Study of HARQ Techniques for Satellite Network Support of 5G Wireless System

Politecnico di Torino

Yaseen Mohammed Salim Alghawi

Supervisor: P. Guido Montorsi

2020

Abstract

Integrating the satellite network with a 5G network to equip the latter with world-wide coverage and high availability is an innovative and interesting idea, however, due to the large distance of the physical link between the two networks, which is in order of hundreds or thousands of kilometers, the communications conducted over this link may not be reliable. Depending on link adaptation is a good idea to increase the link reliability and better exploit the wireless communications channel, but, in this particular link, the adaptation process alone is not enough since the channel information reported to the transmitter is outdated (the channel changes much faster than the round trip time (RTT) of the non−terrestrial system). In this thesis, The implementation of HARQ (hybrid automatic repeat request) in cooperation with ACM (adaptive coding and modulation) is considered to better handle the problem. Using HARQ control loop of feedbacks and transmissions, the reliability of the system is secured, and, by introducing a suitable margin to optimally compromise the throughput efficiency and the frame error rate (FER), and applying that margin to the parameters decided by ACM, the channel can be exploited as well as possible. As for the large RTT and outdated CSI (channel state information), the prediction will be used to forecast the channel variations and tune the ACM parameters properly to maximize the throughput efficiency of the system. Through comparing several linear and nonlinear prediction models, the prediction models proposed to replace the one−tap filter (already used) are linear prediction filter model that uses AR (autoregressive) autocorrelation computation for prediction, nonlinear ARX (autoregressive with exogenous variables) model, and a more modern model based on deep learning. MATLAB simulations are made, and based on the results obtained, using HARQ substantially increases the system performance as compared to a single−shot transmission system, HARQ can substantially increase the throughput efficiency of the system, sometimes up to 70%, therefore HARQ is recommended to be used in the transmission system of the 5G− satellite integrated network (the performance gain is worth tolerating the additional HARQ delay). Furthermore, the system performance gained by using HARQ and ACM with the one−tap filter can be further increased, where, in the results obtained via simulation in this thesis, the throughput efficiency of the system is increased at some points, by 15% in the system using the nonlinear model, and by 25% in the systems using the linear and the deep learning models (the performance is improved even more when considering lower FER), finally, comparing the three proposed prediction models, it turns out that among them the deep learning model and the linear model have the highest performance (both have a very close performance with a slight advantage to the deep learning model). The results suggest that, using the mentioned techniques, a reliable 5G−satellite link with good performance can be achieved, and that the performance can be further improved using more sophisticated prediction models.

Dedication

For only having someone so great guarding my back, that I could dash forward without any concerns. Dedicated to my brother Yaser.

Acknowledgements

I want to thank my supervisor prof. Guido Montorsi, for all of his efforts and helping me countless times along the construction of this thesis, without him, it wouldn't come out as it is now. I would also like to thank the professors both in my country and here in Politecnico di Torino for the knowledge and skills I acquired during my time at the university. Finally, I thank my family and friends for their continuous support.

Abbreviations

NGMN	next generation mobile networks
GEO	geostationary earth orbit
LEO	low earth orbit
HTS	high throughput satellite
5G	fifth generation of mobile communications
NTN	non terrestrial network
SDN	software defined network
NFV	network function virtualization
NR	new radio
HARQ	hybrid automatic repeat request
LTE	long term evolution
LDPC	low density parity check
CSI	channel state information
MEO	medium earth orbit
ARQ	automatic repeat request
FEC	forward error correction
RTT	round trip time
ACM	adaptive coding and modulation
TB	transport block
TTI	transmission time interval
CB	code block
RV	redundancy version
CBG	code block group
TPC	transmit power control
FER	frame error rate
MI	mutual information
LOS	line of sight
NLOS	non line of sight

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Upon the vision of next−generation mobile networks (NGMN) alliance, The communications systems' fifth−generation (5G) should be able to provide worldwide services and access from anywhere on the globe, which means worldwide coverage and high availability, unfortunately, this cannot be achieved through the terrestrial 5G network on its own due to the financial inefficiency of installing and operating stations to cover extremely large areas which at any given time contain a very limited number of users such as deserts, difficult terrains, and oceans. The 5G network is unable to cover such environments with reasonable cost as doing so would require stations with high availability, powerful broadcast capability, and huge cell coverage area, however, such characteristics are available in non−terrestrial network stations, the satellites do have a large footprint with the mentioned requirements and an example of that are the K−band satellites used for television broadcast in the geostationary earth orbit (GEO) where the footprint of one of those is capable of covering an entire continent and the broadcast is available anywhere inside that footprint, they also are popular in the field of long distances telephony and disaster relief. In last years satellite networks have witnessed an increasing interest especially those in the low earth orbit (LEO) because of their improved performance and diminished cost, also augmented capacity corresponding to putting higher frequency bands into use with optical transmissions cooperation. The satellite system flexibility can be improved by applying advanced onboard processing which enables techniques such as interference management and flexible beamforming, these are implemented in what is known as high throughput satellites (HTS). the HTS possess multi−spot beam and frequency reuse techniques, multiple frequency reuse is achieved by using small beams (increasing directivity) with high gain, these abilities result in broadband connectivity with reduced cost[1].

During the next years, it is predicted that more than 100 GEO and HTS based LEO constellations will have been launched with estimated capacity in the order of Tbps. At the same time, on an altitude inferior to that of space networks another non−terrestrial network (NTN) has been developed, Aerial networks composed of various aircrafts using sensors and transceivers for communications, civil and other purposes, however, applying standards of terrestrial networks on space and aerial networks is quite complicated since the two networks have evolved separately over the years due to their focus on different targets. The opportunity to integrate NTN with mobile networks arose once the 5G networks had been introduced, the two networks realized the need to integrate, adding a third dimension

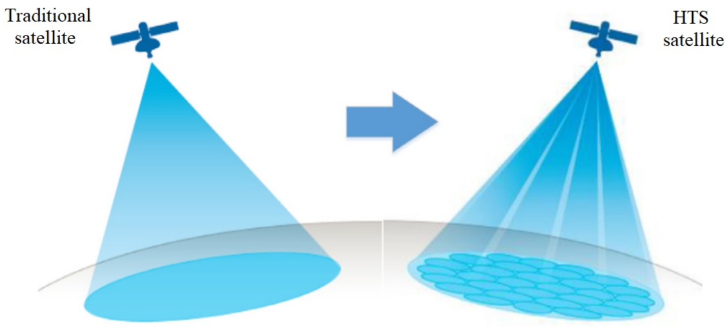

Figure 1.1: Traditional satellite and HTS satellite[1]

to the 5G network and opening the door to a new type of networks, nevertheless, because of weak links, topologies changing over time, special characteristics and restraints on the payload, many problems are yet to be dealt with before being able to achieve 5G and space networks full integration. Recently, many academic and industrial works were conducted on various topics related to satellite and 5G integration such as multi−antenna technologies of space−ground networks, exploiting software−defined networking (SDN) and network function virtualization (NFV) to design a 5G satellite integration ,and essential techniques on multi−satellite cooperation transmission in 5G, unfortunately, most of these studies failed to touch the topic from an industrial point of view. Therefore, much more study and work are needed to address further issues such as security and to narrow the gap between theoretical and industrial levels of development.

1.1 Thesis Structure

The first chapter is an introduction, chapter 2 discusses the 5G NR basics and channel related techniques, the satellite orbits and delays are illustrated in chapter 3, chapter 4 displays the techniques used to handle wireless communications channels, in chapter 5 the key parameters in the NR to NTN channel implemented in the software are explained concentrating on HARQ, chapter 6 is an overview of the software used to simulate the wireless system, the simulation results are shown and discussed in chapter 7, finally, chapter 8 states the conclusion of the thesis.

Chapter 2

5G NR Terrestrial Network

Communications systems are one of research and industry evergreen branches which witnesses new technologies and massive development over the years, particularly mobile communications systems, whose networks beheld development in the form of a new generation approximately every 10 years for around half a century to meet users expectations and keep up with increasing demands. The fifth-generation mobile network is foreseen to extend the capabilities of wireless communications networks and push them to the next level, connecting users with speed and reliability superior to previous generations. With modern techniques and utilities equipped; 5G NR can provide Gigabit services to all connected users, it is also expected to provide global coverage, very high reliability and efficiency, low latency, and a huge number of diverse connections. Whether it be virtual reality, online gaming, remote surgery, or massive IoT, 5G is used for various human users and machine applications.

2.1 5G Use Cases

The requirements of human and machine−type users of the 5G network are different and some of which stand in contrast with one another depending on the services desired. Unlike its preceding generations, 5G takes a new approach to serve these requirements by targeting 3 different use cases.

Enhanced Mobile Broadband (eMBB) This is the extension of the natural mobile broadband upgraded to meet the requirements of the current standards. The eMBB use case targets human−type users concentrating on connectivity and multimedia and data services, this can be achieved by providing very high data rates that can tolerate this type of service.
Regarding cell coverage and supported data rates, there are two possible cases, low mobility and dense environments with very high data rates, and a high mobility environment with lower users' density and data rates. The scenarios are known as hotspot connectivity and wide area coverage respectively, both cases are supported in the 5G eMBB use case.

Massive Machine−type Communications (mMTC) The massive number of devices demanding connectivity and IoT services are supported in this category of 5G. This use case focuses on providing connectivity to a large number of devices, these are expected to have low traffic nature transmissions and no critical constraints on delay, on the other hand, they are expected to have a long lifetime of the battery and low energy consumption to enable remote utilities.

Ultra Reliable Low Latency Communications (URLLC) Applications with strict restrain in terms of reliability and latency are supported in this use case. It essentially targets machine−type communications and wireless control applications such as remotely driven vehicles and manufacturing processes.

2.2 NR Physical Layer

While other layers like RLC and MAC handle segmentation, scheduling, and resources assignment, the physical layer in 5G NR is responsible for the coding, decoding, modulation, and demodulation processes as well as handling multiantenna operations and mapping signals to time−frequency physical resources.

2.2.1 Constellation Sets

5G NR backs many possible modulation options, most of which are the same used in LTE, namely, quadrature phase−shift keying (QPSK), quadrature amplitude modulation (QAM) with cardinalities 16, 64, and 256, these are supported by NR in both uplink and downlink. Besides, 5G supports also BPSK only in the uplink for mMTC processes which have low data rates to improve power amplifiers efficiency by reducing the peak to average power ratio. The list of supported modulation schemes may extend to include other schemes such as 1024 QAM. The variety of modulation schemes used in the system add flexibility and enables throughput or robustness increment options depending on the channel condition.

2.2.2 Waveform

One of the techniques exploited in 5G is Cyclic prefix OFDM which is employed in uplink and downlink in NR, while LTE had it employed only in the downlink. Another is DFT−spread OFDM which was used for LTE uplink is also supported in NR uplink for restrained coverage cases. Using the same OFDM technique in both links enables the same waveform in the two directions, which reduces the complexity of the overall design. All receiver transparent operations are permitted to be used on top of NR waveform like filtering to enhance the shape of the spectrum.
NR uses flexible OFDM frequency options by allowing a scalable value for the subcarrier spacing parameter, for LTE, subcarrier spacing is fixed to 15 kHz, this is respecified in NR to have a value of 15×2^n kHz, where n=0,1,2,3,4, giving possible subcarrier spacing of 15 kHz, 30 kHz, 60 kHz, 120 kHz, and 240 kHz, this is done while maintaining the same 7% of CP overhead in LTE. As for the spectrum of the OFDM signal, NR can use 94 to 99% of the transmission band compared to 90% in LTE, using operations such as filtering and windowing to suppress the spectrum of the OFDM signal outside the allocated frequency band.

2.2.3 Multiple Antennas

To meet the performance requirements of 5G NR, the spectrum used by the network is extended to include the millimeter−wave bands, this enables the deployment of beamforming in NR to obtain the required coverage.
The multiantenna techniques are used for low frequencies to improve the functionalities used in LTE, this enhances the spectral efficiency of the system allowing higher data rates and increasing capacity in the assigned spectrum. The augmentation in spectral efficiency is introduced by the improvement in the spatial resolution which in turn is introduced by a large number of antenna elements in the antenna array, also known as massive MIMO (Multiple Input Multiple Output).
In contrast, for high frequencies, the issue is to acquire coverage rather than higher spectral efficiency, since the higher frequencies give rise to transmission losses, this can be solved by using beamforming which also helps to obtain the required coverage. thus, beamforming is supported in gNB and UE in NR for data transmission, access, and broadcasting.

2.2.4 Channel Coding

NR is equipped with low−density parity check (LDPC) channel codes for the transmission of data and polar codes for signaling and control transmissions. LDPC codes have desirable characteristics in terms of implementation, especially at Gbps level data rate applications. The LDPC codes used in NR differ from the ones used for other systems since NR uses rate compatibility which allows different code rate transmission and enables the use of HARQ incremental redundancy processes.
For the short length blocks used for control signaling at the physical layer and HARQ is not enabled, NR implements polar codes that achieve good performance for blocks of similar length.

LDPC Codes

Classified under the category of linear block codes, LDPC is a family of codes that are characterized by a sparse parity check matrix, which means the number of nonzero entries is small. The performance of LDPC codes is not far from the capacity of a big collection of communication channels, thus, they are used in many standards such as WiMAX and DVB−S2. The low density of nonzero elements permits the implementation of low complexity decoding methods, this usually results in performance close to that of maximum likelihood[2].
The channel coding in 5G NR is depending on LDPC codes, they are constructed in such a way that they can support the performance of IR−HARQ processes. Based on the CSI feedback, the ACM control process sets the adequate modulation and coding schemes, and depending on those, the TB size is selected for the set of resources allocated in the used slot, then the TB size and the selected rate are used to decide the LDPC base graph to be used for coding.
There are 4 redundancy versions available, each of which corresponds to a starting point from which the reading of the created LDPC code is done, so, for 4 RVs there are 4 unique positions in the circular buffer that contains the LDPC codeword. The 5G NR enables the selection of the order of the RVs indices, those are used in the selected order during the HARQ process and the maximum allowed number of

transmissions.
During the HARQ process, after the transmission of the first RV, the process holds expecting an ACK/NACK (feedback) from the receiver. Once the feedback reaches the transmitter the next step is triggered, in case a NACK is received a new transmission is triggered with the next RV, otherwise, (ACK is received or the maximum number of transmissions is reached) HARQ process is released. A HARQ process requires one buffer at the transmitter to keep the generated LDPC codeword and another at the receiver to keep the soft information.

2.2.5 Physical Channels

The physical resources in NR are represented by a time− frequency grid known as physical resource block, a resource block consists of 12 subcarriers by 14 OFDM symbols. An OFDM symbol is the smallest time resource unit, 14 of them in 15 kHz subcarrier spacing form a slot. One slot forms a subframe of 1 ms duration (the number of slots in a subframe depends on the subcarrier spacing, for example, in 30 kHz spacing, a subframe contains 2 slots and still has a 1 ms duration), and 10 subframes form a frame which is the largest unit with a duration of 10 ms.
The resources are utilized as channels that are used to carry higher layers' data. The physical channels used in NR uplink and downlink are listed as follows:

- Physical downlink shared channel (PDSCH), it is the channel used for downlink data transmission.
- Physical uplink shared channel (PUSCH), the channel used by UE for uplink data transmission.
- Physical downlink control channel (PDCCH), this channel is used for the various tasks of control information, such as scheduling required for the reception of PDSCH, and grant−giving scheduling to permit PUSCH transmissions.
- Physical uplink control channel (PUCCH), the channel used for uplink control information such as HARQ ACK/ NACK feedback and downlink CSI for link adaptation.
- Physical broadcast channel (PBCH), it is the channel used for system information broadcast to allow UE to access the network.
- Physical random access channel (PRACH), exploited by UE to ask for the connection setup known as random access.

