

Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN
INGEGNERIA ELETTRONICA, TELECOMUNICAZIONI E
TECNOLOGIE DELL'INFORMAZIONE

Ciclo 35

Settore Concorsuale: 09/F2 - TELECOMUNICAZIONI

Settore Scientifico Disciplinare: ING-INF/03 - TELECOMUNICAZIONI

NB-IOT VIA NON TERRESTRIAL NETWORKS

Presentata da: Carla Amatetti

Coordinatore Dottorato

Aldo Romani

Supervisore

Alessandro Vanelli Coralli

Co-Supervisore

Alessandro Guidotti

Esame finale anno 2023

To my beloved family.

“Remember to look up at the stars and now down at your feet. Try to make sense of what you see and wonder about what makes the universe exist. Be curious. And however difficult life may seem, there is always something you can do and succeed at. It matters that you don’t just give up”

Stephen Hawking

Contents

List of Figures	vii
List of Tables	xi
List of Acronyms	xiii
1 Introduction	5
1.1 Motivations	5
1.1.1 NB-IoT	5
1.1.2 Satellite as an enabler of massive machine type communications	7
1.1.3 The need for new algorithms	8
1.1.3.1 Low Cost and Complexity	8
1.1.3.2 Energy Efficiency	9
1.1.3.3 Support for Massive Number of Devices	10
1.1.3.4 Extreme Coverage	10
1.1.3.5 Latency	10
1.2 Thesis contributions and organization	11
List of Publications	13
2 Narrowband IoT	15
2.1 Physical layer	15
2.1.1 Downlink	17
2.1.2 Uplink	19
2.2 Medium Access Control layer	22
2.2.1 Random Access procedure	22
2.2.1.1 4-step RA	23
2.2.1.2 Early Data Transmission	25
2.2.1.3 2-Step Random Access	26
2.2.2 HARQ	27
2.2.3 Downlink and Uplink Scheduling	28
3 Non Terrestrial Networks	31
3.1 SatCom System Architecture	31
3.1.1 Space Segment	32
3.1.2 Ground Segment	33
3.1.3 Communication Links	34

3.2	The NTN path in 3GPP standardization	35
3.3	NTN system architectures for 5G	39
3.3.1	Direct access	40
3.3.2	Integrated Access and Backhaul	41
3.3.3	IoT NTN architectures	42
3.4	NTN impairments	43
3.4.1	Delay	44
3.4.2	Doppler shift	46
3.4.3	Path Loss	47
3.5	NB-IoT Integration with NTN: State of the Art Analysis	48
3.5.1	Link Budget	49
3.5.2	DL and UL synchronization	50
3.5.3	Timing enhancement	52
3.5.4	RA congestion and E2E delay	53
3.5.5	User scheduling	54
3.5.6	NOMA in NTN	55
3.5.7	ESA funded projects on NTN NB-IoT	56
3.5.8	Conclusion of the SoA Analysis	57
4	Link Budget	59
4.1	Channel Modeling	59
4.2	Link Budget Characterization	59
4.2.1	Antenna Pattern	59
4.2.2	Downlink	61
4.2.3	Uplink	63
4.3	Link Budget for NB-IoT via NTN	65
4.3.1	System Architecture	66
4.3.1.1	Single- and Multi-Satellite Scenarios	66
4.3.2	Numerical Results	68
4.3.2.1	Single-Satellite Scenario	69
4.3.2.2	Multi-Satellite Scenario	71
5	Link level analysis	75
5.1	Preamble detection in NB-IoT via Satellite: a Wavelet-based approach	75
5.1.1	System model	76
5.1.2	Transmitted signal	77
5.1.3	Receiver design	77
5.1.3.1	Detection process	78
5.1.4	ML decision	79
5.1.4.1	Estimation of the synchronization parameters	80
5.1.5	Simulation and results	82
5.1.5.1	Detection performance	82
5.1.6	Delay and CFO estimation performance	83

5.2	Neural Network based Non-Orthogonal Random Access for 6G NTN-IoT	85
5.2.1	System Architecture	86
5.2.2	Transmitted and received signals	87
5.2.3	NN approach to preamble detection	88
5.2.3.1	Neural Network implementation	88
5.2.3.2	Collision classification	89
5.2.3.3	Delay estimator	89
5.2.4	Numerical Results	91
5.2.4.1	Collision classification performance	93
5.2.4.2	Delay estimator performance	94
5.3	SCMA performance in NB-IoT NTN system	95
5.3.1	Overview of NOMA	96
5.3.2	NOMA techniques proposed in 3GPP	97
5.3.3	Sparse Code Division Multiple Access for NB-IoT	99
5.3.4	Scenario	103
5.3.4.1	Simulator design and evaluation	104
5.3.5	Discussion and future works	106
6	System level analysis	107
6.1	NB-IoT random access procedure via NTN: system level performances	107
6.1.1	System architecture	107
6.1.2	Analysis of the access parameters	108
6.1.3	Simulation Model	111
6.1.4	Performance evaluation	111
6.1.4.1	Parameters configuration	114
6.1.4.2	Impact of the user density	115
6.1.4.3	Impact of the number of repetitions	115
6.1.4.4	Impact of the beam visibility window	115
6.1.5	Discussion	116
6.2	A Joint Access and Data Phases Design for NB-IoT via Non-Terrestrial Networks	116
6.2.1	System architecture & assumptions	119
6.2.1.1	NTN peculiarities that compromise the NB-IoT performance	119
6.2.2	Problem formulation	121
6.2.2.1	Mathematical Formulation	121
6.2.2.2	Success Probability	122
6.2.3	Solution to spectral efficiency maximization	127
6.2.4	Numerical results	130
6.2.4.1	User Barring performance	130
6.2.4.2	Impact of beam dimension	131

6.2.4.3	Impact of the requested capacity	132
6.2.5	Conclusion and future work	133
7	Conclusions	137
A	Appendix Localization systems for railway	141
A.1	European Rail Traffic Management System	144
A.2	Towards the Future Generation of Railway Localization and Signaling	
	Exploiting sub-meter RTK GNSS	146
A.2.1	State of the Art and Background	146
A.2.1.1	Related work	146
A.2.1.2	Background: N-RTK	148
A.2.1.3	Background: Dead reckoning	149
A.2.2	HW and NTRIP network selection	149
A.2.2.1	Wireless sensor node	149
A.2.2.2	GNSS module selection	150
A.2.2.3	NTRIP network	150
A.2.3	Experimental Results	150
A.2.3.1	Static Tests	151
A.2.3.2	Dynamic Tests	152
A.2.3.3	Dynamic Tests with RTK and IMU	155
A.2.4	Discussion	155
A.3	Towards the Future Generation of Railway Localization and Signaling	
	exploiting sub-meter RTK GNSS: extension	156
A.3.1	Sensor node	156
A.3.2	GNSS with a Shielded Antenna	157
A.3.3	Dynamic Tests on a Train	158
A.3.3.1	Results	160
A.3.4	Discussion and Conclusion	161
	Bibliography	167
	Acknowledgements	179

List of Figures

2.1	Uplink and Downlink channels.	17
2.2	NPRACH preamble and periodicity.	20
2.3	From the left to the right: 4-step RA procedure, EDT, and 2-step RA.	23
2.4	Example of the channel structure for msgA: mapping between preamble IDs and PRU IDs.	26
2.5	TBS of NPUSCH channel [32].	29
3.1	SatCom System high-level architecture.	31
3.2	Types of payload [33].	32
3.3	Types of coverage [33].	34
3.4	NTN 3GPP roadmap [34].	35
3.5	Available functional splits.	39
3.6	Architectural options with direct user access with transparent (top) and regenerative (bottom) payload [44].	40
3.7	Architectural option with direct user access, regenerative payload, and functional split. Applicable to B, D1, and D2. [44].	40
3.8	Architectural options with IAB with transparent (top) and regenerative (bottom) payload. [44].	42
3.9	Reference system for delay and Doppler in u - v coordinate system.	43
4.1	Antenna radiation pattern [33].	60
4.2	Frequency Reuse scheme.	61
4.3	Geometry for DL LB computation [74].	62
4.4	Geometry for UL LB computation [74].	64
4.5	High-level system architecture: access based on regenerative payload [74].	66
4.6	Example of coverage (u - v coordinates): Low Earth Orbit (LEO) system in S-band; $h_{sat} = 1200$ km; FR3; 6 tiers (127 beams) [74].	67
4.7	Downlink Link Budget [74].	70
4.8	Uplink Link Budget [74].	71
4.9	Example of satellite constellation coverage (1200 km)[74].	72
5.1	High-level system architecture [80].	76
5.2	Stationary dyadic wavelet transform, with g_j and h_j high pass and low pass filters, respectively [80].	78
5.3	Output of multi-resolution analysis [80].	80

5.4	Detection rate with N_{rep} varying from 1 to 32 [80].	82
5.5	CDF of the normalized CFO estimation error [80].	84
5.6	CDF of the normalized ToA estimation error, $SNR = 3\text{dB}$ [80].	85
5.7	CDF of the normalized ToA estimation error. $N_{rep} = 32$. $SNR = 0\text{dB}$. [80].	85
5.8	Link budget analysis of the proposed scenario [80].	86
5.9	High-level system architecture [92].	87
5.10	Structure of the CNN for collision classification [92].	90
5.11	Structure of the CNN for ToA estimation [92].	91
5.12	Real and Imaginary part of the DFT symbols. Set 3. [92].	92
5.13	Outcome of the classification: Set 3. [92].	96
5.14	Outcome of the classification: Set 4 [92].	96
5.15	Example of 4×6 SCMA encoding scheme [101].	101
5.16	Factor graph for 4×6 [102].	101
5.17	Block diagram of the simulator.	105
5.18	SCMA performance. Homogeneous scenario.	105
6.1	System architecture for the RA analysis [103].	108
6.2	RTD of LEO satellite varying the beam center elevation angle. Set 3 [103].	109
6.3	Performances of the random access procedure. Set 3, $\alpha = 0.1$ [103]. . .	116
6.4	Time analysis. Set 3, $\alpha = 0.1$ [103].	117
6.5	Number of transmissions. Set 3, $\alpha = 0.1$ [103].	117
6.6	System architecture [106].	119
6.7	Access and data phases in NB-IoT [106].	120
6.8	Definition of the satellite admittance region	122
6.9	Success probability in Set 3 and Set 4. $\alpha = 0.002[\text{users}/\text{km}^2/\text{channel}]$ [106].	134
6.10	Average interference power in Set 3 and Set 4 [106].	134
6.11	Number of uplink repetitions varying \bar{D} for Set 3 without (a) and with EDT (b). $\alpha = 0.01$ [106].	134
6.12	Number of satellite passes varying \bar{D} for Set 3 without (a) and with EDT (b). $\alpha = 0.01$ [106].	135
6.13	Number of uplink repetitions varying \bar{D} for Set 4 without (a) and with EDT (b). $\alpha = 0.001$ [106].	135
6.14	Number of satellite passes varying \bar{D} for Set 4 without (a) and with EDT (b). $\alpha = 0.001$ [106].	135
A.1	Reference locations for GNSS static tests rural, suburban and urban environments. [115].	152
A.2	References to dynamic test environments including the satellite map and the tracked path (in red). (d) The custom sensor node used for reported tests [115].	153

A.3	CDF of RTK modes (0 = <i>Stand-alone</i> , 1 = <i>Float</i> , 2 = <i>Fix</i>) in dynamic scenario [115].	153
A.4	References to dynamic test environments including the satellite map and the tracked path with RTK + DRK (blue line) and only DRK (purple line) [115].	154
A.5	Overview of the custom localization sensor node where: (A) ZED-F9P GNSS module, (B) STM32L452CEU MCU, (C) SARA-R410M cellular module, (D) ASM330LHH IMU [135].	156
A.6	Static antenna shielding test setup (a) and results as box plot showing static measurement accuracy without RTK at different angles (b) [135].	158
A.7	Dynamic tests: Sensor node antenna measurement setup for Sensor Node 1 (a) and Sensor Node 2 (b) [135].	159
A.8	Number of Satellites received by Sensor Node 1 (a) and Sensor Node 2 (b) on a measurement run between Formigine and Modena [135].	163
A.9	Position Error reported by the GNSS module and the actual error relative to the ground truth for the entire measurement run of Sensor Node 1 between Formigine and Modena, including cellular connectivity status [135].	164
A.10	Position Error reported by the GNSS module and the actual error relative to the ground truth, without the tunnel section, for Sensor Node 1 (a) and Sensor Node 2 (b) [135].	165
A.11	Maps showing position error near Modena Fornaci in areas of interest under a highway overpass (a) and a region with potentially inaccurate ground truth (b). The ground truth is indicated in yellow, Sensor Node 1 in blue and Sensor Node 2 in orange [135].	166

List of Tables

2.1	NPUSCH resource units [7]	16
2.2	NPRACH preamble formats for FDD mode	20
3.1	NTN reference scenarios per system type [39]	40
3.2	NTN reference scenarios for IoT [23]	43
3.3	ESA projects addressing NTN NB-IoT	56
4.1	Satellite parameters for S-band (Set-2 [39])	68
4.2	NarrowBand IoT (NB-IoT) terminal parameters for S-band [7], [77]	68
4.3	Constellation parameters for S-band [76]	69
4.4	Single- and Multi-Satellite results: Carrier-to-Interference plus Noise Ratio (CINR) μ and σ in dB.	73
5.1	Simulation parameters	82
5.2	Normalized RMSE of CFO and ToA	82
5.3	Satellite parameters [23]	92
5.4	Mean and variance of CNR	95
5.5	RMSE of ToA [μs]	95
5.6	NOMA schemes	98
5.7	Simulation parameters	105
5.8	Average BER of each user after the approximation of channel coefficients at the beam center	106
5.9	Average BER of each user in TDL channel	106
6.1	Satellite parameters [23]	112
6.2	Access configuration parameters [105]	112
6.3	Access parameters for satellite configurations. $s = success, f = failure,$ $W = remaining\ time\ of\ beam\ visibility\ window$	113
6.4	Satellite parameters	130
6.5	Results of P_1 , for $\bar{D} = 166\ bytes, \alpha = 0.03[users/km^2/channel]$ Set 3	130
6.6	Results of P_1 , for $\bar{D} = 166\ bytes, \alpha = 0.003[users/km^2/channel]$ Set 4	131
A.1	Localization accuracy for static tests w and w/o RTK	151
A.2	Localization accuracy for dynamic tests w and w/o RTK.	152
A.3	Localization accuracy for dynamic tests w and w/o RTK and DRK.	153
A.4	Position Error Summary Statistics	161

List of Acronyms

3GPP	Third Generation Partnership Project
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BLER	Block Error Rate
BO	Back-Off
CBRA	Contention Based Random Access
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CE	Coverage Enhancement
CFO	Carrier Frequency Offset
CINR	Carrier-to-Interference plus Noise Ratio
CIR	Carrier-to-Interference Ratio
CNR	Carrier-to-Noise Ratio
CP	Control Plane
CRT	Contention Resolution Timer
DL	Downlink
DVB	Digital Video Broadcasting
DVB-S2X	Digital Video Broadcasting - Satellite Second Generation Extension
DVB-S2	DVB - Satellite Second Generation
DWT	Discrete Wavelet Transform
EDT	Early Data Transmission
EIRP	Equivalent Isotropically Radiated Power
eMBB	enhanced Mobile BroadBand
FDMA	Frequency Division Multiple Access
FEC	Forward Error Correction
FFR	Full Frequency Reuse
FFT	Fast Fourier Transform
FoV	Field of View
FR	Frequency Reuse
FR3	Frequency Reuse 3
FSL	Free Space Loss

GEO	Geostationary Orbit
GNSS	Global Navigation Satellite Systems
GS	Ground Station
GW	Gateway
HAPS	High Altitude Platform Systems
HARQ	Hybrid Automatic Repeat reQuest
IoT	Internet of Things
ISL	Inter Satellite Links
KPI	Key Performance Indicator
LB	Link Budget
LEO	Low Earth Orbit
LoS	Line of Sight
LPWAN	Low Power Wide Area Network
MA	Multiple Access
MPA	Message Passing Algorithm
M-S	Multi-Satellite
MAC	Medium Access Control
MCS	Modulation and Coding Scheme
MEO	Medium Earth Orbit
mIoT	massive IoT
mMTC	massive Machine Type Communications
NB-IoT	NarrowBand IoT
NGC	Next Generation Core network
NN	Neural Network
NOMA	Non Orthogonal Multiple Access
NR	New Radio
NTN	Non-Terrestrial Network
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OMA	Orthogonal Multiple Access
PER	Packet Error Rate
PHY	Physical Layer
PPP	Precise Point Positioning
QoS	Quality of Service
RA	Random Access
RAO	Random Access Opportunity

RAR	Random Access Response
RE	Resource Element
RRC	Radio Resource Control
RRM	Radio Resource Management
RTD	Round Trip Delay
RTK	Real Time kinematic
RU	Resource Unit
SatNav	Satellite Navigation
SCMA	Sparse Code Multiple Access
S-DWT	Stationary Discrete Wavelet Transform
SIC	Successive Interference Cancellation
S-IoT	Satellite-Internet of Things
S-S	Single-Satellite
SatCom	Satellite Communication
SDR	Software Defined Radio
SINR	Signal-to-Interference plus Noise Ratio
SIR	Signal-to-Interference Ratio
SLS	System-Level Simulation
SNR	Signal-to-Noise Ratio
SoA	State-of-the-Art
SSP	Sub-Satellite Point
TA	Timing Advance
TBS	Transport Block Size
TDMA	Time Division Multiple Access
ToA	Time of Arrival
UAS	Unmanned Aircraft System
UE	User Equipment
UL	Uplink
UP	User Plane
UT	User Terminal
VLEO	Very LEO
VSAT	Very Small Aperture Terminal

Abstract

The Third Generation Partnership Program (3GPP) has recently defined the content of its Rel.18, also known as New Radio Advanced (NR-A), which will foster the transition from 5G to 6G. Within this release, new services enabling vertical applications, including future factories and personal Internet of Things (IoT) (*e.g.*, wearable, smartphone, car, and handheld devices) networks, have been identified.

Indeed, massive IoT (mIoT) is expected to play a crucial role in NR-A wireless communications, thanks to its inherent capacity to provide seamless communications among heterogeneous devices without human intervention. This will allow the whole society's digitalization using intelligent and remotely controlled management and tracking systems. However, due to the rapid expansion of IoT networks and the exponential proliferation of smart devices, NR-A alone might not be able to completely satisfy the demanding IoT requirements, *e.g.*, ultra-large-scale, highly dynamic, and fully intelligent services. Indeed, we are facing a rapid growth in the number of devices offering existing services and a rising of new IoT applications which most likely will exceed the capabilities of the terrestrial NR-A systems. Indeed, both in terms of connections and bandwidth, terrestrial networks alone might not be able to satisfy the onerous IoT requirements shortly. Moreover, another vital requirement for an infrastructure providing IoT services is that of guaranteeing ubiquitous connectivity to the low-cost, low-powered devices distributed all over the globe. This again might be a challenge for terrestrial IoT, since there are vast areas where terrestrial infrastructures are unfeasible or not economically viable. Therefore, to unleash the full potentiality of the Beyond 5G (B5G) networks providing seamless connectivity everywhere and every time, the 3GPP envisions the integration of Non-Terrestrial Networks since its Rel.17. Indeed, the NTN component fully integrated into the terrestrial infrastructure is not only capable of complementing and extending terrestrial networks, both in densely populated and rural areas, but also providing increased resilience, improved sustainability, high spectrum availability, and greater flexibility. To this aim, the B5G NTN network leverages multi-dimensional, multi-layer architecture, consisting of both space and airborne flying nodes, and encompasses novel enablers, such as Artificial Intelligence (AI).

However, the integration process of NTN in 5G and B5G networks foresees modifications to terrestrial standards to allow reliable communications despite typical satellite channel impairments, *i.e.*, high delay, large Doppler shifts, and large path losses. At the same time, also the NTN component will be modified to accommodate this integration process, for example by embarking the terrestrial Base Station

on board.

In this framework, this thesis aims at proposing techniques at the Physical and Medium Access Control layers that require minimal adaptations in the current NB-IoT standard via NTN. In particular, new algorithms have been studied and proposed.

Thus, the first part of the thesis is devoted to the analysis of the satellite impairments and a detailed link budget analysis is provided.

Then the work follows two paths: analysis at the link layer and the system layer. In detail, the detection of orthogonal NB-IoT preambles was first analyzed, proposing a new algorithm leveraging time-frequency analysis and allows estimating the signal Arrival Time without estimating the frequency offset. The results show that, exploiting the time/frequency decomposition, it is possible to reach $1\mu s$ of accuracy for the estimation of the signal's arrival, at the expense of the high number of signal repetitions. Then the work focuses on detecting the preamble in the presence of collisions. To estimate the number of users and their access parameters, a Non-Orthogonal Multiple Access (NOMA) approach based on Neural Networks was proposed, which is able to classify up to 2 users per preamble with an accuracy of 99%. The classification performance decreases up to 60% in the case of four devices per preamble. The last section of this part focuses on analyzing the performance of a NOMA algorithm applied to users transmitting their data to the satellite. NOMA is a powerful tool, however, an in-depth analysis of the pilot design needs to be carried out to estimate or update the estimation of the channel coefficients.

The system analysis aims at evaluating the Random Access's performance when multiple users perform this procedure, generating collisions. Thus, the different access parameters, such as the number of uplink repetitions, the Random Access Periodicity, and the Back-Off have been tested in different satellite scenarios. The performances are reported in terms of access probability and time taken to conclude the procedure. The extensive simulations show that medium-length periodicity provides a good trade-off between the access rate and the time to successfully conclude the RA. A critical parameter is represented by the number of uplink repetitions, which heavily influences the access probability. By increasing the number of repetitions, this probability increases. However, if the number of repetitions is overestimated, data transmission resources are reduced.

Finally, a novel and heuristic algorithm to design an effective solution, meticulously stitched around the technological constraints of the NB-IoT protocol and the peculiarities of the satellite channels has been proposed. The main objective of the algorithm is to jointly design the RA phase and the data transmission by defining the number of satellite passages, the RA periodicity, and the number of uplink repetitions that maximize the spectral efficiency of the system. The results show that in the case of large beams, where is possible to send only few bits per Resource Unit due to the low Signal to Noise Ratio, it is better to enable the Early Data Transmission (i.e., sending the data in Msg3 of the RA). Indeed, in this case, it is possible to have

frequent RA opportunities, with a lower number of satellite passages, thus reducing the cost of the service.

Chapter 1

Introduction

By 2025, forecasts suggest that there will be more than 75 billion Internet of Things (IoT) connected devices in use. This would be a nearly threefold increase from the IoT installed in 2019 [1]. This enormous growth results from the inclusion of devices in the telecommunication domain, causing a paradigm shift from Human-to-Human (H2H) communication to Machine-to-Machine (M2M) interactions. Examples of these terminals are daily tech devices (*e.g.*, smartphones, wearables, etc.), sensors or actuators in a smart home/office or smart city, as well as industrial devices (*i.e.*, smart machines). Such devices or things are interconnected in a network and exchange data through internet, creating the so-called IoT concept [1]. IoT finds use in different applications and vertical sectors, from smart transportation (including vehicles and trains) to smart homes and health care. This will allow the full society's digitalization by means of intelligent and remotely controlled management and tracking systems. One of the enablers of this paradigm shift is the Low Power Wide Area Network (LPWAN), which includes the NarrowBand IoT (NB-IoT) protocol (conceived as part of the terrestrial cellular networks, including the 4th and 5th generations of mobile communication systems) [2], [3], LoRa [4] and Sigfox [5].

1.1 Motivations

This section aims at answering three fundamental questions that drove the thesis: why NB-IoT? Why Satellite? Eventually, why the need for new algorithms to enable NB-IoT via NTN?

1.1.1 NB-IoT

NB-IoT is a licensed technology introduced by the Third Generation Partnership Project (3GPP) in Rel-13. This is an independent radio interface, tightly connected with Long Term Evolution (LTE), which also shows up in its integration in the current 5G New Radio (NR) specifications. Compared to other technologies, it offers massive Machine Type Communications (mMTC) features at low cost, such as [6]:

- **Massive connections:** NB-IoT is based on Orthogonal Frequency Division Multiple Access (OFDMA) [7]. This allows several users to transmit simultaneously without inter-carrier interference (ICI), since the time-frequency resource allocation is managed by the base station. Further, NB-IoT supports two uplink schemes and two sub-carriers spacing (SCS), *i.e.*, multi-tone and single-tone transmission at 3.75 kHz and 15 kHz, respectively. This offers the flexibility to schedule up to 12 sub-carriers with 15 kHz of SCS or 48 sub-carriers with a spacing of 3.75 kHz with single-tone scheduling. Therefore, the base station may support a large number of users in parallel.
- **Energy efficiency:** the NB-IoT battery can last up to 10 years [8]. Generally, IoT devices belonging to LPWAN, are equipped with small batteries. However, in remote locations or in harsh environments, where it is difficult to change the battery, alternative sources of energy or energy harvesting are required. The radio interface is typically the most power-hungry component for the transmission of data when dealing with wireless devices [9]. For this reason, several mechanisms are applied to extend the life of the device. For example, extended Discontinuous Reception (eDRX) and Power Saving Mode (PSM) reduce the battery consumption by periodically monitoring the paging channel and letting the terminal go to the sleep state, respectively. Besides, NB-IoT employs Data over Non-Access Stratum (NAS) and Radio Resource Control (RRC) suspend/resume for signaling optimization. The transmission over NAS allows the user to send data without activating the User Plane (UP) and supports sporadic data transmissions also. The RRC suspend/resume is a UP optimization procedure that introduces an efficient way to disable and restore it.
- **Low-cost devices:** the enhancement of IoT technology owes a great deal to the substantial decrease in the cost of IoT devices and development boards. The cost of an IoT device is proportional to the computational features it incorporates. If an IoT device has advanced computational features like memory and processing power, its price increases accordingly. Due to their small size and affordability, IoT devices necessitate compact embedded technology. One way to decrease the transceiver cost is to reduce the bandwidth; indeed, in NB-IoT, the Uplink (UL)/Downlink (DL) bandwidth is restricted to 180 kHz, which allows for closing the link budget considering the low transmission power. Reduced bandwidth implies also restricted Transmission Block Size (TBS), which reduces the channel encoding and decoding complexity and allows for a smaller Hybrid Automatic Repeat reQuest (HARQ) buffer size. Moreover, the device operates in Half Duplex mode, and, thus, it is not able to receive and transmit simultaneously. This further decreases the device cost, as no duplexer is required. Additionally, the cost is reduced by limiting the data rate. Indeed, only $\pi/2$ -BPSK, $\pi/4$ -QPSK, and QPSK modulations are allowed. Besides, this allows to reduce the Peak to Average Power Ratio (PAPR).

- **Extended coverage:** NB-IoT supports up to 164 dB of Maximum Coupling Loss (MCL). To provide such wide coverage, the transport block can be transmitted multiple times, *i.e.*, 128 times in UL and 2048 in DL [7]. Thus, the introduction of the repetition improves the Signal-to-Noise Ratio (SNR) and enables proper decoding of the signal. These repetitions are combined at the receiver side in order to increase the SNR ratio. Along with repeating the same transmission several times, other techniques have been used to extend the coverage such as frequency hopping. Both the effects of the repetitions and the frequency hopping are extensively analyzed in the thesis.

1.1.2 Satellite as an enabler of massive machine type communications

Until today, Satellite Communication (SatCom) systems have mainly consisted of large Geostationary Orbit (GEO) satellites serving a large portion of the Earth and providing services such as broadcast, broadband, and backhaul to fixed, portable, and mobile vehicle-mounted terminals. Notably, some spectrum portions are dedicated only to SatCom, *e.g.*, L/S (mobile), C/Ku/Ka (fixed) [10]. Starting from the 90s, several constellations emerged for narrow-band global services via Iridium and Globalstar, which are now in their second generation. More recently (2014), a Medium Earth Orbit (MEO) constellation (O3b/SES) for broadband has been added to cover the equatorial region, and nowadays there is a race to conquer orbital positions at low altitudes to provide global coverage for high data rate services. Starlink and OneWeb are two examples of mega-constellations deployed in Low Earth Orbit (LEO). From a system perspective, satellite operators have remained separate from terrestrial telecom operators [11].

5G New Radio (NR) represents the opportunity for a satellite/terrestrial integration and to provide services across these two domains. Indeed, satellites can extend the coverage into regions not economically viable for terrestrial infrastructures and to provide resilient backup to terrestrial services. Moreover, thanks to the frequency reuse, dynamic resource allocation, and onboard processing, both GEO and LEO satellites have increased to up to 1 Terabit/s the capacity that they can offer. Despite the large coverage and the increased capacity, it is crucial to define some commonalities of standards between satellites and terrestrial networks, in order to allow the former to play an integrated role in 5G and 6G communication systems. In the past, satellites used their own air interface standards, such as Digital Video Broadcasting (DVB), DVB - Satellite Second Generation (DVB-S2), DVB-S2 Extension (DVB-S2X), [12]. However, in recent years, satellites have joined the 3GPP group, which is responsible for developing terrestrial communication standards, *e.g.*, 5G and 6G standards. The current rollout of 5G networks is based on 3GPP Rel. 16, which does not include satellite technology. To address this, a Non-Terrestrial Network (NTN) group has been established within 3GPP as part of Rel. 17 and continues into Rel. 18+, with the goal of integrating satellite technology into 5G and 6G standards by 2025 and 2030, respectively. If by 2025, the aim is to integrate the NTN component

into the terrestrial network, by 2030 the goal is to establish the NTN as an inclusive element in the unified 6G standard. Indeed, this component is not only able to provide very high throughput and low latency services by means of Very LEO (VLEO), that effectively and smoothly integrates with future 6G terrestrial telecommunication networks providing growing capacity per system, as well as flexibility and agility, but also to feed the massive IoT (mIoT), boosting the IoT services beyond the boundaries imposed by the current terrestrial infrastructures. Along these lines, within Rel.17, a new Study Item (SI) started at the beginning of 2021 with the aim to study the integration of mMTC and NTN. In addition, several start-ups have arisen in recent years intending to provide NB-IoT connectivity on a global scale. OQ Technology [13] is one of them, which has already launched its first satellite at LEO [14]. Furthermore, Sateliot [15] is another company that has recently signed a partnership agreement with GateHouse Telecom for delivering 5G IoT services from space [16], [17]. Moreover, Mediatek [18] has successfully completed a field trial of 5G IoT connectivity utilizing the Inmarsat's Alphasat L-band satellite [19]. The above-mentioned examples clearly highlight that satellite-based technologies, able to be incorporated into the existing IoT terrestrial networks, are the way to go [20].

1.1.3 The need for new algorithms

Considering the large number of IoT verticals, *e.g.*, industrial IoT, transportation and logistics, smart cities, smart agriculture, smart buildings, smart oil and gas, and connected healthcare, the requirements are manifold and diversified. Based on the nature of the deployed application, IoT can be divided into two main categories with very distinguishable requirements: Massive and Critical IoT. The former is characterized by billions of low-cost and low-energy devices spread throughout the globe, able to generate small traffic volumes reporting information to the core network on a regular basis (*e.g.*, sensors in a smart home, smart metering, environmental sensor, etc.). The latter has to cope with ultra-high reliability, availability, and extremely low latency communications (*e.g.*, remote surgery, health care, tactile Internet, traffic safety *etc.*). In this thesis, the mIoT has been prioritized, thus, in the following, its requirements are summarized.

1.1.3.1 Low Cost and Complexity

Lowering the device complexity and using low-cost hardware components (targeting a total cost of a few tens of euros) are key enablers for mass-market applications. Design characteristics, such as single receive RF chain, restricting supported peak data rates to the maximum required by IoT applications, and reducing supported data bandwidth, help cutting down the production and deployment costs.

However, these key requirements for IoT applications could result in some limitations for the integration with satellite systems and the terminal design should be developed according to the target SatCom scenario to serve. For example, in

order to achieve these goals, the use of integrated circuits for signal processing, digital/analog conversion, and possibly RF functions is critical. The User Equipment (UE) local oscillator must be cheap while still providing acceptable performance in terms of phase noise and frequency stability. The power amplifier range typically varies from a few milliwatts to less than 1 W of radio frequency peak power depending on the frequency band and type of service. Finally, omnidirectional patch antennas are used whenever possible to reduce manufacturing and installation costs, [21]; however, this could lead to some difficulties in closing the link budget due to the lower gain. Another possible source of price increase could be the presence of a Global Navigation Satellite Systems (GNSS) receiver for the delay and Doppler pre-compensation. Software Defined Radio (SDR) can help improve the functionalities and performance of devices extending their life through over-the-air updates, thus reducing their cost.

1.1.3.2 Energy Efficiency

Since most IoT devices are battery-powered and are expected to be operational for a very long period without human intervention, achieving a high level of energy efficiency for IoT UEs is of paramount importance, in particular when they are deployed in remote areas. The main sources of battery drain in the IoT UE can be identified in: i) power consumption during the activity period, *i.e.*, transmission and reception phases; and ii) Idle/Sleep mode power consumption. While the latter is strictly related to the hardware characteristics, for the former different optimization solutions could be followed. Ideally, the device should only be activated for the minimum time required to send the information bits, minimizing packet retransmissions. This implies that signaling information during idle/sleep status is minimized. When switched to active status, the UE should have a fast (re-)acquisition time and exploit a protocol that minimizes the energy required for transmitting the information bits.

However, keeping the cost of the device low, using low-quality oscillators, and reducing battery consumption by minimizing signal processing makes synchronization more difficult. To overcome these issues and adapt to the satellite scenarios, further information, (*e.g.*, GNSS information, like satellite constellation ephemeris, to compute satellite visibility windows) should be made available for the UE. Besides that, the adoption of an energy-efficient air interface could lead to a reduction in terms of battery consumption. Despite these optimizations, the required transmission power for closing the link could lead to increased energy consumption, which must be considered. This is particularly true for GEO scenarios, which would require UEs equipped with specific outdoor units to aggregate and transmit sensors' data.

1.1.3.3 Support for Massive Number of Devices

For what concerns the UE density, depending on the location and on the type of application, some cells will be more densely populated than others (*e.g.*, up to thousands of devices per square kilometer) and, therefore, IoT connectivity solutions should be able to simultaneously handle most of these connected smart devices [22]. Notably, satellites, thanks to their inherently large footprint, can provide access to a vast number of users. However, due to the sporadic uplink data reporting of NB-IoT applications, incomplete LEO constellations are the baseline approach. In this context, all the devices within the coverage area must be served within a relatively short time (2-4 min), due to the fast movement of the satellite. Since the time/frequency resources are limited, this results in a critical and problematic situation, as the number of IoT devices within the LEO coverage is very large due to the very wide coverage of the satellite. The congestion generated during the signaling phase of message flows between the devices and the base station can lead to the deterioration of the system performance or even to the unavailability of the service. Thus, congestion management techniques and scheduling algorithms are mandatory to enable the protocol to work properly.

1.1.3.4 Extreme Coverage

The extension of the coverage is a major design requirement for mIoT connectivity when considering applications in remote areas and, in general, in areas characterized by a very poor coverage (*e.g.*, assets tracking, solar, oil&gas harvesting, environment monitoring, mining, *etc.*). Satellites can provide worldwide coverage for IoT devices, in densely populated areas as well as in remote ones, providing, especially in the latter, a cost-efficient solution with respect to other terrestrial technologies for UEs interconnection and communication with the rest of the network. According to the scenario and the satellite constellation parameters (*e.g.*, GEO with regional/global coverage, LEO mega-constellation with global coverage or LEO constellations with discontinuous coverage), different type of IoT services could be enabled. However, an in-depth link budget analysis must be conducted to verify that the requirements in terms of Block Error Rate (BLER) are satisfied to permit reliable communications.

