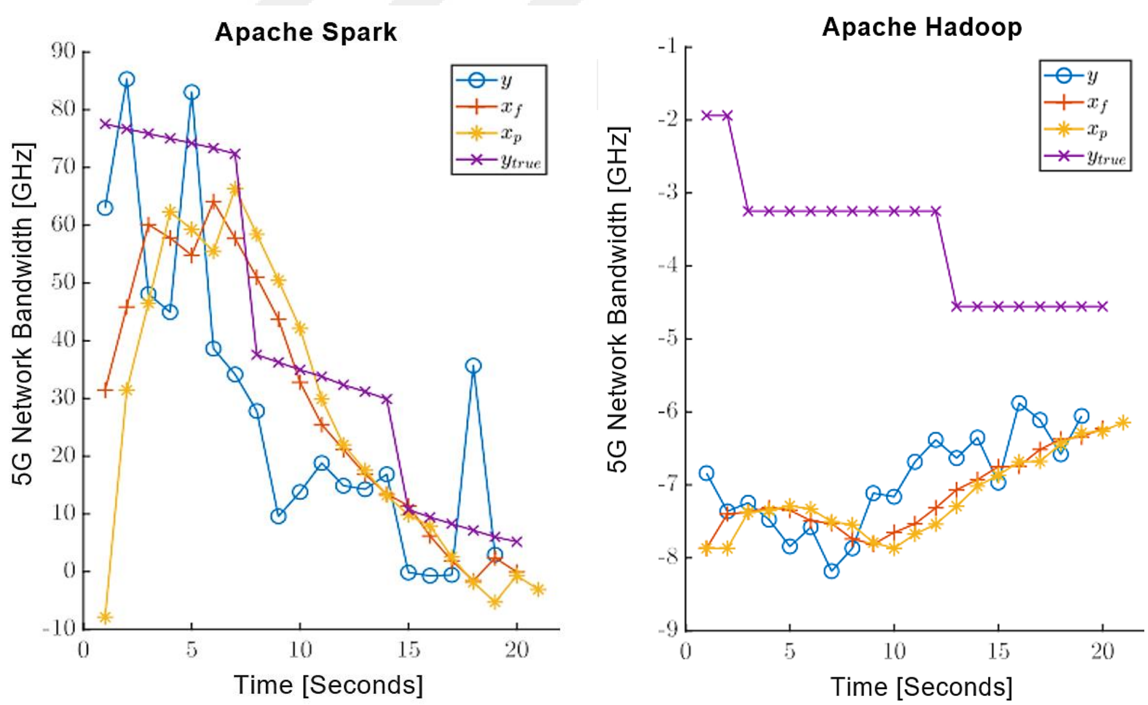


Figure 3.6 : The performance evaluation of big data based frameworks in terms of 5G network bandwidth.

The differences between the throughput values at 50 and 90 GHz for the same bandwidth may have to do with the fact that the algorithm keeps the cell-radius reducing process while the radius is above 100 m. Since the antenna patterns for 50 GHz is smaller than for 90 GHz, this results in existing a higher SNR value for transmitter (TX) and receiver (RX), which yields a higher throughput per mmWave at the given cycle/process radius. The same logic can be applied to explain the higher throughput values at TX since the mmWave is higher than at RX, yielding the same logical conclusion of a higher throughput per mmWave from 50 to 90 GHz and thus presenting higher throughput values for the same frequency and bandwidth configurations.

In the course of the simulations, it has been proved that the throughput value per frequency and bandwidth configuration is proportional to the share percentage of bandwidth allocated to the cell edge, hence if one seeks to find the values for 20% the only necessary step is to multiply 5 throughput values by 2. The apache spark presents a comprehensive perspective on how throughput varies according to the minimum and maximum possible spectrum configurations (50 -10 GHz vs 90 GHz) for both TX and RX, with different antenna capacity loads. The percentage of dedicated bandwidth at antenna edge is limited to the point of only guaranteeing the required throughput level, which is set to 2 Mbit/s, default value. That is to say, a value of mmWave is set, but if that yields a much higher throughput at antenna edge than the required, the model reduces this percentage

3.6 ADVANTAGES OF BIG DATA BASED APACHE SPARK IN 5G

The services, such as MEC can be used to process or store the data close to the end-user with low latency. This service can be highly beneficial to examine sensor data gathered by an autonomous communication. The 5G protocol also offers new services to support connected 5G wireless networks for vehicular communication such as vehicle platooning and remote driving. These communication nodes might lead to exposure of the end-user’s location, thus communicating the privacy feature. The lack of resources also could permit a fake base station to impersonate a genuine base station towards the end-user leading to compromise of end-to-end encryption. Extensive data throughput with the combination of high device density in the network gives the possibility for more extensive communication. The lack of data scrubbing and filtering for unwanted traffic needs to be addressed before implementing public 5G to its full capacity. The 5G protocol is still a work in progress and is about to be finished during the year 2021. Wireless communication and security is the biggest challenge that needs to be addressed before implementing the network into full use. End-users and critical services such as smart city applications need to be provided with the required privacy and security level. The network layout needs to be tested thoroughly before taking into full use with the focus on security and privacy that needs to be the top priority when introducing new technologies.