2.3 5G Challenges

In the previous generations of mobile communications, the main concentration was aimed in the direction of improving the data rates which reduced the focus on achieving global coverage to some extent. 5G NR despite aiming for global coverage in addition to data rate boosting has implemented technologies which are not in the best interest for universal coverage such as ultra cell densification which reduces the coverage of cells, massive MIMO, and beamforming a technique that requires

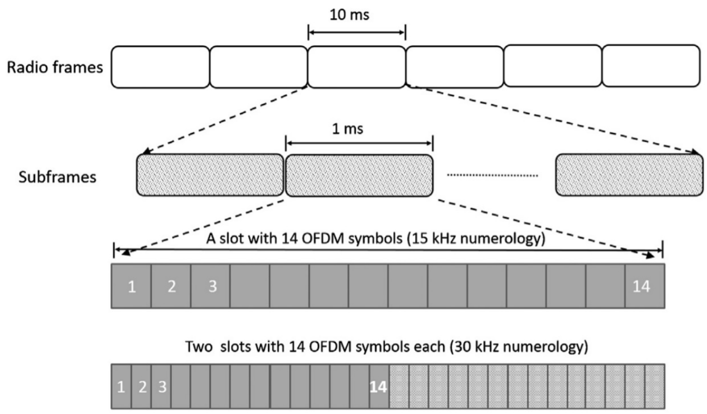

Figure 2.1: 5G NR time frame structure[2]

small cell size, and millimeter−wave meaning higher frequencies and therefore higher losses[3].
Furthermore, the attempt to cover sparse areas and difficult terrains such as mountains is financially consuming, and the sparseness of those terrains deems such expenses sort of wasted since the return from the few users is not cost−efficient. Thus, achieving global coverage under these circumstances may be the greatest challenge facing 5G NR terrestrial network.

2.3.1 Possible Solutions

The problem of lack of coverage and availability could be solved by the integration of a 5G terrestrial network with nonterrestrial networks, namely, satellite networks. Using satellites' availability and taking advantage of their large footprints can cover both large sparse rural areas and difficult terrains in which it would've been extremely difficult to cover via terrestrial stations. Using satellites for these purposes, nevertheless, comes with its own set of problems as large physical delays (further discussed in chapter 3).

Chapter 3

Non−Terrestrial Satellite Networks

A satellite used for communications is a station located in space that has some functionalities, of which the most important is its similarity in task with a communications microwave repeater. The satellite enables two or a higher number of users with the proper equipment to transfer information and exchange them in more than one way, the satellite receives a signal from a communications terrestrial station, amplifies it, and may process it, and afterward, it transmits the signal to earth again to be received by one or more terrestrial stations.
Satellite communications are used and play a huge role in our daily lives, connecting the world and covering major events almost instantaneously, especially in cases of global crisis.[4].

3.1 Satellite Communications Advantages

The satellites are widely used in communications due to some of their well−known merits, the most important of which is the ability of the satellite to see a huge geographical area at any given time. Some of the key advantages of satellite communications are:

- **Cost Efficiency** The cost of the capacity of the satellite is disinterested with the number of users or sites associated with it, nor is it sensitive to the distance between itself and the location of the receiver whether it was in the same region or a different continent[5].

- **Availability** The satellites are known for their large footprints with enough capacity to serve users within that area, so, in remote areas in which the terrestrial networks fail to provide high−speed access or access at all, users tend to rely on satellite availability.

- **Reliability** Unlike terrestrial networks, satellites can work without the intervention of terrestrial infrastructure. In scenarios where terrestrial networks witness outages, satellite links stay operational.

- **High Performance** Satellite has the upper hand against other communications systems in terms of broadcast applications. Taking, for example, TV broadcasting where the satellite serves large areas with unparalleled coverage and availability.

- **Scalability** Receivers or nodes can be added to the network and become ready for deployment almost instantly compared to other networks. Satellites can provide instant infrastructure since all is needed to add a new receiver or node is proper ground equipment.

- **Versatility** globally, all communications forms are supported by satellites. Satellites are used in a very flexible manner; they can manage their tasks independently or contribute to larger networks.

3.2 Satellite Orbits

There are many possible orbits to be considered in the satellite design, out of them, a few are well−known and are the most common in communications applications, these are discussed in the following sections.

3.2.1 Geostationary Earth Orbit (GEO)

The radius of the orbit and its corresponding time interval of revolution have a fixed relationship described in Kepler's third law. Accordingly, there exists an orbit with a certain radius that has a similar angular velocity to that of earth. The orbit is with circular circumference and is located in the same plane as the earth's equator, this makes the satellite appear as if it is soaring motionlessly above a corresponding point in the equator. This very important orbit is known as the geostationary earth orbit. The altitude of the orbit above the surface of the earth is 35786 km and takes an interval of 23 hours and 56 minutes to revolve around the earth.
GEO hosts most of the communications satellites especially in the applications of data transfer between two or higher number of nodes on earth, in which, the satellite operate somewhat like a repeater. The view of the satellite at that altitude is comprehensive towards the side of the planet it is facing which gives it a footprint that covers a substantial amount of land (approximately one−third of the surface of the earth), also, the fact that it has a fixed motion concerning earth reduces to a great extent or even annihilates the need for ground tracking. With that being said, GEO satellite communications have its disadvantages such as the huge path loss due to the large distance and the immense delay for the same reason, thus, reducing the use of GEO satellites in delay critical applications.

3.2.2 Low Earth Orbit (LEO)

LEO satellites operate at altitudes between 160 and 2500 km, they are popular in the applications of mobile satellite communications. The area covered by LEO satellites is obviously smaller than that of GEO, so it would take multiple satellites for global coverage, however, since the link is much shorter, the path loss is lower which yields lower power and smaller antennas, also, the delay reduced to a tolerable value for many applications. LEO can also cover certain areas that cannot be reached by GEO, namely, polar areas.

One disadvantage of LEO satellites is their constrained duration of operation since they are not at the same location in space with respect to earth. LEO satellite moves across a fixed point on earth for an approximate duration of 10 minutes, therefore, it first requires a tracking terminal on the ground, and second, a constellation of them are needed to achieve continuous coverage (12, 24, and 66 satellites constellations are used in some applications).

3.2.3 Medium Earth Orbit (MEO)

The set of satellites operating above LEO and below GEO ranges are referred to as MEO, their specific range extends from 10000 to 20000 km. The advantageous characteristics of MEO include the repeatable traces of repeating ground coverage, flexible number of revolutions per day, and a suitable relative motion with respect to earth allowing accurate location measurements.
MEO satellite can provide an observation interval from 1 to 2 hours for a fixed point on earth, they are similar to LEO only at higher altitudes. MEO acquire the appropriate characteristics for navigation, remote sensing, and positioning applications, the most famous of which is undoubtedly the Global Positioning System (GPS) which goes without saying that it is used worldwide in our everyday lives, GPS normally uses a constellation of 24 satellites or more in half−day duration orbits at an altitude of 20184 km.

3.2.4 Highly Elliptical Orbit (HEO)

Used to cover areas of high latitudes that cannot be reached by GEO and need longer duration connections than those provided by LEO, HEO satellites take elliptical orbits for this purpose. The most famous example of HEO satellites orbit is the one named Molniya orbit used to cover high northern latitudes for the Soviet Union back in the days. The orbit has a perigee of 1000 km altitude and apogee of almost 40000 km, it has a revolution period of 12 hours, of which, almost 10 hours are spent over the northern hemisphere due to the high elliptical circumference of the orbit. Using enough satellites, continuous coverage is achieved since at least one of the HEO satellites used would be noticeable at any given time in the day.

3.2.5 Polar Orbit

A vertical circular orbit is useful for data collection and sensing applications. The characteristics of the orbit are selected such that it can scan the entire globe on a periodic cycle. An example of polar orbit satellites is Landsat, it has an altitude of 912 km and revolution time of 103 minutes, making 14 revolutions per shifting west on the equator by 160 km to return to its original plane in 18 days and 252 revolutions.

3.3 Satellite Delays

Transmission of data via satellites starts by transmission from the user terminal to the satellite and then back to the terrestrial terminal, this process accumulates delay from the following sources:

- Delays in the terrestrial network itself.
- Delay due to propagation over satellite links.
- Processing delay of the baseband signal.
- Protocol induced delay.

The summation of the delay from these sources results in the total delay for an end to end satellite connection. Each of them is briefly discussed in the following sections.

3.3.1 Delay in Terrestrial Networks

This is the delay introduced by the terrestrial network without any latency from the satellite, it includes switching and propagation delays and can be approximated as illustrated in Eq. (3.1).

$$t_{TN}(ms) = 12 + 0.004 \times distance(km) \tag{3.1}$$

3.3.2 Propagation Delay over Satellite Links

The physical delay due to the distance between the terrestrial equipment and the satellite. The calculation of such delay is straightforward $t_{SL}(s) = R/c$, where R is the range from the terrestrial equipment to the satellite for one link or range from terrestrial equipment to the satellite and from that back to terrestrial equipment for the full link. And c is the speed of light.
In satellite communications, this delay has a major contribution to the overall delay and naturally, the delay becomes larger as the considered orbit becomes farther from the surface of the earth.

3.3.3 Signal Processing Delay

The delay is produced by baseband signal processing in the terrestrial station and onboard of regenerative satellites. Typical sources of processing delay are encoding, multiplexing, demodulation, and decoding, and buffering in switching and multiple access.

3.3.4 Protocol Induced Delay

Depending on the protocols implemented for increasing link reliability and reducing errors, a delay is introduced corresponding to the process of that protocol. An example of that is the protocol of Hybrid ARQ, in which the process introduces extra delays for error detection and correction and transmission and processing of ACKs or NACKs with corresponding retransmissions[6].

3.4 Satellite Communications Challenges

As mentioned earlier in this chapter, it is quite possible providing global coverage using satellites, this can be extremely beneficial for terrestrial networks like 5G by integrating the two networks, this, however, introduces the disadvantages of the satellite network to the terrestrial network, which is usually not designed to tolerate them.
Out of satellite disadvantages, the one that has the strongest impact on terrestrial networks is the large distances of the satellite links, which result in large RTT, this limits the use of the network in delay−sensitive applications, places physical constraints on used techniques such as the number of process in HARQ and render feedback information such as CSI expired. In table 3.1, the key parameters of the basic orbits in satellite communications are demonstrated.

Table 3.1: The minimum number of HARQ processes ($N_{HARQ,MIN}$) for terrestrial and non−terrestrial network, assuming 1ms slot duration

Orbit	GEO	MEO	LEO
Altitude (km)	35786	7000 to 25000	300 to 1500
Motion (km/s)	stationary	4.9	7.5
Cell Size (km)	200 to 1000	100 to 500	100 to 500
RTT (ms)	270	95 (at 10000 km)	13 (at 600 km)

3.4.1 Satellite Possible Solutions

One way to deal with the problem of large RTT is to limit or disable feedback−based techniques such as HARQ, while extending terrestrial standards to include and tolerate low and medium satellites RTT.
Another way would be to depend on prediction−based solutions to acquire a version of CSI which can be used to tune the parameters of the transmission rather than waiting for the feedback over the large RTT which would be too late against fast−changing channel since the parameters and impairments of the channel will have changed by the time the estimated SCI is received.

Chapter 4

Channel Variation Handling Mechanisms

Wireless communication channels are susceptive to various sources of distortion, for counting but not limited to noise, interference, and fading. Multipath fading plays a huge role in mobile networks' channels where it results in frequency−selective fading in which the channel response varies over frequency span, also fast fading which forces the channel characteristics to change over time, contributes to increasing transmission errors and drop system performance, moreover, the impact of such impairments is significantly augmented when considering a non−terrestrial channel where the substantial delay introduced by long distances and additional impairments present in space link further complicate the situation, therefore, certain techniques should be implemented to counteract the mentioned sources of distortion and give raise to system performance and reliability.

4.1 Handling Channel Variation Before Transmission

By changing some parameters or techniques at the transmitter side before transmitting, a result could arise at the receiver where the distortion introduced by the channel is canceled or reduced, this, however, depends on the choices made at the transmitter.

4.1.1 Link Adaptation

Channel variation effects can be negated by tuning certain parameters at the transmitter side to obtain constant performance after channel characteristics change take place, the channel state information (CSI) must be known at the transmitter side, otherwise, there is no clue on how to modify the parameters, it is possible to communicate the channel information to the transmitter through channel estimation, measurements, and feedback from the receiver. If done accurately, these techniques enable the communications system to make the best out of the channel[7].

Power Control

The signal−to−noise ratio (SNR) seen at the input of the receiver is directly related to the received power which in turn is proportional to the transmitted power. One way to maintain a constant SNR at the receiver is by dynamically controlling the transmitted power to overcome channel variations. Investing more power into the signal when the channel condition is bad and vice versa keeps the SNR at the receiver steady, and that means a constant rate and performance.

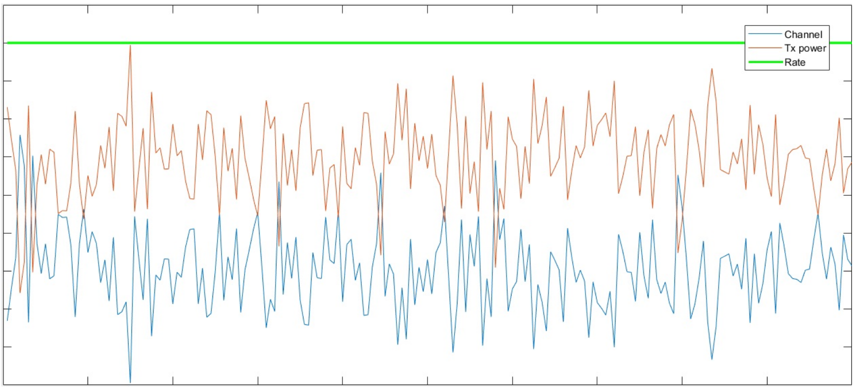

Figure 4.1: Power control constant rate

Power Control Challenges In applications that use satellite links especially GEO, the adaptivity of power control is very restricted due to significant free space loss influenced by a large distance and limited battery power, these constraints reduce the role of power control on GEO satellite links. Other than that, the large round trip time (RTT) makes it very difficult for power control to track fast fading.

Adaptive Modulation and Coding AMC

Adjusting the coding and modulation scheme to cope with the variations of the channel is an efficient way of improving system performance. unlike the power control method, the SNR seen at the receiver side is not constant which means the data rate is not constant over time, however, at each time interval the channel condition is exploited as best as it can be to assure the best performance in terms of data throughput and transmission errors.