1.1.3.5 Latency

Although most of the IoT applications, falling into the Massive IoT category do not require stringent latency requirements (*e.g.*, smart-cities/home, smart power systems, agriculture/environment monitoring, *etc.*), there are some applications, falling into the Critical IoT applications (*e.g.*, remote healthcare, traffic/industrial control, tactile Internet, *etc.*) for which low latency is of paramount importance. For the latter, the delay constraints of GEO could make it impossible to meet the requirements.

Regarding Non-GEO scenarios, it could be possible to meet stringent latency requirements if the satellite constellation and ground segment are properly designed,

such as to allow a continuous connection among the UEs and the Gateways (GWs), either directly or through Inter Satellite Linkss (ISLs). For what concerns these scenarios, it is worth highlighting that the short satellite visibility windows (*e.g.*, few minutes depending on the satellite altitude) and the fast handover (*e.g.*, from few seconds up to few minutes according to the satellite altitude and antenna pattern design) among different cells, in the case of moving beam reference scenario, should be taken into account. This happens in particular when continuous satellite coverage cannot be ensured and the UE connection to the network must be guaranteed within a few opportunity windows throughout the day. Moreover, it is worth highlighting that, according to the chosen air interface and architecture, a proper evaluation of the impact of the propagation delay on protocol procedures and timers should be performed.

From the above-mentioned requirements, it is clear that the design of these extremely complex systems requires manifold analyses at different levels of abstraction, from satellite constellation and ground segment architecture aspects to the evaluation of the air interface behavior, in order to allow the system to operate. To this aim, the main objective of the thesis is to assess the impact of the NTN channel on the Random Access (RA) Procedure and the data phase of NB-IoT and to propose new algorithms to allow the integrated system to properly work, meeting the NB-IoT requirements in terms of RA's success probability, Bit Error Rate (BER), communication latency, and throughput, counteracting the SatCom impairments with minimal modifications to the NB-IoT standard.

1.2 Thesis contributions and organization

The study proposed in this thesis can be mainly categorized into three main activities: i) link budget analysis considering the NB-IoT terminal and the new satellite configuration defined in TR 36.763 [23]; ii) assessment and verification of point-to-point communication links using metrics such as detection probability, estimation of the synchronization parameters, BER for a given SNR or Signal-to-Interference plus Noise Ratio (SINR); iii) assessment of system/network level performance with different deployment and configurations. The Key Performance Indicator (KPI)s include the percentage of users successfully completing the RA procedure and the time taken to conclude such procedure, the spectral efficiency of the system, and the number of satellite passages needed to serve all the users in the considered area.

The work is organized as follows:

- Chapter 2 provides an overview of the NB-IoT protocol, mainly focused on the Physical Layer (PHY) and Medium Access Control (MAC) layers characteristics.
- Chapter 3 describes all the main aspects regarding the SatCom systems architecture, with particular emphasis on NTN architectures proposed in the 3GPP

standardisation process. Later, the main impairments typical of the NTN channel are described. A literature review of the works related to satellite-based NB-IoT systems is provided and the contributions of this thesis in covering the literature gaps are highlighted.

- Chapter 4 describes the mathematical framework for the Link Budget computation. Then, this model is used to compute the downlink and uplink budget analysis in NB-IoT via NTN network both in single-satellite multi-beam and multi-satellite multi-beam scenarios.
- Chapter 5 details the work performed at the link level. In particular, this chapter is divided into two main sub-sections: the first one describes the detection and estimation algorithms of orthogonal random access preambles and their synchronization parameters; while in the second part, collisions are considered due to the fact that multiple users select the same preambles. To overcome the issue of high packet losses caused by the congestion, Non Orthogonal Multiple Access (NOMA) approaches, with and without Neural Network (NN)s algorithms are proposed. In detail, a NN based algorithm is designed to detect the NB-IoT preambles, with and without collisions, classify the number of users using the same resources, and for each of them estimate their Time of Arrival (ToA). Following, the performance of a classical NOMA method is evaluated in the NB-IoT NTN scenario, considering satellite channel coefficients between the users and the flying platform and the different constraints of the NB-IoT standard.
- Chapter 6 provides a system layer perspective of the SatCom-based IoT system: in the first paragraph, the performances of the NB-IoT RA over NTN are thoroughly analyzed varying the number of uplink repetitions, RA periodicity, and the Back-Off (BO) parameters based on the different satellite configurations proposed by the 3GPP. Then, a joint access and data phases design is considered, by designing a heuristic algorithm to achieve the highest spectral efficiency.
- Finally, Chapter 7 concludes the thesis.
- In the Appendix A an application of NB-IoT is reported. The objective of the activity is to track and localize a train with centimeter-level accuracy allowing the development of a more sophisticated signaling system along the railway. For that purpose, a critical IoT setup must be considered to comply with safety-critical regulations usually applied in rail transportation. The communication must provide low latency and worldwide coverage in all possible scenarios and weather conditions. Thus, this specific application scenario has been selected to show current NB-IoT and cellular limitations relying only on terrestrial infrastructure. In the chapter, the performance of the sensor board in different environments is provided in terms of the accuracy of the localization.

List of Publications

- [P1] "Federated Cell-Free MIMO in Non-Terrestrial Networks: Architectures and Performance", A.Guidotti, A.Vanelli-Coralli, C.Amatetti, *IEEE Transaction on Aerospace and Electronic System* (**Submitted**)
- [P2] "O-RAN Based Non-Terrestrial Networks: Trends and Challenges", R. Campana, C.Amatetti, A.Vanelli-Coralli, *EuCNC 2023*
- [P3] "Proof of Concept for Spectrum Sharing Between Terrestrial and Satellite Networks" H.Kokkinen, A.Piemontese, A.Kivinen, L.Kulacz, N.Borios, C.Amatetti, *EuCNC 2023*
- [P4] "Towards the Future Generation of Railway Localization Exploiting RTK and GNSS", D. Mikhaylov, C.Amatetti, T.Polonelli, E. Masina, R.Campana, K. Berszin, C.Moatti, D.Amato, A. Vanelli-Coralli, M.Magno, L.Benini, *IEEE Transactions on Instrumentation Measurement* 2023
- [P5] "Coverage and Interface in-cochannel spectrum sharing between terrestrial and satellite networks", H.Kokkinen, A.Piemontese, L.Kulacz, F.Arnal, C.Amatetti, *IEEE Aerospace Conference 2023*
- [P6] "Neural Network based Non-Orthogonal Random Access for 6G NTN-IoT", C.Amatetti, R.Campana, A.Georganaki, A.Vanelli-Coralli, *IEEE GLOBECOM 2022*
- [P7] "Towards the Future Generation of Railway Localization and Signaling Exploiting sub-meter RTK-GNSS", C.Amatetti, T.Polonelli, E.Masina, C. Moatti, D.Mikhaylov, D.Amato,A.Vanelli-Coralli, M.Magno, L.Benini, *IEEE SAS 2022*
- [P8] "Location-assisted precoding in 5G LEO systems: architectures and performances", A.Guidotti, C.Amatetti, F.Arnal,B.Chamaillard, A. Vanelli-Coralli, *EuCNC/6G Summit 2022*
- [P9] "NB-IoT random access procedure via NTN: system level performances", C.Amatetti, M.Conti, A.Guidotti, A.Vanelli-Coralli, *IEEE ICC 2022*
- [P10] "Preamble detection in NB-IoT via Satellite: a Wavelet-based approach", C. Amatetti, M. Conti, A. Guidotti, A. Vanelli-Coralli, *IEEE GLOBECOM 2021*
- [P11] "NB-IoT over Non-Terrestrial Networks: Link Budget Analysis", M. Conti, A. Guidotti, C. Amatetti, A. Vanelli-Coralli, *IEEE GLOBECOM 2020*

Book chapter

"NTN support for mMTC: Architectural and channel model considerations" C.Amatetti, A.Guidotti, A.Vanelli-Coralli. *Wiley/IEEE "Integrating Machine-Type-Communication (MTC) and Satellites for IoT: Towards 6G" 2023*

Tutorial

"NB-IoT over Aerial and Space Networks: Technology Overview, Challenges, and Potential Solutions", C. Amatetti, M. Conti, A. Guidotti, A. Vanelli-Coralli, M.Kish, M.S., Alouini, *IEEE GLOBECOM 2021*

Chapter 2

Narrowband IoT

This chapter aims at providing an overview of the NB-IoT protocol, mainly focusing on the PHY and MAC layers characteristics. NB-IoT is a cellular communication standard conceived on extensive reuse of the LTE system. Indeed, NB-IoT reuses the LTE modulation schemes for downlink and uplink transmissions, *i.e.*, the Orthogonal Frequency Division Multiplexing (OFDM) and the single carrier-frequency division multiple access (SC-FDMA), respectively [7]. The occupied bandwidth is 180 kHz, divided into 12 subcarriers, each of 15 kHz, in downlink; while in the uplink, the number of sub-carriers depends on the transmission mode, *i.e.*, single tone or multi-tone, as reported in Table 2.1. With the aim to ensure a harmonized co-existence with the LTE carriers, the 3GPP defined three operation modes for NB-IoT:

- In-band mode: the NB-IoT signal uses one Physical Resource Block (PRB) of the available LTE bandwidth.
- Guard-band mode: the NB-IoT signal takes one PRB from the LTE bandwidth's unused guard band.
- Stand-alone mode: the NB-IoT signal is designed to occupy the GSM system's spectrum. In this case, the NB-IoT signal occupies 180 kHz of the 200 kHz GSM carrier, with a band guard of 10 kHz on both sides of the spectrum.

Within Release 13, the NB-IoT system was first conceived to operate in a frequency-division duplexing (FDD)-based mode. This means that downlink and uplink transmissions take place in distinct frequency bands at the same time. However, due to the low power, low complexity, and battery life constraints of the NB-IoT modules, the system at the terminal side operates in a type B half-duplex FDD (HD-FDD) mode [24]. In this way, the UE can either transmit or receive data, but not at the same time. Furthermore, a guard time interval is added between transmissions to allow the device to switch between transmission and reception and vice versa.

2.1 Physical layer

NB-IoT foresees three channels and five signals in downlink and only two channels and one signal in uplink. Before describing each of them in detail, an overview of

TABLE 2.1: NPUSCH resource units [7]

NPUSCH format	Δf	N_{sc}^{RU}	N_{slot}^{RU}	Total Slot duration (T_{data})	Modulations	N_{symbol}^{UL}
1	3.75 KHz	1	16	32	$\frac{\pi}{2}$ BPSK, $\frac{\pi}{4}$ QPSK	7
		16	8	8		
	15KHz	3	8	4	QPSK	
		6	4	2		
2	3.75 KHz	1	4	8	$\frac{\pi}{2}$ BPSK	
		12	2	1		
	15	1	4	2		

the operations performed at the PHY layer is provided, *i.e.*, cell selection/reselection, System Information (SI) acquisition, uplink synchronization, downlink and uplink data transmission. Cell selection is the process of identifying, synchronizing with, and evaluating the suitability of an NB-IoT cell for communication. It includes initial cell selection, non-initial cell selection, and cell re-selection. From a physical layer perspective, one of the main differences between initial and non-initial cell searches is the amount of Carrier Frequency Offset (CFO) that the device has to cope with when synchronizing to a cell [25]. Indeed, the initial cell selection is carried out by the device before it has any knowledge of the network or has established any prior synchronization with it. This can occur, for example, when the device is switched on for the first time and needs to search for a cell in a public land mobile network. In this case, cell selection has to be performed in the presence of a large CFO due to the potential initial oscillator inaccuracy of the device, which for a low-cost device module may be as high as 20 ppm. The NB-IoT cell selection procedure generally consists of the following steps:

1. Searching for the NPSS (Narrowband Primary Synchronization Signal) to detect the presence of an NB-IoT cell.
2. Synchronizing in time and frequency with the NPSS to determine the carrier frequency and the subframe structure within a frame.
3. Identifying the PCID (Physical Cell ID) and the three least significant bits (LSBs) of the SFN (System Frame Number) using the NSSS (Narrowband Secondary Synchronization Signal). The relative subframe between the NSSS and NPSS is used to determine whether the cell is an FDD (Frequency Division Duplex) or TDD (Time Division Duplex) cell.
4. Acquiring the MIB-NB (Master Information Block - Narrowband) to determine the complete SFN as well as the two LSBs of the H-SFN (Hyper-System Frame Number), and to resolve the frequency raster offset. The MIB-NB also provides information on how the SIB1-NB (System Information Block 1 - Narrowband) is transmitted. In a Time Division Duplex cell, the MIB-NB includes information on whether the SIB1-NB is transmitted on the anchor or non-anchor carrier. In an FDD cell, the SIB1-NB is always transmitted on the anchor carrier.

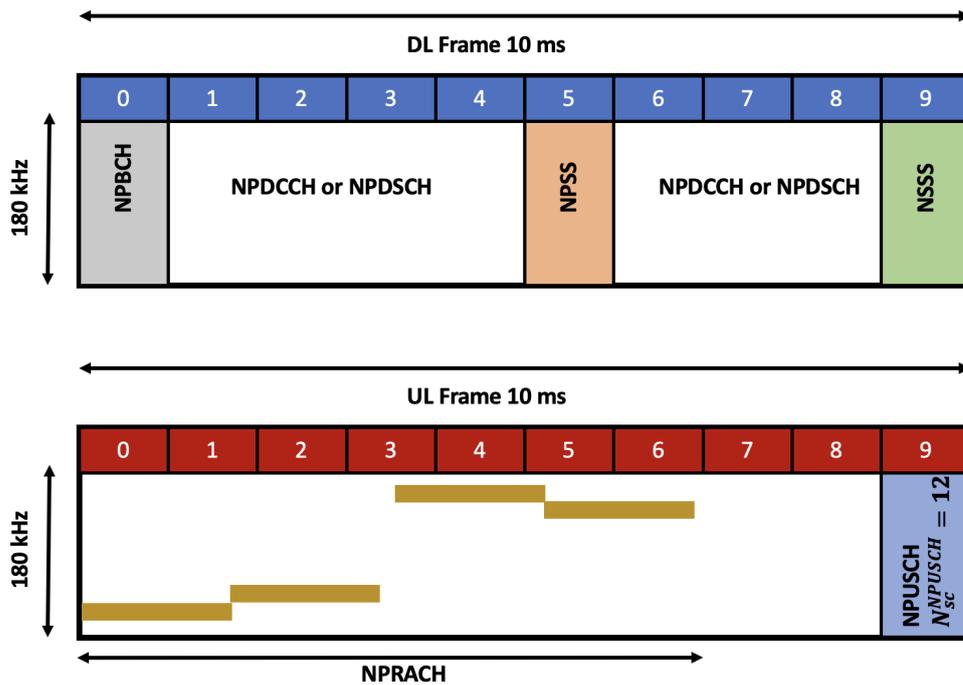


FIGURE 2.1: Uplink and Downlink channels.

5. Acquiring the SIB1-NB to identify the complete H-SFN, the PLMN (Public Land Mobile Network), tracking area, and cell identity, and to prepare for verifying the suitability of the cell.

After the downlink synchronization, the BS (base station) is not aware of the users within its cell. Therefore, to be identified with a permanent identification (ID), users must initiate the access phase, known as the RA procedure, using the Narrowband Physical Random Access Channel (NPRACH). Moreover, this procedure is essential to establish strict time synchronization among various UEs (by means of the NPRACH preamble) and to obtain the resource for uplink data transmission by receiving the control information on the Narrowband Physical Downlink Control Channel (NPDCCH). Finally, after the synchronization and the access phases, the UE is able to establish the data phase, *i.e.*, it can transmit its uplink data using the narrowband physical uplink shared channel (NPUSCH), or receive the downlink data utilizing the narrowband physical downlink shared channel (NPDSCH). The UL and DL channels are illustrated in Fig. 2.1. The organization of the UL frame is only an example.

2.1.1 Downlink

As previously highlighted, the NB-IoT system's PHY layer is derived from LTE. However, to fit the constraints and requirements of the NB-IoT system (for example, channel coding, modulation, mapping, and so on), many LTE PHY blocks that make up the channels and signals were modified.

It is worth emphasizing that the BS can only use two antenna ports. Furthermore,

the quadrature phase-shift keying (QPSK) modulation format is the only one that can be employed for NB-IoT downlink transmissions. The specific characteristics of each channel and signal are presented in the following.

1) Narrowband Primary Synchronization Signal: this is the BS's first essential signal and it enables the UE to perform time and frequency synchronization *i.e.*, to determine the start of the frame and remove the frequency offset caused by its low-cost oscillator. For this reason, the NPSS is based on a Zadoff-Chu (ZC) sequence, which is characterized by a high correlation property. The NPSS is always transmitted in the sixth subframe (*i.e.*, subframe #5) of each frame. The mapping to resource elements is performed between subcarriers $k = 0$ and $k = 10$, and between OFDM symbols $l = 3$ and $l = 13$.

2) Narrowband Secondary Synchronization Signal is the BS's second essential signal. It only contains cell identity (cell ID) information and is carried in the 10th subframe, which means subframe #9, of each even frame. This signal is based on a 132-ZC elements sequence multiplied by a Hadamard sequence, which correspond to 11 OFDM symbols multiplied by 12 subcarriers.

3) Narrowband Reference Signal (NRS), also known as "pilot," is used to estimate channels in the frequency domain. Except for those dedicated to NPSS and NSSS, it is always transmitted in all subframes. The NRS is mapped over the 6th, 7th, 13th, and 14th OFDM symbols of each subframe. The frequency position of NRS Resource Element (RE) is determined by the cell ID. The NRS is made up of QPSK-like complex elements with values $\frac{1}{\sqrt{2}}(\pm 1 \pm j)$, generated by means of the "gold sequence", which are pseudorandom sequence. The positions of the REs dedicated to the NRS are critical in NB-IoT, because the mapping process of all downlink channels should avoid them.

4) Narrowband positioning reference signal (NPRS) is used for downlink Observed Time Difference of Arrival (OTDOA) based positioning accuracy. The sequence is generated as follow:

$$r_{l,n_s}(m) = \frac{1}{\sqrt{2}}[(1 - 2 \cdot c(2m)) + j(1 - 2 \cdot c(2m + 1))] \quad (2.1)$$

with $m \in (0, 1, \dots, 2N_{RB}^{max,DL} - 1)$ and $N_{RB}^{max,DL}$ is the resource block number, n_s is the slot number within a radio frame, l is the OFDM symbol number within the slot; finally c_i is a Gold sequence with length 31.

5) Narrowband wake up signa (NWUS) is a type of power saving mechanism which allows the UE to continue to sleep (*i.e.*, No Wake up) even for DRX OnDuration period when there is no data for the UE and gNB notifies the UE of 'No Wake Up'.

6) Narrowband Physical Broadcast Channel (NPBCH) is the first channel to be decoded by the NB-IoT modules. It is always transmitted in subframe #0 and contains the MIB-NB. The latter includes the fundamental information required by UE to receive further system information, such as the SIB1-NB.

7) Narrowband Physical Downlink Shared Channel is used to send system information blocks and user data. This channel can be mapped over any downlink subframe except those assigned to NPBCH, NPSS, and NSSS (*i.e.*, subframes #0, #5, and #9 if the NSSS is used).

8) Narrowband Physical Downlink Control Channel is required to prepare the data transmission phase. The NPDCCH in NB-IoT is responsible for transmitting control information from the network to the UEs. In this regard, the BS can send various types of control information, including i) Scheduling of downlink resources; ii) data acknowledgement; iii) Uplink resource allocation (*i.e.*, uplink resource scheduling); and iv) direct indication and paging. The control information is stored in a logical block known as Downlink Control information (DCI). There are three types of DCI formats defined in NB-IoT: N0, N1, and N2. Each of them identifies the type of data to be exchanged between the BS and the UEs.

2.1.2 Uplink

The uplink channels and signals are based on the LTE standard, but they are optimized to work with NB-IoT modules. As previously described, the SC-FDMA scheme is employed for uplink transmissions; furthermore, most of the LTE UE transmission chain blocks remained unchanged, but the possible configurations to be used have been reduced (*e.g.*, sets of modulation formats). The main differences between LTE and NB-IoT occur at two levels: the random access channel and the data/control channel. Indeed, a new waveform for the NB-IoT preamble was designed at the random access level and data and control information were combined into a single channel rather than having separate channels.

1) Narrowband Physical Random Access Channel is used to carry the NB-IoT preamble, which is the first signal sent by the UE to the BS in order to request access to the network. The physical layer random access preamble is transmitted over 180 KHz bandwidth and it is based on the single carrier frequency hopping symbol group (SG) [7], as shown in Fig. 2.2. This represents the basic unit of the preamble (or Msg1) and it consists of L identical symbols (with length T_{SEQ}) plus the cyclic prefix (CP). In the 3GPP standard, a given number P of SGs is treated as the basic unit of the preamble. This basic unit can be repeated up to $N_{rep} = 2^j, j \in \{0, 1, \dots, 7\}$ times for coverage extension. Accordingly, the length of a preamble equals $P \cdot N_{rep}$ SGs. Three different CP lengths are supported by the standard, each one defining a preamble format (as detailed in Table 2.2). Besides the CP length, the formats differ also for the subcarrier spacing Δf , SG sizes, and repetitions, as described hereafter. The NPRACH formats have been designed to accommodate the requirements of different targeted cell sizes. It is worth highlighting that format 0 and format 1 differ only for the CP length. In fact, the former is designed to accommodate a cell radius up to 10 km, while the latter up to 40 km. Every symbol group is transmitted on different sub-carriers following a hopping rule. It is worth emphasizing that in the case of preamble Format 0 and 1, the maximum number of available preambles is 48. The

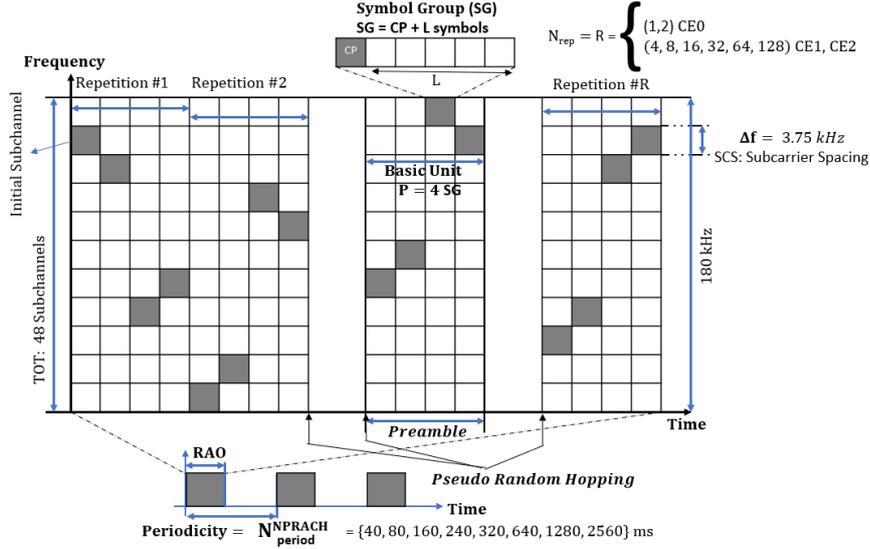


FIGURE 2.2: NPRACH preamble and periodicity.

minimum number of sub-carriers that can be assigned by the BS to send the preamble is 12. In the case of Format 2, being $\Delta f = 1.25 \text{ kHz}$, there are 144 sub-carriers in 180 kHz, thus, 144 preambles are available.

TABLE 2.2: NPRACH preamble formats for FDD mode

Format	Δf	P	L	T_{CP}	T_{SEQ}
0	3.75 kHz	4	5	$66.67 \mu\text{s}$	1.33 ms
1	3.75 kHz	4	5	$266.67 \mu\text{s}$	1.33 ms
2	1.25 kHz	6	3	$800 \mu\text{s}$	2.4 ms

The NPRACH transmission supports either a 3.75 kHz or a 1.25 kHz sub-carrier spacing (SCS, *i.e.*, Δf). The frequency location of the SG follows an **hopping pattern**. The latter is fixed within the basic unit of P SGs and it is constrained in twelve consecutive sub-carriers for Format 0 and 1, while for Format 2, the hopping is constrained in 36 consecutive sub-carriers (this block of consecutive sub-carriers is indicated as N_{sc}) [7]. Symbol groups in preamble format 0 and 1 (with 3.75-kHz SCS) hop by one or six sub-carriers in frequency, whereas symbol groups in format 2 (with 1.25-kHz SCS) hop by one, three, or eighteen sub-carriers in frequency [7]. Note that when repetitions are configured, the hopping between the basic units is no longer fixed, but it follows a pseudo-random selection procedure defined in [7]. The mechanism of frequency hopping consists of computing the pseudo-random sub-carrier index, *i.e.*, $n_{sc}(i)$. In [7], $n_{sc}(i)$ has the following expression:

$$n_{sc}(i) = n_{start} + \tilde{n}_{sc}(i) \quad (2.2)$$

where $n_{start} = N_{offset} + \lfloor n_{init}/N_{sc} \rfloor N_{sc}$. N_{offset} is an offset index to reduce the collisions among the users; n_{init} refers to the first sub-carrier chosen by the user among the 48 (or 144) available for the RA. In the following, the formula to compute

$\tilde{n}_{sc}(i)$ for preamble Format 0 and 1 is provided :

$$\tilde{n}_{sc}(i) = \begin{cases} (\tilde{n}_{sc}(0) + \frac{f_i}{4}) \bmod N_{sc} & \text{if } i \bmod 4 = 0 \text{ and } i > 0 \\ \tilde{n}_{sc}(i-1) + 1 & \text{if } i \bmod 4 = 1, 3 \text{ and } \tilde{n}_{sc}(i-1) \bmod 2 = 0 \\ \tilde{n}_{sc}(i-1) - 1 & \text{if } i \bmod 4 = 1, 3 \text{ and } \tilde{n}_{sc}(i-1) \bmod 2 = 1 \\ \tilde{n}_{sc}(i-1) + 6 & \text{if } i \bmod 4 = 2 \text{ and } \tilde{n}_{sc}(i-1) < 6 \\ \tilde{n}_{sc}(i-1) - 6 & \text{if } i \bmod 4 = 2 \text{ and } \tilde{n}_{sc}(i-1) \geq 6 \end{cases} \quad (2.3)$$

where $\frac{f_i}{4}$ is used to compute the first sub-carrier of each repetition of the same preamble according to the external pseudo-hopping pattern, which depends on $cell_{ID}$, as specified in [7]; finally $\tilde{n}_{sc}(0) = n_{init} \bmod N_{sc}$. It is worthwhile emphasizing that the hopping pattern depends on the first sub-carrier, *i.e.*, n_{init} , which identifies the preamble. This means that if two users chose the same n_{init} their preambles are completely overlapped, leading to a collision.

2) Narrowband Physical Uplink Shared Channel is used for transmitting data and control information from the UEs to the network. There are two formats available for this purpose: format 1 for data and format 2 for control information, such as acknowledgement (ACK) and non-acknowledgement (NACK) messages. The NPUSCH can be configured in different ways to adjust the robustness of the signal according to the conditions of the transmission environment. These transmission configurations are summarized in Table 2.1. Note that Δf and N_{sc}^{RU} stand for the subcarrier spacing and the number of subcarriers, respectively. The different combinations of the number of sub-carriers and slots define the Resource Unit (RU), *i.e.*, the smallest unit that can be sent by the UE.

The main difference between NPUSCH format 1 and 2, apart from the purpose, is that in the latter, a repetition code is used instead of turbo coding. This code involves repeating the ACK/NACK bit (1 for ACK, 0 for NACK) 16 times. After this, the resulting bits are passed to the channel interleaving and scrambling to add randomness to the transmission. The resulting binary message is then mapped to complex numbers using only phase shift keying (PSK) with a shift of 90° degrees between the two phases ($\frac{\pi}{2}$ BPSK) and transmitted over a RU consisting of four slots.

The mapping of the complex numbers to physical resources for both formats follows a "frequency-first" rule, where all the available consecutive sub-carriers of a given symbol are mapped from lowest to highest frequency before moving on to the next non-pilot symbol.

3) Demodulation Reference Signal (DMRS), also known as "uplink pilot", is used for estimating the channel in the frequency domain for uplink transmissions. Unlike the NRS used in the downlink, DMRS elements are not multiplexed within a symbol but are instead assigned to the N_{sc} sub-carriers of an entire symbol (one or three symbols per slot can be used for uplink pilots). These pilot symbols are also referred to as "midambles."

The values of the DMRS sequence depend on the NPUSCH configuration. Thus, when the single-tone transmission is employed, the DMRS is generated by means of the following sequence:

$$\bar{r}_u(n) = \frac{1}{\sqrt{2}}(1 + j)(1 - 2c(n))w(n \bmod 16) \quad (2.4)$$

where $n = 0, \dots, N_{seq} - 1$ with N_{seq} indicating the size of the base sequence. $c(n) \in \{0, 1\}$ is created with a pseudo-random gold sequence and $w(n)$ is a Hadamard sequence of length 16. The length of N_{seq} corresponds exactly to the number of slots to be transmitted for a given NPUSCH message (including all the repetitions of the message). The $w(n)$ values are chosen as a function of the values of the u parameter (where $u = N_{cellID} \bmod 16$). Therefore, from the base sequence in Eq. 2.4, the DMRS sequence is obtained in the following way: in the case of NPUSCH format 1, $r(n) = \bar{r}_u(n)$, i.e., each pilot value is defined from a unique base sequence value. In case of NPUSCH format 2, $r(3n + m) = \bar{r}_u(n)\bar{w}(m)$ with $m = 0, 1, 2$. This means that the three pilots in a given slot are defined from the same base sequence value. Regarding the multi-tone transmission, the base sequence is:

$$r_u n = e^{j\alpha n} e^{j\frac{\pi}{4}\phi(n)} \quad (2.5)$$

where $n = 0, \dots, N_{sc}^{RU} - 1$. $r_u n$ generates the values of each pilot symbol carried over the used subcarriers in a given slot. The α parameter is called ‘‘cyclic shift’’ and varies for $N_{sc}^{RU} = 3$ or 6 and it is equal to 0 when $N_{sc}^{RU} = 12$. The other parameter $\phi(n)$ depends on the number of sub-carriers, i.e., if $N_{sc}^{RU} = 3$ or 6 or 12.

2.2 Medium Access Control layer

The MAC layer interfaces directly with the PHY layer. As such, it performs its functions in almost real-time [26]. The main functions of the MAC layer include:

- Random access procedure;
- HARQ;
- Scheduling of signalling and data, and allocation of resources.

2.2.1 Random Access procedure

The RA procedure is a crucial procedure at the MAC layer and at the UE, as it is the first procedure that is initiated by the terminal when it has data to transmit. The main purpose of the RA procedure is to establish uplink synchronization and obtain an uplink grant for sending the RRC connection request and NAS Attach procedures. In order to reduce the signaling exchange between the UE and the BS, three different RA procedures are supported, named 4-step RA, Early Data Transmission, and 2-step RA. In the following each of them is described.

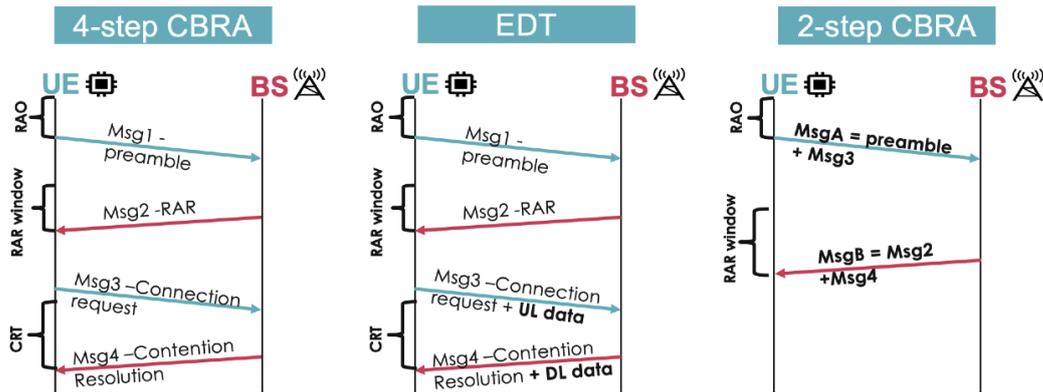


FIGURE 2.3: From the left to the right: 4-step RA procedure, EDT, and 2-step RA.

2.2.1.1 4-step RA

This is a handshake procedure between the UE and the BS and consists of four messages: Msg1 (RA Preamble), Random Access Response (RAR) (Msg2), RRC connection request (Msg3) transmitted in the UL grant, and RRC Contention Resolution message (Msg4). Basic information about the RA procedure is broadcast to the UE in the SIB2-NB (System Information Block 2 - Narrowband) [26].

The RA procedure can be initiated by i) the UE (through its MAC layer) when it needs to move from idle to connected state, or ii) by the BS that needs to connect to a UE in idle mode by sending a message on the NPDCCH ordering the UE to initiate an RA procedure.

The RA procedure can be either contention-based (CBRA), if started by the UE, or contention-free (if initiated by the BS orders) and can be performed on the anchor (*i.e.*, where downlink synchronization happens) or non-anchor carrier. It consists of two sub-procedures: random access and contention resolution phases. Each cell is divided into one to three Coverage Enhancement (CE) levels, which are designed to accommodate devices with different levels of channel quality. The device determines to which coverage area it belongs based on its measured reference signal received power (RSRP). A single NPRACH resource on the anchor carrier and zero or one NPRACH resource on each non-anchor carrier are assigned to the defined coverage enhancement level. The number of CE levels is equal to one plus the number of RSRP thresholds in the `rsrp-ThresholdsPrachInfoList` [27], which can be 1 or 2. An NB-IoT device with the best channel quality selects enhanced coverage level #0, while an NB-IoT device with the poorest channel quality selects enhanced coverage level #2.

Since the simultaneous accesses to the channel by a large number of NB-IoT devices can congest the RA channel, more NRACH resources are provided on the anchor or non-anchor carriers, or the UE is ordered to perform contention-free access. In order to perform the RA, the following parameters are defined at the MAC layer:

- the NPRACH resources;
- the RAR window size;
- the Contention Resolution Timer (CRT).

The NPRACH resources can be transmitted periodically in a given period of time and the periodicity indicates how often this pattern with resources is repeated. Clearly, the number of RA Occasion (RAO), *i.e.*, *NPRACH*, is inversely proportional to the length of the RA periodicity. Within this period, the uplink resources are dedicated to the NPRACH transmission for the configured period of time. Then, the remaining resources till the next opportunity are dedicated to the NPUSCH. The scheduling parameters of NPRACH (*i.e.*, the NPRACH resources) consist of: I) the periodicity (N_{period}^{NPRACH} , shown in Fig. 2.2) which determines the radio frame where an NRACH resource is available; ii) the starting time within the period; iii) the number of selected sub-carriers (N_{sc}) over which the UE can perform its preamble transmission; iv) the starting sub-carrier index within N_{sc} ; v) the number of preamble repetitions during one attempt N_{rep} (*i.e.*, within one period); finally iv) the maximum number of preamble transmission attempts that can be made in the corresponding CE level. Once the UE understands the CE class that it belongs to, it selects and transmits the NPRACH to the BS. The latter, if detects the incoming message, is able to estimate the Round Trip Delay (RTD) for each UE based on the ToA of the received preamble, and will use this estimate to determine the Timing Advance (TA) that each UE should apply. This helps to synchronize the uplink transmission from different UEs in time. It is worth noting that, in this step, the UEs will compete for the same NPRACH channel, since each UE is not aware of the other. In particular, if two UEs randomly select the same preamble sequence, a collision will occur, resulting in the failure of the RA procedure.