4. RESULTS

A new kind of big data based apache spark is introduced with optimized version of it for the purpose of investigating the different wireless communication effects on 5th generation network with a greater number of training epochs. This big data based apache spark is used in this work have distinguished size and number of nodes and its formula for the threshold has been modified – a new parameter for different wireless communication is added to it and an optimization is made for this new parameter on MATLAB (Using Parallel Computing Toolbox and Signal Processing Toolbox). Some details, such as activation function choice and apache spark optimization algorithm choice are this work's original contributions for the achievement of goal. We performed the feature extraction and selection process using apache spark is less time just because of optimization before splitting the dataset for training and testing. By separating 80 percent of the dataset for training and 20 percent of the dataset is parted as testing set. As in past there is no specific application been developed for the purpose of investigating the wireless communication using the apache spark as well as investigating the wireless communication effects on 5th generation networks. The application which is designed and developed solely for fulfilling the identifying the communication conditions and investigating the performance of 5th generation networks communication for the operational purpose using the big data. This application demonstrates that 5th generation network being firstly used for wireless communication based on the dataset. As it would be expected, within the same environment type the maximum cell radius is higher for the 50 GHz band and lower for the 100 GHz band, and as the numerology configuration increases the cell radius is reduced using apache spark. For this analysis, all numerology configurations are considered respecting the limitations imposed by 5G network, and the total number of cells using the model reference parameters is obtained for each bandwidth value. Furthermore, this application shows that the accuracy of apache spark on 5th generation network, which can have mitigated to some degree of optimization due to a greater number of nodes in the apache spark and achieving a high accuracy for wireless communication. We have used an effective and efficient application of MATLAB R2019a because it provides parallel processing toolbox, machine learning toolbox as well as signal processing toolbox for the implementation of this research work. **Figure 4.1 show the 5G network convergence time**

expressed in seconds on x-axis while the 5G network bandwidth is depicted on y-axis with maximum number of bandwidth equal to 82.75 GHz after training on 70% of data. And Figure 4.2 show the 5G network convergence time expressed in seconds on x-axis while the 5G network bandwidth is depicted on y-axis with maximum number of bandwidth equal to 85.40 GHz after training on 80% of data. And Figure 4.3 show the 3D graphs represent the 5G-BER on x-axis while 5G bandwidth on y-axis varies from -10 to 50 GHz however the rate of transmission is depicted on z-axis on 70% of data. And Figure 4.4 show The 3D graphs represent the 5G-BER on x-axis while 5G bandwidth on y-axis varies from -10 to 100 GHz however the rate of transmission is depicted on z-axis on 80% of data. And Figure 4.5 show the prediction of 5G network bandwidth maximum recoded at 70 GHz on 5 nodes of apache spark at 70% of data trained with total 16 nodes.

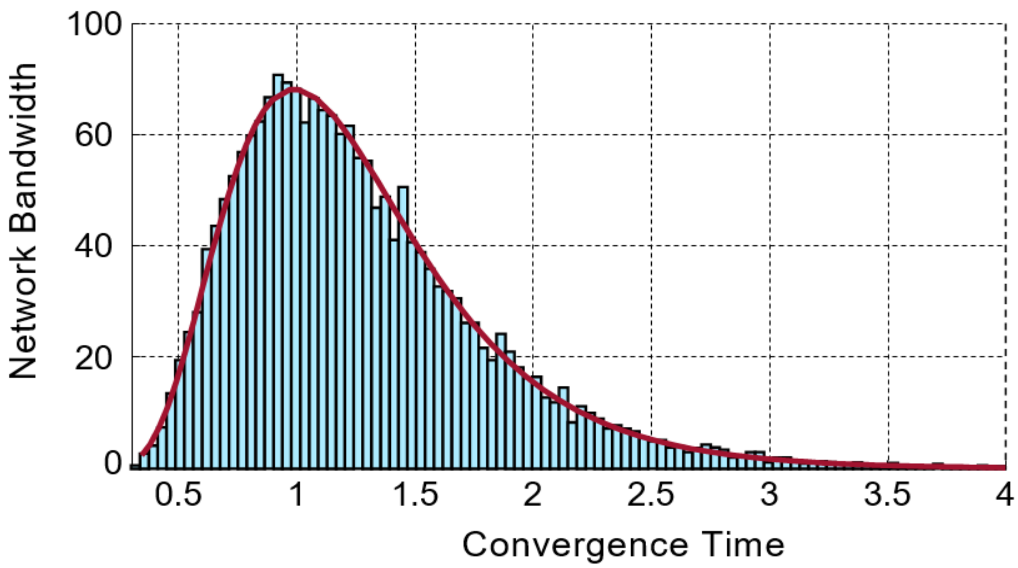


Figure 4.1 : The 5G network convergence time expressed in seconds on x-axis while the 5G network bandwidth is depicted on y-axis with maximum number of bandwidth equal to 82.75 GHz after training on 70% of data.

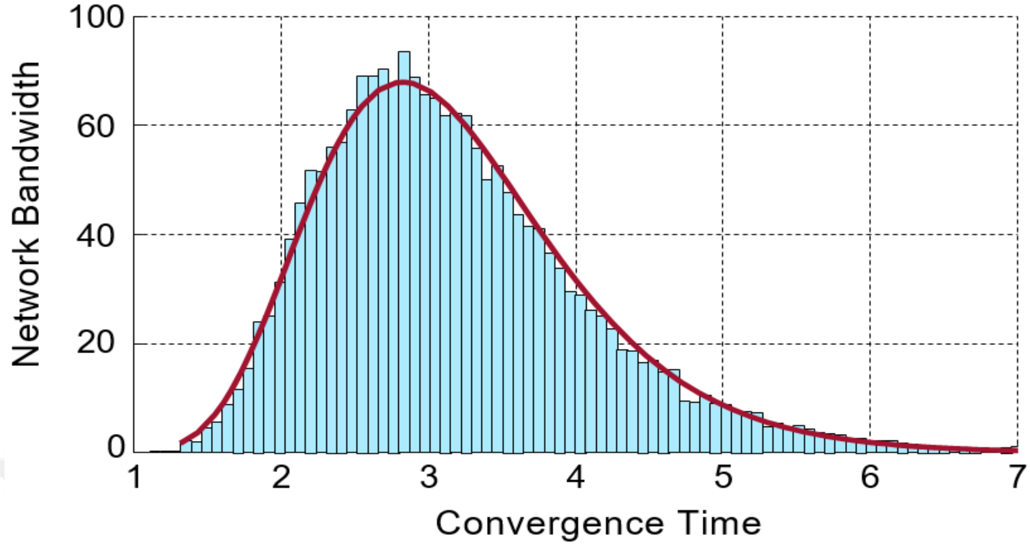


Figure 4.2 : The 5G network convergence time expressed in seconds on x-axis while the 5G network bandwidth is depicted on y-axis with maximum number of bandwidth equal to 85.40 GHz after training on 80% of data.