Channel Coding Channel coding is widely used in wireless communication systems to detect and correct errors, by adding redundant data to the original data to be transmitted, the receiver can detect whether the data codeword has been subjected to errors or not as well as correcting errors in the range of its capabilities, nonetheless, this improvement to the system performance on handling errors point of view is met by data throughput deterioration since the extra bits added are not part of the original data and, therefore they occupy a band which could have been used to increase data throughput. The amount of redundant information added is governed by the parameter known as code rate which is the ratio of the useful data to be transmitted to the overall data after coding operation, for example, a coding scheme with a coding rate of 1/ 3 means that the data generated by that coding scheme is 3 times that of the useful data, that would be considered a waste of recourses in very good channel condition where using a code rate close to 1 would nearly triple the data throughput, on the other hand, using code rate close to 1 will not enable the receiver to correct errors in bad channel condition the throughput is going to drop due to poor system performance against errors. There are many families of codes, the ones used recently in mobile networks are LDPC codes in 5G NR, which substituted the turbo codes used in 4G LTE for data channels, and polar codes in 5G, which replaced convolutional codes used for control channels in 4G.

Modulation Carrying digital data through analog waveforms, modulation has the merit of transmitting a number of bits as a symbol, that number of bits is known as spectral efficiency and, it depends on the cardinality of the modulation scheme used. The total amount of data transmitted is increased by a factor of the spectral efficiency of the modulation scheme in the same available band as opposed to a modulation scheme with spectral efficiency of 1, in return, the signal loses robustness against channel impairments proportional to the spectral efficiency. QPSK and QAM modulations are very popular throughout communication systems.

The quality of the channel is estimated at the receiver side and communicated to the transmitter, which responses with proper coding and modulation. Depending on the feedback from the receiver, AMC tunes the coding and modulation used at the transmitter to achieve the highest possible rate. The transmitter selects lower modulation cardinality and lowers the code rate in case of poor channel condition, which is translated to more resistance against channel impairments and stronger error detection and correction capability at the cost of reducing the rate, meanwhile, if the channel condition becomes good, the modulation cardinality is increased, and a higher code rate is selected (closer to 1), and this is observed at receiver side as an increased data rate.

AMC Challenges For AMC to be effective and acquire the desired results, the feedback must be delivered to the transmitter fast enough to appropriately handle channel variations, in the scope of delay, if the feedback is slower than the changes which the channel is currently undergoing, then it is out of date. The RTT for LEO satellites is much more than that of terrestrial networks like 5G and, the RTT of GEO satellites is substantially larger, as a result, the multipath fading present in space links, which is much faster than half RTT, cannot be reported fast enough and, therefore AMC will not be able to keep up with fast fading, it is still able to track slower variations such as shadowing and free space loss variation.

AMC Possible Solutions AMC tries to remain steady on one appropriate modulation and coding scheme during fast fading by keeping a power margin, the margin to tolerate long delays for AMC control loops and power loops still needs to be defined through more studies. Using prediction techniques to avoid long delays is a promising idea, prediction−based link adaptation and prediction−based channel quality indicator reporting seem to enhance throughput, but they still require further study.

4.2 Handling Channel Variation After Transmission

Reducing the effects of the channel after the data is received is another way to handle channel variation. The receiver attempts to correct errors and/ or asks the transmitter to retransmit the data if they weren't received correctly.

4.2.1 Hybrid Automatic Repeat Request (HARQ)

A technique used at the physical layer that takes advantage of feedback and retransmissions to increase transmission resistance over channel variations and improve system reliability. At the receiver, if the data block contained no errors, the block is accepted. and an acknowledgment (ACK) is sent to the transmitter to confirm correct data reception, however, if the block did contain errors, HARQ attempts to correct those errors through coding error correction capabilities, if all errors are corrected, then the block is accepted, and an ACK is sent to the transmitter, otherwise, the block is not accepted, and a non−acknowledgment (NACK) is sent to the transmitter indicating that the block hasn't been received correctly and requesting data retransmission, therefore, HARQ has two techniques under its disposal.

Forward Error Correction (FEC) Uses coding capabilities to correct errors, FEC is a powerful telecommunications tool that corrects errors depending on the type of codes used and the selected code rate, also, it doesn't require feedback, which means maintaining low additional latency when compared to retransmission based techniques.

Automatic Repeat Request (ARQ) Exploits feedback and retransmission to increase the probability of correct reception. At the receiver, the data block is inspected for errors, if none are found, then the receiver accepts the block and sends an ACK, else, it rejects the block. Unlike FEC, ARQ does require feedback, thus, add large delays due to the propagation delay required for sending feedback and retransmission.

HARQ is implemented to increase link reliability at the cost of additional delay, HARQ RTT is a physical delay which cannot be avoided. For satellite links due to large distance, the RTT surpasses the maximum HARQ timer specified in 5G networks, that is, the time waited by the receiver before is can expect retransmission of the same data block, it also exceeds the minimum number of HARQ processes, which can be computed by Eq.(4.1).

$$N_{MIN,HARQ} = \frac{T_{HARQ}}{TTI} \tag{4.1}$$

$$T_{HARQ} = 2T_P + 2T_{slot} + T_{proc,UE} + T_{proc,BS} \tag{4.2}$$

Where $N_{MIN,HARQ}$ is the minimum number of parallel HARQ processes. T_{HARQ} which can be calculated using Eq.(4.2) is HARQ RTT, which is the time interval between sending a transmission block (TB) and receiving the corresponding ACK/ NACK. transmission time interval abbreviated TTI is the duration of 1 TB transmission, TTI is 1ms for 15kHz subcarrier spacing, and this value changes as subcarrier spacing change for 5G. T_P is propagation time, which depends on physical distance. T_{slot} is block transmission time, which depends on block size. $T_{proc,UE}$ and $T_{proc,BS}$ are the processing time for user equipment (UE) and base station (BS) respectively. The minimum number of processes for satellites is much more than the one required for the 5G network, where GEO satellites require approximately 600 HARQ parallel processes for terrestrial UE and BS link through GEO satellite, this number of processes has a huge impact on the satellite and UE buffer. The same applies to LEO satellites where the RTT is 15 to 63 times greater than terrestrial RTT, therefore extending the number of processes for terrestrial networks to accommodate the requirements of satellite links is unlikely to be seen for some UEs because of buffer and memory restraints, and due to the maximum number of available parallel processing channels. At the early stages of HARQ utilization in systems, the HARQ process operated in a method such that the first block to be transmitted is more protected (lower code rate and modulation cardinality), which meant a reduced number of retransmissions and, as a result, transmission delay and HARQ RTT are reduced as well, but, the HARQ process was treating each block independently and discarding those that were unsuccessfully decoded, this processing method is what was known as HARQ type−I. In type−II HARQ, the block, if decoded unsuccessfully, is now stored in a memory at the receiver so that in case the retransmitted block is also mis−decoded, it is combined with the stored block to attempt decoding the whole block successfully. The most important versions of type−II HARQ are Chase Combining (CC) and Incremental Redundancy (IR).

Chase Combining HARQ (CC−HARQ) Applies diversity gain to increase Signal to Noise Ratio (SNR) gain. The transmitter sends the TB, and in case it wasn't correctly decoded, the transmitter retransmits the same block upon the retransmission request, and the receiver adds them together and decodes.
first transmission

$$Y_1 = S_1 + N_1$$

second transmission

$$Y_2 = S_1 + N_2$$

the receiver combines them as

$$Y = (Y_1 + Y_2)/2$$

the result is SNR gain of 3dB.

Incremental Redundancy HARQ (IR−HARQ) Unlike CC−HARQ, when the first transmission fails, IR−HARQ doesn't retransmit the same block but instead changes the coding used to provide more redundancy, this provides both diversity gain and coding gain.
first transmission

$$Y_1 = S_1(1) + N_1$$

second transmission

$$Y_2 = S_1(2) + N_2$$

the receiver combines them as

$$Y = [Y_1, Y_2]$$

introduce coding gain.

HARQ Implementation

HARQ mechanism is currently used in many communication systems to augment their single transmission reliability and the spectral efficiency of the overall system by paying in return more complexity, buffering, and processing for the devices used, also, a delay is introduced as mentioned earlier in this chapter. Once the block is received, the receiver checks for successful reception, if so, it is ACKed otherwise, it is combined with retransmissions of the same or other versions of the same block to finally decode the correct original data, thus, augmenting the reliability of the transmission system by exploiting time and frequency diversity and coding gain.
For the HARQ process to work correctly, the receiver has to be able to verify the integrity of each received TB. The most used and reliable technique is Cyclic Redundancy Check (CRC), which after FEC and hard decision checks whether or not the block is correctly received. A less reliable technique that comes with reduced complexity is to make decisions based on statistics of soft−symbols before demapping or based on the statistics of soft−bits before or after decoding, this, can reduce the latency by reducing processing time this comes at the cost of decreasing the reliability since that it may indicate an incorrectly received block as successful, or decrease system throughput as a result of potentially indicating a correctly received block as unsuccessfully received.
After verifying the integrity of the block, the receiver sends feedback to the transmitter. The TB used in 4G and 5G NR is the base of data transmission, it contains all the bits to be transmitted within a transmission opportunity. TBs are segmented into Code Blocks (CBs) each of which is encoded separately. Out of possible feedback versions, the one with the least overhead is that which spends one bit in the feedback to indicate the entire TB reception outcome, however in case only some CBs have been wrongly received then, it would be a waste of resources to retransmit the whole TB, this is avoidable through using CB based feedback, for example, using one overhead bit for each CB which allows efficient resource usage by retransmitting only the necessary CBs, but on the other hand, it increases the overhead of the TB.
It is possible to implement HARQ in two manners, synchronous and asynchronous. Synchronous HARQ is implemented such that the time between transmission/ retransmission and the corresponding ACK/ NACK is fixed, for example, 4G uses synchronous HARQ in the uplink (UL) where the time between transmission and feedback is fixed to 4 TTIs which is equal to 4 ms with the maximum allowed number of retransmissions is set to 4. Asynchronous HARQ nevertheless, adds the merit of flexibility by enabling the BS to decide and set up the time between transmission and feedback within a given degree of freedom, and by doing so, asynchronous HARQ is more flexible and has higher adaptivity, this, however, comes at the cost of increasing the overhead because the BS has to communicate its decisions. Taking into account the resources used for transmissions/ retransmissions, another possible set of implementations arise, adaptive vs. non−adaptive HARQ. The latter forces the retransmission to be forwarded using the very same resources used for the initial transmission, also the transmission configuration such as used coding and modulation scheme must be the same, this manner has the advantage of lower overhead since the preconfigured techniques are previously known, on the other hand, it diminishes the choices of resources and configuration which means a lower degree of freedom[8].

HARQ in MAC Layer

HARQ is a technology stretching across physical and MAC layers. Regarding the MAC layer, HARQ operation consists of an entity that supervises several parallel HARQ processes separately for uplink and downlink (DL), each of which has a process identifier. The MAC entity used in DL contains a HARQ entity for each serving cell, and based on the inclusion or not of spatial multiplexing in the physical layer, the involved HARQ processes will support the transmission of one or two TBs. The transmission of data within a HARQ process is managed through information forwarded to it by the HARQ entity. The HARQ information contains New Data Indicator (NDI), Transport Block Size (TBS), Redundancy Version (RV), and HARQ process ID.
The HARQ entity in DL delivers to each HARQ process the received TBs through the physical layer and the related HARQ information belonging to it. According to the result of the received block decoding, the HARQ process forwards MAC Protocol Data Units (PDUs) for demultiplexing or updating the soft buffer with new data and in either way prepare the corresponding ACK/ NACK. System Information Blocks (SIBs) are managed using a dedicated HARQ process, where SIB1 and the other SIBs delivered by system information message are soft−combined through the HARQ process, an acknowledgment is not generated because those are broadcast blocks acquired by the UE utilizing Broadcast Control Channel (BCCH).
In UL, the MAC entity contains the HARQ entity for serving cells with uplink configured. It manages the UL grants received based on UL transmission, which is granted via Physical Dedicated Control Channel (PDCCH) or configured almost persistently through Radio Resource Control (RRC). The related HARQ information is delivered through the physical layer, and the HARQ entity recognizes the HARQ process corresponding to the grant and requests the HARQ process to forward a MAC PDU, UL grant, and related HARQ information to start a new transmission or in case of a failed transmission, generate retransmission. HARQ processes support only one TB linked to a HARQ buffer that is used to generate retransmissions, emptying the HARQ buffer of the process is the task of the HARQ entity[9].

HARQ implementation in 5G NR

Unlike LTE (Long Term Evolution), which uses asynchronous HARQ in DL and synchronous HARQ UL, 5G NR uses asynchronous HARQ in both DL and UL, thus, the multiple HARQ processes are handled in any order, DCI contains a HARQ process number which identifies the HARQ process, and since both the transmitter and receiver are well aware of each HARQ process number, they can manage HARQ processes even in case the processes are running out of order. DCI includes the HARQ process number only in DL in LTE because UL uses synchronous HARQ. In contrast, DCI includes process numbers in both links. NR also uses adaptive HARQ in both links (at least in eMBB and URLLC) as opposed to LTE where non−adaptive HARQ is used for UL.
The two generations of mobile communications networks follow different methods also for setting the timing between transmission and the corresponding response. LTE sets that time to 4 ms, and it is fixed, in contrast, 5G NR utilizes a more flexible method to set this timing through the collaboration between RRC and DCI, where RRC provides a table of possible timing between transmission/ retransmission

and its ACK/ NACK through an RRC message, from those, DCI states one timing from the communicated table, this process permits more flexibility and the possibility of latency reduction. The number of HARQ processes used is configurable in NR, if not, it is set to 8. That number is extended in Rel. 15 to 16. The minimum number of HARQ processes to accommodate the HARQ RRT can be computed as in Eq.(4.1).
5G NR supports HARQ feedback using a single bit for the whole TB, but, as mentioned, this turns out to waste resources by retransmitting the entire TB sometimes just for errors in few CBs, this issue manifests more clearly in 5G rather than 4G since the former has a larger size TB. To have more efficient resource usage and increase transmission flexibility, NR supports Code Block Group (CBG) transmission with one or more bits ACK/ NACK feedback. A CBG contains a number of CBs it can be as small as one CB or as large as an entire TB, this way, the ACK/ NACK feedback can be based on CBGs instead of TB, and, as a result, only the CBGs with corrupted CBs will be retransmitted instead of the whole TB. Using CBGs allows more flexibility in the retransmission operation and enables efficient use of resources, however, using multiple bits for HARQ ACK/ NACK feedback applies more overhead. The segmentation of CBGs inside of a TB is configurable through a higher layer[10]. The HARQ information needed by the HARQ process to function

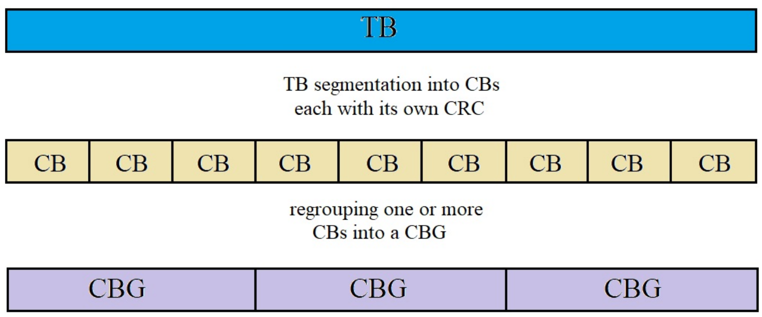

Figure 4.2: CB and CBG segmentation

correctly is embedded within the DCI, tables 4.1 and 4.2 indicate the different fields within HARQ information as specified by DCI format 1_0 and 1_1 in NR.