Once sent the Msg1, the UE waits for the RAR message from the BS. Indeed, it looks for a response in the NPDCCH for a certain number of frames after the end of the preamble transmission. This response window starts at the subframe that contains the end of the last preamble repetition plus 41 subframes (if $N_{rep} \geq 64$) or 4 subframes (if $N_{rep} < 64$) and has a length of RAR window size [28]. The UE looks for an NPDCCH DCI scrambled with RA-RNTI, where RA-RNTI is code to decode the RAR message calculated using the SFN_{id} (the index of the first radio frame of the NPRACH resource) and the k_{id} (the index of the UL frequency carrier ID), as follow:

$$RA - RNTI = 1 + \text{floor}\left(\frac{SFN_{id}}{4}\right) + 256 \cdot k_{id} \quad (2.6)$$

For the anchor carrier, the k_{id} is 0. If the UE receives a RAR in the response window scrambled with RA-RNTI, it means that the message is designated for this UE. This message contains an UL grant indicating available UL resources for the UE to transmit its Msg3. The RAR also comes with a Temporary C-RNTI, which is used to encode all transmissions on NPDCCH. If the UE successfully receives its designated

RAR, it concludes that the RA procedure has been completed successfully and begins the contention resolution process. If no RAR is received during the response window, the UE will start a BO procedure in which it will select a BO value at random and re-transmit a new RA. If a RAR contains a BO index message, the UE will set its BO value accordingly, otherwise, the UE will set it to zero.

After the UE successfully completes the random access procedure, it initiates the contention resolution process, by sending the Msg 3 on the resources scheduled in the RAR message. At the same time, it starts the CRT, during which, it looks for an NPDCCH transmission addressed to its temporary C-RNTI. If such a transmission is received, the UE decodes the corresponding data message on the DL-SCH. If the latter contains a contention resolution identity that matches the transmitted Msg3, the UE stops the timer and concludes that the contention resolution process has been successful, moving to the Connected state. If the message does not contain its identity, it re-starts the procedure after a BO time. It is worth highlighting that if two users transmit the same preamble, which has been detected by the BS, they will receive the same RAR (since the RA-RNTI is the same for both the users) and, therefore, will transmit the Msg3 on the same resources. Only one of the two terminals will receive Msg4, *i.e.*, the contention resolution message. It should also be noted that, with respect to the legacy LTE, the length of the two timers depends on the NPDCCH period, which, in turn, depends on the coverage enhancement zones to which the user belongs. The location of the NPDCCH in the resource grid is called search space, whose length is determined by the parameter R_{max} , which defines the maximum number of sub-frames where the search space extends. The interval between two consecutive NPDCCH is given by $R_{max} \cdot G$, where G is a system parameter which depends on the number of users to be served. This interval is called the NPDCCH period. The combination of R_{max} and G is critical for the scheduling of the resource [29]. For both the timers, the standard poses an upper bound to 10.24s [27]. The possible values of RAR window size and CRT are:

$$\begin{aligned} T &= G \cdot R_{max}, G \in \{1.5, 2, 4, 8, 16, 32, 48, 64\} \\ T_{RAR} &= T \cdot \{2, 3, 4, 5, 6, 7, 8, 10\} \\ T_{CRT} &= T \cdot \{1, 2, 3, 4, 8, 16, 32, 64\} \end{aligned} \quad (2.7)$$

where T is the NPDCCH period. The picture on the left in Fig. 2.3 shows the message exchange of the 4-step RA.

2.2.1.2 Early Data Transmission

The classical four-step random access procedure implies large signalling overhead due to the high number of signals exchanged by the BS and the UE. In Release 15, a new feature called Early Data Transmission (EDT) was introduced to improve power consumptions for UEs that only need to send infrequent small data packets [30] (shown in Fig. 2.3). During the random access procedure, EDT allows these UEs

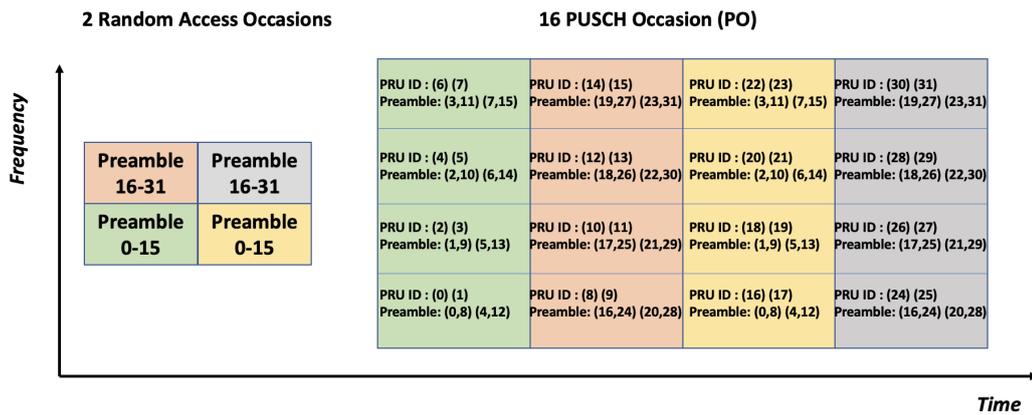


FIGURE 2.4: Example of the channel structure for msgA: mapping between preamble IDs and PRU IDs.

to transmit data in the Msg3. Naturally, this is beneficial for devices that have low power consumption requirements and only need to send small amounts of data infrequently. Indeed, EDT can be enabled only if the UE has a data packet smaller or equal to the broadcast (in the system information blocks) maximum data to be sent in uplink.

In order to start this mechanism, the UE indicates to the BS that it intends to transmit data in Msg3 by selecting a special preamble for the NPRACH channel. Then the procedure continues as in the legacy RA: if the BS detects the Msg1, it transmits the uplink grant with an appropriate Transport Block Size (TBS) in the RAR message. The device, after receiving the Msg2, replies to the base station by sending its identity and its own data. In Msg4, the BS can indicate the devices to be in idle or connected mode if there are other communications.

2.2.1.3 2-Step Random Access

A further enhancement in terms of power consumption and latency reductions is achieved with the 2-step RA procedure, introduced into Rel.16 for 5G NR [31]. As it is visible from Fig. 2.3, the 2-step RA combines the Msg1 and Msg3 into a one-shot transmission, named MsgA, and the BS feedback includes both the TA command and C-RNTI to complete the random access into MsgB (equivalent to the Msg2 + Msg4). After receiving MsgA, the BS can act differently based on the detected signal:

- if it detects the preamble from the UE and successfully decodes the payload, the BS notifies the UE of contention resolution by sending a successful RAR with a TA command.
- If the BS detects a single preamble but fails to decode the payload, it sends back a fallback RAR to the UE with the TA command and an uplink grant for the payload re-transmission, by using the preamble reception time.

- If the BS detects multiple identical preambles from UEs, there is no fallback RAR because the BS is unable to specify the preamble reception time of each UE. Therefore the it transmits a backoff indication to UEs that will attempt random access again.
- if the BS fails to detect the preamble, there is no RAR to the UE.

The channel structure of the MsgA needs to be defined and includes the mapping between the preamble on a PRACH resource and the time-frequency resource for the PUSCH payload, the size of the time-frequency resource for the PUSCH, and other information. 2-step Contention Based Random Access (CBRA) uses a different set of preambles than the legacy 4-step RA, but both sets of preambles can be transmitted in the same or separate Random Access Opportunity (RAO). The payload transmission in 2-step RA consists of PUSCH Occasions (POs) that span multiple OFDM symbols and Physical Resource Units (PRU). Each PO is made up of multiple PUSCH RUs that contain the following fields: i) PRU ID; ii) multiple OFDM symbols and PRBs for uplink transmission; iii) association with preamble(s) of a PRACH resource; iv) modulation and coding scheme; v) Uplink power control-related parameters; and vi) DMRS port and DMRS sequence. An example of channel structure for MsgA is provided in Fig. 2.4.

One PRU is linked to one or more preambles of a PRACH resource, meaning it is identified by the preamble ID(s) of a specific RAO. There are two ways in which the resource mapping can be established between the preambles of a PRACH resource and a RU: many-to-one mapping, *i.e.*, more than one PRACH resource can point to the same RU in the PUSCH channel; and one-to-one mapping, where there is a unique association between the PRACH and data resources. It is worth highlighting that, for the moment being, this procedure is only defined for 5G.

Notably, by reducing the number of messages between BS and UE, the signalling overhead, power consumption as well as latency, can be significantly reduced. However, the main issue of this procedure is that there is no prior information (timing and resource allocation) for MsgA, resulting in a contention-based asynchronous transmission when the cell size is relatively large.

2.2.2 HARQ

The HARQ is a mechanism used in NB-IoT to improve the reliability of the communication link. If the data transmitted over the channel (either downlink or uplink) is corrupted beyond the ability of the Forward Error Correction (FEC) code to correct the errors, NACK message is sent back to the sender, requesting them to re-transmit the same packet. If the receiver can decode the packet, an ACK is sent and the sender can proceed with transmitting the next packet in the queue This should not be confused with the repetition number, which refers to the number of times the same packet is transmitted. Indeed, HARQ is a combination of re-transmissions and error correction. On one hand, where possible, errors are corrected; on the other hand,

where correction is not possible, errors are detected and packet re-transmission is requested. The receiver tries to decode the packet based on current and previous transmissions. HARQ operates at both MAC and PHY layers: re-transmissions occur at the MAC layer, while the PHY layer at the receiver combines one or more transmissions to increase the chances of correct decoding. To improve system efficiency and maintain the same data rate, multiple parallel HARQ processes can be used, but this increases the complexity and cost of the devices. Therefore, a maximum of two parallel processes are allowed for NB-IoT devices.

2.2.3 Downlink and Uplink Scheduling

In NB-IoT, the transmission of NPDSCH only occurs after the receipt of an NPDCCH transmission containing a DCI that provides scheduling information for the NPDSCH. The latter can contain either system information data (such as SIB1-NB and other SIBs-NB) or user data (*e.g.*, random access messages and other data). If the NPDSCH carries data, the UE must first receive an NPDCCH transmission containing an N1 DCI to obtain the scheduling information for the NPDSCH. The scheduling parameters include: i) scheduling delay, *i.e.*, a delay between the end of the transmission of the DCI and the beginning of the NPDSCH transmission; ii) the resource assignment to obtain the number of subframes of the NPDSCH message; iii) the Modulation and Coding Scheme (MCS) to define the TBS size; finally, iv) the number of repetitions of the NPDSCH transmission. However, if the NPDSCH holds SIB1-NB data, the scheduling information can be acquired directly from the MIB-NB and the scheduling of other SIBs-NB can be obtained from SIB1-NB.

In uplink, the scheduling of NPUSCH transmissions is done through the NPDCCH using DCI formats N0 or N1, depending on the type of NPUSCH content to be transmitted (*e.g.*, data or control information). If the UE has data to transmit in the uplink, the BS will send a DCI format N0 to schedule the NPUSCH transmission. On the other hand, when the UE has control information to transmit, the BS will use the DCI format N1 to schedule the transmission, which typically occurs when the BS sends an NPDSCH message and needs feedback on the successful or unsuccessful reception of the message by the UE. The DCI format N0 carries six scheduling parameters related to NPUSCH transmission: i) scheduling delay, which represents the number of subframes the UE has to wait after receiving the NPDCCH. Its minimum value is 8 ms; ii) the sub-carrier indication which defines the sub-carriers where the NPUSCH data should be sent (N_{sc}^{NPUSCH}), depending on the transmission mode (*i.e.*, single or multi-tone); iii) resource assignment, which refers to the number of resource units required for data transmission; iv) repetition number, v) MCS, which allows obtaining other two information, such as the modulation order and TBS size. It is worth emphasizing that higher TBS values (shown in Fig. 2.5) are usually assigned to users in good channel conditions since their signal propagation path losses are smaller; vi) the redundancy version used for the HARQ.

<i>TBS Index</i> (I_{TBS})	<i>Resource Assignment (I_{RV})</i>							
	<i>0</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>
0	16	32	56	88	120	152	208	256
1	24	56	88	144	176	208	256	344
2	32	72	144	176	208	256	328	424
3	40	104	176	208	256	328	440	568
4	56	120	208	256	328	408	552	680
5	72	144	224	328	424	504	680	872
6	88	176	256	392	504	600	808	1000
7	104	224	328	472	584	712	1000	1224
8	120	256	392	536	680	808	1096	1384
9	136	296	456	616	776	936	1256	1544
10	144	328	504	680	872	1000	1384	1736
11	176	376	584	776	1000	1192	1608	2024
12	208	440	680	1000	1128	1352	1800	2280
13	224	488	744	1128	1256	1544	2024	2536

FIGURE 2.5: TBS of NPUSCH channel [32].

Chapter 3

Non Terrestrial Networks

SatComs are the outcomes of the joint effort of communications and space technology research with the aim of achieving ever-increasing ranges and capabilities at the lowest possible cost. This chapter describes the main aspects of SatCom systems, the typical impairments of the satellite channel, the standardization path of NTN, and the State-of-the-Art (SoA) of SatCom based NB-IoT systems.

3.1 SatCom System Architecture

A SatCom system, shown in Fig. 3.1, is organized in the following way: i) a *space segment*, containing one or different active and spare satellites organized into a constellation; ii) a *control segment*, including all ground facilities for the control and monitoring of the satellites, known as Tracking, Telemetry, and Command (TTC) stations, and for the management of the traffic and the associated resources on board the satellite; and iii) a *ground segment*, including all the traffic Earth stations.

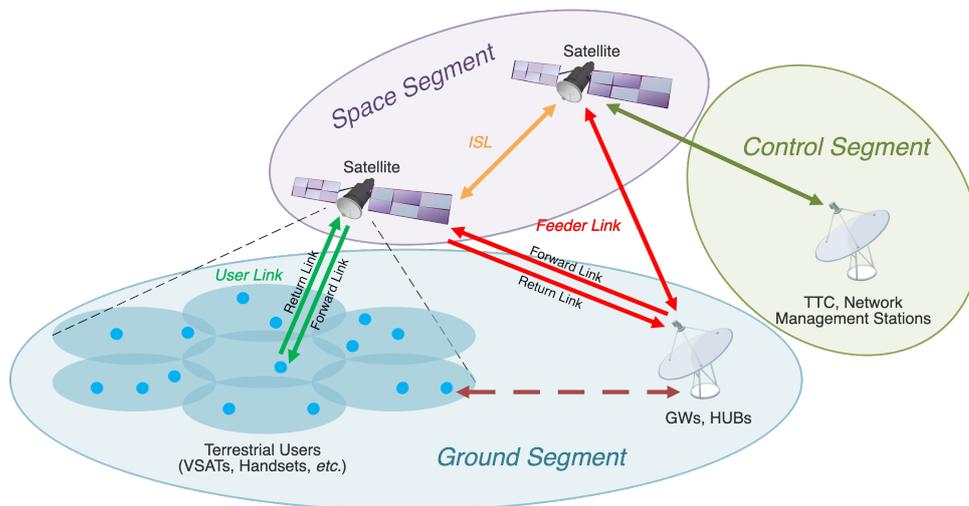


FIGURE 3.1: SatCom System high-level architecture.

Satellite networks are distinguished by their topology, the types of links they support, and the connectivity they offer among earth stations. In terms of configurations, a SatCom system could be organized: i) in a meshed network, where every

node¹ is capable of communicating with every other node. A meshed satellite network is made up of a set of earth stations that are able to communicate with one another using satellite links consisting of radio-frequency carriers. ii) In a star network, where each node is able to communicate only with a single central node, named hub. iii) In a multi-star topology, various hubs (*i.e.*, central nodes) are identified. The other nodes can communicate only with those central nodes.

3.1.1 Space Segment

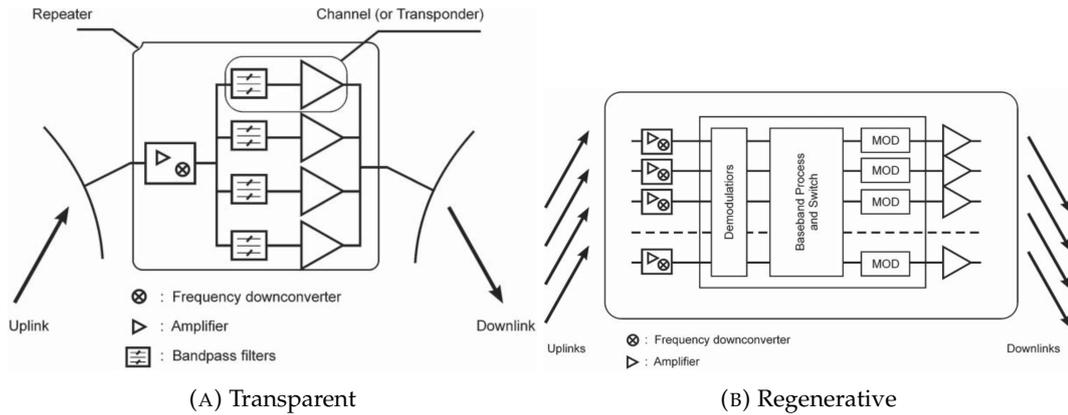


FIGURE 3.2: Types of payload [33].

As mentioned above, in the space segment one or several active and spare satellites orbit the Earth. Each of them is composed of the *payload* and the *platform*. The former is equipped with receiving and transmitting antennas and all the electronic equipment supporting the signal transmissions. Fig. 3.2 shows the two types of payload: *transparent* and *regenerative*.

A transparent payload, shown in Fig. 3.2a, (also known as bent pipe) only acts as a mirror between the elements of the ground segment, since it only amplifies the carrier power and down converts the central frequency. The power gain is of the order of 100-130 dB to increase the power level of the received carrier from a few tens of picowatts to the power level of the carrier fed to the transmitting antenna. To improve the separation between the receiving and the transmitting signals, frequency conversion is necessary. Moreover, since the payload is power limited, the overall bandwidth in the satellite is split into different sub-bands, and the carriers in each sub-band are amplified by a dedicated power amplifier. The amplifying chain associated with each sub-band is called a satellite channel, or *transponder* [33]. The sub-bands are obtained by means of a series of filters called the *input multiplexer* (IMUX). The amplified carriers in the output are then recombined in the *output multiplexer* (OMUX). The transparent payload, shown in Fig. 3.2a, could be carried either on i) a single-beam satellite, where each antenna generates one beam only; or ii) in a multi-beam satellite, where multiple-beam antennas generate multiple beams.

¹A node is either a redistribution point or a communication endpoint. A satellite, as well as a Ground Station (GS) or a ground terminal, can act as network nodes.

Referring to Fig. 3.2b, a multiple-beam regenerative payload is depicted. Differently from the bent-type payload, the uplink carriers are demodulated and re-routed to the output. Having the baseband signals allows *on-board processing* and routing of information from upbeam to downbeam through *on-board switching at baseband*. The frequency conversion is performed by modulating carriers, generated on board, at downlink frequency, which are then amplified and delivered to the destination in downlink.

Each beam identifies a beam coverage area, also known as *footprint*, on the Earth's surface. In the case of multi-beam, the aggregate beam coverage areas define the coverage area. If a satellite has several multiple-beam antennas, its coverage defines the satellite access area, which is directly related to the Field of View (FoV) of the satellite. Moreover, there are two types of beams: "Earth-fixed" and "Earth-moving". "Earth-fixed" beams, also named steerable beams, are used to cover stationary areas on the surface of the Earth that have fixed boundaries. These beams can be directed to different locations on the Earth. In contrast, "Earth-moving beams" are used in systems where the beam generated by the payload moves constantly as the spacecraft moves along its orbit.

In terms of coverage, it is possible to distinguish between *instantaneous* and *long-term* system coverage (shown in Fig. 3.3). The first one includes the aggregated coverage areas at a given time of the individual satellites participating in the constellation. The latter defines the area on the Earth scanned over time by the satellites in the constellation. When considering real-time services, the instantaneous system coverage should at any time have a footprint covering a geographical region where at least an element belonging to the Ground Segment (GS) is present. For store-and-forward services, this condition is not necessary, however, long-term coverage of the service zone is necessary. In order to provide continuous global coverage, a large number of satellites is required for LEO and VLEO satellite constellations. With GEO systems, three satellites are enough to achieve global coverage (except for the polar areas).

The platform consists of all the subsystems that allow the payload to operate, such as: i) the fuel system which is responsible for making the satellite run for years; ii) the solar panels that provide the required energy for the satellite's operation; iii) the satellite and telemetry control system used to transmit commands to the satellite as well as to send the status of the onboard systems to the ground stations.

3.1.2 Ground Segment

Earth stations are categorized into the following classes: i) *user stations*, e.g., handsets, mobile stations, and Very Small Aperture Terminals (VSATs), which allow the user to directly access the space segment; ii) *interface stations*, also named gateways, which link the space segment to a terrestrial network; and iii) *service stations*, like

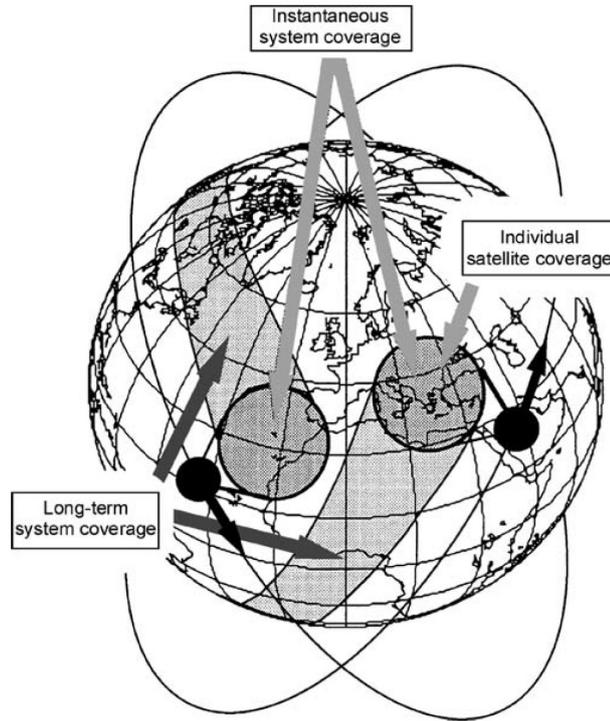


FIGURE 3.3: Types of coverage [33].

hub or feeder stations, which are in charge of collecting or distributing information from and to user stations through the space segment.

The ground segment includes all the earth stations connected to the end user's terminal by a terrestrial network or, in the case of small station VSATs, directly connected to the end user's terminal. According to the volume of traffic on the satellite link and the type of traffic, Stations have different sizes.

3.1.3 Communication Links

All the entities involved in SatCom system communicate by means of the following links:

- DLs from the satellites to the GSs and UEs;
- ULs from GSs and UEs to the satellites; and
- *ISL* between the satellites.

ULs and DLs consist of radio-frequency modulated carriers, while *ISL* can be either radio frequency or optical. Some large-capacity data-relay satellites also use optical links with their ground stations. Carriers are modulated by baseband signals conveying information for communications purposes. Connections among end users include an UL and a DL, and possibly one or several *ISLs*.

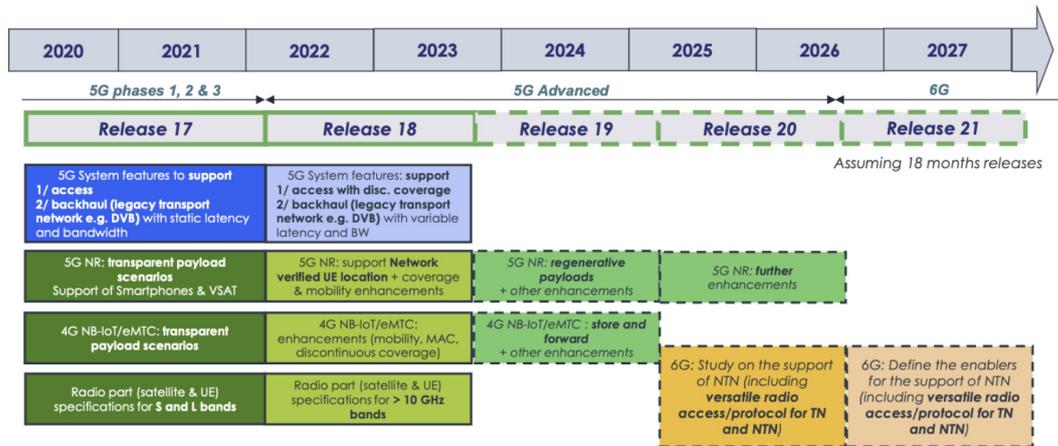


FIGURE 3.4: NTN 3GPP roadmap [34].

A satellite network can establish two types of links: i) *unidirectional links*, where one or more stations only transmit and other earth stations only receive; and ii) *bidirectional links*, where earth stations both transmit and receive. In satellite broadcast-oriented networks, unidirectional links are typically associated with a star topology, whereas bidirectional links can be associated with a star or meshed topology and are required to transport two-way telecommunication services. Furthermore, different types of connectivity can be defined in a SatCom network depending on the type of service (e.g., point-to-point for unicast, point-to-multipoint for multicast/broadcast, multipoint-to-point for multiplexing and concentration, and multipoint-to-multipoint).

A link can be categorized based on the direction of the communication, *i.e.*, can distinguish *forward* and *return* links, as shown in Fig. 3.1. A connection between a service provider and a user is established via a hub (for collecting services) or a feeder station (for broadcasting services, for example). A forward connection is a connection made from a gateway, hub, or feeder station to a user terminal. The return connection is the reverse connection. Both forward and return connections require a UL, a DL, and possibly one or more ISL. The latter can be classified in the following way: i) Inter-Orbital Links, e.g., links between GEO and LEO satellites; ii) Intra Orbital Links, e.g., links between GEO satellites and links between LEO.

3.2 The NTN path in 3GPP standardization

The 3GPP recognized the importance of NTN and included it in its Rel.17 to explore how the inclusion of NTN component enables planned 5G services and to push toward 5G-Advanced that will lead to Sixth-generation (6G) systems. Fig. 3.4 shows the roadmap of NTN standardisation, proposed in [34], covering up to Rel. 21.

The 3GPP started the work on NR over NTN in 2017 with two SIs entitled i) "Study on NR support for Non-Terrestrial Networks", and ii) "Study on using the satellite access in 5G". The former, led by the Radio Access Network (RAN) group with the support of RAN1 (*i.e.*, related to Layer 1) focused on the deployment scenario and

adaptation of the 3GPP channel models for NTN, identifying the potential key impact areas on the NR and proposing solutions for the identified impacts on the RAN protocols and architectures. The outcomes of this study were documented in the Technical Report (TR) 38.811 [35]. Following this initial study, the 3GPP focused on the definition of the use cases for Satellite-based NR as part of the SI "Study on NR support for Non-Terrestrial Networks" under the supervision of Systems Aspects (SA) working group with the support of the SA1 (Service). This SI, started in 2017 but finalized into Release 16 and associated with the Work Item (WI) on "Integration of Satellite Access in 5G", led to the definition of three main categories of use cases for satellite-based NTN:

- **Service Continuity:** the terrestrial network alone cannot provide 5G services, thus the 5G system shall support service continuity between 5G terrestrial access network and 5G satellite access networks. The NTN use cases mapped to the Terrestrial Network (TN)- NTN service continuity include stationary UE (enhanced Mobile BroadBand (eMBB)), Pedestrian UE (eMBB), Machine terminals (mMTC), stationary/ vehicular relay UE (eMBB), relay UEs on vehicles, ships, or high-speed train;
- **Service Ubiquity:** the aim of this use case is to extend the terrestrial network coverage into un-served or under-served geographical areas. Typical Examples of ubiquity use cases are mMTC (*e.g.*, agriculture, asset tracking, metering), public safety (*i.e.*, emergency networks), and home access;
- **Service Scalability:** this use case foresees the multicast or the broadcast of content to a large area by leveraging the large coverage area of satellites. An example of this use case is the distribution of rich TV content (*i.e.*, Ultra High-Definition TV).

After concluding Rel.15 on the scenario and channel models for NR to support NTN, 3GPP focused on both the system and architecture aspects and access technologies with a follow-up Rel. 16. With respect to the former, three activities were started within Rel. 16: i) a SI on "Study on architecture aspects for using satellite access in 5G" within the Working Group 2 (Architecture); ii) a Work Item (WI) on "Integration of Satellite Access in 5G", under the supervision of Working Group 1; and iii) a SI on "Study on management and orchestration aspects with integrated satellite components in a 5G network", within SA5 (Management). The SA1 identified the critical areas related to the integration of NTN in NR and provided solutions and requirements when considering the use cases described in TR 22.822 [36]:

- Roaming between terrestrial and satellite networks;
- 5G fixed backhauling between satellite-enabled NR-RAN and the 5G Core.

These requirements were added directly to the existing SA1 5G specification, TS 22.261 "Service Requirements for the 5G System" [37]. In addition, the SA5 addressed aspects related to management and orchestration for NTN, whose results

are reported TR 28.808 [38].

The studies on the access technologies, guided by the RAN3 group (Interface), are captured in the SI "Study on solutions for NR to support NTN" completed at the end of 2019. These activities defined a baseline for NR functionalities aimed at supporting LEO and GEO satellites. Indeed, within this Release, a set of required adaptations enabling NR technologies and operations in the NTN context were identified, covering several issues in RAN1 (Physical layer), RAN2 (Layers 2 and 3), and RAN3. In particular, the performance assessment of NR over GEO and LEO satellites was provided at both system and link levels, together with a preliminary set of potential solutions for NR adaptations at Layers 2 and 3. Moreover, some architecture aspects were modified with respect to TR 38.811 [35], which is superseded from this point of view by TR 38.821 [39].

Based on the outcome of the Release 16 study, 3GPP initiated the following activities within Release 17: i) a WI named "Solutions for NR to support non-terrestrial networks" within the RAN2; ii) a SI named "Architecture Aspects for Using Satellite Access in 5G" under the supervision of SA2; iii) SI "Study on PLMN selection for satellite access" led by the Core and Terminals (CT) working groups WG1, WG3, and WG4; and iv) SI on "Narrow-Band Internet of Things (NB-IoT) / enhanced Machine Type Communication (eMTC) support for Non-Terrestrial Networks (NTN)" under RAN1. The activities within the WI are devoted to the normative work and aim at specifying the necessary enhancements for LEO and GEO while also targeting support for High Altitude Platform Systems (HAPS) and air-to-ground networks. This involves the physical layer aspects, protocols, and architecture as well as the radio resource management, Radio Frequency requirements, and frequency bands to be used. The WI is based on the following principles:

- Transparent payload architecture;
- Earth fixed tracking areas with Earth fixed and moving cells;
- FDD;
- The type of terminals supported are handheld devices in Frequency Range 1 (FR1), *i.e.*, below 7.125 GHz, and VSAT with external antenna operating in Frequencies band above 7.125 GHz (FR2). The terminals are assumed to have GNSS capabilities.

The WI specified feature enhancements in RAN1, RAN2, RAN3, and RAN4. In particular, in the first two Working Groups, the objective is to address issues related to long propagation delays, large Doppler Shift effects, and moving cells in NTN. This implies enhancements on the PHY and MAC layers aspects, both on the UP and Control Plane (CP). The last two Working Groups focus on architectural enhancements, such as feeder link switchover and cell-related aspects, as well as aspects related to the UE Radio Resource Management (RRM) and RF requirements. SA WG2, which encompasses the system architecture, started its satellite studies in Rel. 17, with the

creation of the SI on architecture aspects for using satellite access in 5G. The scope of this work was to identify key issues for satellite-based NTN in the 5G system architecture and provide solutions for both direct satellite access and satellite backhaul. The result of the work is captured in the Rel. 17 Technical Report TR 23.737 [40]. In November 2019, SA2 approved the Rel. 17 normative work associated with this TR. The aspects that are being addressed by SA2 in the normative work are as follows:

- Mobility Management with large coverage areas and with moving coverage areas;
- Delay in satellite;
- QoS with satellite access and with satellite backhaul;
- RAN mobility with NGENEO regenerative-based satellite access;
- Regulatory services with super-national satellite ground station;

3GPP Core and Terminals workgroups CT WG1, WG3, and WG4 created a study item to address aspects of 5GC (5G core network) architecture for satellite networks. The work in CT working groups concentrates on aspects related to PLMN selection, and the results are captured in Technical Report TR 24.821 [41]. The study phase is completed and the normative work will follow, with the Stage 2 solutions and PLMN selection for satellite access.

Moreover, Rel.17 also marks the beginning of the standardization of mMTC via NTN. Indeed, the 3GPP approved a SI during the RAN meeting #86 and revised it during RAN meeting #91 which enables the IoT operations via NTN. This SI identified the scenarios applicable to NB-IoT and eMTC use cases under the RAN1 and RAN2, and based on these scenarios, identified the recommended changes to support NB-IoT and eMTC over satellite, reusing as much as possible the conclusions of the studies performed for NR- NTN in TR38.821 [39]. The SI is based on the following assumptions:

- UEs are equipped with GNSS;
- IoT characteristics specified up to Rel. 16 are supported;
- the NB-IoT single-carrier and multi-carrier operation are enabled.

The outputs of the study are reported in TR 36.763 [23]. The latter reports aspects related to RA, general aspects related to timers, mobility, Radio Link Failure, HARQ operation, and time and frequency adjustments. In terms of the Core Network, both the Evolved Packet Core (EPC) and the 5G Core (5GC) are assumed to be supported. The evolution of mMTC via NTN continues within Release 18 and it has started in March 2022 with a WI led by RAN2 named "Enhancement solutions for NB-IoT &

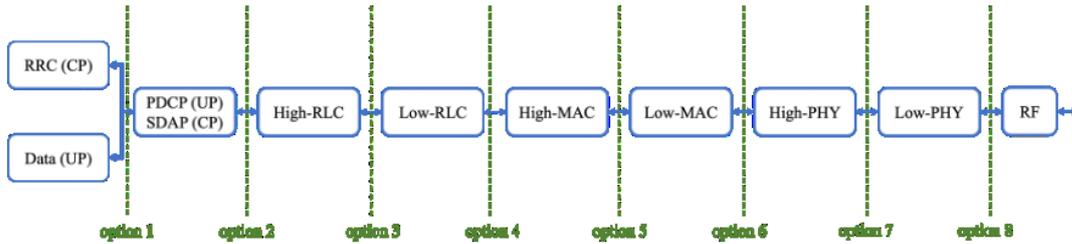


FIGURE 3.5: Available functional splits.

eMTC to support non-terrestrial networks". Besides, Rel.18 foresees also enhancements for eMBB. In particular, the following features will be addressed: i) network-based UE location determination; ii) coverage enhancements; iii) NR-NTN deployment above 10 GHz and support for VSAT/ESIM terminals; iv) NTN-TN and NTN-NTN mobility and service continuity enhancements. Then, for 5G-A, some of the techniques de-prioritized in Rel.18 can be evaluated as candidate features for Rel. 19 and beyond [34].

3.3 NTN system architectures for 5G

When considering the NTN integration into 5G systems, the RAN needs to communicate with the CN through the space segment, which includes the space-borne, *i.e.*, satellite-based communication platforms, and the air-borne, *i.e.*, HAPS. Thus, this heterogeneous system (*i.e.*, NR-NTN) envisages the mapping of these two entities to offer services.

The 5G RAN, *i.e.*, the Next Generation Radio Access Network (NG-RAN), is described by its own interfaces and architectures, which are reported in TR 38.801 [42]. In detail, the 5G NR conceives the functional split of the gNodeB (gNB) in Centralised Unit (CU) and Distributed Unit (DU). The goal of the functional split is to improve scalability, adaptability to various use cases and vertical services, and performance in network management. It is also the baseline approach for Network Function Virtualization (NFV) and Software Defined Networking (SDN). As stated in TS 38.401: i) a gNB can be split into a CU and one or more DU; ii) a DU communicates with only one CU; iii) the DU and CU are connected through the F1 air interface, which is a logical air interface. It is implemented by means of any existing standard, as long as specific signaling operations are ensured [43]. According to TR 38.801, [42], the split of the gNB can be performed in eight different ways shown in Fig. 3.5.