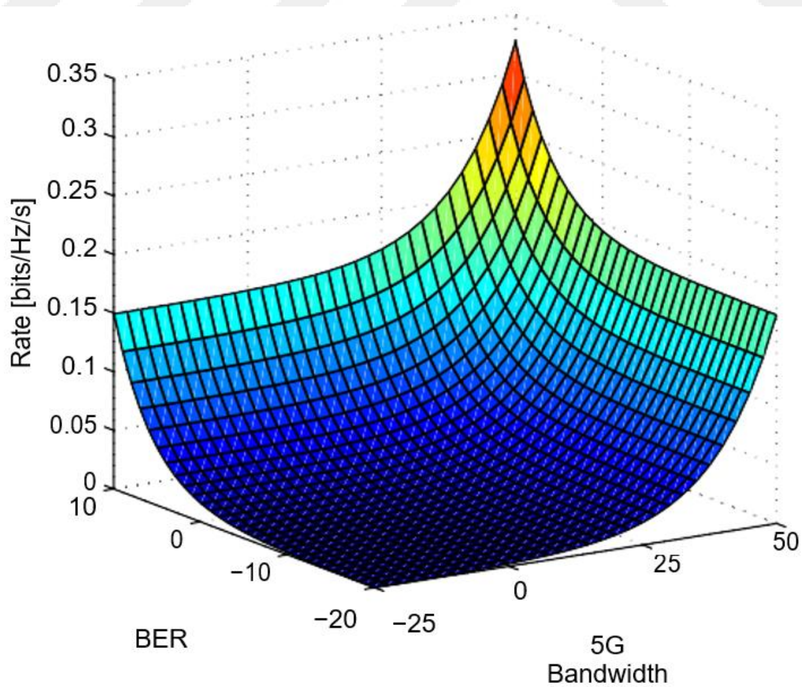


Figure 4.3 : The 3D graphs represent the 5G-BER on x-axis while 5G bandwidth on y-axis varies from -10 to 50 GHz however the rate of transmission is depicted on z-axis on 70% of data.

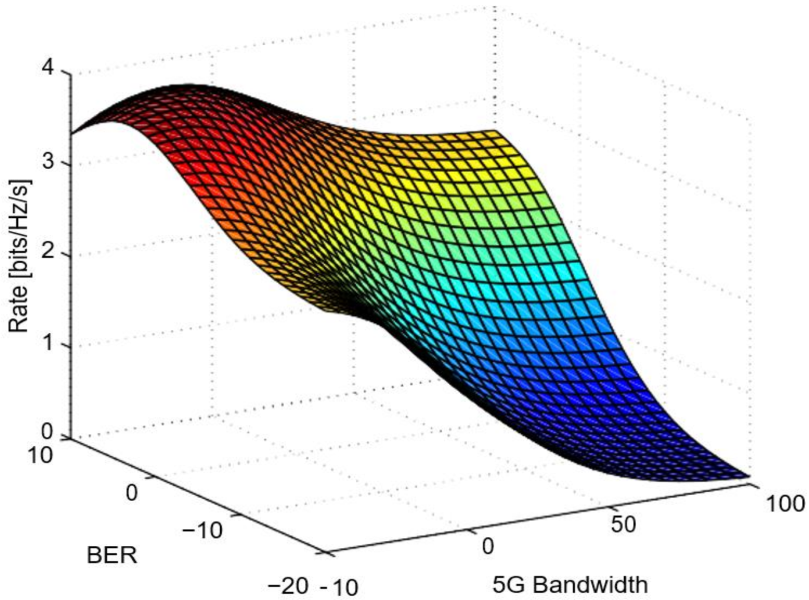


Figure 4.4 : The 3D graphs represent the 5G-BER on x-axis while 5G bandwidth on y-axis varies from -10 to 100 GHz however the rate of transmission is depicted on z-axis on 80% of data.

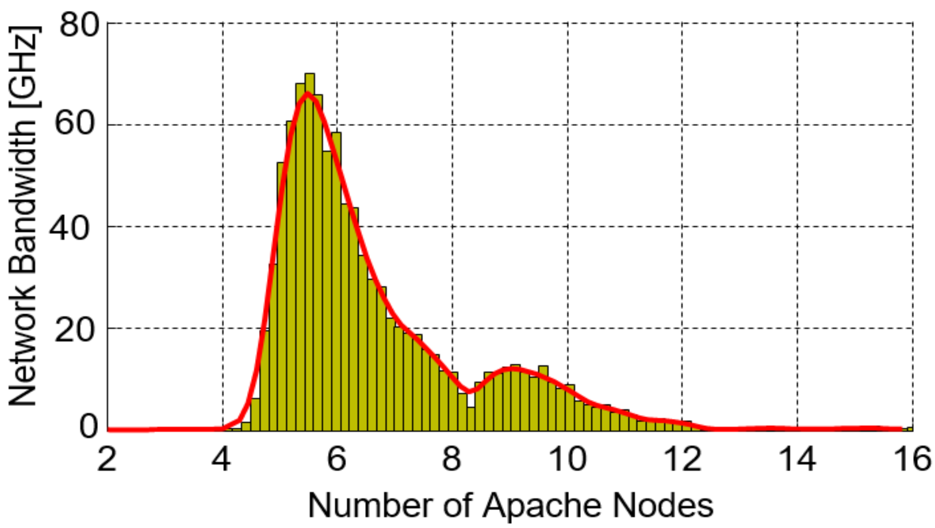


Figure 4.5 : The prediction of 5G network bandwidth maximum recoded at 70 GHz on 5 nodes of apache spark at 70% of data trained with total 16 nodes.