Both formats are used for scheduling of Physical Downlink Shared Channel (PDSCH) in one cell. Other formats like format 0_0 and format 0_1 are used for scheduling of Physical Uplink Shared Channel in one cell, formats 2_0, 2_1, 2_2 and 2_3 for notifying UE groups and Transmit Power Control (TPC) commands.

HARQ challenges

The number of processes has been extended to 16 in NR Rel. 15. For terrestrial networks, the number of processes serves the purpose, but, when considering NR NTN this number needs to be extended, even more, pointing to the fact that satellite transmissions have large delays, also, it should be flexible since the delays introduced by physical satellite links vary according to the type of satellite (GEO, MEO or LEO)

Table 4.1: DCI format 1_0

Field	Bits	Reference
DCI format identifier	1	Constantly set to 1, which means this is for DL
Frequency domain resource assignment	Variable	
Time domain resource assignment	4	
VRB−to−PRB mapping	1	0 : Non−Interleaved 1 : Interleaved
Modulation and coding scheme	5	
NDI	1	
RV	2	
HARQ process number (ID)	4	
Downlink assignment index	2	
TPC command for scheduled PUCCH	2	
PUCCH resource indicator	3	
PDSCH−to−HARQ feedback timing indicator	3	Possible k1 values from 1 to 8

and the angle of elevation. The minimum number of HARQ processes needed to accommodate satellite transmission is illustrated in table 4.3.

Extending the number of processes further would impact the UE in terms of memory. Also, the HARQ supported techniques and parameters already existing in terrestrial networks are impacted by the NTN large delays.

Possible Solutions Some possible principles can be further studies to resolve the addressed problem, one of which is to extend the number of HARQ processes to accommodate low and medium NTN RTTs. While for longer RTT delays, limiting the functions of HARQ and/ or disabling HARQ should be considered.

Table 4.2: DCI format 1_1

Field	Bits	Reference
Carrier indicator	0,3	
DCI format identifier	1	Constantly set to 1, which means this is for DL
Bandwith part indicator	0,1,2	
Frequency domain resource assignment	Variable	
Time domain resource assignment	4	
VRB−to−PRB mapping	0,1	0 bits if not configured by higher layers
PRB bundling size indicator	0,1	0 bits if not configured or set to 'static'
Rate matching indicator	0,1,2	Determined by higher layers
ZP CSI−RS trigger	0,1,2	
Modulation and coding scheme for TB1	5	
NDI for TB1	1	
RV for TB1	2	
Modulation and coding scheme for TB2	5	
NDI for TB2	1	
RV for TB2	2	
HARQ process number (ID)	4	
Downlink assignment index	0,2,4	4 bits if more than one serving cell are configured 2 bits if just one serving cell is configured 0 bits otherwise
TPC command for scheduled PUCCH	2	
PUCCH resource indicator	2	
PDSCH−to−HARQ feedback timing indicator	0,1,2,3	index of k1
Antenna ports and number of layers	4,5,6	
Transmission configuration indication	0,3	
Sounding Reference Signal (SRS) request	2	
CBG transmission information (CBGTI)	0,2,4,6,8	
CBG flushing out information (CBGFI)	0,1	
DMRS sequence initialization	1	

Table 4.3: The minimum number of HARQ processes ($N_{HARQ,MIN}$) for terrestrial and non−terrestrial network, assuming 1ms slot duration

Physical link	Max HARQ RTT (ms)	$N_{HARQ,MIN}$ for 1ms slot	UE feasibility
Terrestrial	16	16	Feasibile in Rel. 15 NR
LEO	50	50	Feasibile with HARQ extension
MEO	180	180	Impact on UE
GEO	600	600	Impact on UE

Chapter 5

Software Implementation and Key Parameters

Using HARQ in cooperation with the other techniques is the key to obtain good performance for the communications system and avoid wasting resources, the implementation of such techniques in the software and system performance evaluation process is explained in this chapter.

5.1 HARQ and ACM Implementation

HARQ along with ACM and power control, are interdependent techniques. ACM constantly changes the coding and modulation used for transmission to adapt to the changes introduced by the channel and provide a reliable communications link. As for power control, the adaptation to the time−varying channel is achieved through controlling the power level to maintain the required throughput. Clearly, both techniques will have to acquire the channel state information in time for the controlling and parameters changes to be effective, otherwise, if the channel state was inaccurately reported or it was outdated (the channel critically changed before the transmitter received the feedback), then the performance of ACM and power control will be impacted and, in turn, the system performance will be worsened. In such a case, the optimal solution is to set a margin to properly handle this situation considering the quality of the channel state reported, the margin optimally compromises the FER and the available capacity, this margin is practically translated into decreasing the size of the TB (or increasing it if the margin has a negative value) to force it to become likely within the capacity of the channel and therefore reduce frame error probability. HARQ is specifically effective in such a case since it enables coding rate control according to the real channel state using a loop of transmissions and feedbacks controlled by the receiver[11].
As mentioned in previous chapters, HARQ maintains the received information from one transmission to the next, storing them in the memory on the receiver side, this enhances the decoding process performance and reduces the latency of the HARQ control loop as it tends to allow faster decoding (fewer transmissions) by combining the stored information and the received one. The two HARQ categories take advantage of the stored information in different ways, CC−HARQ (chase combining HARQ) is retransmitting the same version of TB accumulating SNR over each transmission, on the other hand, IR−HARQ (incremental redundancy HARQ) transmits

a different redundancy version (RV) reducing the code rate with each retransmission, this has the effect of accumulating the mutual information of the TB. As a rule, IR−HARQ is superior to CC−HARQ, this advantage though becomes negligible at the condition of low SNR.

5.2 Terrestrial−satellite Channel State

The NTN physical link is one with a significant distance, which results in large attenuation in signal strength and large delays, other than that, the link is subjected to the effects from the earth's upper atmosphere especially, the ionosphere. Also, the parameters of the receiver (SNR, speed, and whether or not there is LOS) affect the channel variations as seen in figures 5.1, 5.2, 5.3 and 5.4.

For a receiver with low mobility and has a LOS with the satellite, the channel

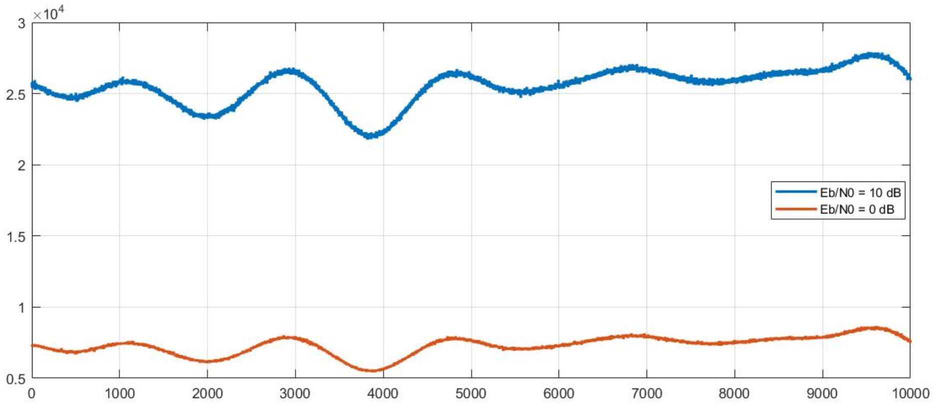

Figure 5.1: Channel capacity variation, LOS, speed = $1km/h$.

seems fairly stable and changes slowly after many slots, the available capacity is much more in the case of 10 dB SNR when compared to that provided by 0 dB SNR as shown in figure 5.1.

As the speed of the receiver increase, so does the channel variations, figure5.2 illus-

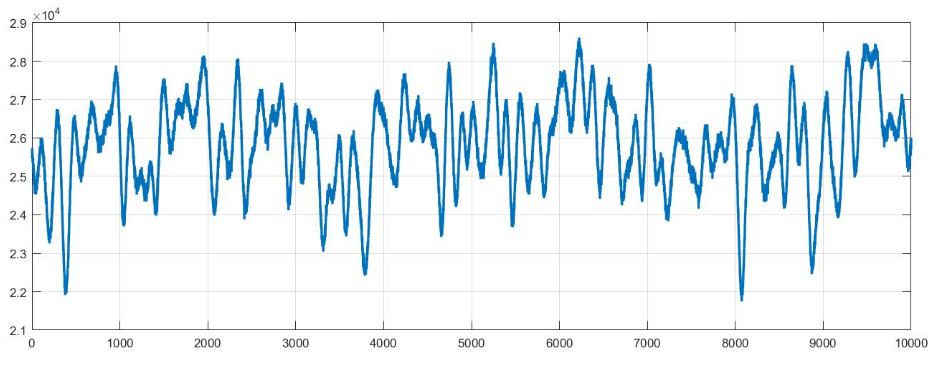

Figure 5.2: Channel capacity variation, LOS, speed = $10km/h$.

trates channel variation in LOS condition with a speed of $10km/h$, which exhibits strong variation when compared to the previous case.

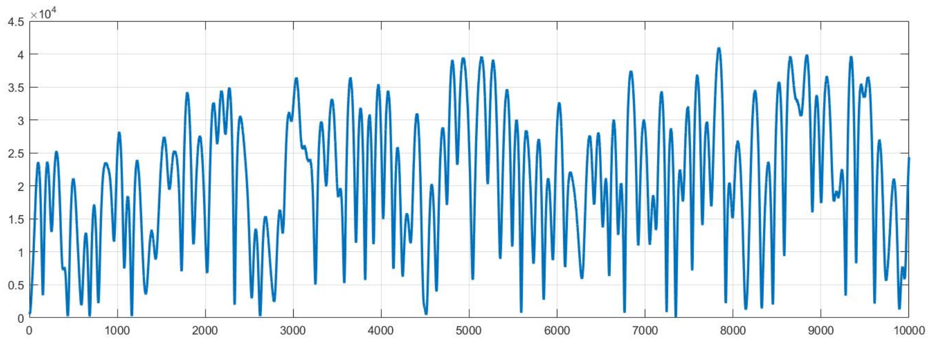

Figure 5.3: Channel capacity variation, NLOS, speed = $10km/h$.

In the NLOS scenario, the channel's behavior becomes more aggressive, the channel capacity changes larger in magnitude, and faster in time, this would make the channel less predictable and increase the impact of outdated feedback.

The worst scenario is when the receiver moves with high speed while it is in NLOS

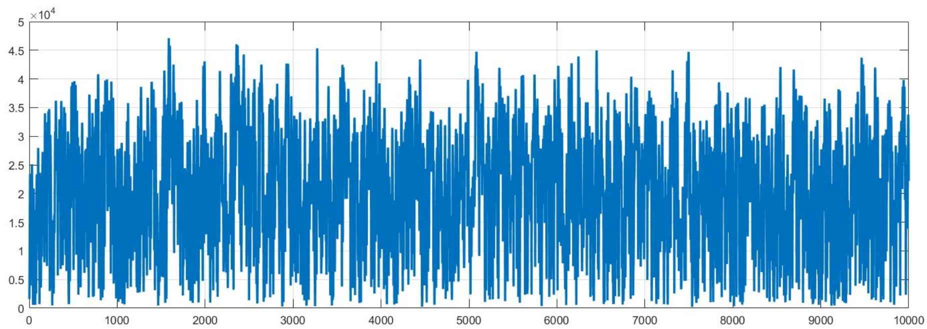

Figure 5.4: Channel capacity variation, NLOS, speed = $100km/h$.

from the satellite, as seen in figure 5.4, the changes are critical and, they occur extremely fast, these conditions will pose a challenge to the prediction process, inaccurate prediction, in this case, can only mean bad performance. The introduction of HARQ proves substantial usefulness in this case, particularly since the performance of single−shot transmission will likely result in low system performance, HARQ can improve the performance of the system under these conditions as its control loop helps to handle the channel better.

5.3 Software Procedure

Upon assigning the desired parameter values (among which the most important are the speed, elevation angle, channel type, RTT, and SNR), the software loops over the determined range of RTT and another over the range of SNR values. During each slot, the following steps take place:

1. Acquire the allocated resources per slot, and the maximum allowed number of transmissions.

2. Set the number of transmissions, number of transmitted blocks and bits, number of received blocks, and bits to zero.

3. Check HARQ process status through the feedback, if the feedback is an ACK or empty, a new HARQ process is allocated with the initial RV, otherwise (NACK), a new RV value is determined as long as the maximum number of transmissions is not reached.

4. For a new HARQ process, the TB size to be transmitted is determined via predicting the channel state (Mutual Information (MI)) and selecting optimal margin.

5. Transmit the TB over the allocated resources.

6. Compute FER and CSI.

7. In case the TB is correctly received, the values of transmitted and received number of blocks and bits are augmented, after which, if maximum transmissions are reached, the feedback is emptied, if not, an ACK is returned.

8. For block reception failure, if maximum transmissions are reached, only transmitted blocks and bits are augmented, and the feedback is emptied, otherwise, a NACK is sent back as feedback.

9. The number of transmissions, blocks, and bits are returned.

10. Repeat for the next slot.

For a TB of size K transmitted over a number (n) of independent Gaussian channels, the FER can be approximated using Eq. (5.1).

$$FER \approx Q\left(\frac{I-K}{\sqrt{V}}\right) \tag{5.1}$$

The equation is acquired under the assumption of ideal Gaussian distribution input, which means other performance decreasing factors such as the suboptimality of the coding process and receiver are not considered in the computation. Where I and V are the aggregated mutual information and dispersion over n independent channels, respectively.

$$I \triangleq \sum_{i=1}^{n} c(x_i) \tag{5.2}$$

$$V \triangleq \sum_{i=1}^{n} v(x_i) \tag{5.3}$$

$$c(x) \triangleq log_2(1+x) \tag{5.4}$$

$$v(x) \triangleq log^2(e).\left(1-\frac{1}{(1+x)^2}\right) \tag{5.5}$$

Where x is the SNR of the channel. The FER equation holds for incremental redundancy HARQ communication systems as well. In step 6 of the procedure, FER is computed, if not measured during simulation, it is approximated for the k transmission as follows:

$$FER_k \approx Q\left(\frac{I_k-K}{\sqrt{V_k}}\right) \tag{5.6}$$

Where I_k is the aggregated MI and V_k the aggregated dispersion over the allocated resources up until this point. The TB size K is compared against the MI of the slot, if less, then the block should be received correctly, if not, the MI from the next transmission is aggregated and K is compared against the sum of the MI until the block is through or the maximum number of transmissions is reached and in this case if K is still larger there will be a frame error. This is an advantage of using HARQ as compared to using a single transmission system. The figure 5.5 illustrates the K value against slot MI.