In 5G-NTN, the 3GPP identified six different macro scenarios (shown in Table 3.1) for the NTN providing access to the UE based on the type of orbit, payload, and its capabilities. In all of them, the link between the UE and the satellite (*i.e.*, the user service link) works either in S-band (*e.g.*, 2 GHz) or in Ka-band (*e.g.*, 20 GHz in the downlink and 30 GHz in the uplink). Moreover, LEO satellites can be distinguished based on whether they provide fixed or moving beams for coverage on the ground.

It is worth emphasizing that only scenarios D1 and D2 include ISL, as shown in Table 3.1.

TABLE 3.1: NTN reference scenarios per system type [39]

System	Transparent	Regenerative
GEO	A	B
LEO steerable beams	C1	D1
LEO fixed beams	C2	D2

Therefore, based on i) the location of the gNB, thus, based on the type of payload, either transparent or regenerative, and ii) the type of user access link, either direct or relay-based, it is possible to distinguish several architectural options, described in the following sub-sections.

3.3.1 Direct access

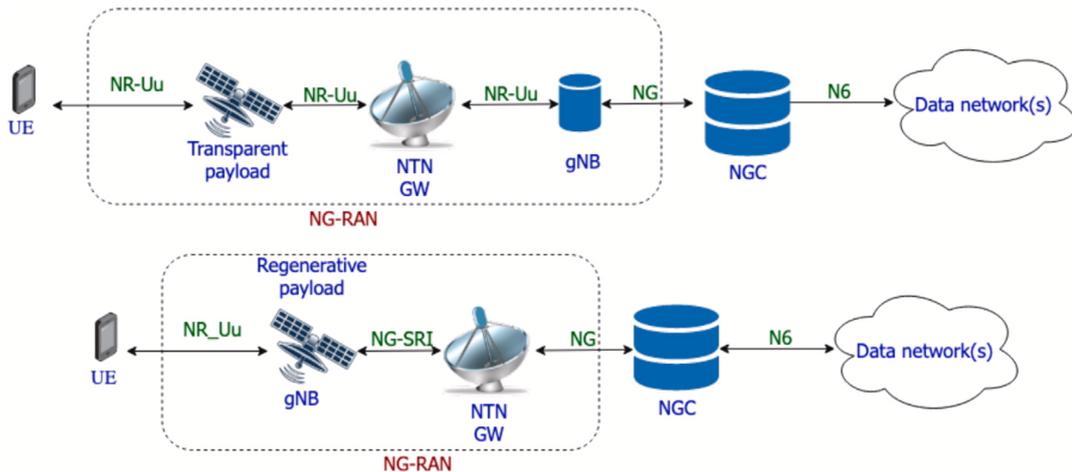


FIGURE 3.6: Architectural options with direct user access with transparent (top) and regenerative (bottom) payload [44].

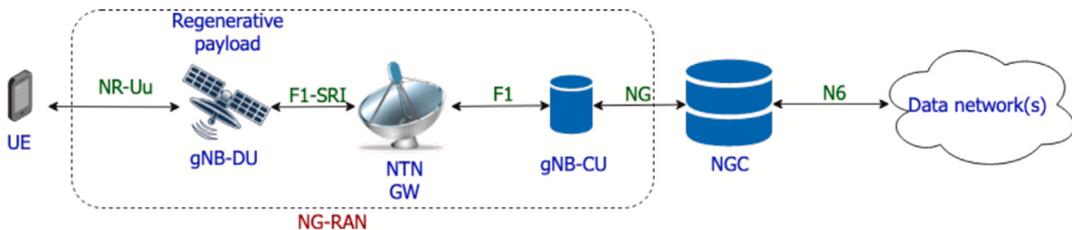


FIGURE 3.7: Architectural option with direct user access, regenerative payload, and functional split. Applicable to B, D1, and D2. [44].

In Fig. 3.6 the architectures with direct access are shown. In particular, the top diagram depicts a transparent payload and the bottom one a regenerative payload. Clearly, in this architecture, the users are directly connected to the satellite. In the transparent architecture, the platform only relays the NR signal between the

gateway and the user terminal. Here, the gNB is conceptually located at the gateway, which means that all NR procedures are completed on the ground. The NR-Uu air interface is implemented on both links. Since it is a terrestrial interface, the typical impairments of the satellite channel, such as the large Round Trip Delay and the Doppler shift, must be taken into account on both the feeder and user links. In terms of the number of connections, each gNB is capable of managing a few tens of beams. Since in multi-beam NTN systems, each satellite might be able of generating even hundreds of beams, depending on the mission requirements, multiple gNBs might be needed in order to manage the NTN node. In case the satellites generate a reduced number of beams, then a single gNB can serve multiple payloads.

The regenerative architecture foresees the gNB or part of it on the flying platform. Even if it is a more expensive and complex solution, it allows for greater flexibility in adapting to the NTN channel.

When the full gNB is implemented on board, the NR-Uu protocols are entirely terminated on-board, *i.e.*, this Air Interface is only present on the user service link. The GW basically acts as a Transport Network layer node, terminating all transport protocols and connecting to the 5GC and the onboard gNB via the NG interface. This Air Interface is logical, *i.e.*, it can be implemented by means of any Satellite Radio Interface (SRI), as, for instance, the DVB-S2, DVB-S2X, or DVB-RCS2, [45].

As previously mentioned, the regenerative payload also allows for the possibility of implementing a functional split between the DU on the satellite and the CU on the ground (shown in Fig. 3.7). Currently, the adopted solution for the split is option 7.2, in which the user access link uses the Uu air interface and the DU and CU are connected by the F1 air interface on the feeder link. Usually, one DU is responsible for managing multiple CUs. However, it is worth noting that a single satellite can generate hundreds of beams, so when multiple satellites in a swarm serve the same area, multiple CUs are necessary to manage the swarm. This architecture poses a challenge related to the F1 interface (*i.e.*, the interface between the CU and the DU) since it requires a persistent connection between the gNB-DU and the gNB-CU and it cannot be closed and re-activated on-demand; as such, with moving satellites as in a non-stationary scenario, all of the connections towards the served UEs would be dropped once the satellite is outside of the visibility of the current gNB-CU. Thus, smart implementations of the F1 interface and/or the functional split in NTN shall thus be designed.

3.3.2 Integrated Access and Backhaul

In a relay-based access solution, the UEs do not directly connect to the gNB, but to an Integrated Access and Backhaul (IAB) node. The IAB allows NR access technologies to be used not only on the link between the gNB and terminals but also on wireless backhaul links. The architecture, in Fig. 3.8, includes one IAB donor node and multiple IAB nodes. The donor node provides connectivity to the CN, to which it is connected by means of conventional methods, such as fiber or microwaves, and

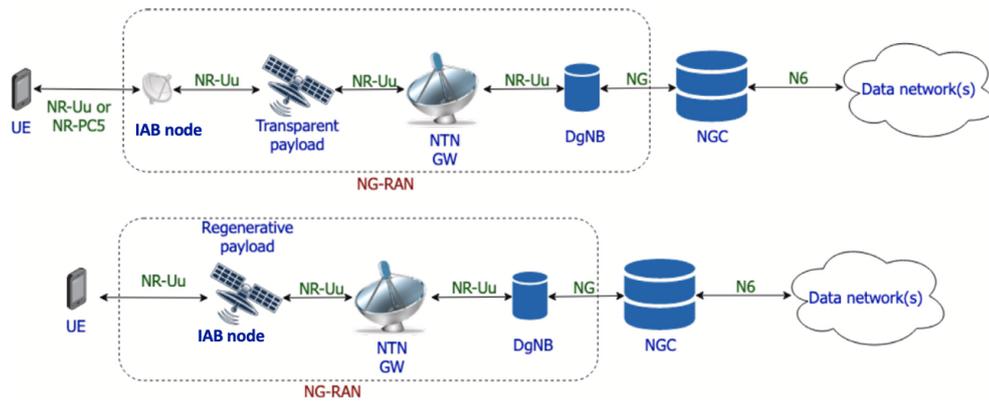


FIGURE 3.8: Architectural options with IAB with transparent (top) and regenerative (bottom) payload. [44].

serves the IAB nodes and the terminal directly linked to it. Each IAB node consists of a Mobile Termination (MT) part and a DU part. The former connects the IAB node to a parent DU, which could be the IAB donor or the DU part of another node, while the DU part communicates with UEs or the MT part of child IAB nodes. It is meaningful to highlight that: i) the MT part acts as a UE when communicating with the parent DU. ii) From the UE perspective, the DU part works as a normal gNB. In fact, it is not able to differentiate between normal gNBs and IAB nodes, because the IAB node's architecture is transparent to the user.

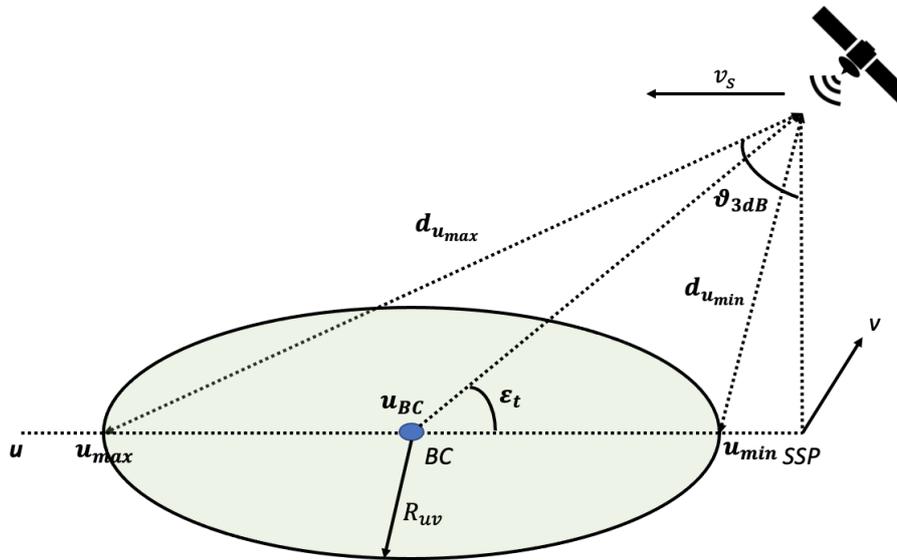
When considering an NTN element, the IAB can be implemented using either a transparent or regenerative payload. The architecture with relay nodes and transparent payloads (top diagram in Fig. 3.8) is more complex compared to direct access scenarios due to the introduction of a potentially large number of on-ground IAB nodes acting as gNBs. Because the IAB node can terminate protocols up to layer 3, no modification is foreseen on the user service link. In this case, the impact of typical satellite channel impairments only needs to be assessed on the forward link. When an onboard IAB node is implemented, *i.e.*, the regenerative solution is used, and the payload cost increases but there is the advantage of terminating protocols up to layer 3 on the satellite, as in the regenerative direct access architecture. For the feeder link, the connection between the IAB node and the donor gNB must be implemented using an NR-Uu air interface [46]. However, these architectures are currently under investigation for NTNs [39].

3.3.3 IoT NTN architectures

According to the latest specification of the 3GPP, in the mMTC scenario, the device is directly connected to the transparent payload, with the user access link always in S-band (*i.e.*, 2 GHz). The scenarios identified by the 3GPP for NB-IoT via NTN are reported in Table 3.2. In addition to the two orbits already proposed for 5G NR, *i.e.*,

TABLE 3.2: NTN reference scenarios for IoT [23]

System	Transparent
GEO	A
LEO steerable beams	B
LEO fixed beams	C
MEO fixed beams	D

FIGURE 3.9: Reference system for delay and Doppler in u - v coordinate system.

LEO and GEO, this scenario foresees also the inclusion of MEO satellites. Moreover, compared to the satellite configurations proposed for NR-NTN in the TR 38.821, the other two configurations namely set 3 and set 4, for LEO and GEO satellites are considered, where the satellite beam diameter is extended. This is due to the fact that mMTC, and in particular NB-IoT, does not support RRC-connected mode mobility. Therefore, when small beams are considered, the expected frequent beam switching will force a device to move back and forth between RRC idle and RRC connected states. A further solution to decrease the beam switching frequency is to consider steerable beams at the expense of the SNR.

3.4 NTN impairments

This section reports some considerations about the delay, Doppler, and path losses elaborated in "NTN support for mMTC: Architectural and channel model considerations" [47].

Notably, compared to a terrestrial channel, satellite links pose challenging issues related to large delays, Doppler shifts, and path losses.

In the following, the system shown in Fig. 3.9 is considered as the baseline, where the following assumptions hold: i) a beam center located on the uv -plane at (u_{BC}, v_{BC}) ; ii) a beam radius $r_{uv} = \sin \vartheta_{3dB}$, where ϑ_{3dB} is the 3 dB half beam width, defined based on the NTN configuration sets in TR 38.821 [39]; and iii) a satellite orbiting the spherical Earth at an altitude h_{sat} . The conversion between (u, v) coordinates and satellite look angles is given by:

$$\begin{cases} u = \sin \vartheta \cos \varphi \\ v = \sin \vartheta \sin \varphi \end{cases} \quad (3.1)$$

$$\begin{cases} \vartheta = \arcsin \left(\sqrt{u^2 + v^2} \right) \\ \varphi = \arctan \left(\frac{v}{u} \right) \end{cases} \quad (3.2)$$

For the sake of clarity, only a single beam is considered, but this assumption does not impact the generality of the proposed mathematical framework. In fact, when assessing the impact of the channel impairments on the NB-IoT procedure, it shall be noticed that such procedures involve users from the same beam.

3.4.1 Delay

To characterize the satellite channel in terms of latency, it shall be noticed that: i) compared to terrestrial communications, the propagation delay is predominant; and ii) the feasibility of most of the NB-IoT procedures is impacted by the RTD and the differential delays (the maximum, in particular) of users belonging to the same beam. In the following, the processing delay at the network elements is considered negligible and the propagation delay on the user link is considered the only contribution to the overall latency. Thus, the goal of this paragraph is to compute the maximum differential delay, *i.e.*, the maximum differential slant range, between two generic users. By definition, the differential delay is the difference in delay between a generic user and the user experiencing the minimum delay in the same beam.

Indeed, the one-way propagation (T_{ow}) delay of two or more terminals in the same beam can be split into two parts:

$$T_{ow} = \Delta\tau + T_{ul} = \quad (3.3)$$

where T_{ul} describes the delay component common to all the UEs in the same beam (please, note that the feeder link and the ISL are not considered, otherwise they belong to the common delay in addition to the delay on the user link) and $\Delta\tau$ defines the differential component of the delay.

It is worth emphasizing that all sources of the common delay can be pre-compensated, [44]. By means of orbital and geometrical considerations, the maximum difference in slant range between any two beam users is obtained when: i) the beam center is located on the satellite ground track, *i.e.*, $v_{BC} = 0$; and ii) the two users are located at

the intersections between the ground track and the beam edge, *i.e.*, $v_{min} = v_{max} = 0$, where *min* and *max* define the users located closest and farthest from the satellite, respectively. In a given time instant along its orbit, the satellite is seen at an elevation angle ε_t from the beam center; thus, from Eq.(3.1), it is possible to write:

$$u_{BC} = \frac{R_E}{R_E + h_{sat}} \cos \varepsilon_t \quad (3.4)$$

where R_E is the Earth's radius. Since the beam radius in uv coordinates is known, it is possible to obtain the coordinates of the two users as:

$$\begin{cases} u_{min} = u_{BC} - r_{uv} = \frac{R_E}{R_E + h_{sat}} \cos \varepsilon_t - r_{uv} \\ u_{max} = u_{BC} + r_{uv} = \frac{R_E}{R_E + h_{sat}} \cos \varepsilon_t + r_{uv} \end{cases} \quad (3.5)$$

Moreover, it is worth highlighting that: i) when the Sub-Satellite Point (SSP) is in the beam, the minimum slant range location is given by the SSP itself; and ii) when the maximum slant range point is over the satellite's FoV, the maximum slant range location is on the FoV. From these observations, it is possible to write:

$$u_{min} = \begin{cases} \frac{R_E}{R_E + h_{sat}} \cos \varepsilon_t - r_{uv}, & u_{BC} > r_{uv} \\ 0, & u_{BC} \leq r_{uv} \end{cases} \quad (3.6)$$

and

$$u_{max} = \begin{cases} \frac{R_E}{R_E + h_{sat}} \cos \varepsilon_t + r_{uv}, & u_{BC} + r_{uv} \leq \frac{R_E}{R_E + h_{sat}} \\ \frac{R_E}{R_E + h_{sat}}, & u_{BC} + r_{uv} > \frac{R_E}{R_E + h_{sat}} \end{cases} \quad (3.7)$$

From the u -axis coordinates, the corresponding elevation angles are:

$$\varepsilon_i = \arccos \left(\frac{R_E + h_{sat}}{R_E} u_i \right) \quad (3.8)$$

with $i = min, max$. Knowing the elevation angle ε_i and the nadir angle $\vartheta_i = \arcsin u_i$, the Earth central angle can be computed as $\lambda_i = \pi/2 - \varepsilon_i - \vartheta_i$. This allows obtaining the slant range as:

$$d_i = R_E \frac{\sin \lambda_i}{\sin \vartheta_i} \quad (3.9)$$

Thus, the maximum differential slant range, Δd_{max} , is given by:

$$\Delta d_{max} = R_E \left(\frac{\sin \lambda_{max}}{\sin \vartheta_{max}} - \frac{\sin \lambda_{min}}{\sin \vartheta_{min}} \right) \quad (3.10)$$

From the above equation, it is straightforward to obtain the maximum differential delay as follows:

$$\Delta\tau_{max}(\varepsilon_t, h_{sat}, r_{uv}) = \frac{R_E}{c} \left(\frac{\sin \lambda_{max}}{\sin \vartheta_{max}} - \frac{\sin \lambda_{min}}{\sin \vartheta_{min}} \right) \quad (3.11)$$

with c being the speed of light. Eq. 3.11 clearly highlights that the maximum differential delay is a function of the elevation angle at the beam center, the beam radius, and the satellite altitude since both the nadir and Earth central angles depend on these parameters from Eq.(3.6), (3.7), and (3.8).

3.4.2 Doppler shift

The Doppler shift consists of the change in the carrier frequency due to the relative motion between the satellite and the user terminal. Similarly to the above analysis for the propagation delay, only the mathematical framework of the maximum differential Doppler shift is discussed since all terms introducing a common shift can be pre-compensated assuming, as per 3GPP current analyses, that: i) the terminals are equipped with GNSS capabilities, and ii) the satellite ephemeris are known.

By means of orbital and geometric considerations, the worst-case scenario (in terms of maximum differential shift between any two users) arises when the beam's major semi-axis lies on the satellite ground track, [48]. For a terminal not located at the beam center, but on the beam major semi-axis, the same Doppler curve applies with a horizontal shift given by the time instant at which that UE will see the satellite at $\varepsilon = \pi/2$. The differential Doppler between any two users is obtained by evaluating the Doppler shift at the corresponding elevation angles and computing the difference. Therefore, the maximum variability is obtained when the UEs are at the two beam edges on the beam major semi-axis. In [49], the authors proposed a simplified formula for the Doppler shift experienced at locations on the satellite orbit projection on the ground as a function of the elevation angle:

$$f_d(\varepsilon_i) = f_c \frac{\omega_s R_E \cos \varepsilon_i(\varepsilon_t)}{c} \quad (3.12)$$

where ω_s is the satellite's angular speed and f_c the carrier frequency. It is worth emphasizing that in the above formulation, the elevation angle at the user location ε_i is a function of the elevation angle at beam center ε_t , as a consequence of the dependency of the u -axis coordinate from the beam center elevation angle, discussed above. To provide the maximum differential Doppler shift, the elevation angles at the minimum and maximum Doppler shift locations, *i.e.*, Eq. (3.6) and (3.7), must be computed. From these, it is possible to write:

$$\begin{aligned} \Delta f_{d,max} &= f_d(\varepsilon_{max}) - f_d(\varepsilon_{min}) \\ &= 2f_c \frac{R_E + h_{sat}}{c} \omega_s r_{uv} \end{aligned} \quad (3.13)$$

From Eq. 3.13, it is clear that the maximum differential Doppler depends on the satellite altitude and the beam radius only, and not on the elevation angle at the beam center.

3.4.3 Path Loss

For a satellite communication link between a UE on ground and the satellite, the overall losses, L , can be computed as:

$$L = PL + L_E = \underbrace{L_B + L_A + L_{POL}}_{PL} + \underbrace{L_F + L_D}_{L_E} \quad (3.14)$$

where PL represents losses due to the channel impairments while L_E are the losses related to the equipment configuration. In details: *i*) L_B is the basic path loss, which combines free space, clutter, and shadowing losses; *ii*) L_A represents the losses due to atmosphere; *iii*) L_{POL} is the polarisation mismatch loss; *iv*) L_F represents the losses in the equipment; and *v*) L_D represents the depointing losses.

The *basic path loss* is the combination of Free Space Loss (FSL), L_{fs} , clutter loss, L_{cl} , and log-normal shadowing, L_σ :

$$L_B = L_{cl} + L_{fs} + L_\sigma \quad (3.15)$$

The clutter loss models the attenuation of signal power caused by surrounding buildings and objects on the ground. It depends on the elevation angle, ε , computed in Eq. 3.8, the system operating frequency, f_c , and the environment. Typical values for this parameter can be found in [35] for different scenarios and it can be always assumed null in Line of Sight (LoS) conditions. For a generic user located at slant range d from the satellite, the FSL is given by:

$$L_{fs} = 20 \log_{10} \left(\frac{4\pi d f_c}{c} \right) \quad (3.16)$$

where c is the speed of light. As for the shadowing loss, L_σ is modeled as a log-normal random variable with zero mean and variance related to the harshness of the shadowing environment, *i.e.*, $L_\sigma \sim (0, \sigma_s^2)$, and the values of, σ_s^2 are provided by 3GPP for dense urban, urban, and rural scenarios as a function of the elevation angle in [35].

Atmospheric losses take into account the atmospheric gases absorption, L_{gas} , the rain/snow fall and cloud attenuation, L_{rain} , and the scintillation losses, L_s :

$$L_A = L_{gas} + L_{rain} + L_s \quad (3.17)$$

Atmospheric gas absorption depends mainly on frequency, elevation angle, altitude above sea level, and water vapor density (absolute humidity). In particular, in order to be compliant with 3GPP standardization, losses are computed as provided in

Annex 2 of ITU-R P.676 for slant paths. The atmosphere is modelled with temperature 288.15 K, pressure 1013.25 hPa, and water vapour density 7.5 g/m³ [50]. Rain and cloud attenuations are dependent on the geographical location of the ground terminal. Section 2.2 of ITU-R P.618-13 describes a method to estimate the long-term statistics of attenuation due to rain [10]. For 3GPP System-Level Simulation (SLS), the baseline is to consider clear sky conditions only and in any case, rain attenuation is considered negligible for frequencies below 6 GHz [35]. Scintillation is a variation of the amplitude of received carriers caused by variations in the refractive index of the troposphere and the ionosphere. The tropospheric scintillations, impacting signals in Ka-band, are modeled as a fixed term depending on the user elevation angle [35]. These values are obtained by means of the procedure described in ITU-R P.618 [10]. The ionospheric scintillations, impacting signals in S-band, are modeled as a fixed term of 2.2 dB as from [35], and described in ITU-R P.531-13 [51].

It is also necessary to consider the *polarisation mismatch loss* observed when the receiving antenna is not oriented with the polarisation of the received wave because propagation through the atmosphere can also affect the polarization. In fact, the ionosphere introduces a rotation of the plane of polarization of an angle, $\Delta\psi$ which is inversely proportional to the square of the frequency. This rotation is particularly dangerous for linear polarization. Furthermore, with linear polarisation, the receiving antenna may not have its plane of polarisation aligned with that of the incident wave. In general, the polarization mismatch loss can be defined as [33]:

$$L_{POL} = -20 \log_{10}(\cos \Delta\psi) \quad (3.18)$$

The *equipment losses* represents the losses in the transmitting and receiving equipment respectively, in the feeder between the power amplifier and the antenna:

$$L_F = L_{FTX} + L_{FRX} \quad (3.19)$$

In particular, L_{FTX} is the feeder loss between the transmitter and the antenna, while L_{FRX} is the feeder loss between the antenna and the receiver.

Finally, the *depointing losses* are functions of the transmission and reception angles misalignment, θ_T and θ_R respectively, with respect to the antenna boresight. The result is a fallout of antenna gain with respect to the maximum gain on transmission and on reception, which can be formulated as a function of the θ_{3dB} , later explained in this chapter:

$$L_D = L_T + L_R = 12 \left(\frac{\theta_T}{\theta_{3dB}} \right) + 12 \left(\frac{\theta_R}{\theta_{3dB}} \right) \quad (3.20)$$

3.5 NB-IoT Integration with NTN: State of the Art Analysis

As seen so far, the added value brought by the NTN component into the terrestrial system architecture has been recognized by the 3GPP integrating NTN into the NR

starting from Rel.17. At the same time, 3GPP decided to study the feasibility of the NB-IoT air interface on the satellite channel, starting the study phase for NB-IoT support for NTN from 2021. TR 36.763 [23] identified the necessary adaptations and evaluated the performance of NB-IoT on NTN. Since then, various aspects of this integration process have been addressed in several scientific publications. Hereafter, the main findings in the SoA are summarized.

3.5.1 Link Budget

The performance of a SatCom system must be evaluated on all the links participating in the connection among the two end terminals. In fact, they all impact the QoS for the entire end-to-end (E2E) system, specified in terms of BER or Packet Error Rate (PER). Due to the limited MCS available in the NB-IoT standard, it is of paramount importance to assess the link performance by evaluating the signal power budget and the noise contribution budget necessary to allow reliable communications. For these reasons, a consistent number of works studied the link budget in the NB-IoT NTN system.

In [52], the link budgets of a satellite-based NB-IoT system under different parameters have been addressed. In particular, the link and device parameters were chosen in accordance with the latest 3GPP specifications, while the satellite parameters were left open for design. The achievable spectral efficiency as a function of satellite antenna EIRP (Effective Isotropic Radiated Power) and G/T (Gain to noise Temperature) are shown through numerical simulations for both, LEO and GEO satellites, and under different transmission modes. The authors demonstrate that, in the downlink case, to enable an NB-IoT system capable of achieving the highest possible spectral efficiency, a minimum EIRP of 25 dBW for a LEO satellite at 600 km altitude is needed and 57 dBW for a GEO satellite are considered. In the uplink, for a 12-carrier transmission mode, a minimum G/T of -2 dB/K for an LEO satellite and 28 dB/K for a GEO satellite are required. In case of lower values of G/T, the link can still be closed by using the other transmission modes (e.g., 1,3, or 6 subcarriers) for the Single-Carrier Frequency Division Multiple Access signal or reducing the spectral efficiency.

In [53], the authors assessed a thorough link budget analysis of the proposed satellite scenarios, which were different from the one presented in 3GPP. In particular, two scenarios of IoT satellites have been considered: LEO satellites, orbiting at an altitude of 770km with an inclination of 84.6°, and GEO satellites. The complete surface of the Earth requires three GEO satellites to achieve Earth constant coverage, while 66 satellites would be required for the LEO scenario. For DL communication the bandwidth is fixed to 180 kHz. While, for UL communication, all the possible UL signal bandwidths are tested, i.e., 3.75, 15, 45, 90, and 180 KHz. For the GEO scenario, the SNR varies from -17.7dB, when 180kHz bandwidth is considered in the edge of the spot, to 2.4dB in the center of a spot when 3.75kHz bandwidth is considered. For LEO, the SNR varies from -8.8dB to 13.8dB depending on the bandwidth.

Thus a proper choice of the number of uplink and downlink repetitions allows to close the link.

In [54], the authors computed the SNR and the SINR, compliant with the 3GPP methodology, considering different satellite orbits, such as GEO, LEO, and Nanosatellites with the aim to evaluate the NB-IoT capacity and the energy consumption. By comparing the performance achieved in the three different systems, the authors observed that the LEO full coverage system is the one providing more capacity in terms of messages per second, but at the same time, it is the most expensive one although more cost-effective per message compared to the GEO and the Nano. The use of a sparse LEO system can be a suitable solution since it reduces the overall system cost and at the same time it keeps its efficiency in relation to the cost per message. The GEO system is more power efficient, as one watt in orbit will provide more throughput over the lifetime of the satellite than LEO or Nanosystems. Whereas the latter appears to be the cheapest per year. For the analysis, the authors considered 27 Physical Resource Block (PRB) packed up in 5Mhz bandwidth and simultaneous users occupying all the available resources for the data, thus no congestion is foreseen in the system.

3.5.2 DL and UL synchronization

In order to minimize the changes to NB-IoT devices for NTN support, the synchronization needs to re-use legacy physical channels and existing procedures for initial cell access. The large Doppler shift and RTD typical of the NTN represent an obstacle to the synchronization phase. The authors in [55] propose to pre-compensate the propagation delay and Doppler to perform UL synchronization. The UE pre-compensation can be based on the following pre-requisites:

- the UE position is known using embedded GNSS capability;
- DL synchronization is achieved first through the legacy synchronization channels (NPSS/ NSSS/ NRS) in NB-IoT.
- UE decodes the relevant SIB carrying the satellite position and velocity.

The authors recommended to use only the satellite position and velocity, instead of the complete satellite ephemeris to reduce the signal overhead and to increase the accuracy of the synchronization.

A general overview of the integration process of the NTN into NR is presented in [56], where, based on the architecture options proposed in the standardization fora, the authors discussed and assessed the impact of the satellite channel characteristics, such as high latency, Doppler effect, and link budget on the PHY and MAC layers, both in terms of transmitted waveforms and procedures for enhanced mobile broadband and NB-IoT applications. They concluded that for the NB-IoT scenario, the system architecture does not include Relay Nodes and, thus, the impact of large Doppler shifts is a limiting factor. With respect to the large delays, the TA procedure

might pose technical challenges when the information transmission is scheduled outside a specific time window.

To overcome the issue of the high Doppler shift in the uplink communications, the authors in [57], focusing on the implementation of NB-IoT on LEO/VLEO satellite systems, propose a solution for Doppler shift compensation based on the frequency advance technique.

In [58], different solutions to make RACH signal robust to residual frequency offset have been evaluated:

- use of beams with a diameter not wider than 50 Km;
- restrict the preamble sequences that can be used to avoid sequence spacing of 2 SCS apart for Format 1 and 6 SCS apart for Format 2. Such a solution will come at the cost of reduced RACH capacity and degraded spectral efficiency;
- GNSS aided solution: in this way, the terminal is able to determine its location and estimate propagation delay to satellite (knowing satellite ephemeris) and pre-compensate Doppler and delay before transmitting PRACH;
- low impact change of the preamble sequence design by introducing a fractional frequency hopping sequence per preamble.

The authors deeply explored the last solution and observed that: i) the fractional frequency hopping can be applied post-Fast Fourier Transform (FFT) with no impact on RACH capacity or spectral efficiency; ii) Frequency offset ambiguity is resolved for all possible residual frequency offsets thus allowing to support all potential beam sizes.

Considering the same problem of uplink synchronization achieved with the RA, in [59] a joint detection and estimation technique robust to long delays and significant Doppler effects, without requiring any modification to the current NB-IoT RACH waveform is proposed. In particular, the UE has not GNSS capabilities, therefore the satellite performs the pre-compensation of the Doppler shift at the center of the on-ground beam and broadcasts the common delay to all UEs inside the concerned beam to be taken into account for the uplink transmission. In this case, the differential delay and the differential Doppler are assumed to be handled by the protocol. The proposed method foresees the extension of the Cyclic Prefix length (within the preamble) by considering the current one ($266.67 \mu s$ long) plus additional symbols from the same Symbol Group. The method can be useful in terms of ToA estimation accuracy by eliminating the CFO as well as the Doppler rate effects when estimating the ToA. On the other hand, the actual performance of the method depends on the satellite configuration. Indeed, the method (*i.e.*, without GNSS) cannot work if the differential delay exceeds the maximum RTD that the system can handle, meaning the length of one symbol group.

3.5.3 Timing enhancement

A comprehensive overview of the integration of satellite and NB-IoT standard is proposed in [58], [60], [61]. In all the studies, the authors analyze the effects of the typical satellite propagation delay and Doppler on the uplink synchronization not only considering aspects related to the physical layer, but also to the preamble ambiguity, uplink scheduling, and MAC timers. Indeed, due to the large differential delay within a satellite cell, preambles from different Random Access Opportunities (RAO) from near/far devices can overlap when received by the network. This results in an erroneous transmission of the RAR since the RA-RNTI depends on the corresponding RAO. To overcome the issue, the authors in [61] proposed to extend the resources dedicated to the RACH transmission in each RAO by a quantity equal to the maximum differential delay in the beam. This method is simple and effective but on the other hand, fewer resources are dedicated to data transmission due to the differential delay. Regarding the uplink scheduling, both in [55] and [60] the authors proposed to include a UL scheduling offset when indicating UL grant in the RAR in Message 2 or in the DCI carried on NPDCCH in order to consider the large satellite delay. At the same time, also the RRC Connection Request message (*i.e.*, the message 3 in the RA) can be updated to include the propagation delay estimated by the device, which the network may use for device-specific scheduling of UL and DL transmissions. With respect to the timers, the device needs to delay the start of the following timers for IoT NTN by the device estimated RTD [55], [60]:

- RAR window timer: during the RA procedure, after the device transmits the preamble, it waits for a RAR message from the network during a period defined by the RAR window timer. As a response from the network cannot be received during the satellite RTD, the device needs to delay the start of the RAR window by the estimated RTD. If the timer expires and no RAR is received, the device re-attempts the preamble transmission.
- the contention resolution timer: after the device transmits Message 3, it waits for a contention resolution message from the network during the contention resolution timer. Similar to the RAR window, the start of the contention resolution timer needs to be delayed by the estimated RTD to receive Message 4 from the network. If the timer expires and no Message 4 is received, the device reattempts the preamble transmission.
- the scheduling request prohibit timer: if a Buffer Status Report (BSR) is triggered (UL data available) but there are no UL resources for transmitting the BSR, Scheduling Request (SR) is triggered. The SR is pending until a UL grant is received and a BSR or all available data is transmitted. If the timer expires and the SR is still pending, a new SR is transmitted. To avoid re-transmitting an SR before the network has the opportunity to respond to the SR, the device needs to extend the SR prohibit timer by the estimated RTD.

- HARQ RTD Timer/ UL HARQ RTD Timer: the HARQ RTD Timer/UL HARQ RTD Timer specifies the minimum number of sub-frames before a DL assignment / UL grant for HARQ re-transmission is expected. However, the DL assignment/ UL grant for HARQ re-transmission cannot be received during the satellite RTD by the UE and this can be taken into account by extending the RTD timers by the estimated RTD

Expanding the RA procedure timers and including an offset before the NPUSCH transmission increases the network access time, as demonstrated in [62]. Here, the authors designed a testbed based on Open Air Interface (OAI) implementation for the 3GPP users and base station, and Hardware (HW) implementation for the NTN channel emulating the experienced delay in the communication link between the users and the base station. As outcomes of the activity, they measured the single-user access time during the RA procedure over a GEO satellite, showing that this value is higher in the NTN case than in the terrestrial one, due to the adaptation of all the timers to the RTD and the inclusion of the offset before the Message 3 transmission. It is assumed that the UE is equipped with the GNSS to perform the synchronization.