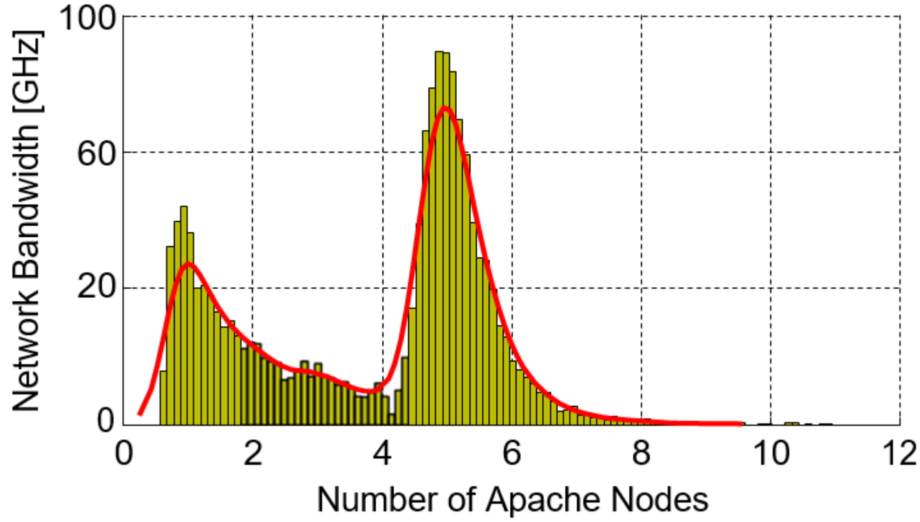


Figure 4.6 : The prediction of 5G network bandwidth maximum recoded at 90 GHz on 5 nodes of apache spark at 80% of data trained with total 12 nodes.

In Figure 4.6 show The prediction of 5G network bandwidth maximum recoded at 90 GHz on 5 nodes of apache spark at 80% of data trained with total 12 nodes. This research work presented in the literature contribution, where it is assumed that each entry of the MIMO channel is a composition of multiple scatters summed at the output to form a complex gain, and which vary faster than the symbol interval, or more specifically, assuming the symbol rate is less than the coherence bandwidth of the channel for communication in 5G using big data-based apache spark. Nevertheless, most systems that operate in broadband need to deal with frequency-selective MIMO channels, since the previous assumption is no longer valid. Modern wireless applications require faster data rates where the flat fading assumption cannot be assumed, enhancing the study of techniques to overcome this problem with lower BER in 5G network Visualization of 5G BER in terms of SNR from 25 GHz to 100 GHz as we show in Figure 4.7.

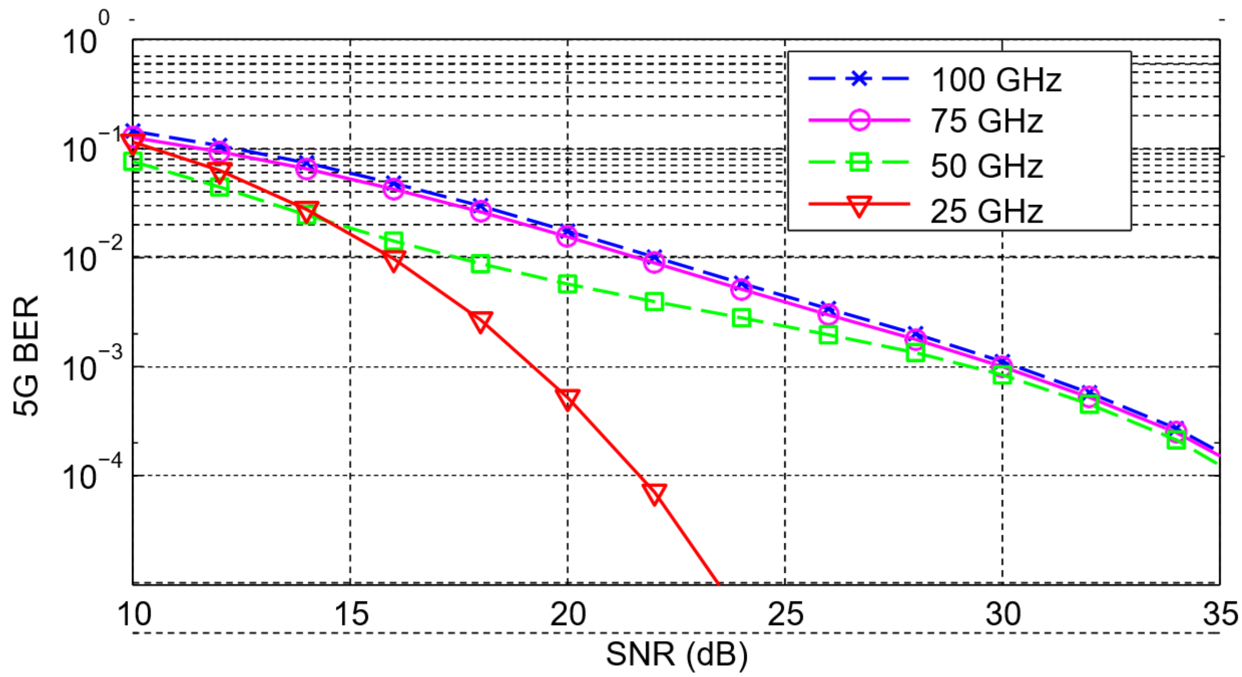


Figure 4.7 : Visualization of 5G BER in terms of SNR from 25 GHz to 100 GHz.