For example, the FER for the first transmission and HARQ retransmission are:

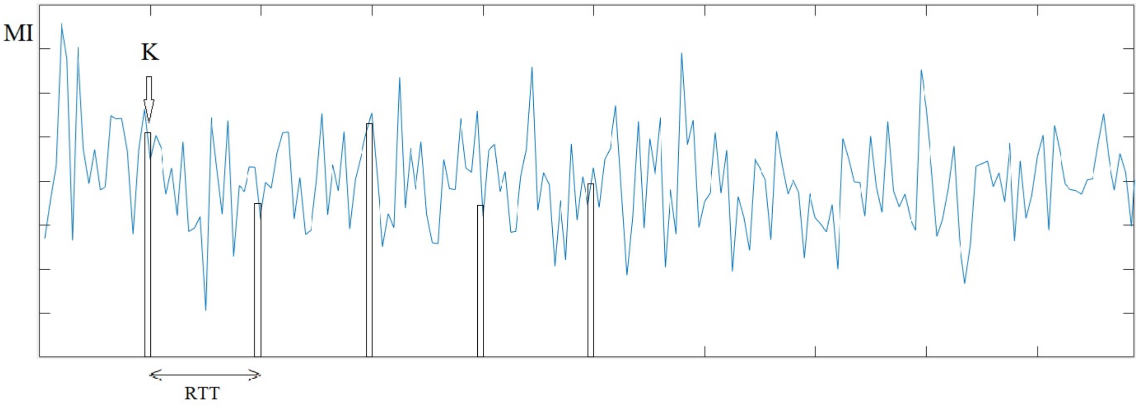

Figure 5.5: TB size impact on channel MI

$$FER \approx Q\left(\frac{I_1 - K}{\sqrt{V_1}}\right), FER \approx Q\left(\frac{I_1 + I_2 - K}{\sqrt{V_1 + V_2}}\right) \tag{5.7}$$

In the first term, if the transmission satisfied the required FER then the transmission is deemed successful for both single transmission system and HARQ system, on the other hand, if the transmission failed to meet the required value then, for single transmission that is a failed transmission, while the HARQ system uses retransmission which in IR−HARQ is seen as aggregating the capacity from the first and second transmissions, that increases the probability of achieving the desired value of FER. A symbol level simulation of the allocated resources and channel type allows MI and dispersion sequence creation at slot level.

$$I_t = \sum_{i=1}^{n} c(x_{t,i}) \tag{5.8}$$

$$V_t = \sum_{i=1}^{n} v(x_{t,i}) \tag{5.9}$$

Where i and t are resources and slot indices. Escalating the process to a number (k) of slots used in a HARQ running process, the MI and dispersion are computed as:

$$I_k = \sum_{j=1}^{k} c(x_{tj}) \tag{5.10}$$

$$V_k = \sum_{j=1}^{k} v(x_{tj}) \tag{5.11}$$

The mutual information and dispersion over slots can be used as channel state information to be exploited in the prediction process instead of more common parameters such as SNR. The prediction process mentioned in step 4 uses a single tap filter with a bandwidth corresponding to the delay $D/2$ which is the time elapsed before the transmitter acquire the channel state, the filter does not need second−order channel statistics, it does increase the ability of prediction through removing high frequencies from I_t, the mathematical representation of the filter is:

$$\hat{I}_t = \hat{I}_{t-1} + \frac{2}{D}(I_{t-D/2} - \hat{I}_{t-1}) \tag{5.12}$$

Where, $\hat{I}$ is the predicted MI, D is the RTT delay. The TB size is then set according to the following equation:

$$K = [\hat{I}_t + n.M] \tag{5.13}$$

The margin M is chosen through these steps:

$$f = E(Q(x + FB - MI)) - TFER \tag{5.14}$$

The target frame error rate (TFER) is set among the initial parameters, while FB is the feedback, and the Q function is approximated by the step function in the software. After finding the zeros in the resulting values f, a proper range of margin values is found, from which the value that maximizes the throughput is selected.

5.3.1 Performance Indicators

There is more than a way to measure the performance of a communications system, the indicators measured in the software are:

BER evaluated as the number of not received bits over the number of transmitted bits. $BER = \frac{K_T - K_R}{K_T}$, where K_T is the number of transmitted bits, K_R is the number of the received bits. The BER is a measure of how many bits were received incorrectly compared to the total number of bits transmitted, it is a very common and popular performance indicator in the field of communications systems.

FER the number of not received frames over transmitted frames. $FER = \frac{B_T - B_R}{B_T}$, where B_T is the number of transmitted frames, B_R is the number of the received frames. FER indicates the amount of unsuccessfully received frames taking into consideration the total number of transmitted frames, it is similar to BER but works on frames instead of bits. The lower the FER value, the better the system performance, the same applies to BER.

Goodput received bits over used resources. $T = \frac{K_R}{N}$, where N is the number of employed resources.

Throughput Efficiency received bits over channel capacity (Shannon bits). $T_e = \frac{K_R}{\sum I_t}$, where I_t is the MI per time slot. The throughput efficiency is a measure of how well the channel is exploited, in other terms, the number of bits correctly delivered compared to the theoretical limit. Obviously, the higher the throughput efficiency the better. T_e is used as the main performance indicator in this report.

Average Latency time taken for transmitted data until the reception. $\bar{L} = D \times \frac{N}{n.B_T}$ slots.

The performance indicator used for the system performance assessment in this report is the throughput efficiency $T_e = K_R/\sum I_t$ where K_R is the number of received bits and I_t is the Shannon bits for the used resources. The throughput efficiency of the system is governed by the targeted FER, by accurately using prediction, selecting optimal margin, and determining the proper TB size, the system can achieve the maximum possible throughput under the constrains of desired FER value and suboptimality of the prediction process. Another way to select the margin is to perform another loop for a range of margin values, doing it this way, one would observe the diminishing values of both FER and throughput efficiency as the value of the margin increases, this is quite understandable since a larger margin results in being cautious and setting a smaller TB size value which in turn leads to transmitting and receiving fewer bits, the reduced number of received bits will also reduce the throughput efficiency since the two are proportional. On the other hand, being ungenerous with the TB size increases the probability of receiving the block without errors, therefore, if HARQ is disabled, it would be a good idea to choose a sufficiently large margin for better reception chances, otherwise, a low margin is advised to exploit the resources and depend on the HARQ process to achieve the targeted error threshold.

Chapter 6

3GPP New Radio SW Implementation and Performance Assessment Satellite Networks Software

The software is implemented to simulate the behavior of NR 5G network communication through satellite networks. Taking into account the channel models of satellite networks and its distinctive parameters such as long RTT, the software can simulate an integration of 5G NR and satellite networks.
The structure of the software is constructed using various examples from the 5G toolbox section in the MathWorks website, so obviously, the software is implemented in a MATLAB environment. Many "advanced" toolboxes functions are utilized in the software which internally resorts to "ordinary" functions, therefore, to understand the explicit body of the software, one would do well to comprehend the task of the "advanced" functions used[12].

6.1 Software Requirements

The software was originally implemented within MATLAB 2018b release and in 2019a release later, it depends on the 5G toolbox in the following form:

- 5G toolbox 1.0 (2018b release) for PDSCH, PDCCH and SSB simulation.
- 5G toolbox 1.1 (2019a release) for PUSCH simulation.

However, the version used for this thesis requires at least MATLAB 2019b release to work without problems due to the fact that some options used were not configured in earlier releases. Other than that, the software requires the following toolboxes to be installed:

5G Toolbox This toolbox supplies functions and examples for simulation and modeling of 5G NR communications systems. It provides the means for link simulation and other verification and testing functionalities[13].

Communications Toolbox Supplies functions for communications systems analysis, design, and simulation. The toolbox contains the most important functions to simulate communications systems' physical layer, some of which are channel coding, modulation, and other essential functions, also the most popular output forms of communication systems like, constellation, eye diagrams, BER, and many analysis tools are supported within this toolbox.

DSP (Digital Signal Processing) Toolbox Provides functions and applications for analyzing and simulating signal processing systems. DSP applications for radar, audio, IoT, and many communications applications can be modeled using this toolbox, it also supports other tools such as FIR, IIR, and other filters, as well as scopes to analyze them like time scope and spectrum analyzer.

Signal Processing Toolbox supplies functions and applications to analyze and process samples signals. Tools for resampling, filter design, power spectrum estimation, and other related purposes are supported in signal processing toolbox. It also provides the means for measurements of SNR, distortions, and other parameters.

6.2 Main Features

PDSCH/ PUSCH Simulation uplink and downlink shared channels simulation using time and frequency resources for slot−based transmission, in which, data payload, pilot symbols, and synchronization burst are configurable according to 5G NR standards.

Physical Control Channel Simulation forcing the control channel bits to fill up the entire slot, the performance of control channel and polar coding used can be simulated and evaluated through measuring parameters such as BER.

Transport and Physical Channel Level Processing takes advantage of 5G toolbox nrPDSCH and nrPUSCH encoding and decoding functions in order to implement physical channel level processing which includes modulation, scrambling, and layer mapping. While transport channel level processing such as data segmentation, CRC, and LDPC encoding and rate matching, are implemented in DLSCH and ULSCH encoder and decoder objects.

Channel Estimation can be performed by taking the channel parameter directly from channel generating function, this, of course, is not possible in practical communications systems. Instead, channel estimation using pilot symbols is the method used in practical systems, it is also an alternative way for channel estimation in the software. After either method of estimation, equalization is performed.

Satellite Payload uses nonlinear amplifier and sampling and multiplexing options to implement satellite payload.

Phase Noise Tracking by using one or more subcarriers as pilot symbols, what is known as phase tracking reference signals can be implemented to track the phase noise and measure phase noise error for medium or high SNR values.

Synchronization through SSB bursts, the synchronization of satellite links can be investigated. The receiver searches for synchronization blocks within half a frame, in case the signal is synchronized correctly, the DCI block is decoded which was encoded using polar codes.

6.3 Software Architecture

The software is built using various functions and files, from which, the most important are:

NR_params The parameters of the simulated communications system are adjusted within this file. It contains the parameters for Monte Carlo simulation, used channels (PDSCH, PUSCH, PDCCH or SSB), BS and UE, pilot symbols, and HARQ. Some of the parameters set in this file for counting but not limited to:

- Simulation length and SNR parameters such as, the maximum number of bits and errors, SNR type, and channel estimation method.
- Cell ID, Tx, and Rx number of antennas, cyclic prefix, and symbol rate for gNodeB configuration.
- Target code rate, modulation cardinality, and the number of layers of PDSCH/ PUSCH.
- HARQ and ACM enable/ disable options.

Other parameters such as RTT, number of frames, RV for HARQ, subcarrier spacing, speed, and channel type are set and can be adjusted at the beginning of NR_main file.

NR_initialize The initialization of the variables is done through this file. The file uses the parameters previously set in NR_params file to prepare the variables and objects which will be used later for the simulation. Some of those variables can be altered directly such as channel parameters and HARQ sequence type. In general, this file shouldn't be adjusted except for software upgrades.

NR_main Simulation execution is performed by executing this file. As the name suggests, this file is the main file in the software in which other files and functions are utilized. It generally consists of some parameters' initialization followed by two loops; the outer represents SNR values change while the inner is performed around the slots to be transmitted. The simulation stops once the maximum number of bits or errors is reached and afterward, the results can be displayed.

NR_step The processing is handled in this file. HARQ processes and SSB burst generation is done within this file as well as generating modulation symbols mapping grid, calls physical and transport−level functions for channel processing and performs channel creation, and at receiver, it inverts the process using estimated channel parameters (ideal or simulated).

NR_TX It is responsible for the transmitter side operations. The file calls the necessary functions to perform channel coding and generating the transmission signal waveform, according to the feedback received regarding the HARQ process and channel information, the modulation and coding for ACM are set, also, the transmitter selects the proper action to perform depending on the HARQ feedback in order to generate the TB for transmission.

NR_RX This file handles the operations performed by the receiver, namely, demodulation, decoding, channel estimation, and CSI related processes, HARQ and ACM receiver side processes including generating feedback and ACK/ NACK. the main function assumes knowing additional channel information such as noise and path gains as they are extracted directly from the channel generation process (this is not the case in practical systems).

Chapter 7

Results and Discussion

The system has been simulated various times applying different scenarios, namely, changing channel type (A, B, C, D) for LOS and NLOS cases, adjusting the equipment speed, and setting a target frame error rate. The performance of the system is indicated through the throughput efficiency parameter, computed over a range of RTT values, and a range of SNR values for a set of possible HARQ process number of transmissions.

7.1 Prediction

RTT values for satellite communications are substantial, the ordinary HARQ process would suffer given the fast−changing channel compared to the RTT. Depending only on the feedback with such delay means that the system would have to rely on outdated information to tune the proper parameters for transmission which is as if the system is navigating blindly. One way to defeat the outdated feedback problem is to use prediction, taking into account the previous channel information and apply it to foretell what would be the channel condition after a given delay, if done correctly, the transmitter can set the befitting values of the parameters to exploit the channel without risking high error probability, However, if the predicted value is far from the actual one (poor prediction), the system performance will deteriorate.

Software Model

The prediction process integrated within the software is a single tap (one pole) filter, the filter uses only the previous MI value to forecast the upcoming slot MI using the parameter alpha to tune the prediction process. The filter removes high−frequency components limiting to a bandwidth related to the lag $D/2$ assuming the CSI can be observed at the transmitter after that specific lag. The general process of prediction is illustrated in figure 7.1. For small RTT, the prediction process is much simpler since the channel wouldn't have changed much by the next slot, the prediction might as well be removed for very small delays, as for such RTT, the feedback would not be outdated. Nevertheless, the prediction is essential when it comes to large RTT values, the prediction becomes that much more difficult though.