3.5.4 RA congestion and E2E delay

The design of the NB-IoT via NTN system foresees not only the adaptation of the NB-IoT air interface to the NTN peculiarities, i.e., the large delay, Doppler frequency, path loss, and high speed of the flying platform but also meeting the requirements of the terrestrial NB-IoT standard, such as the massive connectivity and the low power consumption. For this reason, most recent studies have focused on system-level analyses [63]–[65].

In [63], the authors addressed the design of NB-IoT over satellite service, compliant with 3GPP specifications, which aims at facing the most critical issues arising from the employment of NB-IoT over NTNs in a real application scenario. To offer a real smart agriculture service operating in Europe, the proposed solution exploits 24 LEO satellites, grouped into 8 different orbits, moving at an altitude of 500 km. The configured protocol stack supports the transmission of tens of bytes generated at the application layer by counteracting the issues introduced by the satellite link by means of the GNSS receiver. The authors tested both the legacy 4-step RA followed by the data transmission and the EDT and showed that the latter reduces the communication latency up to 40%. In terms of congestion management techniques for the RA phase, a fixed BO window equal to 65536 ms is considered. Finally, the authors observed that with a reduced number of satellites per orbit (from 3 to 2) is still possible to drain all the generated data, but at the cost of much higher average communication latency.

A similar approach is followed in [64], where the E2E packet delay is measured by varying the number of considered CubeSat (*i.e.*, 4 or 8). Regarding the Random

Access, the number of possible PRACH preambles is 48, the periodicity is fixed to 240 ms, while the BO parameter is set to 2048 ms, in order to mitigate the probability of collisions. The E2E packet delay is computed by considering the time spent on the cell selection, RA Procedure, scheduling decisions, and the actual physical transmission. The authors observed that the most noticeable feature is that the constellation numerosness significantly affects the E2E packet delays. In particular, more CubeSats allow covering NB-IoT terminals for more protracted periods, hence reducing the overall delays. Besides, the time required to complete the RA procedure increases as the number of NB-IoT terminals increases. In fact, when more users perform the random access procedure, the number of collisions increases and packet delays also increase.

In [65], the performance of the IoT LEO NTN utilizing LTE-M is evaluated in terms of connection density, which corresponds to the number of users allowing an outage rate of 1%. The latter defines the fraction of the total number of users N that fail to deliver their packets to the destination receiver within a transmission delay of at most 10 s. The evaluation results showed that a single LEO satellite can support a connection density of 364 devices per km^2 at 600 km of altitude and 78 devices per km^2 at 1200 km. It is worth mentioning that although the NTN IoT connection density is significantly lower than that of terrestrial IoT networks, the NTN system is characterized by extremely large cell sizes.

3.5.5 User scheduling

Due to the fact the SatCom system is dynamic and heterogeneous, there is a need to take proactive decisions regarding the users to schedule in the same frame and how to efficiently distribute the resources among them. Indeed, delaying the scheduling of a certain user would result in a resource allocation based on outdated parameters, thus impacting its efficiency in optimizing the available resources. A solution to overcome this issue is proposed in [66], where the authors designed an uplink resource allocation strategy to select the set of users to transmit their uplink data in the available radio resources which maximizes the sum profit of the selected users. The sum profit assigns different profits to users depending on their data packet sizes, channel conditions, and satellite visibility time. The performance of the system is evaluated through three Key Performance Indicators (KPIs): the throughput (KP1), the fairness of service among users in the geographical area of interest (KP2), and the fairness among different coverage levels (KP3). The authors concluded that maximizing KPI 1 would be favorable for a system with a large number of LEO satellites, while KPI 2 would be advantageous in a system where it is mandatory to save spectrum (NB-IoT carriers) and at the same time keep low the number of satellites in a constellation. Finally, the maximization of KPI 3 provides the best distribution of services among users in different coverage classes, but the lowest throughput (58 kbit/s). Depending on the data demand over a particular area, this scenario would

be beneficial for operators targeting to aggregate the IoT traffic through a lower number of LEO satellites with a wide antenna beam on Earth.

The work in [67] proposed a static resource allocation approach to reduce the high values of differential Doppler under the maximum value supported by the standard. To achieve this goal, the authors re-modeled the coverage area in smaller regions in such a way that the differential Doppler inside each region should be below the allowed threshold, which the authors calculated to be 950 Hz. This work has been extended in [68]. The group scheduling is performed both in the time and in the frequency domain. In the former, the NB-IoT carrier is assigned for a certain amount of time to each group by the Base Station (BS). This is possible because the BS knows the position of all the UEs. In the second approach, the uplink resources are assigned to the users in the frequency domain. This means that several NB-IoT carriers are needed, and not just one, for the uplink transmission, which results in a total of 3.6 MHz of uplink bandwidth of $(B = 20 \cdot 180kHz)$, where 20 is the number of users' groups. The authors concluded that with both techniques the differential Doppler is reduced from 8.5 kHz down to 880 Hz. The scheduling in the time domain outperforms the one in the frequency domain (in terms of BLER) since the latter has a dependency on the number of secondary NB-IoT carriers used for the transmission.

3.5.6 NOMA in NTN

The authors in [69] consider the existing NOMA schemes. They compare the performance of some algorithms, such as the interleave division multiple access (IDMA), pattern division multiple access (PDMA), sparse code multiple access (SCMA), power domain non-orthogonal multiple access (PD-NOMA), and multi-user shared access (MUSA), as typical examples. Among them, the SCMA scheme evolved from the low-density signature for code division multiple access (LDS-CDMA), is taken as the most promising candidate with good performance and high overload rate. This scheme distinguishes different users through the predefined codebooks. Because of the indistinguishable power in one beam, the SCMA is more applicable for LEO satellite communication. The available research on the SCMA mainly focuses on two aspects: one is the codebook design and the other is the multiuser detection. The core of the codebook design is to reduce the number of points with large power in the superimposed constellation. The operations can be summarized as follows. First, the signal points with large power are reassigned to the points with small power to construct a low-PAPR mother constellation. Then, proper rotations are performed on the designed constellation to obtain different users' unique codebooks and maximize the Euclidean distance of the projection points with the largest power among the users colliding in a resource. Therefore, an optimized codebook design with low PAPR is got. Simulation results show that compared with the 16-point SCMA mother constellation (called T16QAM), the proposed constellation called R12 and the constellation R4 can achieve lower PAPR and better performance. Meanwhile, they can lower computational complexity as well.

In [70], the authors consider a scenario where a Satellite Access Point (SAP) provides wireless access for terrestrial UEs, aiming at satisfying the requirement of low power consumption and massive connections. They propose a Superimposed Preamble (SP) code domain NOMA protocol for the uplink mMTC S-IoT system. Then, they analyze and derive the analytical performance expressions in the presence of collisions and channel estimation errors. In the proposed scheme, the UEs randomly choose a pilot sequence, which is superimposed on their data to perform access to the SAP. Each pilot sequence corresponds to a unique codebook. Thus, the SAP can identify the collided pilots and recover the singleton pilots.

3.5.7 ESA funded projects on NTN NB-IoT

TABLE 3.3: ESA projects addressing NTN NB-IoT

Project Name	Topic	Objective	Status
IoT SatBack	Design and development of backhauling solutions for NB-IoT	Flexible bandwidth and data rate Low coordination in the access among the users Good spectral efficiency also the transport layer	Closed
NB-IoT4Space	Design and development of NB-IoT via NTN direct access and regenerative NGSO payload	Validation of the adaptations of the protocol stack Development of an on-board eNB within a small platform (12U)	Closed
Narrowband IoT for SmallSat Network	Demonstration of 5G NB-IoT NTN direct access and transparent LEO satellite	Adaptation of the NB-IoT protocol to the NTN channel Implementation of algorithms for Doppler compensation, synchronization, cell search, paging, access	Closed

The proposed modifications to the NB-IoT air interface to tackle the challenges imposed by the satellite channel have also been investigated and evaluated through field tests and proofs of concept as part of different funded projects. The objective of NB-IoT4Space project [71] is to develop a demonstrator where communications between an NB-IoT UE and an eNB adapted for satellite communications are functionally verified. In particular, the project foresees the validation of the adaptations to the UU's protocol stack (*e.g.*, PHY, MAC, RLC, PDPC, and RRC layers) to make the NB-IoT radio standard viable over satellite and the development of an on-board eNB capable of supporting the adapted 3GPP NB-IoT standard within the mass, power, and volume constraints of a small platform (<12U). The proposed demonstrator consists of:

- NB-IoT Traffic Emulator to emulate the traffic generated by several UEs.
- an NB-IoT module with UU interface adaptations.
- Channel Emulator to emulate all the impairments caused by LEO satellites considering different scenarios. In particular, is able to introduce variable delay, variable Doppler Shift, Phase Noise, and Fading.
- OnBoard eNB Prototype, *i.e.*, an eNB module with the UU interface adaptations.

In Narrowband IoT for SmallSat Networks project [72], the main purposes are: i) to identify areas for which the existing terrestrial NB-IoT requires significant adaptations to allow direct access to a LEO satellite with a transparent payload; ii) to implement critical algorithms for Doppler compensation, synchronization, cell search, paging, and access; and iii) to demonstrate the complete system by means of a space segment that is installed on an SDR payload with applicable antenna systems and a solution for the UE which can be installed either on dedicated HW or be integrated into a suitable purpose-specific chipset.

The IOT SATBACK project [73] aims at designing and developing a prototype backhauling solution for future NB-IoT networks. The project objectives are:

- Study report containing an analysis of the main backhauling scenarios for NB-IoT networks.
- Definition of the requirements to allow efficient integration of terrestrial NB-IoT with satellite backhaul.
- Implementation and test of the prototype.
- Satellite backhaul solution and a corresponding testbed.
- Demonstration of efficient backhauling of NB-IoT in a representative environment using the developed hardware and software.

The targeted improvement is to enable new satellite communication services for backhauling Machine-to-Machine (M2M) and Internet-of-Things communications.

3.5.8 Conclusion of the SoA Analysis

Based on the analysis of the SoA, it can be inferred that the activities both in Academia and industry have focused their effort on adapting the air interface and the procedures of the NB-IoT standard to the peculiarities of NTN. The use of GNSS is the baseline assumption to enable NB-IoT via NTN with the minimum amount of modifications to the terrestrial standard. Indeed, knowing its own position, the UE can be synchronized to the BS before transmitting the preamble, delay the start of the timers in the Random Access procedure, and increase the Scheduling request timer and the HARQ RTD. However, even if GNSS enables the problems of differential Doppler shift and differential delay to be overcome, the large RTD has a negative impact on procedure timers and the procedures themselves. Therefore, it is required to shorten the signaling exchange between the transmitter and the receiver, *e.g.*, by means of the EDT or the 2-Step RA. This is particularly valid when a massive number of users needs to communicate with the network within a short interval, creating congestion. The latter reduces the access probability and increases the E2E latency. Since congestion has only recently been considered, no algorithm has been proposed to handle it considering the low visibility window of satellites. For this reason, the thesis focuses on both system and link-level analyses with and without congestion.

Regarding the link level analysis, the first assessed activities foresee the detection of the NB-IoT preamble and the estimation of the user's ToA leveraging the peculiar structure of this signal and the concept of the repetitions. Afterward, the focus moved to the case of multiple devices performing random access on the same occasion, generating collisions. One of the possible countermeasures to the management of collisions is represented by NOMA schemes, which require the knowledge of the number of users transmitting on the same resources. Thus, a NN algorithm to cope with the uncoordinated random access performed by a massive number of NB-IoT devices was designed. Indeed, the NB-IoT devices transmit their preambles to the gNB, conceptually located at the gateway. Here the NN-based detector classifies the number of colliding users and for each of them estimates the ToA. Following these activities, a classical NOMA approach, named Sparse Code Multiple Access (SCMA) is applied to the data phase of NB-IoT via NTN, in order to verify the applicability of NOMA to NTN. The performance is reported in terms of BER, varying the estimation accuracy of the channel coefficient of each user. Moving to the system-level analysis, the goal of the first work was to establish a baseline for the RA procedure via NTN performed by a large number of devices without proposing any modification. Indeed, the objective was to evaluate the access probability and the mean access time considering different satellite scenarios defined by the 3GPP in the TR 36.763 (giving priority to LEO satellites), several configurable access parameters typical of the NB-IoT standard, such as the back-off values, the number of sub-carriers reserved for the preamble transmission, and the length of RAO, and finally, many values of the user density. After having assessed such performance, both the access and data phase design have been jointly considered. Indeed, given the user density, the user demand in the congested zone, and the satellite parameters, a heuristic algorithm that optimally combines the number of satellite passages, the length and the number of the RAO, and the number of uplink repetitions is designed with the aim to maximize the capacity by serving all the users in the system. The algorithm does not foresee scheduling strategies for the data phase. The idea behind the heuristic approach is to reduce the complexity of the system in order to make it applicable in practical situations, such as the design of a satellite constellation to provide IoT services.

Chapter 4

Link Budget

As described in Chapter 3, the performance for a SatCom system performance should be assessed on all the links involved in the communications, depending on the type of architecture chosen: i) *DLs* from the satellites to the GSs and UEs; ii) *ULs* from GSs and UEs to the satellites; and iii) *ISL* between the satellites (if any). Indeed, the performance of each link participating in the connection affects the Quality of Service (QoS) of the communications among end-users, specified in terms of BLER or PER for digital communications.

In this chapter, the mathematical framework for the Link Budget (LB) computation is described.

4.1 Channel Modeling

In Chapter 3 the main impairments and contributions to the path loss have been discussed. The framework described here is then adopted in all the system-level analyses and for the computation of the channel coefficients between the users and the satellite, which will be described in the following chapters.

4.2 Link Budget Characterization

This section discusses the performance of individual links in satellite communication. It provides a method to evaluate the signal-power budget and noise-contribution budget to determine the SNR level. It also covers how to take into account multiple links and interference contributions to compute the SINR. The approach described follows 3GPP methodologies [35], [39] and applies to both downlink and uplink. The formulation does not consider inter-satellite links as they are outside the scope of this work and not yet considered in the system architectures under evaluation. All values in this section are given in logarithmic units unless otherwise stated.

4.2.1 Antenna Pattern

When defining a link budget, it is of paramount importance to properly understand the main features of an antenna pattern. This discussion will focus on aperture antenna characteristics. The radiation pattern shows the variation in gain with the

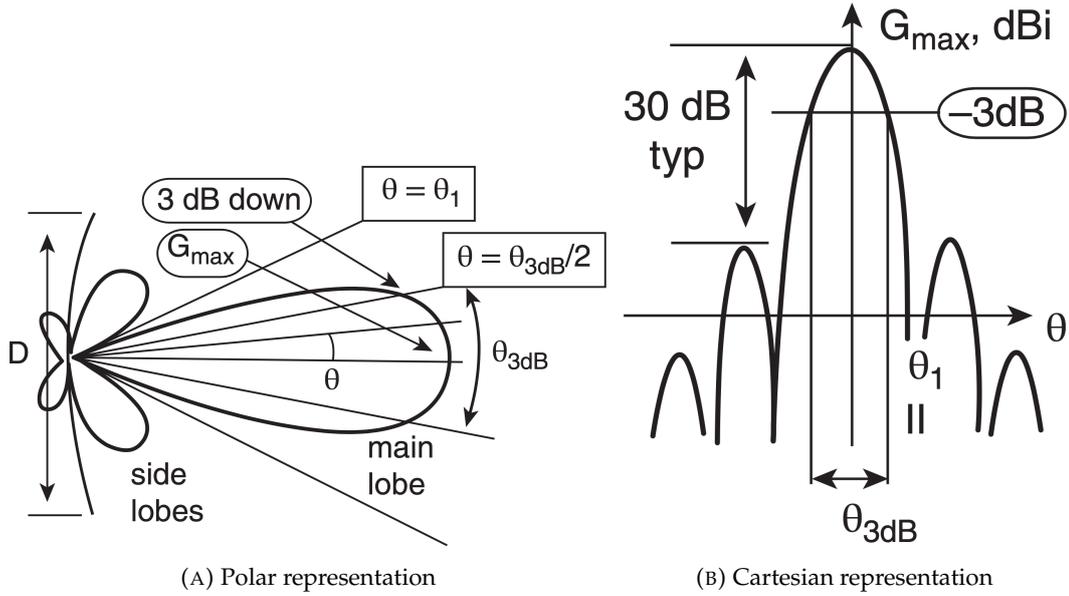


FIGURE 4.1: Antenna radiation pattern [33].

direction and is related to the antenna's characteristics. The antenna gain is defined as the ratio of the power radiated (or received) by the antenna in a given direction to the power radiated (or received) by an isotropic antenna fed with the same power [33]. The gain is highest in the boresight direction and has a value obtained as follows:

$$G_{MAX} = 10 \log_{10} \left[\zeta \left(\frac{\pi D f_c}{c} \right)^2 \right] \quad (4.1)$$

where D refers to the antenna diameter, and ζ represents the efficiency of the antenna, which directly affects the effective aperture area of the antenna $A_{eff} = \zeta(\pi D^2/4)$. The efficiency ζ depends on several factors that take into account the illumination law, spill-over loss, surface impairments, ohmic and impedance mismatch losses, and so on [33].

The antenna characteristics can be completely defined in θ - ϕ coordinates. To avoid cumbersome notation, the following formulation assumes that the antenna pattern has rotational symmetry, meaning that the antenna pattern only depends on the angle θ regardless of the ϕ direction of the transmitted/received signal. This allows the antenna beam to be fully represented as shown in Fig. 4.1.

The angular beamwidth is the angle between the directions in which the gain falls to a specified value relative to its maximum value. The 3dB beamwidth, shown in Fig.4.1a as θ_{3dB} , is a commonly used measure and corresponds to the angle between the directions in which the gain falls to half its maximum value. The antenna gain in a specific direction ϑ can generally be defined for both transmission, G_T , and reception, G_R , as:

$$G(\vartheta) = G_{MAX} + 10 \log_{10} \Omega(\vartheta) \quad (4.2)$$

where $\Omega(\vartheta)$ is the antenna pattern as a function of the ϑ direction. According to the

antenna design, different functions can be provided. The 3GPP focuses on the Bessel beam model, which is defined as follows:

$$\Omega(\vartheta) = \begin{cases} 1, & \vartheta = 0 \\ 4 \left| \frac{J_1(ka \sin \vartheta)}{ka \sin \vartheta} \right|^2, & \vartheta \neq 0 \end{cases} \quad (4.3)$$

where: *i*) a is the antenna aperture radius; *ii*) $k = 2\pi f_c/c$ is the wave number; *iii*) $J_1(\cdot)$ is the Bessel function of the first order; and *iv*) ϑ is the angle measured from the bore-sight of the antenna. It is worth emphasizing that the equation for Ω is valid in both transmission and reception directions; moreover, if the antenna pattern symmetry is not circular, Ω is a function of both θ and ϕ .

Once the antenna gain is defined, the transmitted antenna's Equivalent Isotropically Radiated Power (EIRP) can be characterized by taking into account the losses due to equipment, as given in Equations 3.19 and 3.20:

$$EIRP(\vartheta) = G_T(\vartheta) + P_T - (L_T + L_{FTX}) \quad (4.4)$$

where P_T is the transmitted power.

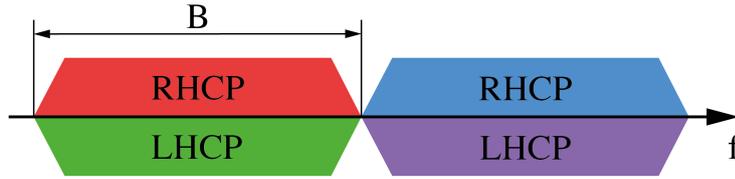


FIGURE 4.2: Frequency Reuse scheme.

An antenna's wave radiation consists of an electric field component and a magnetic field component, which are orthogonal and perpendicular to the wave's direction of propagation and vary at the wave's frequency. The polarization of the wave is conventionally defined by the direction of the electric field, which is not necessarily fixed. The direction of polarisation becomes particularly important in multi-beam antenna patterns. It is common to use frequency coloring schemes where two beams operate in the same frequency band but with opposite polarizations, as shown in Fig. 4.2, to optimize the bandwidth usage.

4.2.2 Downlink

After having define the main elements contributing to the LB, SNR and SINR computations can be performed.

For the sake of clarity, the reference model only shows one satellite and two beams, but this assumption does not affect the generality of the proposed mathematical framework. It should be noted that the link budget computation procedure is applied to each beam and satellite. While numerical simulations will be performed

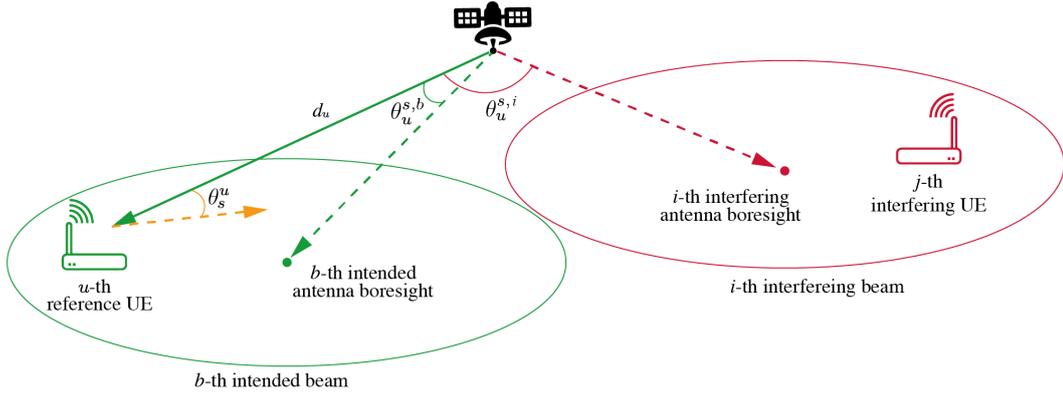


FIGURE 4.3: Geometry for DL LB computation [74].

in a multi-satellite, multi-beam environment, the mathematical framework can be simplified. The reference scheme for DL computation is shown in Fig. 4.3, where: *i*) the intended u -th user (*e.g.*, is randomly positioned in the intended b -th beam and at distance d_u from the satellite. *ii*) an interfering signal is transmitted from the i -th satellite antenna towards the j -th user of the i -th beam; *iii*) $\theta_u^{s,i}$ defines the angle between the i -th antenna boresight of the s -th satellite, and the direction of the u -th user in the b -th beam (when $i = b$, this is the angle from the intended antenna boresight); *iv*) finally, θ_s^u is the angle between the antenna pointing of the u -th user and the s -th satellite, which does not depend on the satellite antenna index as they are assumed to be colocated. The power received by the u -th user in the b -th beam is computed as follows:

$$S_u^{(DL)} = EIRP^{(DL)}(\theta_u^{s,b}) + G_{RX}(\theta_s^u) - L_u^s \quad (4.5)$$

where $G_{RX}(\theta_s^u)$ indicates the receiving UE antenna gain as a function of the angle between the UE antenna pointing and the s -th satellite direction. The transmitted EIRP from the b -th antenna, of the s -th satellite, towards the u -th user is $EIRP(\theta_u^{s,b})$. While L_u^s is the overall loss for the u -th user with respect to the s -th satellite computed as in Eq. 3.14. The receiving gain, G_{RX} is a function of θ_s^u and depends on the antenna pattern of the satellite and device.

The noise power at the receiver is given by [33], [35]:

$$N = 10 \log_{10} \left(\frac{T_a}{10^{0.1L_{FRX}}} + T_0 \left(1 - \frac{1}{10^{0.1L_{FRX}}} \right) + T_0 (10^{0.1N_f} - 1) \right) + 10 \log_{10}(\kappa B) \quad (4.6)$$

where N_f is the receiver noise figure, T_a is the receiver antenna temperature, κ is the Boltzmann constant, and $T_0 = 290$ K is the reference noise temperature.

Therefore, the SNR, $SNR_u^{(DL)}$, is obtained by subtracting Eq. 4.6 from Eq. 4.5:

$$SNR_u^{(DL)} = S_u^{(DL)} - N = EIRP^{(DL)}(\theta_u^{s,b}) + 10 \log_{10}(G/T)_u - L_u^s \quad (4.7)$$

where G/T describes the receiver characteristics through the ratio of the antenna receiving gain, G , in the considered direction and the system noise temperature, T at the input of the receiver¹. It can be written in general as:

$$G/T = G_{RX} - N_f - 10 \log_{10}(T_0 + (T_a - T_0)10^{-0.1N_f}) \quad (4.8)$$

Finally, it is necessary to consider the interference. The interference from the i -th interfering beam of the k -th satellite is computed as:

$$I_{k,i,u}^{(DL)} = EIRP^{(DL)}(\theta_u^{k,i}) + G_{RX}(\theta_k^u) - L_u^k \quad (4.9)$$

while the overall interference at the u -th user in the b -th beam is given by:

$$I_u^{(DL)} = 10 \log_{10} \left(\sum_{k=1}^{N_{sat}} \sum_{i=1}^{N_{int}} 10^{0.1I_{k,i,u}^{(DL)}} \right) \quad (4.10)$$

where N_{int} indicates the number of co-channel, co-polarized and cross-polarized beams, and N_{sat} is the number of satellites in the constellation. It is noteworthy that in the i -th interference term, the satellite EIRP is now computed based on the angle between the i -th interfering antenna boresight and the direction of the u -th intended user in the b -th beam. As for the SNR, from Eq. 4.10, it is straightforward to obtain the Signal-to-Interference Ratio (SIR), $SIR_u^{(DL)}$, by subtracting Eq. 4.10 from Eq. 4.5. Finally, the overall SINR is given by:

$$SINR_u^{(DL)} = -10 \log_{10} \left(10^{-0.1SNR_u^{(DL)}} + 10^{-0.1SIR_u^{(DL)}} \right) \quad (4.11)$$

4.2.3 Uplink

As for the DL case, the mathematical framework for the UL LB computation is presented considering one satellite and two beams. However, all the assumptions hold for all the satellites and beams considered in a multi-satellite, multi-beam system. Fig. 4.4 shows the reference system, where: *i*) the u -th reference user for the LB computation is randomly located inside the b -th reference beam at a distance d_s^u from the s -th satellite *ii*) the interfering i -th devices are randomly located in both the reference b -th beam and the interfering i -th beams at a distance d_s^i from the s -th satellite; *iii*) $\theta_u^{s,b}$ defines the angle between the boresight of the b -th reference antenna on the s -th satellite and the u -th user; *iv*) θ_s^u refers to the angle between the boresight

¹The system noise temperature can be written in linear units as follows:

$$T = T_a + T_{eRX} = T_a + T_0(N_f - 1) = N_f T_0 + T_a - T_0 = N_f \left(T_0 + \frac{T_a - T_0}{N_f} \right)$$

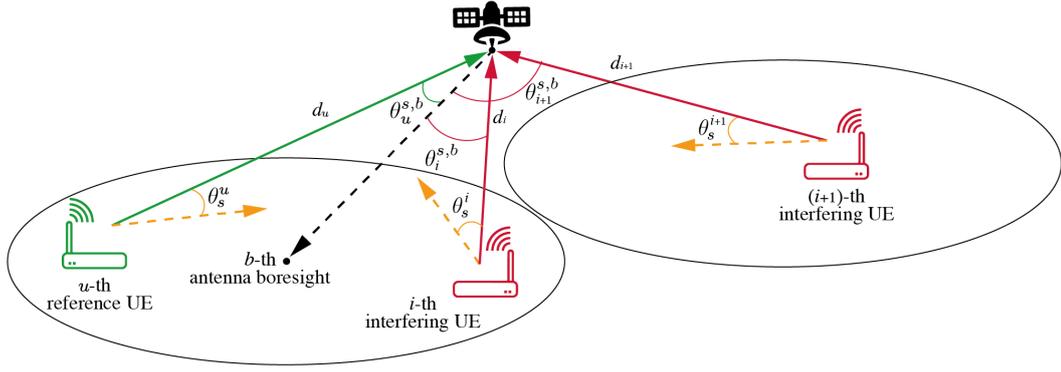


FIGURE 4.4: Geometry for UL LB computation [74].

of the u -th user's antenna and the s -th satellite; v) $\theta_i^{s,b}$ denotes the angle between the boresight of the b -th receiving antenna on the s -th satellite and the i -th interferer; vi) θ_s^i denotes the angle between the boresight of the i -th interferer's antenna and the s -th satellite.

As in the DL case, the power received by the b -th satellite antenna from the u -th user is computed as follow:

$$S_{u,s,b}^{(UL)} = EIRP^{(UL)}(\theta_s^u) + G_{RX}(\theta_u^{s,b}) - L_s^u \quad (4.12)$$

where, this time, the EIRP is a function of the UE antenna radiation pattern with respect to θ_s^u . While the receiving gain of the b -th antenna of the s -th satellite is a function of the angle $\theta_u^{s,b}$ and depends on the satellite antenna pattern. The path loss L_s^u corresponds to the one in Eq. 4.5.

The noise power in the UL is computed by using Eq. 4.6, but N_f refers to the satellite payload noise figure, and T_a is satellite antenna equivalent noise temperature. Thus, UL SNR is obtained as:

$$SNR_u^{(UL)} = S_u^{(UL)} - N = EIRP^{(UL)}(\theta_s^u) + 10 \log_{10}(G/T)_s - L_s^u \quad (4.13)$$

where G/T is the same for Eq. 4.8, but it is now referring to the satellite payload characteristics.

The interference from the i -th interfering User Terminal (UT), towards the b -th beam of the s -th satellite, is computed as:

$$I_{k,i,u}^{(UL)} = EIRP^{(UL)}(\theta_s^i) + G_{RX}(\theta_i^{s,b}) - L_k^u \quad (4.14)$$

where the parameters are the same as in Eq.4.12, however it is worth highlighting that $EIRP^{(UL)}$, and L are related to the position of the i -th interferer with respect to the s -th satellite. Finally, the total interference towards the reference b -th beam of

the s -th satellite is given by:

$$I_{s,b}^{(UL)} = 10 \log_{10} \left(\sum_{i=1}^{N_{int}} 10^{0.1 I_{i,s,b}^{(UL)}} \right) \quad (4.15)$$

with $i = 1, \dots, N_{int}$, being N_{int} the number of interfering terminals. In the uplink, the i -th interference term is represented by either *i*) a device served by the reference b -th beam of the reference s -th satellite; *ii*) a device served by another beam of the same s -th satellite; *iii*) or a device served by another beam-satellite pair of the constellation. Besides, the interference should be considered coming from terminals transmitting towards co-polarized beams and cross-polarized beams. From Eq. 4.15, it is straightforward to obtain the SIR from the u -th UT transmitting towards the b -th antenna of the s -th satellite, $SIR_{u,s,b}^{(UL)}$, by subtracting Eq. 4.15 from Eq. 4.12. Then, the overall SINR is given by:

$$SINR_{u,s,b}^{(UL)} = -10 \log_{10} \left(10^{-0.1 SNR_{u,s,b}^{(UL)}} + 10^{-0.1 SIR_{u,s,b}^{(UL)}} \right) \quad (4.16)$$

The LB formulations, both in UL and DL, are valid for any SatCom system either for the user- or the feeder-link and can be used also for analyses on all the systems foreseen for NTN [35], [39].

4.3 Link Budget for NB-IoT via NTN

The following section is directly derived from the outcomes of the author's publication "NB-IoT over Non-Terrestrial Networks: Link Budget Analysis", [74], in the framework of the 3GPP study item on NTN standardization process. Although many different aspects have been studied in the recent past in order to better understand the requirements for NB-IoT integration over NTN, a complete link budget analysis taking into account the parameters approved in 3GPP study items was not available at the time of the publication. Indeed, to fully support the system design of future NB-IoT over NTN, the link budget analyses shall consider the NTN architecture in its entire complexity, in particular, it shall address the effect of moving non-geostationary satellites, *i.e.*, different elevation angles, actual propagation conditions, and multi-beam and multi-user interference.

In the following, this analysis is presented, providing the link budget evaluation for an NB-IoT system by considering fixed IoT device on Earth and moving LEO satellites in different constellations, accounting for the multi-beam interference, in the downlink, and the multi-user interference, in the uplink.

4.3.1 System Architecture

With reference to Fig. 4.5, the main elements of the high-level system architecture [35] are: *i)* a UE, which is a NB-IoT device, fixed or mobile; *ii)* a regenerative payload satellite or Unmanned Aircraft System (UAS), providing connectivity to the UE through the user link; *iii)* a gNB implemented on the satellite; *iv)* a ground segment GW interconnected to the flying platform through the feeder link; *v)* the Next Generation Core network (NGC).

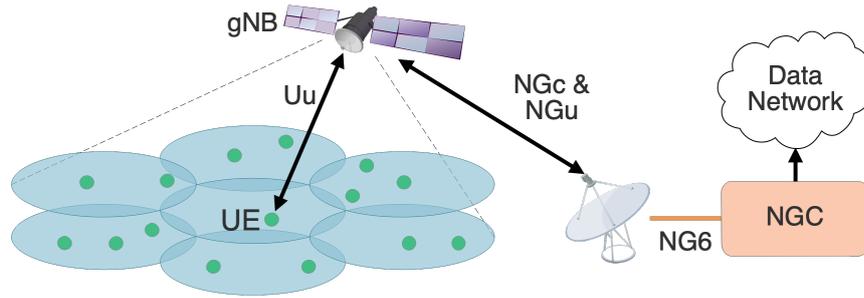


FIGURE 4.5: High-level system architecture: access based on regenerative payload [74].

The aforementioned architecture can be declined into different scenarios, [39], depending on the orbit and altitude of the flying platforms. In the first stage of standardization for NTN, the focus was on GEO platforms and LEO platforms at 600 and 1200 km of altitude. For LEO platforms, the coverage can be achieved with *fixed* or *moving* beams. Later in this analysis, if not otherwise specified, the following system configuration is assumed: direct access, regenerative payload, LEO or GEO multi-beam platform operating in S-band, with moving beams for LEO constellations. This configuration refers to scenario *B* and scenario *D2* in [39], as shown in Table 3.1.

4.3.1.1 Single- and Multi-Satellite Scenarios

For the Single-Satellite (S-S) scenario, the coverage area of the satellite, thus the multi-beam layout, is defined in u - v coordinates. Assuming θ_{3dB} as the antenna 3 dB angle, the beam radius on the u - v plane is defined as $R_{3dB} = \sin(\theta_{3dB})$. The hexagonal tessellation on the u - v plane is then obtained by locating the adjacent beam centers at a distance corresponding to the Adjacent Beam Spacing, $ABS = R_{3dB}\sqrt{3} = \sqrt{3}\sin(\theta_{3dB})$, as it is explained in [39], [75]. An example of a satellite antenna pattern, with hexagonal tessellation, is shown in Fig. 4.6. In [39], the values of ABS for LEO and GEO systems, are provided assuming for the satellite antennas the Bessel radiation pattern of [35] and described in Eq. 4.3. On the other hand, for the UE an omnidirectional antenna has been considered, thus regardless of the angle of departure and arrival defined in Section 4.2 the maximum gain has been considered for the UE antenna. For what concern the UE, it is worth mentioning that the noise power at the receiver is considered independent of the UE, meaning it is the same for every UE and satellite's beam pairs. It is worth highlighting that

the antenna pattern described in [35] is regularly defined with respect to θ - ϕ coordinates, while the beam spacing is defined regular with respect to u - v coordinates, *i.e.*, circular beams with radius R_{3dB} are considered in [39]. Clearly, this choice leads to a mismatch. Indeed, a regular grid in θ - ϕ plane cannot be regular on the u - v plane and vice versa. Therefore moving away from the central beam of the coverage, the surrounding beams begin to deform from the circular shape, leading to an imperfect coverage. However, in this work, perfect coverage, as initially intended from 3GPP documentation, has been used.