The frequency-selective channels may actually achieve higher rates than flat channels. This issue has been analyzed in this work, where the big data based apache spark develop a simulation-based method to compute the achievable rate of MIMO systems operating under different communication conditions. The main result to mention is that the rate increases with the selective channel order with wireless communication. Methods for combating communication problems are usually categorized into single carrier techniques or multiple carrier techniques with apache spark. The first method uses only one carrier frequency to transport information that is filtered, equalized, at the receiver (it can be performed linear or feedback equalizers). The latter, consist of splitting the broadband signal that suffers 5G network into several small narrowband signals that only perceive a flat channel. Once this is done, the previously discussed MIMO techniques can be applied to each flat fading block channel. Single carrier systems are computationally more demanding than multiple carrier systems, and for that reason most MIMO transmissions do not employ single carrier systems to mitigate frequency-selective multipath effects. In the next section, the most well-known multiple carrier method to combat network is presented.

5. DISCUSSION

The forthcoming change of paradigm along with the predicted exponential growth of data flowing through our wireless 5G networks is driving the search for disruptive technologies capable of corresponding with these new demands with big data based apache spark. Among them and with great potential, lies in-band full-duplex communications, which will double the spectral efficiency of current transmissions and will facilitate management of the available spectrum. However, self-inflicted interference poses a great challenge to the deployment of this technology. If not properly dealt with, this so-called self-interference may create unbearable signal-to-interference-plus noise ratios (SINR) for allowing the establishment of reliable communication links. Thus, this thesis starts by reviewing the literature on in-band full-duplex systems for 5G with apache spark and basic concepts of multiple-input multiple-output (MIMO) wireless 5G communications, which allows it to study and propose new techniques and systems where the performance of full-duplex transmissions can be enhanced.

In first place, an in-band full-duplex relay station was formally and fully described, special focus was given to the system impairments that cause residual self-interference with apache spark. In order to mitigate this effect, spatial suppression filters were studied, that take advantage of the extra degrees of freedom brought by MIMO transmissions. The bit error rate (BER) for two different types of these filters, null-space projection (NSP) and minimum mean square error (MMSE), was evaluated and where it was possible to observe a self-interference resilience of approximately 20 dB. However, spatial suppression suffers from distortion when shaping the transmitted and desired signal. For that reason, adaptive feedback cancellation filters were proposed. Namely, a least means square (LMS) and a recursive least squares (RLS) filter were derived for MIMO frequency-selective channels and under OFDM transmissions. Both techniques' performance was measured in terms of BER plots that demonstrate the possibility of obtaining levels up to 35 dB of self-interference resilience, depending on the system impairments that pose the main obstacle for these filters.

Secondly, and considering the methods derived for canceling self-interference, the use of a massive array at the relay station was explored. The orthogonal properties between large dimension matrices can in fact contribute to cancellation of the residual part of this undesirable effect. Therefore, a system with MMSE filtering was proposed combined with zero forcing (ZF) detection and precoding filters. The BER was assessed at the relay and at the final destinations, showing a gain in interference resilience when using several antennas and revealing an optimal relay transmit power. Consequently, an algorithm to find the optimal power allocation scheme that minimizes the system energy efficiency (EE) under a minimum link quality was derived. The comparison of different big data frameworks in terms of working on wireless networks as we show in Table 5.1.

Table 5.1 : The comparison of different big data frameworks in terms of working on wireless networks.

ARTICLE	**TECHNIQUE**	**SUPPORT TO NETWORK**
[67]	Apache Hadoop	Network traffic analysis in 5G network
[68]	Apache Kafka	Big Data Streaming Process in 5G
Proposed	Apache Spark	Wireless communication with high network bandwidth

Finally, bidirectional relaying was brought together with in-band full-duplex transmissions, in the hope of reducing the number of employed resources in a two-way relay channel (TWRC) to the same as machine-to-machine communication (m2m). To that end, techniques from PLNC were studied and adapted to the previous systems. Two protocols were considered for this purpose, namely, de-noise-and-forward (DeF) and compute-and-forward (CF), where it was possible to evaluate the performance of both schemes in terms of BER or symbol error rate (SER). Furthermore, the integration of large antennas is proposed for the CF protocol and the performance degradation of the scheme was assessed for different numbers of antennas and for channel estimation errors. Additionally, the achievable sum rate of this system was derived and an optimization algorithm proposed in order to find the optimal transmit powers that maximize it.

6. CONCLUSION

The fifth-generation (5G) network is the latest generation of cellular mobile communications. Performance-wise, it is designed to achieve high data rate, reduced latency, high adjustability, energy savings, high capacity, and support for industries via massive device connectivity often called Internet of Things (IoT). These increased performance and novel technologies introduced by the 5G standard offer new possibilities for multiple industries and applications such as factories and wireless communication. However, harnessing premature technologies without extensive research is exposing devices for external attacks and risking lives in critical applications such as automotive and transportation. Thus it is crucial to investigate the technologies used in the later phases of the development. Apart from the increased speed and low communication latency, one significant advantage in the protocol is the possibility of network slicing. The 5G networks are to be built in a way enabling logical network slices, which can be modified to allow telecom operators to provide a network on an as-a-service basis. These slices can be highly modified based on the needs of customers to meet a wide range of use cases. Network slicing technology can provide connectivity for emergency services requiring high availability and low latency. At the same time, another network slice offers high throughput connectivity for streaming in a car on the move. A single 5G system can serve multiple slices simultaneously for a single device. While the development of the 5G protocol is reaching closer towards the production version, wireless communications are becoming more popular. Wireless communications are observing the environment via multiple inputs such as high-resolution cameras and optical radars, meanwhile producing an extensive amount of data. The recent progress in real-time computing is already enabling powerful algorithms to take control of networks to execute simple tasks. This Master's thesis examines the new possibilities 5G networks offer for wireless network communication with high bandwidth of 90 GHz. The prediction of 5G network bandwidth maximum recorded at 90 GHz on 5 nodes of apache spark at 80% of data trained with total 12 nodes. The main question to answer is will the 5G technology offers sufficient performance for wireless communication. This thesis also examines what challenges does the 5G platform face and does it offer sufficient privacy and security with big data. This thesis is a literature review with no empirical research. The first chapter examines the technological solutions of the 5G standard and provided new features. The second chapter compares the preconditions that wireless communication pose for a network. The third chapter maps possible implementation challenges for wireless network communication and present ways to secure the connection.