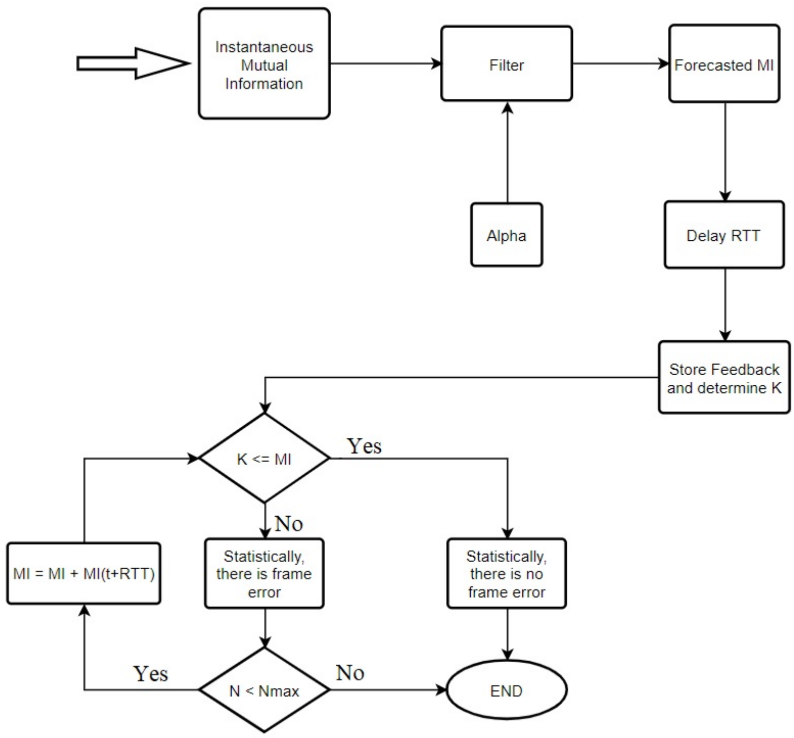

Figure 7.1: Prediction process flowchart

Proposed Non-linear Model

System identification toolbox provides various linear and nonlinear models to be considered when attempting to predict channel variations, since the values to be predicted are following a nonlinear behavior $I = log(1 + \frac{|E_s|^2}{N_c})$, using a linear model to forecast the MI directly is not a good fit, instead, one could use a linear model to predict the value E_s and apply the formula afterward. The values otherwise could be predicted directly using a nonlinear model. By comparing the performance of some models, such as, autoregression and state of space models using MSE as an indicator, the best performance model fitting the MI provided data was a nonlinear autoregressive model with exogenous variables (ARX), ARX models are known for their predictive capability and stability. The command used to estimate the parameters of the nonlinear ARX model is *nlarx*. The model is trained using a set of MI values to estimate the parameters, afterwards, the models are used to forecast the channel variations and better set the parameters for the next transmission.

The structure of the nonlinear model contains model regressors which evaluate the

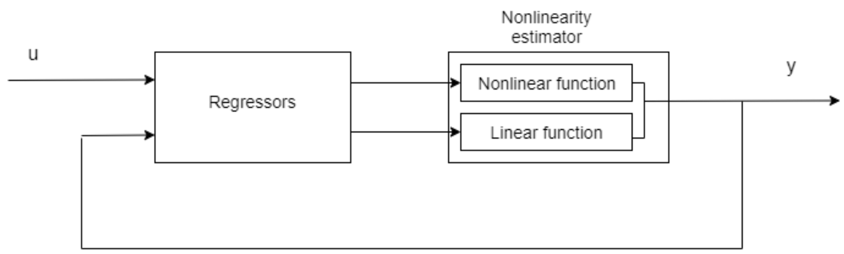

Figure 7.2: Nonlinear ARX channel prediction model

regressor parameters through the previous input and output values, the regressors are used as an input to the nonlinearity estimator which consists of nonlinear block and linear block in parallel, those are then used to estimate the output as illustrated in figure 7.2.

Proposed Linear Model

One of MATLAB prediction tools is the linear prediction filter. The coefficients of the filter can be computed using the 'lcp' command, the FIR filter can then be used to forecast the upcoming time series value based on the past values. The model

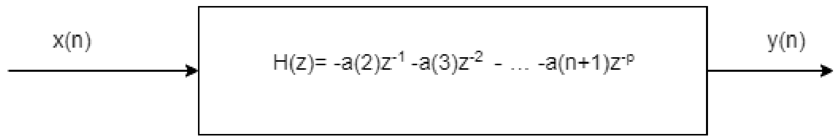

Figure 7.3: Linear channel prediction filter model

computes the parameters for the forward linear predictor through the autocorrelation approach of an AR model and is computed in such a way to minimize the MSE of the prediction process.

Deep Learning Model

Using a deep learning network consisting of an input layer, 100 hidden units lstm layer, fully connected layer, and a regression layer as seen in figure 7.4, the network is trained to predict the time series MI values. From figures 7.5 and 7.6, it is

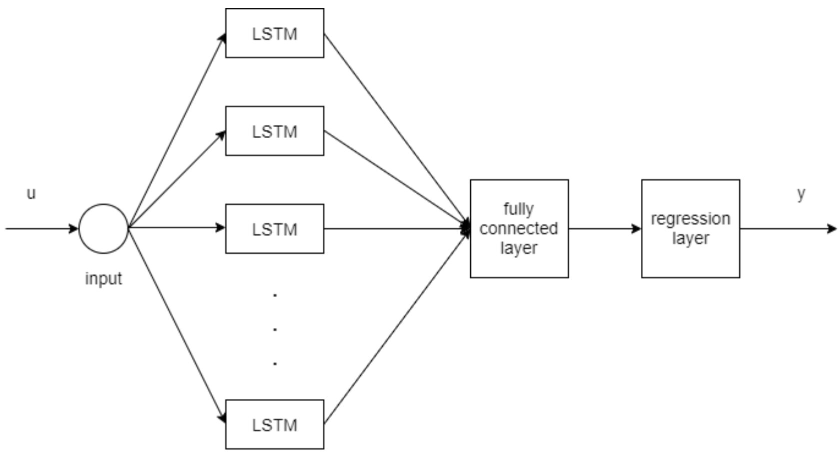

Figure 7.4: Nonlinear ARX channel prediction model

clear that the training process for higher speed channels is more difficult, since the channel is changing more rapidly.

7.2 Simulation Results

Simulating the system by changing the mentioned parameters and experiencing different scenarios, and using a set of prediction models, the following results are

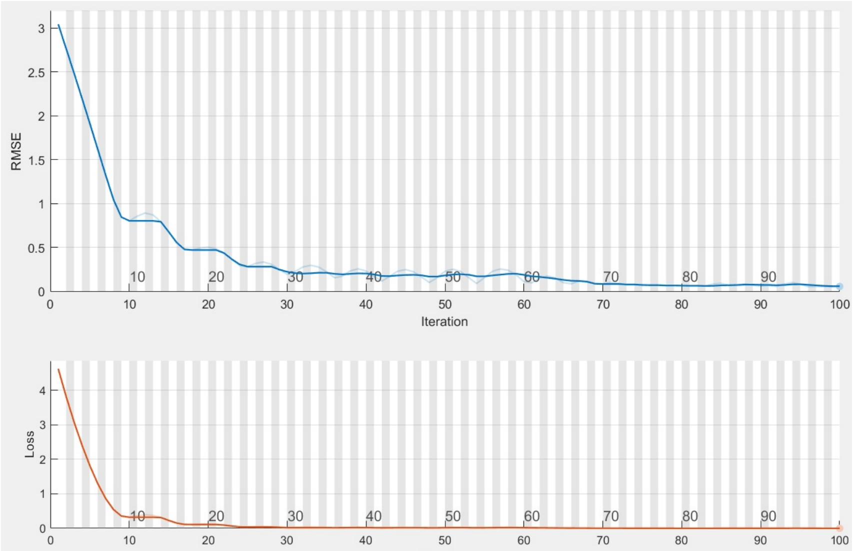

Figure 7.5: Deep Learning model training for $10km/h$ speed, NLOS channel

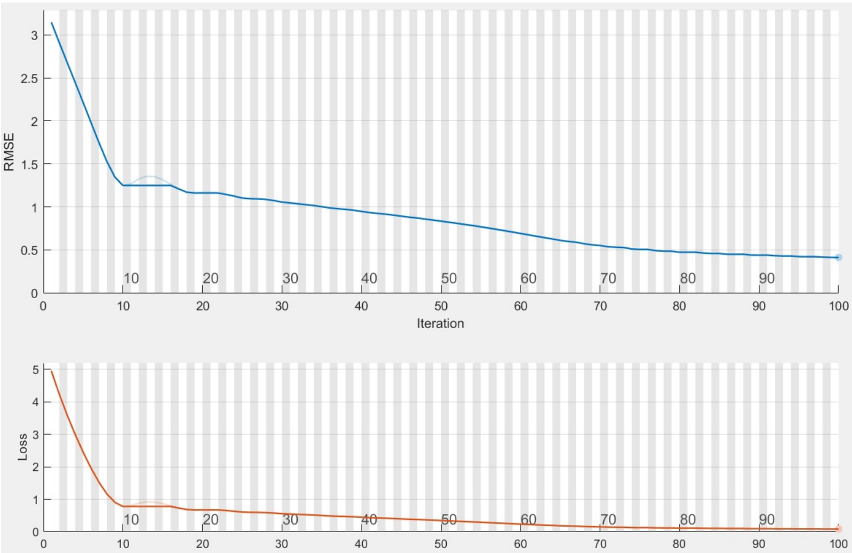

Figure 7.6: Deep Learning model training for $100km/h$ speed, NLOS channel

obtained. As observed from figure 7.7, for a NLOS scenario and low speed ($1km/h$) the performance of the system is almost indistinguishable for 1, 2, or 4 transmissions and E_s/N_0 of 0, 5, 10, and 15 dB at low and mid values of RTT all around 90%. Only at RTT of 52 ms or above the difference of performance becomes noticeable, this is because at high RTT values the feedback becomes outdated and the signal is more likely to experience channel variations along the way. At higher mobility, the advantage of using HARQ becomes manifest as the throughput efficiency is higher as the number of transmissions used increases, it is only natural that the performance is better with increasing SNR, but, the key point here is the superiority of HARQ and the throughput efficiency boost it introduces to the system which reaches 70% at one point. For a speed of ($100km/h$), single−shot transmissions ($Ntx = 1$) denoted by red curves in figure 7.9 show poor performance with less than 15% throughput efficiency for RTT equal or above 13 ms and less than 40% at all times. The performance of the system that uses HARQ with 2 transmissions ($Ntx = 2$) seen as green

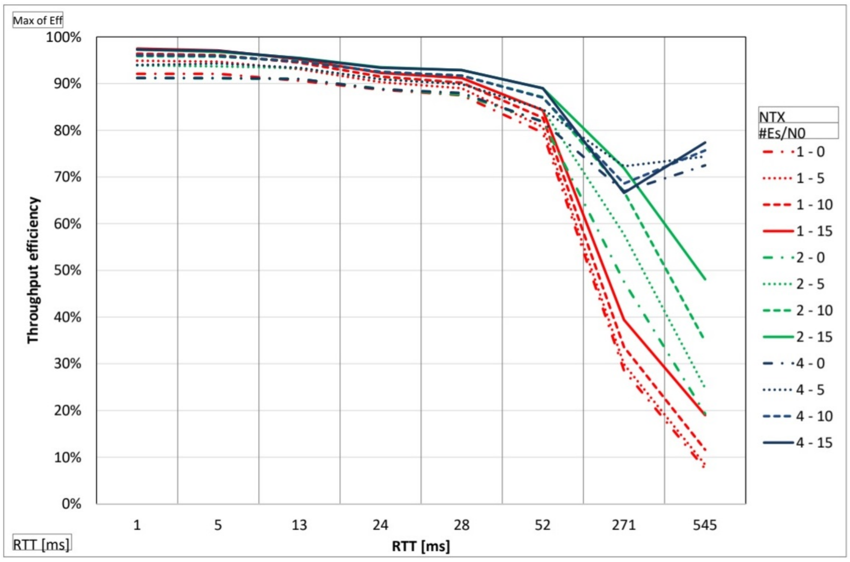

Figure 7.7: Throughput efficiency vs RTT. NTN-CDL NLOS channel, speed $1km/h$, FER $\leqslant 10^{-2}$

curves varies between 10% and 60% depending on the SNR value, while the blue curves show the performance of HARQ equipped with 4 transmissions and throughput efficiency of around 70% or higher for SNR of 5 dB or more. Considering the highest simulated SNR, HARQ with 2 transmissions increases the throughput efficiency by approximately 45%, and HARQ using 4 transmissions adds around 25% on top of that. Instead of fast changes in NLOS channel, the channels are steadier and, the performance of the system is roughly the same for the different number of transmissions and SNR in a LOS channel, as seen in figures 7.10 and 7.11, the throughput efficiency is high in all of the cases with slightly lower performance for 0 dB 1 and 2 transmissions.

7.2.1 Simulated Performance

The performances reported earlier are all acquired through statistical computations performed at the transmitter once the feedback and the transmission parameters are known. The full simulation instead, performs the transmission and applies channel effects, and at the receiver, it performs the decoding and demodulation, thus, the throughput efficiency and error calculations are more practical, it worth mentioning that the simulation process will take much longer (sometimes up to several hours).
From figure 7.12, the general picture is the same with HARQ raising the throughput efficiency as it uses more transmissions, however, comparing the simulated and computed performances, it is noticed that simulated throughput efficiency is slightly below that of the computed one, this firstly verifies the validity of using the computed results, and secondly, illustrated the suboptimality of the actual processes such as coding which is noticed as the slight difference between the two performances where the computed throughput efficiency acts as an upper limit to the simulated one.

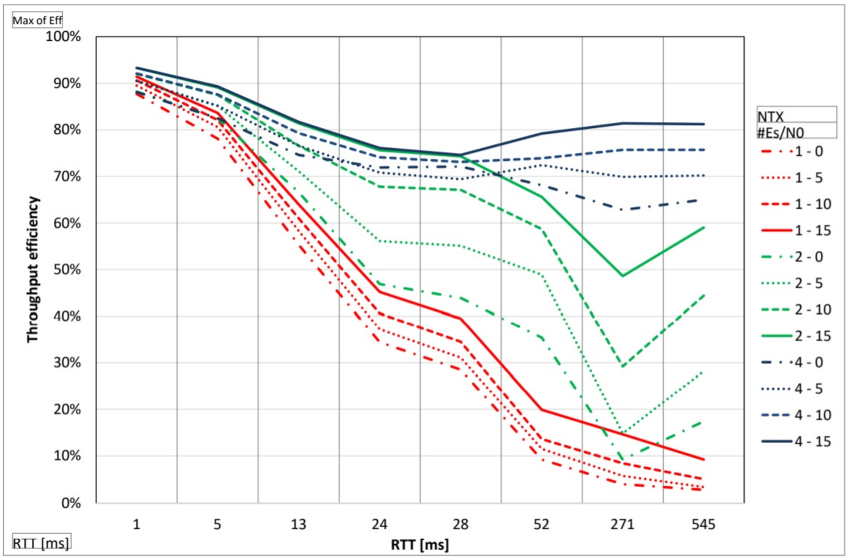

Figure 7.8: Throughput efficiency vs RTT. NTN-CDL NLOS channel, speed $10km/h$, FER $\leqslant 10^{-2}$

7.2.2 Prediction Performance

The performance is also affected by the prediction model used, the previous results are obtained using a one−tap filter, which uses the preceding value along with a tuning factor to forecast the upcoming one. One can argue that the performance of the system can be enhanced using a more sophisticated model, MATLAB provides several models mainly in the system identification toolbox and econometrics toolbox. Depending on MSE performance, three models have been chosen, a nonlinear ARX model, a linear model, and a deep learning model.
The nonlinear model is set up using the 'nlarx' MATLAB function, which uses a vector of data to evaluate the parameters of the model. the performance of the system using the proposed model as a prediction means is reported in figures 7.13 and 7.14.