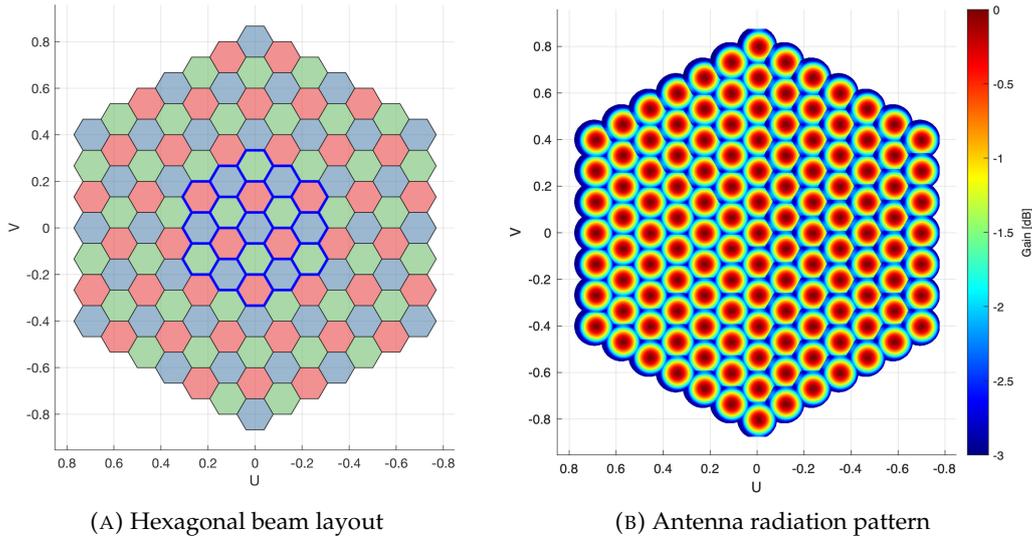


FIGURE 4.6: Example of coverage (u - v coordinates): LEO system in S-band; $h_{sat} = 1200$ km; FR3; 6 tiers (127 beams) [74].

For the Multi-Satellite (M-S) scenario, the methodologies proposed in [39] and [76] have been taken into account. In particular, this work focuses on the constellation-based design, which is based on the definition of a reference constellation, in which all of the orbital parameters, the antenna pattern, and RF characteristics shall be defined. Notably, for a system aiming at achieving global coverage, the satellite altitude, the minimum elevation angle, and inclination, as well as the Half Power Beam Width must be carefully chosen. Considering a specific satellite, with a specific altitude and position, and referring to the beam layout previously defined, it is possible to derive the coverage on the Earth's surface for that flying platform, thus the coverage for the entire constellation. The antenna pattern for each satellite in the constellation is derived with the same procedure detailed for the S-S scenario; however, a different number of beams per satellite have been used, as shown in Table 4.3. It is worthwhile noting that, for this purpose, in the satellite coordinate reference system, the u -axis is pointing towards the West and consequently the v -axis is pointing towards the North, because of the convention adopted by the 3GPP for the NTN standardization process, [39].

4.3.2 Numerical Results

The link budget analyses for single- and multi-satellite scenarios, is performed by means of Monte Carlo simulations, based on the configurations reported in Tables from 4.1 to 4.3 which are derived from [39], [77], [78], and [76]. The results detailed in this section are derived following the mathematical framework described in Section 4.2. The reference satellite-UE geometry, for the LB computations, in both S-S and M-S scenarios, is the one depicted in Fig. 4.3 and 4.4.

Although [39] does not explicitly consider the NB-IoT case, the selection of these parameters is based on the observation that the smaller antenna size at the satellite, as it is in Set-2, is adequate also to the NB-IoT case. As a matter of fact, in an Satellite-IoT (S-IoT) scenario, not only the IoT devices must be low-cost and low-energy, but it is equally important to have small, low-complexity equipment also in the space segment to reduce the cost of the infrastructure and thus the need of a small low-cost antenna at the satellite. Although the parameters chosen for the constellations allow achieving global coverage, the actual antenna implementation generating such a massive number of beams could be impractical, in particular on small and cheap LEO platforms. It is a fair assumption that a lower number of beams will be considered for actual implementations. Furthermore, constellations with partial coverage will be probably considered for IoT services in the near future, and therefore these aspects will be the subject of future studies.

TABLE 4.1: Satellite parameters for S-band (Set-2 [39])

Parameter	GEO	LEO	
		600 km	1200 km
a : equivalent antenna radius [m]	6	0.5	0.5
δ_{EIRP} : EIRP density [dBW/MHz]	53.5	28	34
$G_{TX,max}$: max TX gain [dBi]	45.5	24	24
$G_{RX,max}$: max RX gain [dBi]	45.5	24	24
G/T [dB/K]	14	-4.9	-4.9
N. beams	61-FFR; 127-FR3; 19-retained		

TABLE 4.2: NB-IoT terminal parameters for S-band [7], [77]

Parameter	Value
P_{TX} : transmission power [dBm]	23
$G_{TX,max}$: max TX gain [dBi]	0
$G_{RX,max}$: max RX gain [dBi]	0
T_a : antenna temperature [K]	290
N_f : noise figure [dB]	9
Polarisation	linear
B_{DL} : DL bandwidth [kHz]	180
B_{UL} : UL bandwidth [kHz]	3.75, 15, 45, 90, 180

TABLE 4.3: Constellation parameters for S-band [76]

Parameter	GEO	LEO	
		600 km	1200 km
Orbit type	GEO	LEO, circular	
Orbit inclination	0°	87.5°	
N_o : N. orbits	1	17	10
N_s : N. satellites (per orbit)	4	30	17
Beam diameter (at nadir) [km]	450	190	90
N_b : N. beams (per satellite)	547	127	91

Also, since NB-IoT terminals are not defined in the same 3GPP simulation set, the system bandwidth has been modified in line with the NB-IoT system, by choosing 180 kHz for the DL and 3.75, 15, 45, 90, 180 kHz for the UL, regardless to the adopted Frequency Reuse (FR) scheme, *i.e.*, Full Frequency Reuse (FFR) and Frequency Reuse 3 (FR3). For the sake of synthesis, only the results for the case with $B = 45$ kHz are reported for the UL scenario, *i.e.*, the NPUSCH Format 1 configuration, with 3 tones allocation [7]. The UE parameters refer to [77] and [79], considering NB-IoT Class 3 terminals.

4.3.2.1 Single-Satellite Scenario

For S-S scenario the beams are located on the u - v plane so as to have an equivalent elevation of the satellite at 45° (GEO) or at 90° (LEO) [39] from the coverage center. A variable number of tiers is considered depending on the FR scheme so as to ensure that the same number of interfering beams is considered for all scenarios; in order to properly model the overall interference experienced by each UE, the numerical results are gathered from the 19 internal beams only, which are highlighted by the thick blue line in the example in Fig. 4.6a. The UEs are deployed in the coverage on a regular grid in u - v coordinates. It shall then be noticed that, in the DL, the interference towards any UE is generated by the co-channel transmitting antennas on the satellite; thus, interference is defined by geometry, *i.e.*, it is fixed given the receiving UE location and the beam layout configuration. The only stochastic aspect to be taken into account is the shadow fading, modeled as a log-normal random variable $L_\sigma \sim (0, \sigma^2)$ with the values of σ^2 depending on the UE's elevation angle [35]. As for the UL, interference is clearly depending on the specific interfering UE locations. In this case, then, both the stochastic shadow fading and the interference source location vary through Monte Carlo iterations. Furthermore, the interference is not defined by geometry as in the DL, but it is impacted by the scheduling algorithm, which is assumed to be random, with one UE per beam transmitting at each iteration as it is in [39].

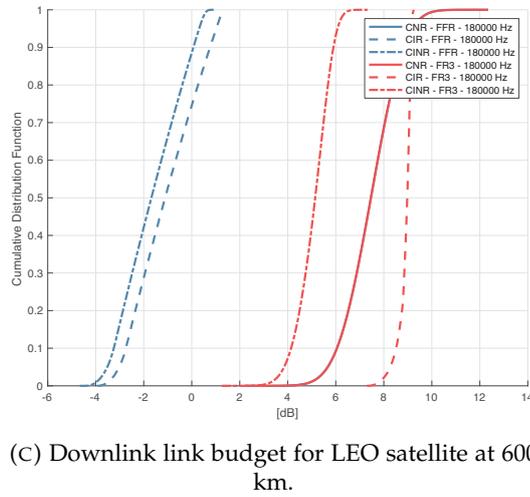
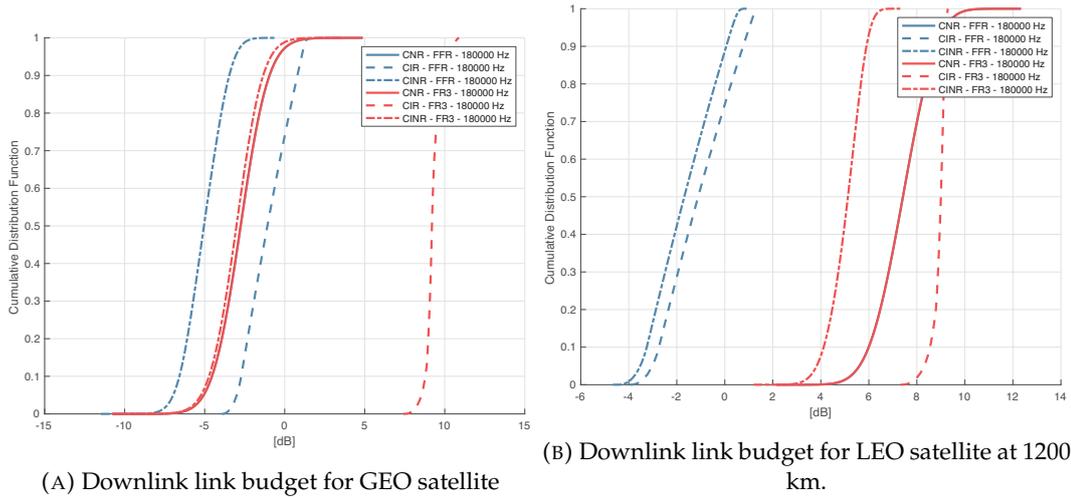


FIGURE 4.7: Downlink Link Budget [74].

Fig. 4.7a to 4.7c, show the Carrier-to-Noise Ratio (CNR)², CIR and CINR Cumulative Distribution Function (CDF) for the DL, while figures from 4.8a to 4.8c for the UL. Table 4.4 provides the mean and standard deviation of the CINR for all cases. As can be seen from the results, the CINR with FFR is quite low for all scenarios; this is related to the severe interfering environment since no interference management technique is implemented. An improvement in the performance can be seen for FR3. The difference in the standard deviation between the DL and the UL scenarios is motivated by the fact that, in the DL, interference is defined by geometry, *i.e.*, the beam layout is fixed and the interfering satellite antennas are always transmitting towards their beam centers; while in the UL, a random scheduling algorithm is implemented to identify the interfering users at each time frame and, thus, an increased variability arises. Regarding the downlink CNR, the curves are identical in FFR and FR3 due to the fact the EIRP/MHz is constant; while the uplink CNR is

²The 3GPP documentation refers to CNR, Carrier-to-Interference Ratio (CIR), and Carrier-to-Interference plus Noise Ratio (CINR), thus in this chapter the same nomenclature has been adopted, in order to be compliant with the specifications. However it is important to highlight that those values correspond to SNR, SIR, and SINR computations presented in Section 4.2.

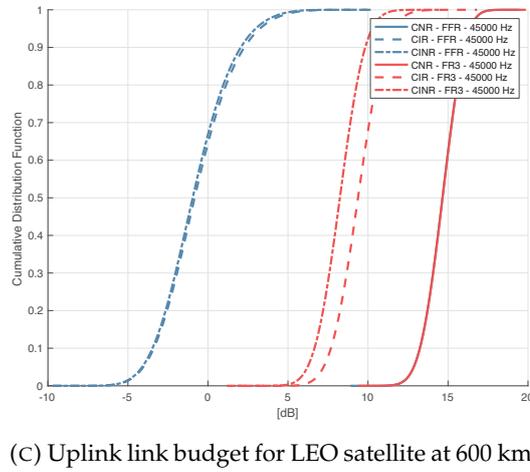
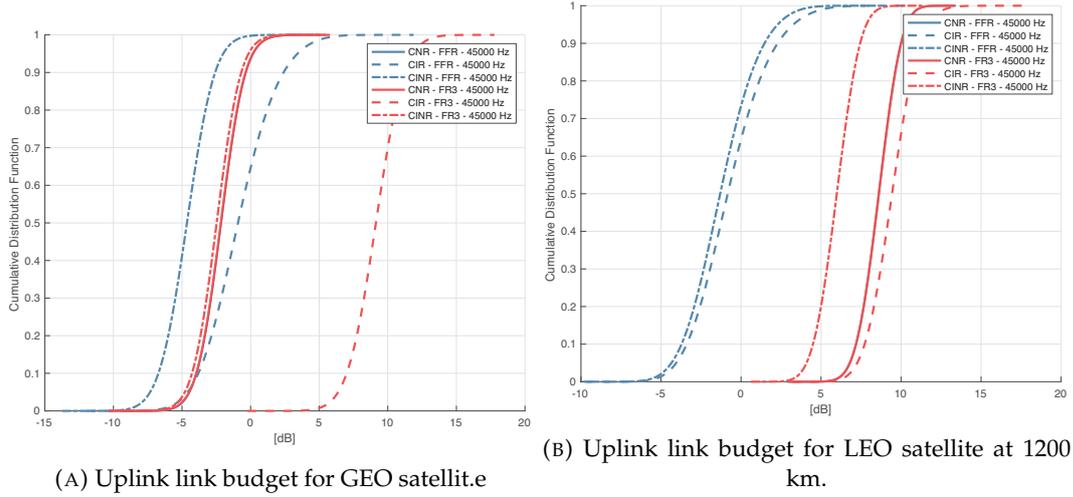


FIGURE 4.8: Uplink Link Budget [74].

the same in FR3 and FFR as the assigned bandwidth is independent of the coloring scheme.

4.3.2.2 Multi-Satellite Scenario

The simulations for the M-S scenario are time-driven, *i.e.*, at each time instant the geometry of the constellation changes, thus the link budget and interference components change. Then both the satellite constellation geometry and the stochastic shadow fading previously described, vary through Monte Carlo iterations. The UEs are deployed on a regular *lat-lon* grid and the statistics are gathered from all the points of the grid. The performance analysis is then computed for the whole evolution in time of the satellite constellation (whose parameters are reported in Table 4.3), considering a uniform user deployment. For each constellation geometry, the UE are assigned to the best satellite/beam pair according to the link budget fixed terms, *i.e.*, the antenna pattern for the constellation is projected in *lat-lon* coordinates, as previously described, then the highest value of $G - L_{fs}$ is chosen for each

UE such to identify the best serving satellite/beam pair. An example of global coverage, showing the link budget fixed terms, is depicted in Fig. 4.9. The minimum elevation angle considered for the UE is 10° . In the DL, the useful received signal, C (Eq. 4.5), is computed with respect to the best serving satellite/beam pair while the interference, I (Eq. 4.10), is considered from all the satellite/beam pairs visible to the UE. As opposed to the S-S scenario, here the interference toward any UE is not only dependent on the beam layout geometry but also from the satellite-UE relative positions, *i.e.*, slant range and elevation angle with respect all the constellation satellites. As for the UL, a random scheduling is performed in the same fashion of S-S scenario, such to select only one transmitting UE for each beam of every satellite. Then the interference towards all the satellite/beam pairs visible from the UE is computed, as previously described, accounting for the satellite-UE relative geometry. In this case, then, the interference is also impacted by the scheduling algorithm.

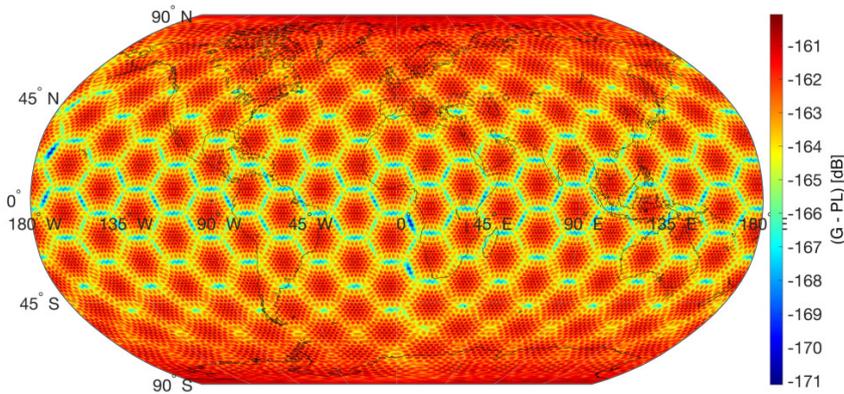


FIGURE 4.9: Example of satellite constellation coverage (1200 km)[74].

For the sake of synthesis, CDF graphs are not shown but a summary of the results, with the mean and the standard deviation of the CINR for all cases, is shown in Table 4.4. Since the constellation provides full coverage on ground, the mean values of all cases are lower with respect to the link budget calibration results shown previously, while the standard deviation is higher. These results are motivated by the fact that the coverage is much bigger with respect to the S-S scenario; focusing on the LEO at 600 km for example, the coverage is composed of $N_{beams} = 127$ beams, opposed to the 19 retained beams for the statistics for the S-S scenario.

Numerical results show that FFR schemes require the implementation of interference management techniques, while for multi-color FR schemes the link budget already provides a good performance.

TABLE 4.4: Single- and Multi-Satellite results: CINR μ and σ in dB.

Case		GEO		LEO1200		LEO600	
		μ	σ	μ	σ	μ	σ
DL (S-S)	FFR	-5.04	1.20	-1.65	1.22	-1.65	1.22
	FR3	-3.03	1.33	5.06	0.69	5.07	0.68
DL (M-S)	FFR	-9.24	6.09	-8.13	2.81	-8.37	2.59
	FR3	-8.50	6.42	-1.20	3.16	-1.34	3.35
UL (S-S)	FFR	-4.65	1.62	-1.20	2.01	-0.82	2.15
	FR3	-2.48	1.42	5.98	1.14	8.24	1.23
UL (M-S)	FFR	-6.45	6.40	-3.81	3.56	-3.50	3.20
	FR3	-5.26	3.85	0.21	3.34	3.55	4.53

Chapter 5

Link level analysis

This chapter reports the analysis performed at the link level of the SatCom based NB-IoT system.

In particular, in the first section, the preamble detection and the estimation of the synchronization algorithm are addressed. Since only orthogonal users are considered, *i.e.*, no collisions are present in the system, these results may act as a baseline. Indeed, in the next Sections, collisions are taken into account. As a matter of fact, when considering a SatCom system, which generates a beam much wider than a terrestrial cell, an enormous number of users can potentially communicate with the flying platform. As the time/frequency resource is very limited in NB-IoT, this high number of devices contending the same resources generate congestion. Thus, in the following, two approaches are proposed to alleviate the packet loss ratio the network by proposing a Neural Network (NN) based RA and NOMA approach for data transmission in NTN.

5.1 Preamble detection in NB-IoT via Satellite: a Wavelet-based approach

This section reports the results obtained in the author's publication "Preamble detection in NB-IoT via Satellite: a Wavelet-based approach" [80].

In the scientific literature, several studies focused on the analysis of the RA procedure in terrestrial environments, e.g., [81]–[85]. In [81]–[84], the classical way of synchronization parameters estimation is performed, *i.e.*, either joint frequency and timing estimation or CFO estimation and compensation before timing estimation, based on the Fast Fourier Transform (FFT) analysis. While in [85], only the ToA is estimated, after having eliminated the CFO. In this work, the performance of a novel NB-IoT RA acquisition procedure is proposed and assessed in a specific scenario defined by the 3GPP in [39]. In particular, the objective is to consider the detection problem of NB-IoT preambles in the presence of CFO and RTD, and the estimation of these two synchronization parameters. A method based on Non-Decimated or Stationary Discrete Wavelet Transform (S-DWT) is proposed for the detection of the random access preamble. The reason that leads to the choice of this algorithm is the peculiar structure of the signal. Indeed, a joint time-frequency analysis allows

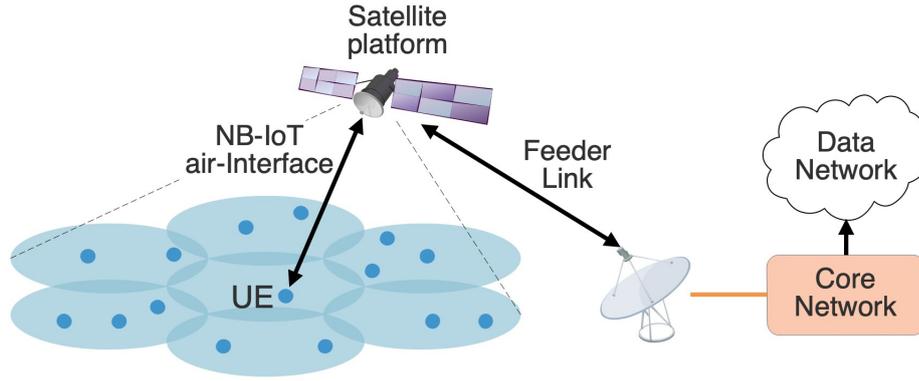


FIGURE 5.1: High-level system architecture [80].

decomposing the signal at multiple time and frequency resolutions, identifying not only the frequency hopping sequence through the Maximum Likelihood (ML) criterion but also the frequency location in the time domain. In particular, the use of the S-DWT is motivated by the fact that the Fourier transform (FT) has a fundamental limit. The latter gives the spectral content of the signal, but it does not provide information about the time localization of those spectral components. Information about local features of the signal, such as changes in frequency, becomes a global property of the signal in the frequency domain. As described in 2, the NB-IoT preamble is a non-stationary signal, transmitted through a single carrier frequency hopping modulation, carrying its information in the frequency hops. For this reason, it is essential to acquire information about the signal variations in time and frequency, such as their localization in time, their frequency of occurrence, and an estimation of the magnitude of these variations. The S-DWT permits to define such variations, allowing to determine the time when the frequency changes, and, therefore, to estimate the ToA efficiently, without having to estimate and compensate the CFO first. The robustness of the S-DWT to the CFO is due to the fact that the bandwidth of each of its decomposition level is always greater than the sub-carrier spacing of the frequency hopping single tone, not causing an amplitude reduction of the desired sub-carrier. While, this effect is more visible with the FFT, because, due to the CFO, the sub-carrier is not sampled at the peak of the sinc function of the FFT itself.

5.1.1 System model

With reference to Fig. 5.1, the following system configurations are assumed: direct access, *i.e.*, the stationary UE is directly connected to the satellite, regenerative payload, and a LEO multi-beams platform operating in S-band with moving beams. This configuration refers to scenario D2 in [39]. In addition, the following assumptions hold: *i)* the feeder link is assumed ideal; *ii)* the user link is an Additive White Gaussian Noise (AWGN) channel; *iii)* a standalone NB-IoT deployment is considered; *iv)* the UEs are equipped with a GNSS receiver so as to pre-compensate the

Doppler shift and the propagation delay due to the satellite distance. Notably, assumption *iv* is the baseline assumption for the 3GPP SI [86]. Accordingly, the uncertainties of the residual parameters that the detector has to deal with are: [87]:

- timing error at the satellite $< \pm 50 \mu s$;
- frequency error at the satellite $< \pm 40 Hz$ in S band.

5.1.2 Transmitted signal

The time-domain waveform of the transmitted preamble can be expressed as [81]–[83]

$$x_{p,n}[l] = \sum_{p=0}^{P-1} X_{p,n}[N_{sc}(p)] e^{j2\pi \frac{N_{sc}(p)}{L} l} \quad (5.1)$$

where $x_{p,n}[l]$ is the l -th sample of the n -th symbol in the p -th SG and $X_{p,n}[N_{sc}(p)]$ represents the n -th symbol on $N_{sc}(p)$ sub-carrier where SG p is transmitted. For each symbol in each SG, $X_{p,n}[N_{sc}(p)]$ has a unitary value. In addition, $l = [L_{p,n} - L_{CP}, \dots, L_{p,n} + L - 1]$, $n = [0, \dots, N - 1]$, and $L_{p,n} = pL_{SG} + nL$, where $L_{SG} = L_{CP} + NL$ is the length of one SG, L_{CP} is the size of the CP, while one symbol is L samples long. At the receiver side, it is reasonable to assume that the baseband signal is affected by a ToA τ , normalized by the symbol duration, and by a CFO η (normalized by the sampling frequency) due either to the imperfect match between the carrier frequency of the user and the carrier frequency of the system or to the differential Doppler. Hence, the received signal can be written as:

$$z_{p,n}[l] = h_p e^{j2\pi\eta(l-\tau)} \sum_{p=0}^{P-1} X_{p,n}[N_{sc}(p)] e^{j2\pi \frac{N_{sc}(p)}{L} (l-\tau)} + w_{p,n}[l] \quad (5.2)$$

where h_p is the channel coefficient between the user and the satellite and $w_{p,n}$ is the additive white Gaussian noise with zero mean and variance N_0 .

5.1.3 Receiver design

The first objective of the receiver is the detection of the preamble. The detection problem is formulated as a binary hypothesis; H_0 represents the only noise hypothesis, while H_1 is the signal plus noise hypothesis:

$$\begin{aligned} H_0 : z_{p,n} &= w_{p,n}[l] \\ H_1 : z_{p,n} &= h_p e^{j2\pi\eta(l-\tau)} \sum_{p=0}^{P-1} X_{p,n}[N_{sc}(p)] e^{j2\pi \frac{N_{sc}(p)}{L} (l-\tau)} + w_{p,n}[l] \end{aligned} \quad (5.3)$$

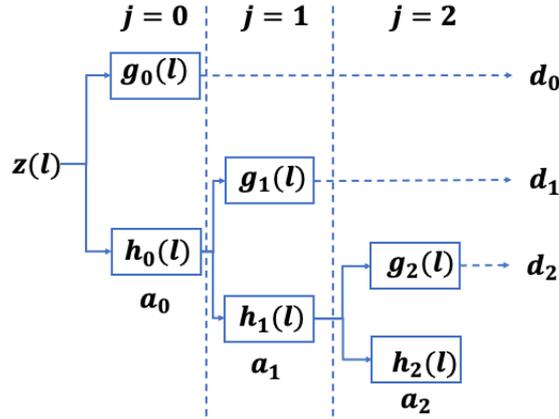


FIGURE 5.2: Stationary dyadic wavelet transform, with g_j and h_j high pass and low pass filters, respectively [80].

For the NB-IoT standard, as for 3GPP specification [88], it is required that $P_{FA}[H_1|H_0]$ (false alarm probability) is lower than 0.1% and $P_D[H_1|H_1]$ (correct detection probability) is greater than 99%.

5.1.3.1 Detection process

For the preamble detection and the hopping frequency estimation, a method based on the stationary dyadic wavelet filter banks is proposed. The choice of the stationary version of the classical Discrete Wavelet Transform (DWT) is due to the fact that, even if the DWT filtering operations are linear and time-invariant, the decimation operation, which is intrinsic in the classical DWT, together with the filtering, produces a time-variant system. As a consequence, if the starting point of the signal is shifted by only a few samples, the DWT coefficients can completely change [89]. Since the received preamble is shifted in time due to the RTD, the resulting decomposition signal mismatches the original one. Such time-variant property makes it difficult to apply the classical DWT on signal detection. A solution to this problem is therefore represented by the stationary dyadic wavelet transform [90], shown in Fig. 5.2.

Similarly to the classical DWT, the received signal is decomposed into J levels, but it is not decimated. The number of sub-carriers falling within the bandwidth of the decomposition filter decreases at each level, up to the last level which contains only a single sub-carrier, *i.e.*, the frequency equals the sub-carrier spacing. In this way, the hopping tones of the signal are not only distinguishable in the frequency domain, but follow the order in which they are transmitted. Every j -th level contains the detailed (or wavelet) coefficients of the incoming signal that are acquired through the use of the wavelet multi-resolution characteristic. In the multi-resolution analysis, the approximations of the original signal are computed at various resolutions with orthogonal projections on different spaces $(V_j)_{j \in \mathcal{Z}}$. Referring

to Fig. 5.2, $z(l)$, projected in V_0 , is decomposed into the details part, d_1 and large-scale approximation part, a_1 , then a_1 can be decomposed further in its detailed and large-scale approximation parts, and so on. To distinguish between the possible frequencies hopping transmitted, a quadrature-matched filter between the detailed coefficients of the incoming preamble (d_j^s) and those of the local replica (d_j^r) is implemented. The number of detector branches is equal to the number of sub-carriers expected in the single decomposition level, which includes the frequency bands of the form $[\frac{1}{2^{j+1}T_s}, \frac{1}{2^jT_s}]$ (T_s is the sampling period).

5.1.4 ML decision

The ML criterion leads to the selection of the frequency tone hypothesis $N_{sc}(p)$ which maximizes the likelihood function given by:

$$\Lambda(d_j^s | N_{sc}(p), \eta, \tau) = \left(\frac{1}{\sqrt{\pi N_0}} \right)^N e^{\left(\frac{-\|d_j^s - d_j^r\|^2}{N_0} \right)} \quad (5.4)$$

under the assumption of equiprobable $N_{sc}(p)$ in the range $[0, \frac{180}{SCS} - 1]$. The unknown ToA and CFO are modeled as parameters set of deterministic elements. For each observation interval, *i.e.*, L_{SG} , the outputs of this filter operation are:

$$S(N_{sc}(p)) = \frac{1}{L_{SG}} \sum_{i=0}^{L_{SG}-1} d_j^s[i] d_j^r[i]^* \quad (5.5)$$

The sufficient statistic to choose between the $N_{sc}(p)$ hypothesis, in each observation interval, is represented by square envelope $|R_p|^2 = |S(N_{sc}(p))|^2$. The one which maximizes $\Lambda(d_j^s | N_{sc}(p), \eta, \tau)$ is chosen and it is indicated as $|R_{p_{max}}|^2$. With the aim to exploit the concept of the repetitions, a non-coherent accumulation along the total number of P SGs is performed, meaning that the decision rule to discern between the hypothesis H_0 and H_1 is:

$$\Lambda = \frac{1}{P} \sum_{p=0}^{P-1} |R_{p_{max}}|^2 \begin{matrix} > \lambda & H_1 \\ < \lambda & H_0 \end{matrix} \quad (5.6)$$

where λ is the threshold. In the only noise case, Λ is the sum of $2 \cdot P$ real Gaussian random variables with identical distribution $\sim \mathcal{N}(0, N_0/2)$. Therefore, Λ is a scaled central χ^2 distribution with $2 \cdot P$ degree of freedom. While, under the H_1 hypothesis, it follows a non-central scaled χ^2 distribution, with non-centrality parameter equal $2P \cdot \sqrt{2S/N_0}$, where S is the signal energy. After the definition of the probability density function (pdf) of the sufficient statistic, taking into account both hypotheses, the Neyman-Pearson detection theory is applied to establish the constant false alarm rate at the receiver. For a given false alarm probability, number of symbol

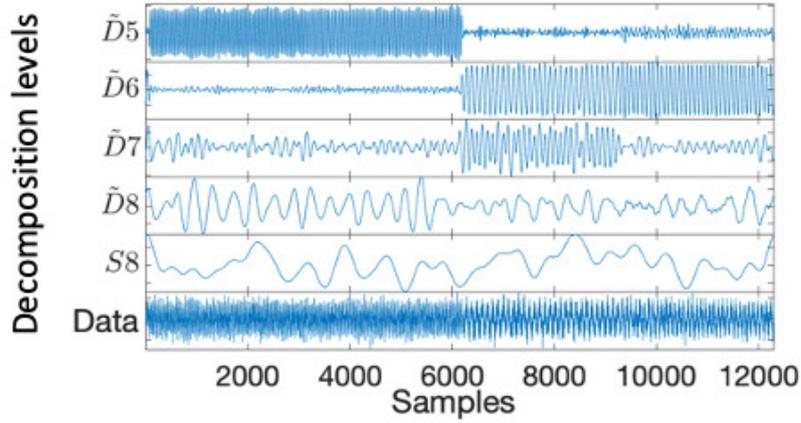


FIGURE 5.3: Output of multi-resolution analysis [80].

groups, and noise variance, the threshold λ is computed, such that it guarantees the satisfaction of the following relation:

$$p_{fa} = \int_{\lambda}^{\infty} f_{na}(x) dx \quad (5.7)$$

where $f_{na}(x)$ is the pdf of the statistics in the only noise case. From eq. 5.7, given P_{fa} , N_0 and P , λ is given by [82]:

$$\lambda = \frac{N_0}{2P} F_c^{-1}(1 - P_{fa}; 2P) \quad (5.8)$$

where F_c^{-1} is the inverse cumulative distribution function of f_{na} .

5.1.4.1 Estimation of the synchronization parameters

Once the transmitted waveform has been detected, from the S-DWT of the signal, the eNB firstly estimates the ToA, and later the CFO. Referring to Fig. 5.3, through the S-DWT based multi-resolution analysis, it is always possible to divide the SGs in two halves, where there is a difference of $\pm 22.5 \text{ KHz}$ between the SGs belonging to them, identifying at least one of the two halves in a certain level j . Indeed, looking, for example, the fifth level of Fig. 5.3, it is visible that the sequence of observations $(y_i)_{1 \leq i \leq N}$ follows two different distributions, $p_{\xi_0}(y_i)$, characterized by the parameter ξ_0 , in the interval $[1, \tau]$, and $p_{\xi_1}(y_i)$ in the interval $[\tau + 1, N]$, characterized by the parameter ξ_1 . Thus, the ToA estimation problem is equivalent to the estimation of the unknown point where occurs a change in the distributions of observed data. Following, again, the classical formulation of the binary hypothesis testing about the sequence $(y_i)_{1 \leq i \leq N}$,

$$\begin{aligned} H_0 : \xi &= \xi_0 \text{ for } 1 \leq i \leq N \\ H_1 : \xi &= \xi_0 \text{ for } 1 \leq i < \tau \\ &\xi = \xi_1 \text{ for } \tau \leq i \leq N \end{aligned} \quad (5.9)$$

H_0 represents the hypothesis of no change in the sequence $(y_i)_{1 \leq i \leq N}$, while H_1 assumes a change in the sequence after the sample τ . Consequently, the likelihood ratio (LR) between the two hypotheses is:

$$\Lambda_1^N = \frac{\prod_{i=1}^{\tau-1} p_{\xi_0}(y_i) \prod_{i=\tau}^N p_{\xi_1}(y_i)}{\prod_{i=1}^N p_{\xi_0}(y_i)} \quad (5.10)$$

Therefore, the log-likelihood ratio (LLR) is:

$$S_\tau^N = \sum_{i=\tau}^N \ln \frac{p_{\xi_1}(y_i)}{p_{\xi_0}(y_i)} \quad (5.11)$$

In this case, the two distributions are Gaussian and they have a common mean μ before and after change point τ , and a different variance, σ_0 and σ_1 , respectively. Thus, every time that a new observation is taken into account, the distance between the two pdfs is computed, through the Kullback-Leibler (KL) divergence [91], defined as:

$$\begin{aligned} E(S_\tau^N) &= -KL(p_{\xi_0}(y_i) || p_{\xi_1}(y_i)) \text{ for } i < \tau \\ E(S_\tau^N) &= KL(p_{\xi_1}(y_i) || p_{\xi_0}(y_i)) \text{ for } i \geq \tau \end{aligned} \quad (5.12)$$

After the change, the expected value of LLR will be the highest one, leading to the estimation of τ :

$$\tilde{\tau} = \max_{1 \leq \tau \leq N} S_\tau^N \quad (5.13)$$

The following step consists in eliminating the residual frequency uncertainty. Now, being synchronized to the slot in which the preamble transmission began, it is possible to remove the CP from each SG. The receiver coherently accumulates over L_{coh} samples for each SG. Then the output is averaged along the number of SGs. L_{coh} is designed to be the maximum affordable value that maintains a negligible degradation due to the overall angle rotation during the integration. In this way, the actual frequency rotation $2\pi l T_s \eta$ is discretized with a step equal to the angle rotation given by $2\pi L_{coh} T_s \eta$. Therefore, we can write:

$$\Lambda(\eta) = \left| \sum_{l=0}^{L_{coh}-1} r_l(\tilde{\tau}) e^{-j2\pi l T_s N_{sc}(p)\eta} \right| \quad (5.14)$$

where r is the reconstructed signal, after having compensated the ToA $\tilde{\tau}$. The uncertainty due to the frequency offset is solved by looking for the value of η which maximizes $\Lambda(\eta)$.