6.1 FUTURE RECOMMANDATION

Due to the nature of this thesis, there are many possible research paths that directly follow the work developed and presented. This section succinctly enumerates some research that would complement this work. The first to mention is the possibility of combining spatial suppression and feedback adaptive cancellation filters, so that higher levels of self-interference mitigation could be achieved. This requires carefully choosing each filter's relative position and understanding their mutual effect in apache spark and apache Hadoop with map-reduce technology in big data. Another meaningful result is to assume more realistic channel estimations when considering large arrays of antennas with apache spark. In fact, this is one of the major problems of large dimension channel matrices, which is neglected in this work by simply assuming the same procedure for the small number of antennas case. Finally, the next step in research is to generalize the bidirectional full-duplex relaying for the big data based techniques to the case of multiple-way relay channel, exploring more advanced big data based techniques with machine learning technique with higher dimension lattices.

REFERENCES

[1] Tran, T.X.; Hajisami, A.; Pandey, P.; Pompili, D. Collaborative Mobile Edge Computing in 5G Networks: New Paradigms, Scenarios, and Challenges. IEEE Commun. Mag. 2017, 55, 54–61.

[2] Tran, T.X.; Hosseini, M.-P.; Pompili, D. Mobile edge computing: Recent efforts and five key research directions. IEEE COMSOC MMTC Commun-Frontiers 2017, 12, 29–34.

[3] Wu, Q.; Ding, G.; Du, Z.; Sun, Y.; Jo, M.; Vasilakos, A.V. A Cloud-Based Architecture for the Internet of Spectrum Devices Over Future Wireless Networks. IEEE Access 2016, 4, 2854–2862.

[4] Wu, C.; Wang, Y.; Yin, Z. Energy-efficiency opportunistic spectrum allocation in cognitive wireless sensor network. EURASIP J. Wirel. Commun. Netw. 2018, 2018, 1–14.

[5] Wieruch, D.; Holfeld, B.; Wirth, T. Wireless Factory Automation: Radio Channel Evolution in Repeated Manufacturing Processes. In Proceedings of the WSA 2016, 20th International ITG Workshop on Smart Antennas, VDE, Munich, Germany, 9–11 March 2016; pp. 1–4.

[6] Montgomery, K.; Candell, R.; Liu, Y.; Hany, M. Wireless User Requirements for the Factory Workcell. Available online: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=928517 (accessed on 10 January 2021).

[7] Mochón, M.A.; Saez, Y. A review of radio spectrum combinatorial clock auctions. Telecommun. Policy 2017, 41, 303–324.

[8] Kalliovaara, J.; Ekman, R.; Paavola, J.; Jokela, T.; Hallio, J.; Auranen, J.; Talmola, P.;

Kokkinen, H. Designing a testbed infrastructure for experimental validation and trialing of 5G vertical applications. In Proceedings of the International Conference on Cognitive Radio Oriented Wireless Networks, Lisbon, Portugal, 20–21 September 2017; pp. 247–263.

[9] Hu, F.; Chen, B.; Zhu, K. Full Spectrum Sharing in Cognitive Radio Networks Toward 5G: A Survey. IEEE Access 2018, 6, 15754–15776.

[10] Ejaz, W.; Ibnkahla, M. Multiband Spectrum Sensing and Resource Allocation for IoT in Cognitive 5G Networks. IEEE Internet Things J. 2017, 5, 150–163.

[11] Doyle, L.; Kibilda, J.; Forde, T.K.; DaSilva, L. Spectrum Without Bounds, Networks Without Borders. Proc. IEEE 2014, 102, 351–365.

[12] Dong, X.; Cheng, L.; Zheng, G.; Wang, T. Network access and spectrum allocation in next-generation multi-heterogeneous networks. Int. J. Distrib. Sens. Netw. 2019, 15, 1550147719866140.

[13] IEEE. 802.11-2016-IEEE Standard for Information Technology–Telecommunications and Information Exchange between Systems Local and Metropolitan Area Networks–Specific Requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications; IEEE: Piscataway, NJ, USA, 2016.

[14] Bhattarai, S.; Park, J.-M.; Gao, B.; Bian, K.; Lehr, W. An Overview of Dynamic Spectrum Sharing: Ongoing Initiatives, Challenges, and a Roadmap for Future Research. IEEE Trans. Cogn. Commun. Netw. 2016, 2, 110–128.

[15] Ashraf, S.A.; Aktas, I.; Eriksson, E.; Helmersson, K.W.; Ansari, J. Ultra-reliable and low-latency communication for wireless factory automation: From LTE to 5G. In Proceedings of the 2016 IEEE 21st International Conference on Emerging Technologies and Factory

Automation (ETFA), Berlin, Germany, 6–9 September 2016; pp. 1–8.

[16] Akyildiz, I.F.; Lee, W.-Y.; Vuran, M.C.; Mohanty, S. NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey. Comput. Netw. 2016, 50, 2127–2159.

[17] Mahmud, R.; Kotagiri, R.; Buyya, R. Fog Computing: A Taxonomy, Survey and Future Directions. In Internet of Everything; Springer: Berlin, Germany, 2018; pp. 103–130.

[18] Li, J.; Li, Y.K.; Chen, X.; Lee, P.P.; Lou, W. A Hybrid Cloud Approach for Secure Authorized Deduplication. IEEE Trans. Parallel Distrib. Syst. 2014, 26, 1206–1216.

[19] Chien, H.-T.; Lin, Y.-D.; Lai, C.-L.; Wang, C.-T. End-to-End Slicing as a Service with Computing and Communication Resource Allocation for Multi-Tenant 5G Systems. IEEE Wirel. Commun. 2019, 26, 104–112.