Looking at figure 7.13, it is clear that the difference in performance is not fixed, nevertheless, it is unquestionable that the throughput efficiency produced by the system using the nonlinear model is better than that with the one−tap model, the nonlinear model enhances the performance by 15% at one point other than that it is 10% or less, the boost in performance is not huge but, at this point, it is already difficult to introduce a slight improvement. The performance of the 2 and 4 transmissions (green and blue curves respectively) based HARQ systems are comparable up to an RTT of 28 ms, only at larger RTTs that the power of 4 transmissions appears, where the 4 transmissions HARQ system maintains a throughput efficiency of approximately 40% higher than the performance of the one using 2 transmissions. Unlike the NLOS channel, the difference between performances of the same model is very close let alone different models, figure 7.14 shows the performance of the two models over the LOS channel. The throughput efficiency from the two models are almost the same and they both achieve high performance, the reason being that the LOS channel is changing slowly and almost constant and the upcoming MI values are nearly the same as the present ones, thus, predicting is easier and the parameters can be tuned accurately to achieve high throughput.

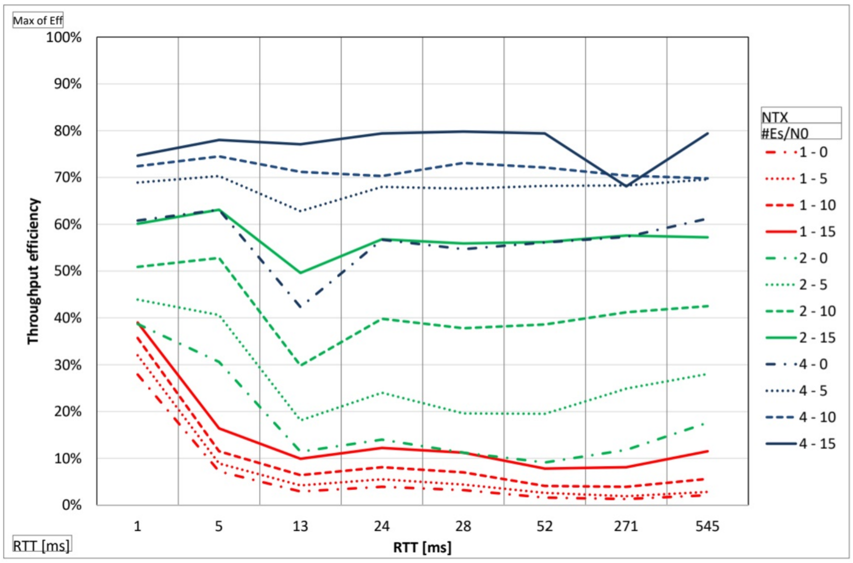

Figure 7.9: Throughput efficiency vs RTT. NTN-CDL NLOS channel, speed $100km/h$, FER $\leqslant 10^{-2}$

Figure 7.15 shows the values the throughput efficiency with respect to the margin, the margin values are set to compromise the targeted FER and the maximum achievable throughput efficiency. Denoted in the figure, the two margin values required to achieve the targeted FER (the values are not optimal but very close), the blue bar corresponding to the one−tap filter HARQ system shows that a margin of 1.8 is needed for the system to achieve the desired FER (in this case 10^{-2}), and the resulting throughput efficiency under this FER constrain was 34%, on the other hand, the nonlinear model HARQ system required only a margin of around 0.8 to reach a throughput efficiency of 39% which is slightly higher than the one achieved by the one−tap model under the same FER condition. According to this, a HARQ system with a better prediction model requires a lower margin to stay under the same FER threshold, the lower margin is then translated to a higher throughput efficiency since a smaller margin means a larger data frame.

As seen in figure 7.16, the ARX nonlinear model is constantly better regardless of the channel condition. The performance is slightly lower for A channel type of $100km/h$ speed (green curve) for low RTT values, but at higher RTT it becomes slightly better. As for other conditions, the proposed model enhances the performance of the system.

Figures 7.17 and 7.18 display the performance boost gained from using a more suitable linear filter than the filter already in use, the performance is slightly better in the case of 4 transmissions, however, it is more apparent in 2 and 1 transmissions cases where the gain reaches 25% at some points. Considering the performance for an FER of e^{-3}, the gain in case of 4 transmissions is better (around 10%) and is substantially better in the case of 2 transmissions where it surpasses 40% at certain RTT values.

Looking in−depth at the two figures 7.17 and 7.18, the performance of ACM with a better prediction model is quite recognizable, especially in the cases of 1 and 2 transmissions (red and green curves), this is natural since better prediction means more accurate estimate, and, therefore, the coding rate and modulation can be set

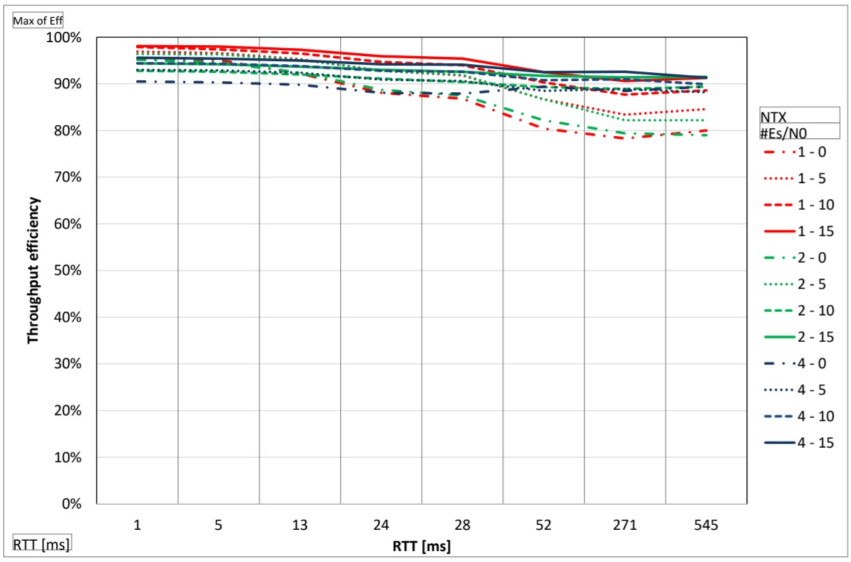

Figure 7.10: Throughput efficiency vs RTT. NTN-CDL LOS channel, speed $10km/h$, FER $\leqslant 10^{-2}$

more accurately by the ACM process. Another thing can be observed from the two figures, which is the performance difference between the two models as a lower FER is targeted, where, in the case of 4 transmissions, the grows from around 2 or 3% to approximately 10%, while in the case of 2 transmissions, the gap widens from around 15% to 30 and sometimes 40%, which indicates that the 4 transmissions are doing a better work compensating for the less accurate prediction than the 2 transmissions HARQ system.

The deep learning model performance also displays an improvement where the throughput efficiency is $10-25\%$ higher most of the time for an FER of e^{-2}, as illustrated in figure 7.19. And for an FER of e^{-3} the performance is further enhanced as shown in figure 7.20.

Unlike the other models, the deep learning model does not resort to computing the autocorrelation of the time series, despite that, the neural network is doing much better than the one−tap filter, as seen in figure 7.19.

Comaring the results obtained from the two different FER values using the linear and deep learning models (figures 7.17, 7.18, 7.19, and 7.20), it can be noticed that the performance gained by using HARQ in the terrestrial−nonterrestrial link grows as we aim for lower FER.

Comparing the three proposed prediction models as illustrated in figure 7.21, one can observe that among the models the one with the worst performance is the nonlinear model (it is still better than the one−tap filter though) where both models have a very slight superiority in case of 4 transmissions, while in the other two cases they are around 10% better than the nonlinear model. The linear model and the deep learning model have a very similar performance with the latter being slightly better, deep learning model also has the advantage of being faster at the training and predicting process since it doesn't need to compute the autocorrelation of the time series (practically speaking).

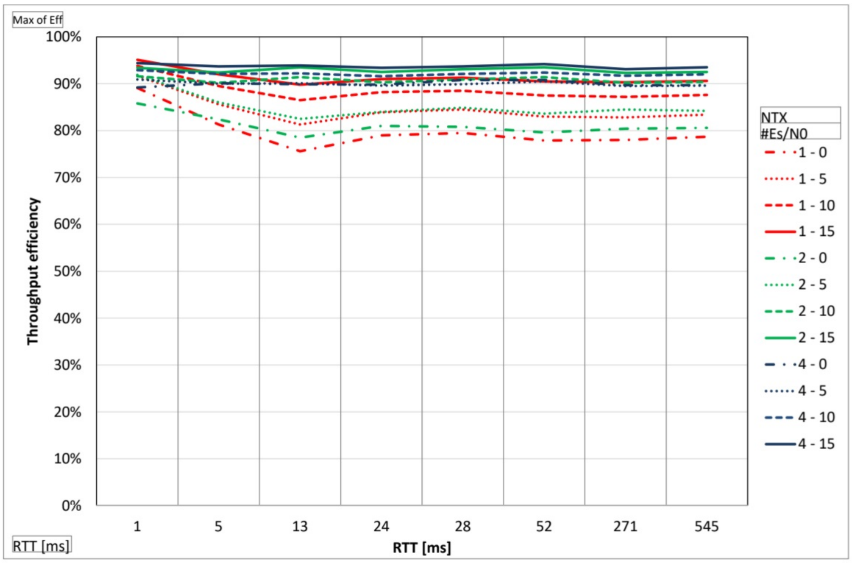

Figure 7.11: Throughput efficiency vs RTT. NTN-CDL LOS channel, speed $100km/h$, FER $\leqslant 10^{-2}$

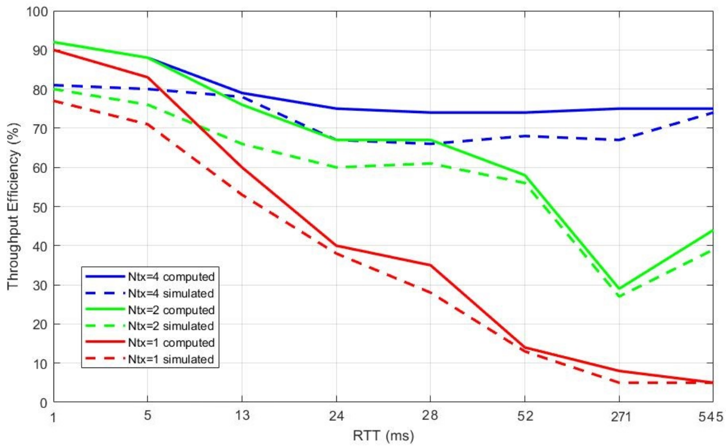

Figure 7.12: Simulated VS computed throughput efficiency vs RTT. NTN-CDL NLOS channel, speed $10km/h$, FER $\leqslant 10^{-2}$

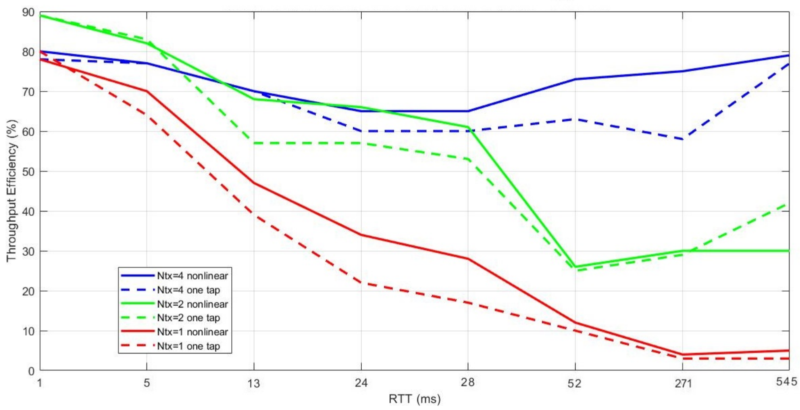

Figure 7.13: Nonlinear model VS one tap filter throughput efficiency vs RTT. NTN-CDL NLOS channel, speed $10km/h$, FER $\leqslant 10^{-3}$

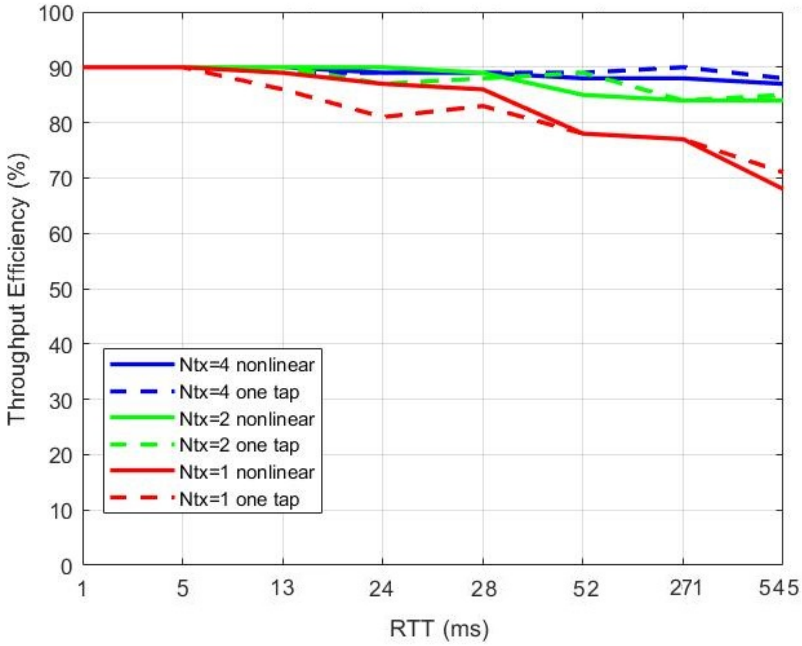

Figure 7.14: Nonlinear model VS one tap filter throughput efficiency vs RTT. NTN-CDL LOS channel, speed $100km/h$, FER $\leqslant 10^{-3}$

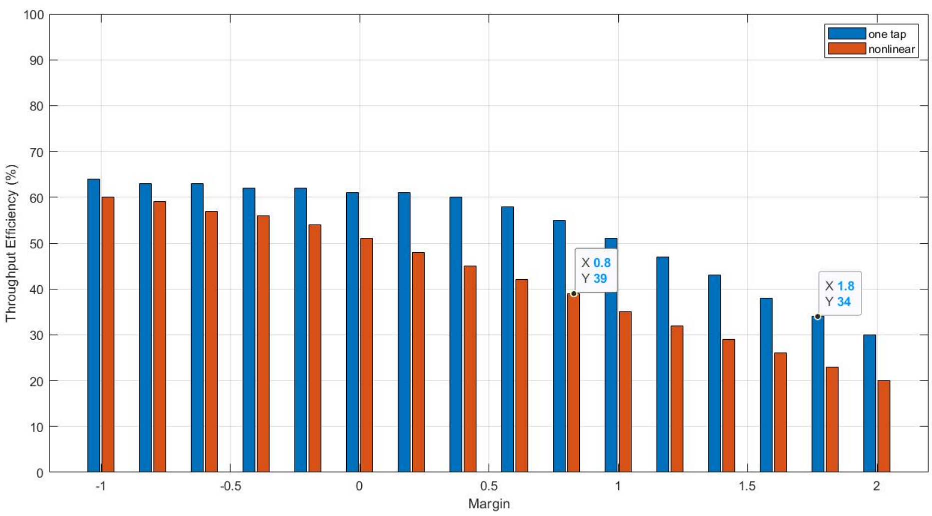

Figure 7.15: nonlinear model VS one tap filter throughput efficiency vs the margin. NTN-CDL NLOS channel, speed $10km/h$, $Ntx = 2$, $RTT = 52$