TABLE 5.1: Simulation parameters

Parameter	Value
Carrier frequency	2 GHz
Satellite altitude	600 km
Elevation angle	90°
Beam diameter	90km
Δf	3.75 KHz
Decomposition levels	8
CP length	266.7 μs
Timing uncertainty	$[-50, 50] \mu s$
Frequency uncertainty	$[-40, 40] Hz$

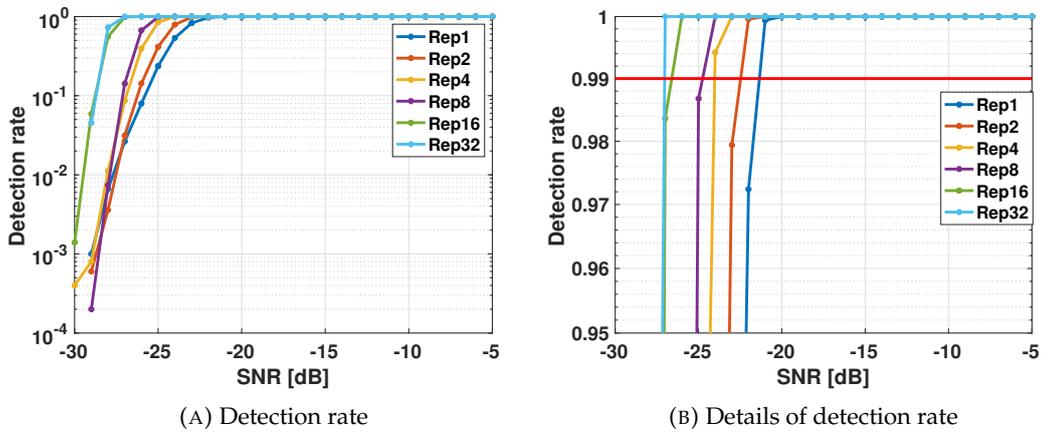
FIGURE 5.4: Detection rate with N_{rep} varying from 1 to 32 [80].

TABLE 5.2: Normalized RMSE of CFO and ToA

N_{Rep}	SNR	CFO	ToA
8	3 dB	$2.16 \cdot 10^{-4}$	0.0058
16	3 dB	$1.44 \cdot 10^{-4}$	0.0052
32	0 dB	$1.36 \cdot 10^{-4}$	0.005

5.1.5 Simulation and results

In this subsection, the performance of the RA procedure, evaluated by means of numerical simulations, is shown in terms of the detection rate and accuracy of the estimation synchronization parameters. The simulation parameters are listed in Table 5.1.

5.1.5.1 Detection performance

Fig. 5.4a reports the detection probability performance in the presence of CFO and ToA. The results are obtained through Monte Carlo simulations for N_{rep} varying from 1 to 32, i.e., for preamble length $P = 4 \cdot N_{rep}$. The SNR values of the UE vary from $-30 dB$ to $-5 dB$. The performance of the detector has been tested with the highest value of CFO and ToA, which are 40 Hz and 96 samples, i.e., 50 μs at the

sampling frequency of 1.92 MHz, respectively.

With reference to Fig. 5.4b, taking into account the minimum number of repetitions, the SNR that guarantees 99% ($SNR_{99\%}$) of detection is -22 dB. While the use of 32 repetitions leads to a gain of 6 dB of SNR. Similar performances are achieved in [82].

5.1.6 Delay and CFO estimation performance

The estimation performance of the synchronization parameters is evaluated in terms of the normalized root mean square error (RMSE) (Table 5.2) and they are shown in the form of CDF in Fig. 5.5-5.7. Both the CFO and ToA are randomly selected in the range $[-40, 40]Hz$ and $[-50, 50]\mu s$, respectively. Initially, the algorithm is evaluated by taking into account the frequency and the timing residual error after the pre-compensation done with the GNSS. Regarding the normalization of the CFO and ToA estimation error, it is done on the sub-carrier spacing and on the symbol duration, respectively. It is observed that fixed the SNR value, the RMSE of both the ToA and CFO becomes smaller with increasing N_{rep} (from 8 to 16), as expected. While decreasing the SNR, it is necessary to keep higher the number of repetitions ($N_{rep} = 32$), to cope with the worst channel conditions. The RMSE of the CFO is always smaller than 1 Hz and the one of the ToA is almost $1\mu s$. Fig. 5.5 shows the CDFs of the normalized CFO estimation error, which confirms the robustness of the algorithm. Then, to understand the goodness of the ToA estimation, the proposed method has been compared with the technique applied in [59]. Here, the NPRACH reception method allows estimating the ToA, in such a way that the CFO, present in the received signal, is eliminated. The authors consider a residual CFO, after the downlink frequency synchronization, equal to 600 Hz. This algorithm works in the frequency domain. Indeed, at the receiver side, the CP is removed and the DFT is applied. After that, the symbols inside the same SG are combined to obtain the SG-sum (SG-S) and the differential processing is performed by multiplying the $p - th$ SG-S with the complex conjugated $(p + 1) - th$ SG-S of the same repetition, obtaining the symbols $Z_{p,1}$. The latter are collected in an array v_u . Their position in v_u corresponds to their hopping step $\Delta(p) = N_{sc}(p + 1) - N_{sc}(p)$. Thus, N_{rep} arrays are built. For each array element, the Rife & Boorstyn (R&B) method is implemented: $V_u[k] = \sum_{n=0}^{NDFT-1} v_u[n] e^{-j2\pi k \frac{n}{NDFT}}$. For the latter 512 points NDFT are used and k indicates the sub-carrier index. After that, $V_u[k]$ is non-coherently combined over N_{rep} : $X[k] = \sum_{u=0}^{N_{rep}-1} |V_u[k]|^2$. Thus, the ToA is estimated as: $D = \frac{k_{max}}{NDFT \cdot SCS}$, where SCS indicates the sub-carrier spacing and $k_{max} = argmax_k X[k]$. In the following, the ToA estimated with this algorithm is indicated as R&B. Referring to Fig. 5.6 - 5.7, the CDFs of the normalized ToA estimation error are reported for the S-DWT and the R&B algorithm for 600 Hz of CFO. It is seen that the proposed algorithm outperforms the R&B one because its CDF curves have a steeper slope. When the number of repetitive transmissions increases from $N_{rep} = 8$ to $N_{rep} = 16$, the ToA

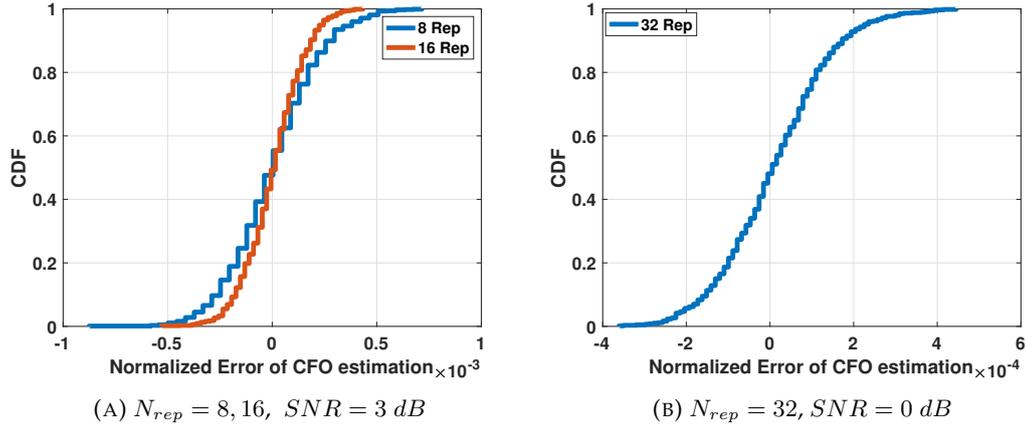


FIGURE 5.5: CDF of the normalized CFO estimation error [80].

estimation of all algorithms becomes better. It is worth noting that, despite the difference between the two simulated CFO, the curves obtained with the designed algorithm follow the same trend, demonstrating its robustness against the CFO. Since this study was carried out by contextualizing it in a SatCom system, defined by the architecture described above, it is interesting to verify whether the detection and parameter estimation performances meet the requirements of the NB-IoT standard based on the link budget analysis of the simulated system. For this purpose, the curves in Fig. 5.8 are considered, which represent the uplink budget of the Single-satellite scenario obtained with the same methodology explored in [74], setting the uplink bandwidth (BW) to 3.75KHz . Since the curves depend on the system bandwidth, it is necessary to extract the equivalent C/N_0 to be used in the end-to-end link simulations: $\frac{C}{N_0} = \frac{C}{N} \cdot BW = BW \cdot \left(\frac{1}{CINR} - \frac{1}{CIR}\right)^{-1}$ and $\frac{C}{N_0} = \frac{C}{N} \cdot BW$, where CIR is the carrier to interference ratio and $CINR$ is the carrier to noise plus interference ratio. Once the equivalent value of C/N_0 is found, it is possible to consider it as a threshold for the proposed end-to-end link-level simulations. It is assumed that the interference is Gaussian and included in the N_0 term. Thus, it is visible that the proposed receiver is able to fully satisfy the requirements of the standard in our NB-IoT SatCom system, being the $SNR_{99\%}$ bigger than the equivalent C/N_0 . The overall performance demonstrates the ability of the proposed algorithm to address the NTN scenario previously described. The advantage of this method is the possibility to estimate the ToA in an accurate manner, without having to compensate the CFO first. This implies that the ToA estimate is not bounded by the CFO estimation errors, decreasing the uncertainty region of the total search space. Whereas, the main drawback is the use of a high number of preamble repetitions to reach the accuracy of $1\mu\text{s}$.

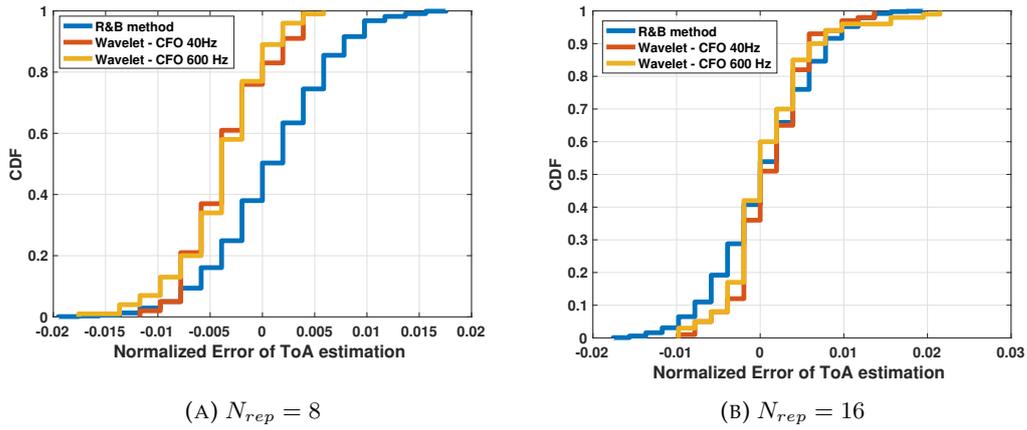


FIGURE 5.6: CDF of the normalized ToA estimation error, $SNR = 3\text{dB}$ [80].

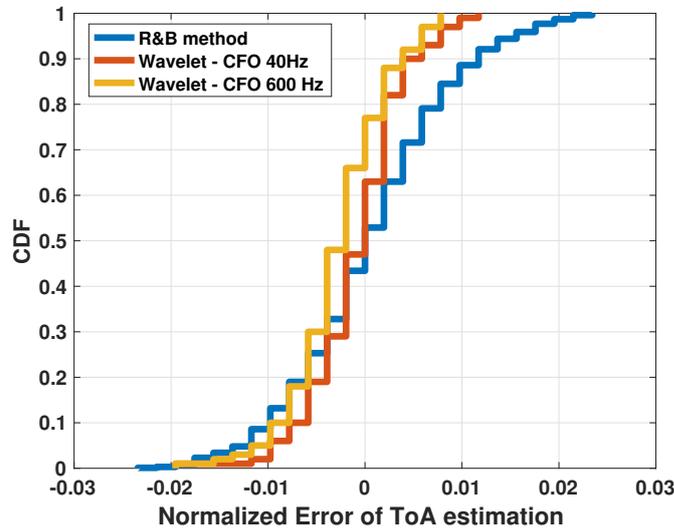


FIGURE 5.7: CDF of the normalized ToA estimation error. $N_{rep} = 32$. $SNR = 0\text{dB}$. [80].

5.2 Neural Network based Non-Orthogonal Random Access for 6G NTN-IoT

In this section, the analysis and the results derived in the author's publication "Neural Network based Non-Orthogonal Random Access for 6G NTN-IoT" [92] are presented. As previously mentioned, IoT communications are mainly characterized by sporadic uplink data reporting, therefore, non-continuous satellite coverage, provided by cost-efficient incomplete constellations, is the baseline approach for most of the foreseen IoT-NTN architectures. Thus, all the terminals within the satellite beam, simultaneously competing to access the network, must be served within the short visibility period of the flying platform, generating congestion. For the moment, 3GPP in its Rel.18 recognizes the congestion has a problem [93], since it may cause the deterioration of system performance or even the unavailability of the service.

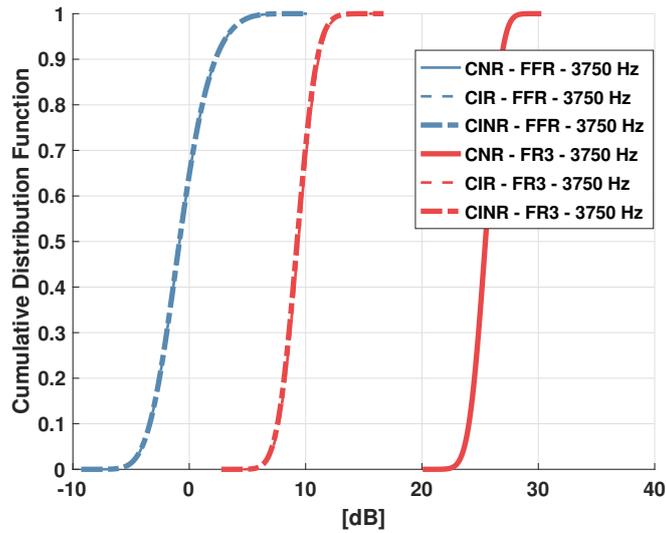


FIGURE 5.8: Link budget analysis of the proposed scenario [80].

However, solutions will be most likely proposed in Rel. 19, where more sophisticated architectures are considered also for IoT scenarios [34]. A possible countermeasure to this issue is represented by the application of NOMA schemes to Message 3 of the RA protocol, [94]. However, in order to fully benefit from the advantages brought by NOMA, it is fundamental to identify and count the terminals transmitting the same time/frequency resources. In this context, this study proposes an innovative method for the detection of non-orthogonal NB-IoT preambles and the estimation of their synchronization parameters, based on Convolutional Neural Network (CNN) algorithms. Indeed, the ultimate goal is to improve the capacity of the critical random access phase by not discarding interfering signals, and thus by increasing the number of served users during the short satellite visibility window. In particular, a novel approach is proposed where first the number of collided users is detected and, then, their ToA are estimated under different propagation conditions and satellite configurations.

5.2.1 System Architecture

With reference to Fig. 5.9, the main elements of the high-level system architecture are: *i*) a plethora of UEs, which are *fixed* NB-IoT sensors; *ii*) satellites providing connectivity to the UEs through the user link; *iii*) ground segment, *i.e.*, the gateways (GW), which provide inter-connectivity between the satellite constellation and the gNBs, by the means of feeder links, and the 5G-A CN through the ground distribution network, in particular with the Operations Support Systems (OSS) entity, in charge of managing the overall system. The following system configurations are valid: direct access, transparent payload, LEO single-beam platform operating in S-band, with moving beam. As for the previous work, the following assumptions hold: *i*) the feeder link is ideal; *ii*) a standalone NB-IoT deployment is considered;

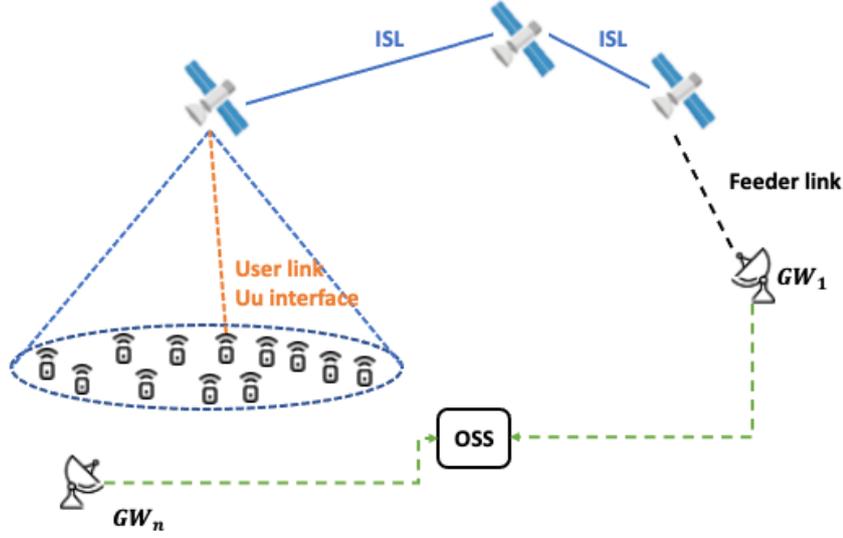


FIGURE 5.9: High-level system architecture [92].

iii) the UEs are equipped with a GNSS receiver so as to pre-compensate the Doppler shift and the propagation delay due to the satellite distance.

5.2.2 Transmitted and received signals

In the time domain, the transmitted preamble is the same one presented in Eq. 5.1: In the considered coverage area, J users are uniformly distributed. Within the beam, each user is described by a complex channel coefficient representing the channel between the j -th user and the satellite, given by:

$$h_j = \frac{\sqrt{G_{TX} G_{RX} \Omega(\theta_j^s)}}{4\pi \frac{d_j}{\lambda} \sqrt{L_{loss,j}} P_z} e^{-j \frac{2\pi d_j}{\lambda}} \quad (5.15)$$

where *i*) G_{TX} represents the transmission gain of the device antenna, which is equal for all users. For the purpose of the work it is assumed that the terminals are equipped with an omnidirectional antenna, being the users NB-IoT devices; *ii*) G_{RX} is the maximum receiving gain of the satellite; *iii*) d_j is the slant range between the j -th user in the beam and the satellite; *iv*) λ is the carrier wavelength; *v*) P_z is the noise power, which depends on the satellite antenna equivalent noise temperature, T , and the user bandwidth, B . *vi*) L_{loss} represents the channel losses, which are user-dependent as they vary with the position of the terminal. These losses include $L_{loss,j} = L_\sigma + L_a + L_s + L_{cl}$, which are the same presented in Chapter 3. Finally, the term $\Omega(\theta_j^s)$ defines the satellite radiation pattern, which is a function of the angle between the antenna pointing of the j -th user and the s -th satellite, *i.e.*, θ_j^s , which does not depend on the satellite antenna index due to antenna co-location. In this study, a Bessel radiation pattern is considered [35]. Based on the above assumptions

and definitions, the signal received by the satellite is:

$$y_p = \sum_{j=1}^U \sqrt{P_{tx}} h_j s_{p,j} + z_p \quad (5.16)$$

where U is the number of concurrent users, *i.e.*, the users transmitting the same preamble identified by the sub-carrier P , P_{tx} is the transmitting power, which is assumed to be the same for all users, and z_p is a complex circularly-symmetric Gaussian random variable with zero-mean and unit variance. The latter is a licit assumption since the noise power is included in the term P_z in eq. 5.15. Finally, $s_{p,j}$ denotes the signal from the j -th user, given by:

$$s_{p,j} = \sum_{p=0}^{P-1} X_{p,n,j}[N_{sc}(p)] e^{j2\pi\eta\frac{L-\tau}{L}N_{sc}(p)} \quad (5.17)$$

where τ is the ToA due and η is the CFO. At the receiver side, the DFT is applied to each symbol of each SG, including the CP.

5.2.3 NN approach to preamble detection

In the proposed scenario, two CNNs are designed to extract the non-linear relationship between the DFT coefficients of each symbol of each SG and their targets and they are used in cascade: the first one performs the collision classification, *i.e.*, it classifies the number of users transmitting the same preamble, and its output is used as input for the ToA estimation network. Please, consider that the description of the dataset is provided in the Section on the numerical results.

5.2.3.1 Neural Network implementation

The CNN architectures for the collision classification and the ToA estimation are depicted in Fig. 5.10 and Fig. 5.11, respectively. Both the schemes follow the same structure up to the last layer, where they differ to perform the two different tasks, *i.e.*, the classification of the number of users and the regression; the latter aims at yielding ToA values close to the original ones in the Mean Squared Error (MSE) sense. With respect to the network's input, both the CNNs take the DFT coefficients of the received symbols and the second network, in charge of the ToA estimation, leverages also the number of colliding users identified by the first network. The first two layers are implemented with convolutional layers followed by max-pooling operations. Thus, the input is processed so as to extract spatially correlated features that are exploited for task learning. The 1D convolution layer is followed by batch normalization, Rectified Linear Unit (ReLU) activation function, and max-pooling. The latter aims at reducing the output dimension of the CNN and making the network invariant to the small changes in the input and, therefore, to the noise in the training samples. It is worth emphasizing that the ToA and the CFO are assumed to be

constant throughout a transmission, therefore the 1D convolution layer is capable of extracting invariant features of the frequency-domain signal. Subsequently, these features are shaped into a single vector that feeds three fully connected layers. The first two are followed by the ReLU activation, while the last one is softmax activation for the classification network and a linear layer for the regression.

5.2.3.2 Collision classification

The collision classification is commonly applied to all the available preambles and separates them into several classes. This network accepts $(Re(y), Im(y))$ of the DFT coefficients of each symbol as an input and it is trained to predict the number of devices transmitting the same preamble. To this aim, the following $U + 1$ different classes are considered:

1. **CLASS 0:** the preamble is unused, meaning there is only noise.
2. **CLASS 1:** the preamble is selected by a single UE, *i.e.*, collision-free.
3. **CLASS u :** the preamble is chosen by u out of U UE for $u \in 2, \dots, U - 1$, therefore, there are u collisions.
4. **CLASS U :** the same preamble has been selected by U terminals.

The label of each $U + 1$ class consists of a one-hot vector $t = OneHot(U)$, where the index u in the truth label $t = [t_0, t_1, \dots, t_U]$ is one and zero everywhere. Therefore, the number of collided users per preamble is given by the L_0 norm of the sparse vector t . This network is trained so as to minimize the softmax cross entropy loss function (or negative log-likelihood), given by:

$$L_{cc}(t, p) = - \sum_{u=0}^U t_u \log(p_u) \quad (5.18)$$

where p_u is the output of the Softmax function, which computes the probability associated with each category by taking the scores of the Fully Connected layer (FCL), as follows:

$$p_u = \frac{e^{a_u}}{\sum_{j=0}^U e^{a_j}} \quad (5.19)$$

where $[a_0, \dots, a_U]$ are the outputs of the FCL. Therefore, the index (u_{max}) associated to the highest probability value, *i.e.*, $p_{max} = arg \ max \ p$, within the vector $p = [p_0, \dots, p_U]$ is selected to obtain the number of colliding users.

5.2.3.3 Delay estimator

After the preamble collision classification, the ToA estimation is performed by the second CNN. For each preamble, the neural network takes as input the $(Re(y), Im(y))$ of the DFT coefficients of each symbol together with the class value that identifies the number of non-orthogonal users, *i.e.*, u_{max} , and it is trained to estimate the ToA

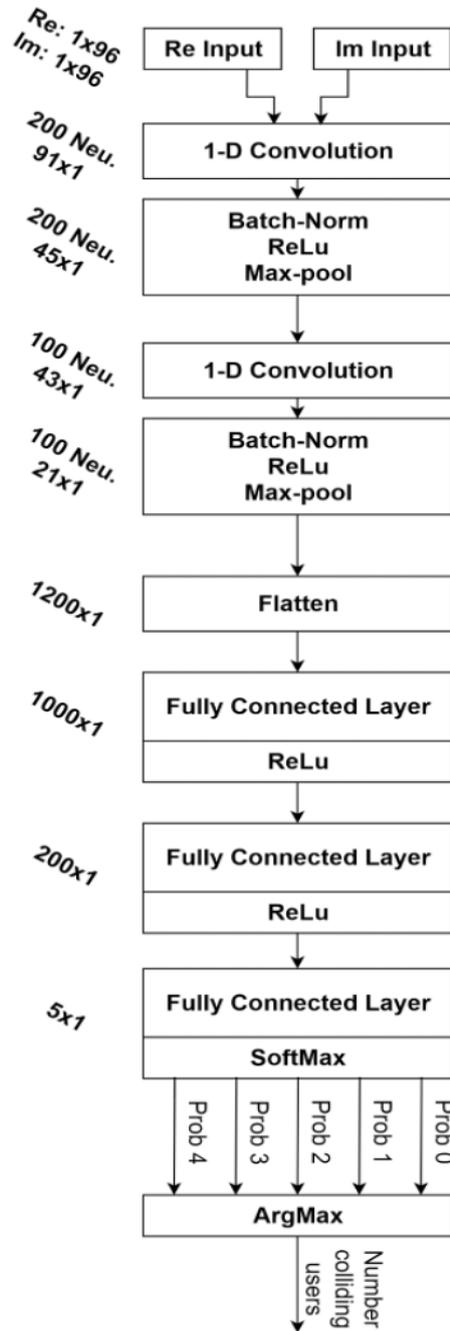


FIGURE 5.10: Structure of the CNN for collision classification [92].

of the colliding terminals in each signal. To this aim, the ToA values are normalized on the highest value, thus enabling the training convergence, and they are collected in the vector $\mathbf{Y} = [\tau_0, \tau_1, \dots, \tau_{U-1}]$. Therefore, the network finds an estimate $\hat{\mathbf{Y}}$ of \mathbf{Y} so as to minimize the MSE loss function *i.e.*, $\mathbf{E} = \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2$.

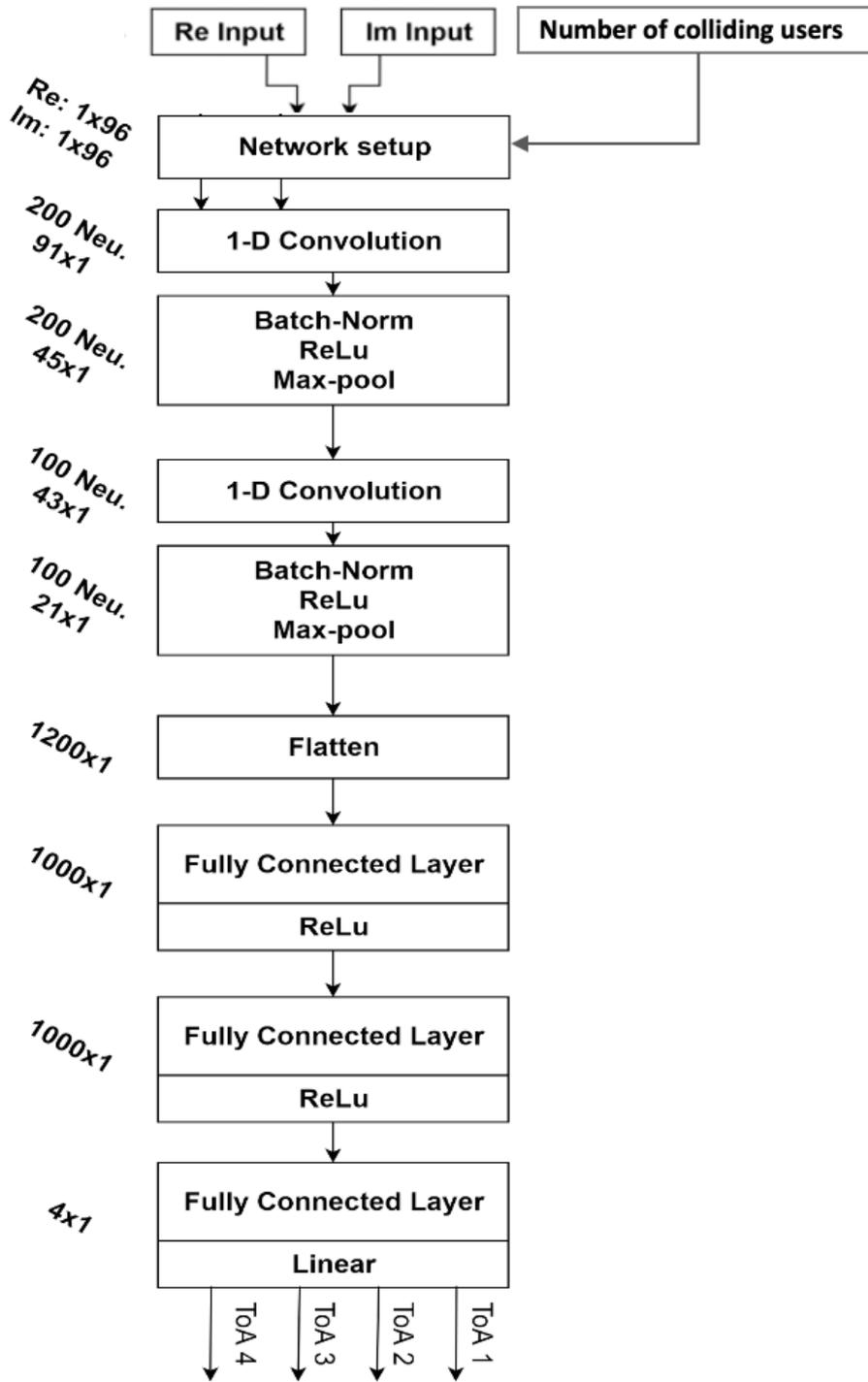


FIGURE 5.11: Structure of the CNN for ToA estimation [92].

5.2.4 Numerical Results

In this section, the performance of the proposed CNN-based Random Access is examined. Firstly, the accuracy of the number of colliding users classification is assessed and then the ToA estimation performance is evaluated. For the first network,

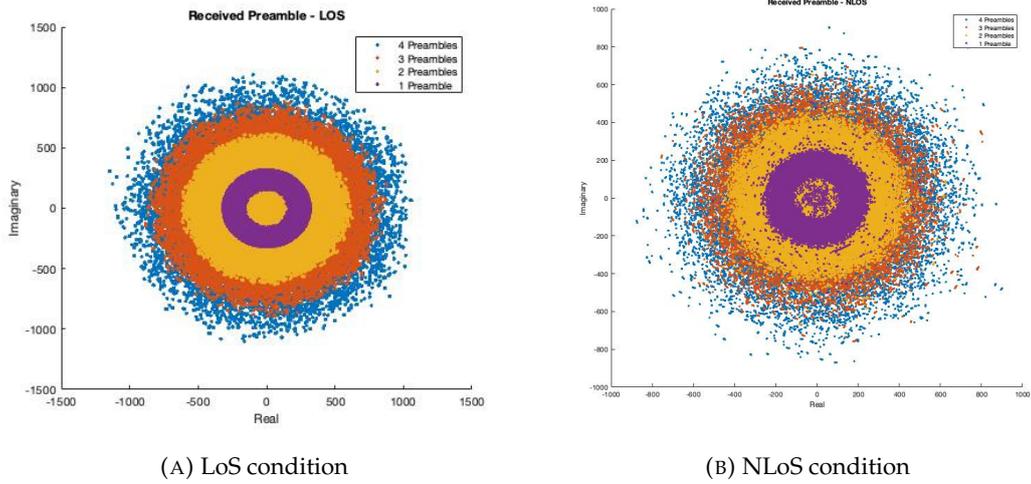


FIGURE 5.12: Real and Imaginary part of the DFT symbols. Set 3. [92].

TABLE 5.3: Satellite parameters [23]

Satellite orbit	Set 3 LEO 600 km	Set 4 LEO 600 km
Equivalent satellite antenna aperture	0.4 m	0.097 m
Sat Tx/Rx max Gain	16.2 dBi	11 dBi
3dB beamwidth	22.1 degree	104.7 degree
Sat beam diameter	234 km	1700 km
G/T	-12.8 dB K ⁻¹	-18.6 dB K ⁻¹

the learning rate is fixed to $5 \cdot 10^{-4}$, while for the second one, it is $1 \cdot 10^{-4}$, which guarantees the convergence of both the CNNs. The dataset is synthetic and consists of $42 \cdot 10^6$ samples generated in Matlab and organized in 4200 batches with 10000 examples each, where every input is the couple of the real and imaginary part of 96 DFT symbols, trained over 20 epochs. Each preamble consists of four SGs, each one with five symbols plus the CP, whose length over time is the same as a symbol, *i.e.*, $266.7 \mu\text{s}$. Therefore, considering four repetitions, the total amount of symbols sent is given by $\{6 \cdot 4 \cdot 4 = 96\}$. For both the collision classification and the ToA estimation, we generated a dataset with DFT symbols containing from 0 to 4 colliding users per preamble; 80% of the samples are used for training with the stochastic optimization method based on adaptive momentum (ADAM), and 20% for testing. The estimation performance of the ToA is evaluated in terms of the RMSE (Table 5.5). For each user transmitting the same preamble, the network estimates the ToA and the estimation accuracy is obtained by averaging all the RMSEs up to the value u_{max}

$$RMSE = \sqrt{\frac{1}{u_{max}} \sum_{u=1}^{u_{max}} \|\mathbf{Y}_u - \hat{\mathbf{Y}}_u\|^2} \quad (5.20)$$

Both the CFO and ToA are randomly selected in the range $[-40, 40] Hz$ and $[-50, 50] \mu\text{s}$, respectively [87]. Indeed, the proposed method is evaluated considering the frequency and the timing residual error after the pre-compensation done with the GNSS.

Moreover, all datasets are generated considering the following scenarios: *i*) one coverage enhancement zone, *i.e.*, all the 48 sub-carriers are available for the preamble transmission; *ii*) NB-IoT transmitting power equal to $P_{tx} = -7$ dB; *iii*) receiver noise power equal to $P_z = -163.8$ dB; *iv*) two different satellite configurations proposed in [23], *i.e.*, set 3 and set 4, whose parameters are shown in Table 5.3; *v*) one propagation environment, *i.e.*, sub-urban, in both LoS and NLoS conditions, whose parameters are available in [39]; *vi*) UEs uniformly deployed in the beam coverage. It shall be noticed that, in UL, the interference towards the satellite is clearly depending on the specific interfering UE location. Therefore, both the shadow fading and the interference source location vary through each example, in addition to the sub-carrier index of the preamble. The shadow fading is modeled as a log-normal random variable $L_\sigma \sim (0, \sigma^2)$ with the values of σ^2 depending on the UE's elevation angle [35]. In addition, the SNR ranges reported in Table 5.4 are used. For each of the proposed scenarios, the network is trained with the minimum achievable SNR and it is tested by randomly choosing users within the beam coverage. As it is visible, the SNR variance in NLoS condition is higher than the one in the LoS, due to the presence of the clutter loss. Notably, the larger the beam size, the greater the variance of σ^2 .