[20] Coll-Perales, B.; Gozalvez, J.; Maestre, J.L. 5G and Beyond: Smart Devices as Part of the Network Fabric. IEEE Netw. 2019, 33, 170–177.

[21] Gringoli, F.; Patras, P.; Donato, C.; Serrano, P.; Grunenberger, Y. Performance Assessment of Open Software Platforms for 5G Prototyping. IEEE Wirel. Commun. 2018, 25, 10–15.

[22] Di Mauro, M.; Liotta, A. An experimental evaluation and characterization of VoIP over an LTE-A network. IEEE Trans. Netw. Serv. Manag. 2020.

[23] 20. Barakabitze, A.A.; Ahmad, A.; Mijumbi, R.; Hines, A. 5G network slicing using SDN and NFV: A survey of taxonomy, architectures and future challenges. Comput. Netw. 2020, 167, 106984.

[24] Chen, B.; Wan, J.; Shu, L.; Li, P.; Mukherjee, M.; Yin, B. Smart Factory of Industry 4.0: Key Technologies, Application Case, and Challenges. IEEE Access 2018, 6, 6505–6519.

[25] Ludwig, S.; Karrenbauer, M.; Fellan, A.; Schotten, H.; Buhr, H.; Seetaraman, S.; Niebert,

N.; Bernardy, A.; Seelmann, V.; Stich, V.; et al. A5G architecture for the factory of the future. In Proceedings of the 2018 IEEE 23rd international conference on emerging technologies and factory automation (ETFA), Torino, Italy, 4–7 September 2018; Volume 1, pp. 1409–1416.

[26] Shi, W.; Cao, J.; Zhang, Q.; Li, Y.; Xu, L. Edge Computing: Vision and Challenges. IEEE Internet Things J. 2016, 3, 637–646.

[27] Gubbi, J.; Buyya, R.; Marusic, S.; Palaniswami, M. Internet of Things (IoT): A vision, architectural elements, and future directions. Futur. Gener. Comput. Syst. 2013, 29, 1645–1660.

[28] Vermesan, O.; Bröring, A.; Tragos, E.; Serrano, M.; Bacciu, D.; Chessa, S.; Gallicchio, C.; Micheli, A.; Dragone, M.; Saffiotti, A.; et al. Internet of robotic things: Converging sensing/actuating, hypoconnectivity, artificial intelligence and IoT Platforms. In Cognitive Hyperconnected Digital Transformation: Internet of Things Intelligence Evolution; River Publishers: Gistrup, Denmark, 2017; pp. 97–155.

[29] Gremban, K. Editorial and Introduction to the Issue: Risk and Rewards of the Internet of Things. IEEE Internet Things Mag. 2018, 1, 2–3.

[30] Hu, Y.C.; Patel, M.; Sabella, D.; Sprecher, N.; Young, V. Mobile edge computing—A key technology towards 5G. ETSI White Paper 2015, 11, 1–16.

[31] Mao, Y.; You, C.; Zhang, J.; Huang, K.; Letaief, K.B. A Survey on Mobile Edge Computing: The Communication Perspective. IEEE Commun. Surv. Tutor. 2017, 19, 2322–2358.

[32] McCann, J.; Quinn, L.; McGrath, S.; O'Connell, E. Towards the Distributed Edge–An IoT Review. In Proceedings of the 2018 12th International Conference on Sensing Technology

(ICST), Limerick, Ireland, 3–6 December 2018; pp. 263–268.

[33] Bockelmann, C.; Pratas, N.K.; Wunder, G.; Saur, S.; Navarro, M.; Gregoratti, D.; Vivier, G.; De Carvalho, E.; Ji, Y.; Stefanovic, C.; et al. Towards Massive Connectivity Support for Scalable mMTC Communications in 5G Networks. IEEE Access 2018, 6, 28969–28992.

[34] Popovski, P.; Trillingsgaard, K.F.; Simeone, O.; Durisi, G. 5G Wireless Network Slicing for eMBB, URLLC, and mMTC: A Communication-Theoretic View. IEEE Access 2018, 6, 55765–55779.

[35] Siddiqi, M.; Yu, H.; Joung, J. 5G Ultra-Reliable Low-Latency Communication Implementation Challenges and Operational Issues with IoT Devices. Electronics 2019, 8, 981.

[36] Gidlund, M.; Lennvall, T.; Akerberg, J. Will 5G become yet another wireless technology for industrial automation? In Proceedings of the 2017 IEEE International Conference on Industrial Technology (ICIT), Toronto, ON, Canada, 22–25 March 2017; pp. 1319–1324.

[37] Orlosky, J.; Kiyokawa, K.; Takemura, H. Virtual and Augmented Reality on the 5G Highway. J. Inf. Process. 2017, 25, 133–141.

[38] Cheng, J.; Chen, W.; Tao, F.; Lin, C.-L.; Da Xu, L. Industrial IoT in 5G environment towards smart manufacturing. J. Ind. Inf. Integr. 2018, 10, 10–19.

[39] Baharom, B.; Ali, M.T. Multiple-element PIFA MIMO antenna system design for future 5G wireless communication applications. In Proceedings of the 2017 IEEE Asia Pacific Microwave Conference (APMC), Kuala Lumpur, Malaysia, 13–16 November 2017; pp. 743–746.

[40] Gante, J.; Falcão, G.; Sousa, L. Deep Learning Architectures for Accurate Millimeter Wave

Positioning in 5G. Neural Process. Lett. 2019.

[41] Wang, C.; Zhao, Z.; Sun, Q.; Zhang, H. Deep learning-based intelligent dual connectivity for mobility management in dense network. In Proceedings of the 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall), Chicago, IL, USA, 27–30 August 2018; pp. 1–5.

[42] Santos, J.; Leroux, P.; Wauters, T.; Volckaert, B.; De Turck, F. Anomaly detection for smart city applications over 5G low power wide area networks. In Proceedings of the 2018 IEEE/IFIP Network Operations and Management Symposium, Taipei, Taiwan, 23–27 April 2018; pp. 1–9.