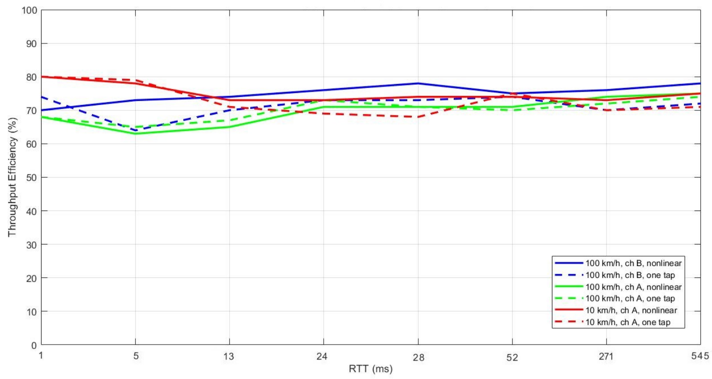

Figure 7.16: Nonlinear model VS one tap filter throughput efficiency vs RTT. NTN-CDL LOS channel, $Ntx = 4$, FER $\leqslant 10^{-2}$

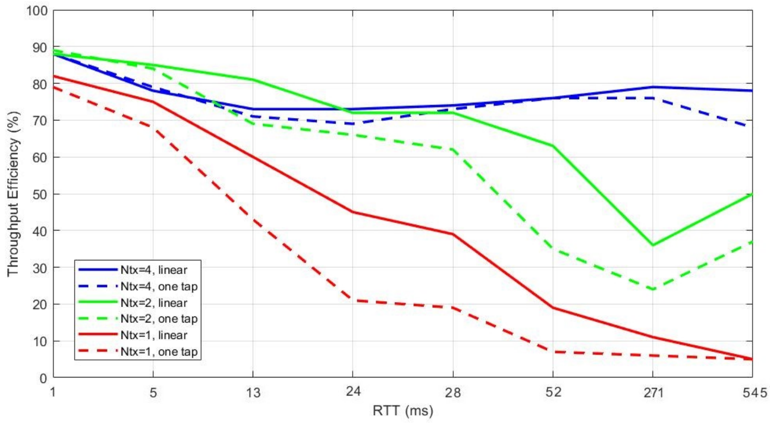

Figure 7.17: Linear model VS one tap filter throughput efficiency vs RTT. NTN-CDL NLOS channel, speed $10km/h$, FER $\leqslant 10^{-2}$

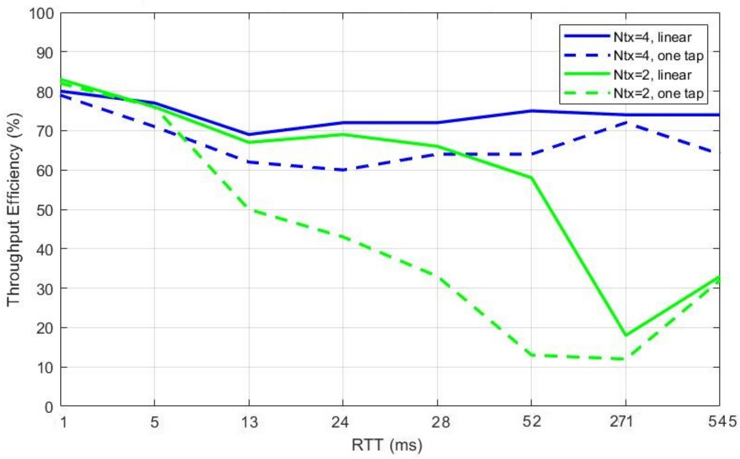

Figure 7.18: Linear model VS one tap filter throughput efficiency vs RTT. NTN-CDL LOS channel, speed $10km/h$, FER $\leqslant 10^{-3}$

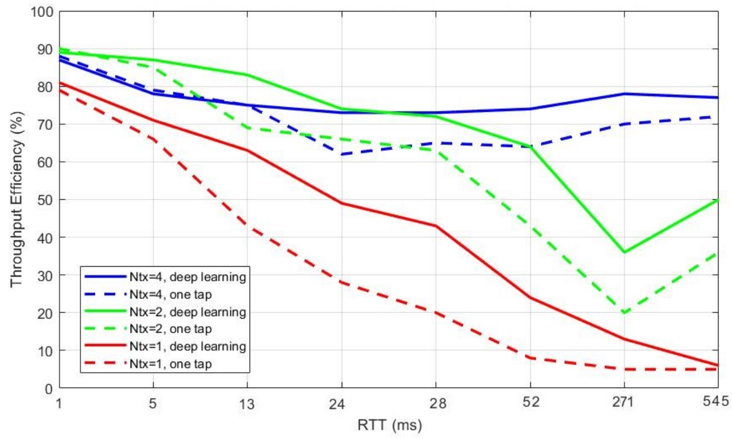

Figure 7.19: Deap learning model VS one tap filter throughput efficiency vs RTT. NTN-CDL NLOS channel, speed $10km/h$, FER $\leqslant 10^{-2}$

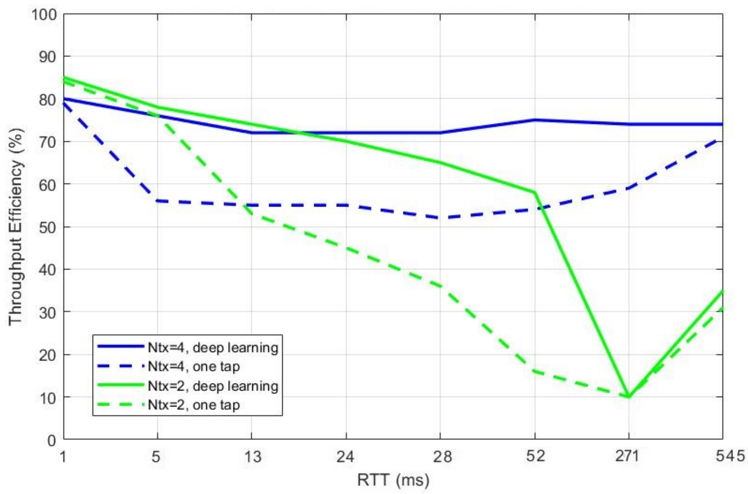

Figure 7.20: Deep learning model VS one tap filter throughput efficiency vs RTT. NTN-CDL LOS channel, speed $10km/h$, FER $\leqslant 10^{-3}$

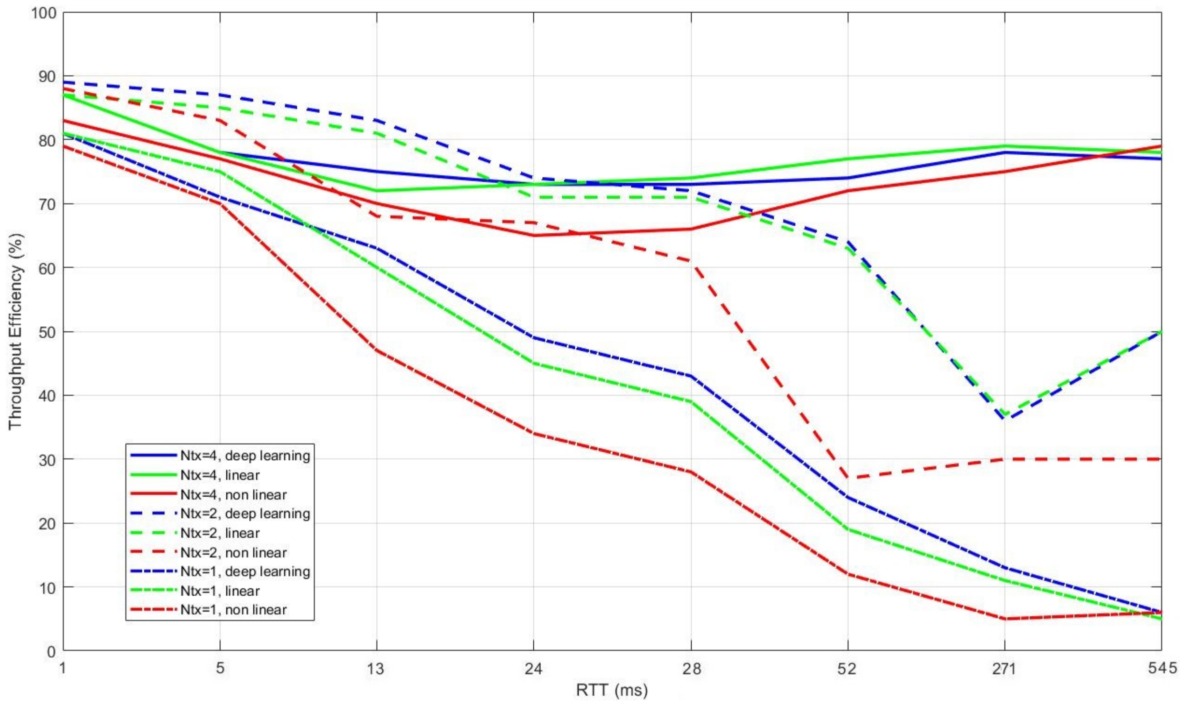

Figure 7.21: Deep learning model VS Linear model VS Non linear model throughput efficiency vs RTT. NTN-CDL LOS channel, speed $10km/h$, FER $\leqslant 10^{-3}$

Chapter 8

Conclusion

Taking into account the various simulations and comparisons made, it may be concluded that although using HARQ does increase the latency and requires additional memory resources, HARQ significantly improves the system performance (throughput efficiency), this improvement tends to show in LOS channel at high RTT and speed values, however, it is much clearer and more significant in NLOS channel cases. So, based on the obtained results, it is strongly recommended to use HARQ in the NTN version.
The problem of outdated feedback CSI can be solved by using filters or models to predict channel variations and enable setting the correct values for the parameters at the transmitter side. The used one−tap filter does a good job when it comes to that.
The results obtained through simulating the entire transmission and reception process are more practical than those acquired through statistically−based computations due to the consideration of the suboptimality of some of the processes, an indicator of the performance impact of such suboptimality is the slightly worse performance of the simulated version as opposed to the computed one.
The performance of the system can be further enhanced using a better prediction model such as the proposed nonlinear ARX model, which improves the performance of the system due to the better forecasting capabilities. The model improved the HARQ system throughput efficiency by 15% at some points and maintained a slightly better performance throughout the simulated range of RTT.
The linear model displays better performance than both the current single pole filter and the nonlinear model. Optimal linear filters have proven a high prediction ability and, as a result, a higher system performance. The performance of the system is only slightly improved for 4 transmissions, but for the other cases, the throughput efficiency is increased by 15 to 25% for an FER of 10^{-2}, and up to 40% for a 10^{-3} FER.
Among all proposed prediction models, the deep learning model is the one with the best performance (the linear model has a comparable performance), the deep learning model is faster as well due to the lack of need to compute the autocorrelation function for the time series representing channel MI.

8.1 Recommendations and Future Work

Based on the results obtained throughout the work conducted in this thesis; the following is recommended:

- Using HARQ and ACM techniques in the 5G−satellite integrated network.
- Exploiting the prediction capability of optimal linear filters or neural networks to enhance NTN system performance.

For future work; I would suggest:

- Repeating the simulation using a range of linear and nonlinear filters, comparing their performance since the models suggested in this thesis are selected on an MSE basis.
- Repeating the simulation with a range of machine learning and other deep learning models to further improve the system performance.
- Investigating the impact of LOS and NLOS transitions and\or speed change on the proposed prediction models to determine which is more robust against channel scenario change.

Bibliography

[1] Zhang, S., Zhu, D., & Wang, Y. *A survey on space-aerial-terrestrial integrated 5G networks.* COMPNW Computer Networks, 174. 2020.

[2] Zaidi, A., Athley, F., Medbo, J., Gustavsson, U., Durisi, G., & Chen, X. *5G physical layer : principles, models and technology components.* 2018.

[3] Al-Dulaimi, A., Wang, X., & Chih-Lin, I. *5G networks : fundamental requirements, enabling technologies, and operations management.* 2018.

[4] Ippolito, L. J. *Satellite communications systems engineering : atmospheric effects on satellite link design and system performance.* 2017.

[5] Elbert, B. R. *Introduction to satellite communication.* Norwood, Mass.: Artech House. 2008.

[6] Maral, G., Bousquet, M., & Sun, Z. *Satellite communications systems : systems, techniques and technology.* Chichester: Wiley. 2011.

[7] Duhamel, P., Szczecinski, L., Ciblat, P., & Bassi, F. *Recent advances in HARQ communications.* Paper presented at the International Conference on Telecommunications (ICT 2019), Hanoi, Vietnam. 2019. https://hal-centralesupelec.archives-ouvertes.fr/hal-02172249

[8] Marsch, P., Bulakçi, Ö., Queseth, O., & Boldi, M. *5G System Design Architectural and Functional Considerations and Long Term Research.* Chichester, UK: John Wiley & Sons, Ltd. 2018.

[9] Enescu, M. *5G new radio a beam-based air interface.* 2020.

[10] *TS 138 212 - V15.2.0 - 5G; NR; Multiplexing and channel coding.* 3GPP TS 38.212 version 15.2.0 Release 15.

[11] Guidotti, A., Cioni, S., Colavolpe, G., Conti, M., Foggi, T., Mengali, A., Montorsi, G., Piemontese, A. & Vanelli-Coralli, A. *Architectures, standardisation, and procedures for 5G Satellite Communications: A survey.* COMPNW Computer Networks, 183, 2020-2012. 2020.

[12] Foggi, T., Ugolini, A., Colavolpe, G., Guidotti, A. & Vanelli-Coralli, A. *3GPP New Radio SW Implementation and Performance Assessment for Satellite Networks* 2019.

[13] MATLAB Support. *https://www.mathworks.com* accessed on Aug,2020.

[14] *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study on New Radio (NR) to support non-terrestrial networks* Release 15. 2019.

[15] *3rd Generation Partnership Project; Technical Specification Group Radio Access Network; NR; Multiplexing and channel coding* Release 15. 2018.

[16] Yeo, J., Park, S., Oh, J., Kim, Y., Lee, J., & Workshops, I. G. *Partial Retransmission Scheme for HARQ Enhancement in 5G Wireless Communications.* 1-5. 2017.

[17] Strodthoff, N., Goktepe, B., Schierl, T., Hellge, C., & Samek, W. *Enhanced Machine Learning Techniques for Early HARQ Feedback Prediction in 5G.* IEEE J. Select. Areas Commun. IEEE Journal on Selected Areas in Communications, 37(11), 2573-2587. 2019.

[18] Yeo, J., Bang, J., Ji, H., Kim, Y., Lee, J., & th Ieee Vehicular Technology Conference, V. T. C. F. *Outer code-based HARQ retransmission for multicast/broadcast services in 5G.* IEEE Veh Technol Conf IEEE Vehicular Technology Conference, 2019-September, 2019.

[19] Sahin, C., Liu, L., Perrins, E., & Ma, L. *Delay-Sensitive Communications Over IR-HARQ: Modulation, Coding Latency, and Reliability.* IEEE J. Select. Areas Commun. IEEE Journal on Selected Areas in Communications, 37(4), 749-764. 2019.