[43] Maimó, L.F.; Gómez, Á.L.P.; Clemente, F.J.G.; Pérez, M.G.; Pérez, G.M. A self-adaptive deep learning-based system for anomaly detection in 5G networks. IEEE Access 2018, 6, 7700–7712.

[44] Parwez, M.S.; Rawat, D.B.; Garuba, M. Big data analytics for user-activity analysis and user-anomaly detection in mobile wireless network. IEEE Trans. Ind. Inform. 2017, 13, 2058–2065.

[45] Maimó, L.F.; Celdrán, A.H.; Pérez, M.G.; Clemente, F.J.G.; Pérez, G.M. Dynamic management of a deep learning-based anomaly detection system for 5G networks. J. Ambient Intell. Humaniz. Comput. 2019, 10, 3083–3097.

[46] Hussain, B.; Du, Q.; Ren, P. Deep learning-based big data-assisted anomaly detection in cellular networks. In Proceedings of the 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, UAE, 9–13 December 2018; pp. 1–6.

[47] Li, R.; Zhao, Z.; Zheng, J.; Mei, C.; Cai, Y.; Zhang, H. The learning and prediction of application-level traffic data in cellular networks. IEEE Trans. Wirel. Commun. 2017, 16,

3899–3912.

[48] Ahmed, K.I.; Tabassum, H.; Hossain, E. Deep learning for radio resource allocation in multi-cell networks. IEEE Netw. 2019, 33, 188–195.

[49] Zhang, C.; Patras, P.; Haddadi, H. Deep learning in mobile and wireless networking: A survey. IEEE Commun. Surv. Tutor. 2019, 21, 2224–2287.

[50] Coutinho, E.F.; de Carvalho Sousa, F.R.; Rego, P.A.L.; Gomes, D.G.; de Souza, J.N. Elasticity in cloud computing: A survey. Ann. Telecommun. 2015, 70, 289–309.

[51] Qing, C.; Cai, B.; Yang, Q.; Wang, J.; Huang, C. Deep learning for CSI feedback based on superimposed coding. IEEE Access 2019, 7, 93723–93733.

[52] Jiang, Z.; Chen, S.; Molisch, A.F.; Vannithamby, R.; Zhou, S.; Niu, Z. Exploiting wireless channel state information structures beyond linear correlations: A deep learning approach. IEEE Commun. Mag. 2019, 57, 28–34.

[53] Prasad, K.S.V.; Hossain, E.; Bhargava, V.K. Energy efficiency in massive MIMO-based 5G networks: Opportunities and challenges. IEEE Wirel. Commun. 2017, 24, 86–94.

[54] Cheng, X.; Liu, D.; Zhu, Z.; Shi, W.; Li, Y. A ResNet-DNN based channel estimation and equalization scheme in FBMC/OQAM systems. In Proceedings of the 2018 10th International Conference on Wireless Communications and Signal Processing (WCSP), Hangzhou, China, 18–20 October 2018; pp. 1–5.

[55] Ez-Zazi, I.; Arioua, M.; El Oualkadi, A.; Lorenz, P. A hybrid adaptive coding and decoding scheme for multi-hop wireless sensor networks. Wirel. Pers. Commun. 2017, 94, 3017–3033.

[56] Kang, J.M.; Kim, I.M.; Chun, C.J. Deep Learning-Based MIMO-NOMA With Imperfect SIC Decoding. IEEE Syst. J. 2019.

[57] Kim, M.; Kim, N.I.; Lee, W.; Cho, D.H. Deep learning-aided SCMA. IEEE Commun. Lett. 2018, 22, 720–723.

[58] Xue, S.; Ma, Y.; Yi, N.; Tafazolli, R. Unsupervised deep learning for MU-SIMO joint transmitter and noncoherent receiver design. IEEE Wirel. Commun. Lett. 2018, 8, 177–180.

[59] Kim, M.; Lee, W.; Cho, D.H. A novel PAPR reduction scheme for OFDM system based on deep learning. IEEE Commun. Lett. 2017, 22, 510–513.

[60] Yildirim, Ö. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Comput. Biol. Med. 2018, 96, 189–202.

[61] Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A search space odyssey. IEEE Trans. Neural Netw. Learn. Syst. 2016, 28, 2222–2232.

[62] Bai, S.; Kolter, J.Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv 2018, arXiv:1803.01271.

[63] Jiang, M.; Liang, Y.; Feng, X.; Fan, X.; Pei, Z.; Xue, Y.; Guan, R. Text classification based on deep belief network and softmax regression. Neural Comput. Appl. 2018, 29, 61–70.

[64] Barlacchi, G.; De Nadai, M.; Larcher, R.; Casella, A.; Chitic, C.; Torrisi, G.; Antonelli, F.; Vespignani, A.; Pentland, A.; Lepri, B. A multi-source dataset of urban life in the city of Milan and the Province of Trentino. Sci. Data 2015, 2, 150055.

[65] Mohammed, T. A., Alhayali, S., Bayat, O., & Uçan, O. N. (2018). Feature reduction based on hybrid efficient weighted gene genetic algorithms with artificial neural network for machine learning problems in the big data. Scientific Programming, 2018.

[66] Mohammed, T. A., Bayat, O., Uçan, O. N., & Alhayali, S. (2019). Hybrid Efficient Genetic Algorithm for Big Data Feature Selection Problems. Foundations of Science, 1-17.

[67] Yang, Jie & He, Haiyang & Qiao, Yuanyuan. (2014). Network traffic analysis based on Hadoop. 1-5. 10.1109/VITAE.2014.6934437.

[68] Hiraman, Bhole & M, Chapte & Abhijeet, C. (2018). A Study of Apache Kafka in Big Data Stream Processing. 1-3. 10.1109/ICICET.2018.8533771.