5.2.4.1 Collision classification performance

Fig. 5.12a and 5.12b depict the inputs of both the CNNs, *i.e.*, the $\{Re, Im\}$ parts of the DFT symbols in LoS and NLoS conditions varying the number of colliding users per preamble. The datasets are equally distributed among the classes. It is worth emphasizing that the annulus in the figures, which correspond to the outer edge of an interfered preamble, gains a different dimension based on the number of colliding terminals (shown in different colors). Indeed, this area hosts symbols that have been constructively interfered by 1, 2, or 3 more users in a received signal and the values of the $\{Re, Im\}$ parts are key in the learning process. It is worth emphasizing that the plotting order is from one to four users per preamble. Clearly, the dataset includes all the classes previously explained. Given that the symbols of an interfered signal set belong to a circle and not an annulus, the larger the number of collisions per signal the less symbols will be accommodated in the *Learning Range (LR)* of the given group, *i.e.*, the real and imaginary parts of the signal with constructive interference. This means that destructively interfered symbols might cause a miss classification if there are no other symbols in the preamble that have acquired sufficient amplitude to fit within the *LR* annulus. From the distribution of the inputs in Fig. 5.12a and 5.12b it is understandable that, with the same number of examples per class, the CNN is able to correctly classify 0, 1, and 2 interferers per preamble, as reported in Fig. 5.13a - 5.14b. Indeed, with a maximum of two colliding terminals, the network is able to correctly classify the number of colliding users with a probability of about 0.99. Indeed, when there are one or two users per preamble, the network has more examples to learn. On the other hand, classes 4 and 5 are the most difficult to correctly classify. The classification accuracy decreases due to the fact that

the network has fewer examples in the corrected LR . In NLoS condition, edges of the LR have less outlined boundaries compared to the LoS condition, because symbols of a collision class are placed outside the LR of the corresponding group. This leads to a degradation of the classification performance, shown in Fig. 5.13b and 5.14b. Due to the large beam sizes and, consequently, the low elevation angles, resulting in large values of CL, the classification of users within Set 4 in NLoS condition shows the worst performance.

5.2.4.2 Delay estimator performance

Referring to Table 5.5, it is observed that, despite low CNR values and uncompensated CFO, the RMSE of the ToA is around $5 \mu s$ in the absence of collision, while the performance decreases by increasing the number of superimposed symbols, as expected. It should be noted that, fixed the number of interferes, the estimation accuracy remains constant across the two different satellite sets in both the LoS and NLoS conditions. Subsequently, to understand the goodness of the CNN-based ToA estimation, we compare it with the technique applied in [59], which is the R&B algorithm described in section 4.1.6. However, it should be noted that with the traditional methods, it is possible to estimate only a single value of ToA, regardless of the number of interferes. Therefore, the accuracy of the R&B estimation algorithm is measured as follow: $\mathbf{E}_{\text{R\&B}} = \min \| \mathbf{Y} - \hat{\mathbf{Y}}_{\text{R\&B}} \|$. Considering only one UE, the proposed CNN approach provides higher accuracy under the NLoS condition with respect to the one achievable with the R&B method, as shown in Table 5.5. However, the CNN-based estimator is not able to reach the accuracy of the R&B estimator with one user in LoS. The reason is twofold. First of all, the number of examples needs to be higher and the second reason lies in the choice of the error function. Indeed, a custom error function, where Y is a local copy of the transmitted signal delayed by the true ToA and \hat{Y} is the received signal delayed by the estimated ToA better resembles the approximation of the maximum likelihood function, which is exactly what the R&B algorithm does.

Clearly, the CNN-based ToA estimation provides advantages in case of collisions, as it is able to estimate the ToA for each colliding user also in the presence of CFO as well as with low CNR values.

The proposed method allows to identify the colliding preamble and to know their ToA. This information enables to implement NOMA schemes both in the Msg3 (with or without data) or in the data phase, *i.e.*, when the user is in RRC connected state, as the receiver is aware of the number of users in the network. The algorithm is designed to work with the GNSS but it could also be adopted in cases where the UEs do not have GNSS capabilities. On the other hand, to provide a more accurate estimation, the network needs as input a large number of examples with different types of channel coefficients. In fact, the ToA estimation in set 4 is slightly better, since the NN can learn from different types of coefficients due to the wider beam dimension. Future works foresee the use of real data in order to assess how the degradation of

TABLE 5.4: Mean and variance of CNR

Satellite Set	Propagation condition	SNR dB
Set 3	LoS	$\mu = 9.3$
		$\sigma = 1.42$
	NLoS	$\mu = -6.9$
		$\sigma = 7.03$
Set 4	LoS	$\mu = 0.1$
		$\sigma = 2.14$
	NLoS	$\mu = -15.05$
		$\sigma = 10.04$

TABLE 5.5: RMSE of ToA [μs]

Satellite Set	Prop. condition	N.users	CNN Output	R&B Output
Set 3	LoS	1	5.3 μs	0.27 μs
		2	16.4 μs	26.7 μs
		3	19.8 μs	30.3 μs
		4	22.5 μs	32.7 μs
	NLoS	1	5.7 μs	5.8 μs
		2	17.0 μs	74 μs
		3	20.5 μs	142.2 μs
		4	23.1 μs	165.1 μs
Set 4	LoS	1	4.8 μs	1.4 μs
		2	15.3 μs	27.2 μs
		3	18.1 μs	57.3 μs
		4	21.7 μs	112.6 μs
	NLoS	1	4.2 μs	4.89 μs
		2	14.7 μs	27.8 μs
		3	17.3 μs	87.8 μs
		4	20.9 μs	120 μs

the accuracy of the CNN. In particular, an analysis of the variation of channel coefficients in a real environment should be done to efficiently design the length of the preamble.

5.3 SCMA performance in NB-IoT NTN system

This section is the output of the research study performed in the context of "Study of 6G enabling key technologies based on 3GPP standard", which was supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2021-0-00847, Development of 3D Spatial Satellite Communications Technology). The objective is to identify and assess the performance of an uplink NOMA technique suitable to accommodate a large number of NB-IoT devices via NTN.

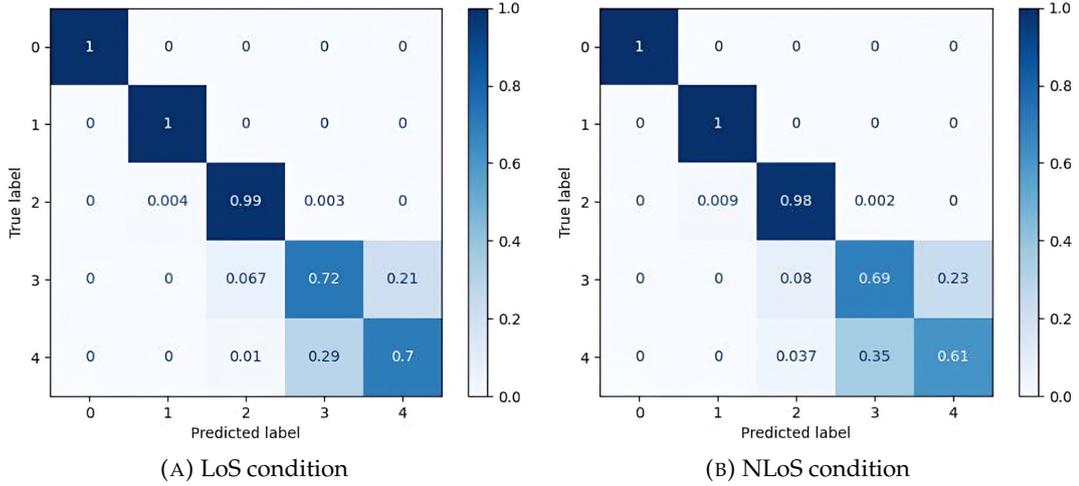


FIGURE 5.13: Outcome of the classification: Set 3. [92].

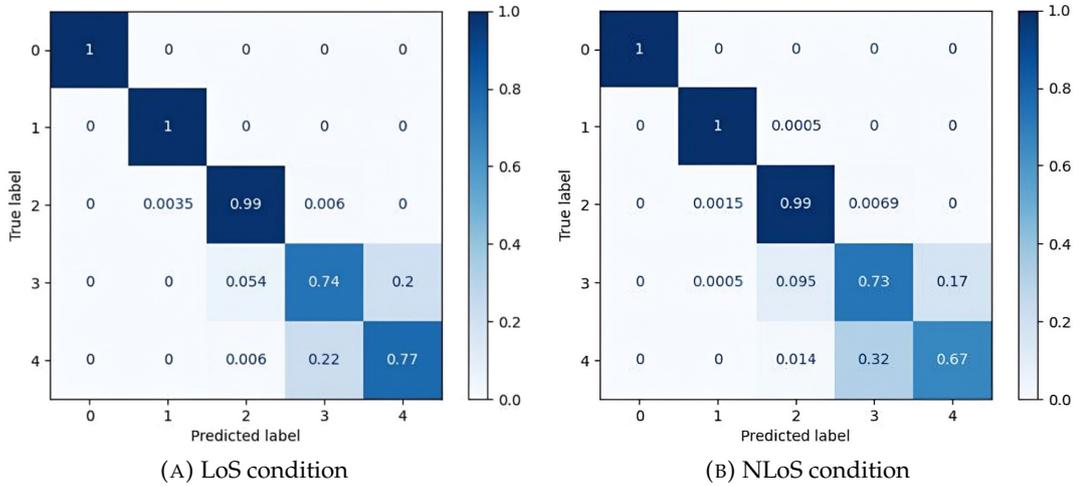


FIGURE 5.14: Outcome of the classification: Set 4 [92].

5.3.1 Overview of NOMA

In wireless communication systems, Multiple Access (MA) schemes are used to allow multiple users to access and share the same system resources simultaneously. MA techniques have been considered the landmark of each generation of cellular mobile systems. For example, Frequency Division Multiple Access (FDMA) was introduced in 1G (the first generation of analog cellular systems) Time Division Multiple Access (TDMA), and Code Division Multiple Access (CDMA) were used for 2G and 3G respectively. Conversely, OFDMA is utilized in 4G and 5G. All aforementioned MA schemes can be regarded as Orthogonal Multiple Access (OMA) solutions because orthogonal resources in the time, frequency, or code domains are used for users' separation. However, since the demand for cellular mobile communications increases rapidly, OMA technology will probably not meet the requirements of 6G, which needs to support a massive number of connections and higher system capacity. Therefore, NOMA schemes have attracted a lot of attention for future terrestrial and satellite communications [95]. In NOMA technology, more than one layer

of data for more than one UE can be transmitted simultaneously without frequency, time, and/or spatial domain separation. At the receiver side, different layers of data can be separated by using interference cancellation or iterative detection. NOMA schemes can be applied to all the 5G use cases.

In a non-orthogonal multiple access scheme, both the transmitter and the receiver must be jointly optimized in order to transmit data layers of different UEs simultaneously over the same transmission resources. At the transmitter, the signal of multiple UEs can be superimposed over the same time, frequency, and/or spatial resource. The transmitter is characterized by MA signature and features, which are typically used to differentiate the users. At the receiver, the signal from different UEs can be detected correctly by proper receivers, such as Successive Interference Cancellation (SIC) receivers. Compared to OMA communications, the main advantages of NOMA be summarized as follows [96]:

- Support overloaded transmission (*i.e.*, support a larger number of users with respect to available transmission resources);
- Achieve multi-user capacity;
- Enable reliable and low-latency grant-free transmission.
- Increased system's spectral efficiency.

5.3.2 NOMA techniques proposed in 3GPP

Several NOMA techniques have been proposed in 3GPP by different companies in order to meet the diversified requirements towards 5G. Until 3GPP TSG-RAN WG1 (RAN-1) #86, at least 15 candidate NOMA schemes have been proposed for 5G. On RAN-1 #86b, a general framework of NOMA schemes was agreed upon [97], which helps to categorize existing operations in NOMA schemes into bit-level operations and symbol-level operations. The former focuses on the design of NOMA algorithms related to channel coding and bit-level interleaving, while symbol-level NOMA mainly lays emphasis on symbol spreading and mapping. Bit-level NOMA schemes exploit low-rate FEC codes to enhance the detection accuracy, and/or take advantage of user-specific interleaving within the multiuser interference. The latter plays with symbols and mainly focuses on bit-to-symbol mapping. A large portion of symbol-level NOMA schemes utilizes short sequence-based spreading to enhance connectivity. These schemes can be further divided into two subcategories according to the densities of the spreading sequences. Some other symbol-level NOMA schemes make use of long sequence-based scrambling/spreading, where the receiver exploits the difference between these sequences. In addition to where the NOMA scheme is applied, *i.e.*, at the bit or symbol level, the difference among these schemes is mainly on UE's signature design, *i.e.*, whether the scrambling sequence, interleaver or spreading code is used to differentiate UEs [98]. Therefore, these schemes can be classified into the following three categories:

TABLE 5.6: NOMA schemes

	Scheme	Features	Receiver
Scrambling based	Resource Spread Multiple Access (RSMA)	Low rate channel code and scrambling codes with good correlation properties	SIC
	Low Code Rate and Signature-based Shared Access (LSSA)	Each UE's data is bit or symbol level multiplexed with UE-specific signature pattern	SIC
Interleaved based	Interleave Division Multiple Access (IDMA)	Bit-level interleavers	ESE
	Interleave Grid Multiple Access (IGMA)	Bit level interleavers or grid mapping pattern	ESE
Spreading based	Sparse Code Multiple Access (SCMA)	Multi-dimensional constellation and LDS	MPA
	Pattern Division Multiple Access (PDMA)	Code linked to the sparse mapping	MPA
	Low Density Sequence - Signature Vector Extension (LDS-SVE)	Transformation and concatenation of two element signature vectors into a larger signature vector	MPA
	Multiple User Shared Access (MUSA)	Random complex spreading codes with short length	SIC
	Non-Orthogonal Coded Access (NOCA)	Low correlation sequence	SIC
	Non-Orthogonal Coded Multiple Access (NCMA)	Spreading codes based on Grassmannian line packing problem	PIC
	Low Code Rate Spreading (LCRS)	Application of direct spreading of modulation symbols	SIC

- Scrambling-based NOMA schemes where the users are distinguished by different scrambling sequences. At the receiver side, a Successive Interference Cancellation (SIC) algorithm is applied to separate different UE's data.
- Interleaving-based NOMA schemes: different interleaves are used to distinguish different UEs, and a low code rate FEC can be jointly applied. As a receiver, an Elementary Signal Estimator (ESE), and a Maximum a Posteriori (MAP)/Message Passing Algorithm (MPA) can be employed.
- spreading-based NOMA schemes: different spreading codes are used to distinguish different UEs. The MPA, SIC, or Parallel Interference Cancellation (PIC) can be used depending on the density of the signature.

The general features of the proposed candidate schemes are given in Table 5.6.

NOMA schemes can be categorized based on the spreading density [99]:

- "Full-length spreading" (UEs transmit on full set of resources);
- "Sparse spreading" (Low Density Signature-based spreading)

Moreover, under each of the above categories, the schemes can be further divided based on the spreading schemes:

- Sequence-based non-orthogonal spreading;
- Low code rate spreading;
- Combinations of the two.

Regarding the spreading density, the schemes are categorized based on whether the UEs transmit on the entire set of available resources or using a sparse set of

resources. Thus, for each case, it is possible to observe that with the "full-length spreading", the UEs transmit on the full set of resources. Thus, the latter experience a similar Interference level, subject to channel selectivity in time/frequency. Moreover, both sequence-based spreading and low code rate spreading are possible. Regarding the "Sparse spreading", it is based on Low-Density Signature, therefore, the actual spreading factor for a particular user is smaller than the total resources the transmissions are spread over. It targets overloaded conditions, exploiting LDS-based spreading to realize different interfering signals on different sets of resources. Its spreading nature facilitates close-to-optimal Multi-User Detection (MUD) receivers based on MPA-based iterative receivers. However, the overall receiver complexity is still significantly high. In addition, both sequence-based spreading and low code rate spreading are possible. Further categorization is also possible based on the spreading approach: non-orthogonal (or quasi-orthogonal) spreading sequences or low code rate spreading. In the former case, symbols are spread using binary or M-ary sequences and they may be QAM symbols or directly mapped from information bits (*e.g.*, SCMA employing multi-dimensional constellation points). In the latter case, the "spreading" is realized by employing a low-rate code to encode the information bits and it aims at exploiting coding gains beyond repetitions. UE-specific scrambling (RSMA) or interleaving (IDMA) can be applied further. At the receiver side, MPA-based receiver structure is necessary to distinguish UEs based on UE-specific interleaver. Besides, it can be combined with signature-based non-orthogonal spreading schemes. In UL, NOMA schemes based on low code rate spreading can exploit the benefits of long/pseudorandom code structures, which make these schemes more robust to partial overlaps between transmissions. An important consideration is the support of partially overlapped transmissions, either in frequency or time domain. For instance, in the time domain, it would be beneficial to support UEs with a different number of time domain repetitions on partially overlapped resources. To this aim, the spreading should ensure low cross-correlation between sequences of different lengths.

5.3.3 Sparse Code Division Multiple Access for NB-IoT

As previously described, NB-IoT supports sporadic uplink data transmission with a low data rate performed by an enormous number of terminals. Thus, to support the congested traffic and massive connectivity, along with QoS requirements, NOMA has been considered an enabling technology [100].

In NTN scenarios, the mMTC requirements are exasperated by the low visibility window of the satellite, which further underlies the necessity of serving also the users that transmit on the same time/frequency resources. Therefore, NOMA is not only an emerging paradigm for terrestrial wireless networks but also for NTN.

Thus, the main objective of this section is to assess the performance of NOMA along the satellite channel. Among all the possible techniques previously cited, the Sparse Code Multiple Access (SCMA) has been selected and different reasons lead to this

choice. SCMA is a code domain NOMA technique, where the data of each user is mapped to codewords of SCMA codebooks, and a low-complexity iterative MPA algorithm is exploited at the receiver to decode the received data. Therefore, multiple symbols of the different devices are superposed non-orthogonally on the same sub-carrier achieving a higher overloading gain. Thanks to its sparsity nature, SCMA is robust to channel variations and reduces the interference on each RE. In this way, the receiver might be able to achieve the near Maximum Likelihood performance with lower complexity. Moreover, since the nonzero spread modulation symbols of the same data are different, the constellation of SCMA provides additional coding gain and higher spectral efficiency than the constellation that simply applies the repetition scheme on the nonzero symbols. Besides, it is a short spreading code technique as the constellation generated by the algorithm has the same size as the original constellation of the users; therefore, no delay is added with a further encoding of the message. Last but not least, SCMA can be applied to the uplink contention-based transmissions, such as the RA procedure thanks to the blind decoding capability of the MPA.

In the following the SCMA encoding and decoding procedures are briefly described [101].

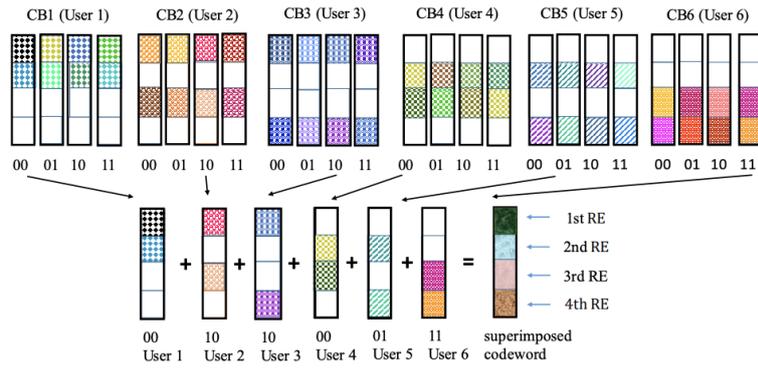
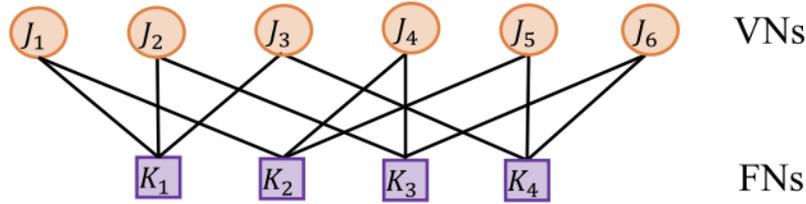
Encoding. Considering an uplink SCMA system, there are J users that transmit data to the gNB using K REs. The input bits of each user are mapped to a complex codeword using the SCMA encoder. User j intends to transmit b_j bits and the encoder will map these bits to a codeword m_j selected from a predefined codebook CB_j as follows:

$$m_j = CB_j(b_j) \quad (5.21)$$

where $m_j \in A_j \subset C^K$, with A_j denoting the set of codewords of the j -th user. The encoder has J layers and each layer has a specific CB dedicated to each user. Therefore, layer and user are interchangeable. The codebook of each user has its own sparsity pattern and can be written as a matrix of size $K \times M$, where M defines the number of codewords per user. In a CB, each column vector (*i.e.*, the codeword) is sparse and consists of d_v non-zero elements at a certain fixed RE pertinent to a specific user. For the j -th user, the codebooks can be expressed as:

$$CB_j = V_j \Delta_j A'_{MC} \quad (5.22)$$

with V_j denoting the binary mapping, A'_{MC} the multi-dimensional constellation, and Δ_j the constellation operator for the j -th user. The mapping matrix is selected in such a way that each user has an active transmission over a few fixed REs only. The assignment of REs among the users can be represented by a signature matrix $F = [f_1, f_2, \dots, f_J]$, where $f_j = \text{diag}(V_j V_j^T)$. $F_{jk} = 1$ implies that the j -th user has active

FIGURE 5.15: Example of 4×6 SCMA encoding scheme [101].FIGURE 5.16: Factor graph for 4×6 [102].

transmission over the k -th RE. An example of a signature matrix is the following:

$$\mathbf{F}_{4 \times 6} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \quad (5.23)$$

Thus, the binary mapping matrices corresponding to the first two users are:

$$\mathbf{V}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \mathbf{V}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad (5.24)$$

For example, V_2 indicates that user 2 transmits over the first and the third REs. Once the mapping matrices are performed, the next step is to design the mother A'_{MC} constellation and constellation operator, Δ_j .

An illustration of 4×6 encoding scheme is provided in Fig. 5.15.

Decoding. The receiver of SCMA is based on the MPA, which is an algorithm to conduct inference from graphical models by passing belief messages between the nodes. Before detailing the decoding procedure, the factor graph corresponding to 4×6 SCMA system is shown in Fig. 5.16. Looking at the matrix $F_{4 \times 6}$, it is possible to note the first column refers to the first user, and it has non-zero values in the first and second rows, meaning that its data is transmitted on the first and second RE.

Therefore, in the factor graph in Fig 5.16 there is an edge between Variable Node (VN) J_1 and the Function Node (FN) K_1 and K_2 . The MPA is applied over the factor graph shown in Fig 2 and messages are passed between the VNs and FNs in both directions of the SCMA factor graph. It is worth emphasizing that the factor graph of the SCMA system contains cycles in it. Thus, messages are passed between the nodes for some iterations until the termination criteria is achieved. The objective is to detect the symbols transmitted by each user with the following steps:

1. Initialization;
2. Passing of messages between FNs and VNs;
3. Termination and selection of the codeword.

In the following, each step is described.

Initialization. The main objective of this phase is to compute the Likelihood ratio (LR) of the transmitted codewords at each FN. Let us assume a FN_l where data of three users corresponding to set $\zeta_l = (v_1, v_2, v_3)$ are superimposed. Thus, the LR function at FN_l is $\psi_l(y_l|x_1, x_2, x_3, N_0)$. Here, x_1, x_2, x_3 denote the codewords transmitted by users belonging to set $zeta_l$ and y_l the transmitted signal. Assuming that the set of codeword elements assigned to user v on RE l is denoted as $CB_{l,v}$, then the likelihood function of FN_l is given by:

$$\psi_l(y_l|x_1, x_2, x_3, N_0) = e^{(-\frac{1}{N_0}||y_l - (h_{l,1}x_{1,1} + h_{l,2}x_{1,2} + h_{l,3}x_{1,3})||^2)} \quad (5.25)$$

where $x_{l,v}$ defines the codeword element transmitted by the v -th user on the l -th RE and $h_{l,v}$ denotes the channel coefficients for the v -th user. In total, KM^{df} values are stored for the likelihood function. Each codeword has an a priori probability equal to $\frac{1}{M}$. Therefore, the first message from VN to the l -th FN is this probability:

$$p_{v1 \rightarrow l}^{init}(x_1) = p_{v2 \rightarrow l}^{init}(x_2) = p_{v3 \rightarrow l}^{init}(x_3) = \frac{1}{M} \quad (5.26)$$

Messages' passing between FNs and VN. In order to send the message from the FN to one user, information received on the FN from the other two users may be regarded as extrinsic information. The message passed from the FN l to VN v_1 is given as:

$$p_{l \rightarrow v_1} = \sum_{x_2 \in CB_2} \sum_{x_3 \in CB_3} \psi_l(y_l|x_1, x_2, x_3, N_0) \cdot p_{v2 \rightarrow l}(x_2)p_{v3 \rightarrow l}(x_3), \quad x_1 \in CB_1 \quad (5.27)$$

In eq. 5.27, messages from two nodes, *i.e.*, $p_{v2 \rightarrow l}(x_2)p_{v3 \rightarrow l}(x_3)$ are multiplied by the local likelihood function of the l -th FN and then marginalized with respect to v_1 . In the same way, it is possible to define the messages from l -FN to v_2 and v_3 . The aim of this phase is to send the guess of what signal is received at FN l for all possible values of VN v .

Then, the message is sent in the opposite direction, *i.e.*, from VN to FN. Referring to

Fig. 5.16, l_1 and l_2 are the FNs connected to VN v . Thus, the message from VN v to FN l_1 and l_2 are:

$$\begin{aligned} p_{v \rightarrow l_1}(x_v) &= \frac{P_a x(v) p_{l_2 \rightarrow v}(x_v)}{\sum_{x_v} p_{l_2 \rightarrow v}(x_v)} \\ p_{v \rightarrow l_2}(x_v) &= \frac{P_a x(v) p_{l_1 \rightarrow v}(x_v)}{\sum_{x_v} p_{l_1 \rightarrow v}(x_v)} \end{aligned} \quad (5.28)$$

P_a is the a-priori probability for the user v and $p_{l_2 \rightarrow v}(x_v)$ refers to the updates VN v received from FN l_2 . The denominator normalizes the belief, so that the latter falls in the range $[0,1]$. Due to the fact that the factor graph has cycles, the passing of messages between FNs to VNs is repeated for a few iterations and it stops when a considerable change is observed in the belief computed at each VN.

Termination and selection of the codewords. The final belief is computed at each VN and it corresponds to the product of the a-priori probability and messages from the neighbouring FNs of each VN:

$$I_v = P_a(x_v) p_{l_1 \rightarrow v}(x_v) p_{l_2 \rightarrow v}(x_v) \quad (5.29)$$

Finally, the Log-Likelihood Ratio of I_v is computed.

5.3.4 Scenario

In order to evaluate the performance of SCMA over NTN, the simulated scenario includes a LEO satellite with a regenerative payload generating a single moving beam and a plethora of NB-IoT devices equipped with the GNSS module. The users can transmit their data only if are within the satellite beam. For the moment being, it is assumed that they have already performed the RA procedure, thus, only the data phase is considered. In particular, in order to apply SCMA, only the single-tone transmission with 15 kHz of sub-carrier spacing is foreseen with QPSK modulation.

Uplink transmission

In the uplink, the transmitted signal by the j -th UE to the gNB is the following:

$$s_j(t) = \sum_{k=0}^{K-1} X_{n,k} \Phi_k(t - nT) \quad (5.30)$$

$$\Phi_k(t) = \frac{1}{\sqrt{T}} e^{j2\pi\Delta f k t} \text{rect}\left(\frac{t - T/2}{T}\right) \quad (5.31)$$

where $X_{k,n}$ is the complex input data vector modulating the N subcarriers, where n refers to the n -th OFDM symbol. The sub-carriers depend on the codebook, and, consequently, $X_{k,n}$ on the specific codeword belonging to the chosen codebook.

Due to the sparse nature of SCMA codebooks, non-zero values from d_f out of J users overlap over each RE and also each user data is transmitted on $d_v < K$ REs.

The received signal at the k -th RE is (for the sake of clarity the index of the OFDM symbols is omitted):

$$y_k(t) = \sum_{j \in U_k} h_{k,j}(t) s_{k,j}(t) + n_k(t) \quad (5.32)$$

$h_{k,j}(t)$ is the channel between the j -th user and the satellite (computed as Eq. 5.15), and $n_k(t)$ is the noise which is Gaussian distributed.

In this work, the users are assumed to be stationary; however, the channel coefficients have a dependency on the time based on the position of the satellite, since there is a phase rotation that depends on the slant range.

5.3.4.1 Simulator design and evaluation

The block diagram of the simulator is shown in Fig. 5.17 and the main parameters are listed in Table 5.7. Six users transmit their data on four resource elements following the matrix $F_{4 \times 6}$. For all of the active devices, the classical steps of the NB-IoT transmission chain are computed until the generation of the QPSK symbols. Then, to each user is assigned a codebook, and, thus, each QPSK symbol is mapped to a specific codeword of the codebook. The latter step includes also the sub-carrier mapping since each codebook is identified by a specific pair of sub-carriers. Successively, the Inverse Discrete Fourier Transform (IDFT) is performed, and, after the insertion of the CP, the convolution between the signal and the channel coefficient is performed. For the simulation campaign, two different scenarios have been considered. In the first one, all the users experience the same channel, thus the E_b/N_0 is homogeneous over the K REs. Therefore, the performances obtained can be referred to a specific E_b/N_0 value. Conversely, in the second scenario, referred to as a heterogeneous scenario, the channel conditions are different in the REs, which is more realistic. Fig. 5.18a shows the BER vs E_b/N_0 curves obtained in the AWGN channel, considering a homogeneous scenario. At each Monte Carlo simulation, the channel coefficient varies. With ideal channel estimation, the SCMA decoder is able to provide a $BLER \leq 0.1$ for $E_b/N_0 \geq -2$ dB. As expected, the performance degrades when a Tapped Delay Line (TDL) channel is applied (shown in Fig. 5.18b). This channel foresees a LoS component characterized by a Rician K factor equal to 10 dB and one NLoS component with an attenuation of 20 dB and a delay spread of 5 ns. The NLoS component worsens the performance, in fact, only with an $E_b/N_0 \geq -0.5$ the $BER \leq 10^{-1}$.

Table 5.8 shows the performance when the channel coefficients of each user are approximated with the one at the beam center. Indeed, the goal is to verify the behavior of the SCMA with non-perfect channel estimation. In this case, the channel coefficients are different for each user and vary at each Monte Carlo. Although this approach is simple and computationally effective, the decoder is not able to properly decode the codewords (independently of the considered SNR). Finally, Table 5.9 reports the BER for the six users and each of them has a different channel coefficient, simulating a more realistic scenario. In this case, the channel coefficients

TABLE 5.7: Simulation parameters

Parameter	Values
Number of SCMA layers J	6
Number of REs K	4
Number of bits per symbol	2
Number of Non zero elements d_v	2
Monte Carlo iterations	20
Number of super imposed users d_f	3
Number of iterations in the RX N_{iter}	10

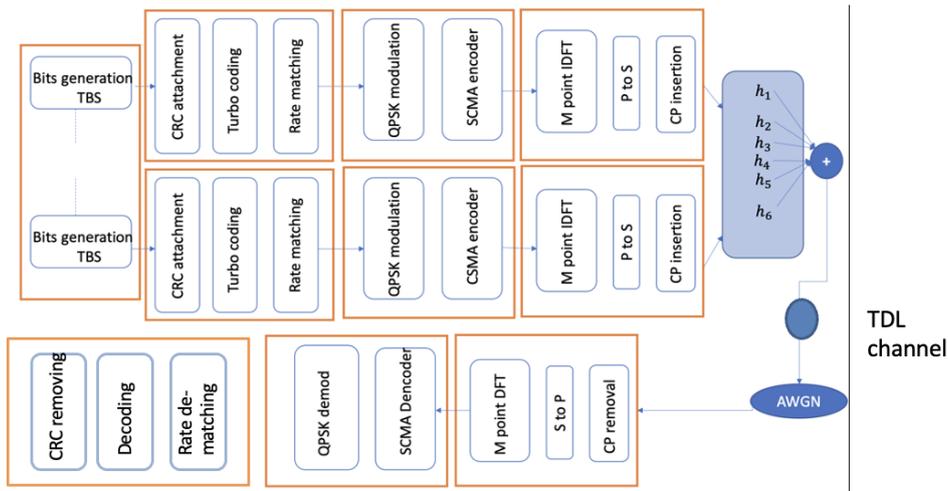


FIGURE 5.17: Block diagram of the simulator.

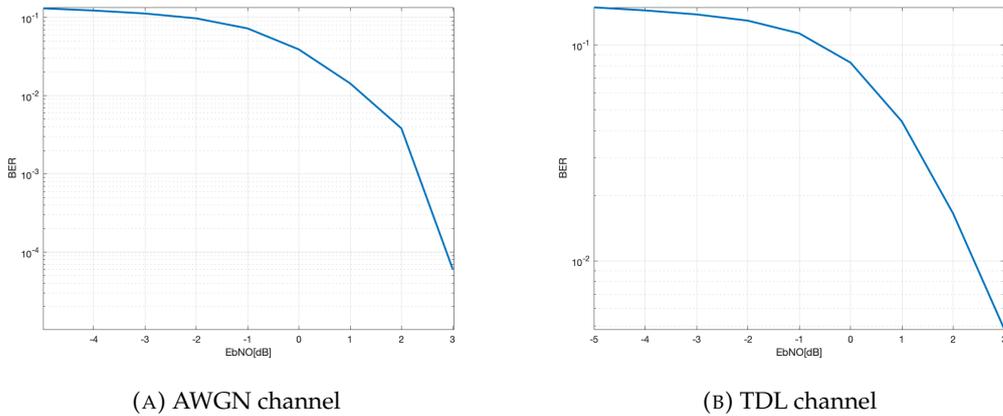


FIGURE 5.18: SCMA performance. Homogeneous scenario.

remain constant over the Monte Carlo Simulations. As it is possible to observe, due to the different channel conditions, only some of the users meet the condition of $BLER = 0.1$, ensuring reliable communications.

TABLE 5.8: Average BER of each user after the approximation of channel coefficients at the beam center

User	BLER
1	0.51
2	0.50
3	0.50
4	0.48
5	0.52
6	0.50

TABLE 5.9: Average BER of each user in TDL channel

User	BLER
1	0.1078
2	0.1991
3	0.0218
4	0.072
5	0.0648
6	0.1701

5.3.5 Discussion and future works

The work reported here is only the first step to evaluate the feasibility of SCMA in the NTN scenario. Indeed, NOMA is a powerful tool for mMTC scenario, especially in satcom-based networks where the satellite visibility window is limited and the users should be served in a crowded environment. The SCMA allows the decoding of all six users at the same time, even with the low channel coefficients typical of NTN. However, the proposed scheme requires an accurate estimation of the channel coefficients in order to properly decode the symbols. Perfect synchronization is considered among the users since only the data phase transmission is taken into account. Future works foresee the implementation of SCMA in a grant-free scenario, where collision among the codebooks could happen. Therefore, an appropriate channel estimation must be performed, considering the movement of the satellite. Furthermore, a proper design of the pilots needs to be carried out to track the phase shift. Moreover, since the detection of the codeword, and, thus, the codebook of each user can be seen as a classification problem, a NN approach will be investigated.

Chapter 6

System level analysis

This chapter evaluates the performance of the RA and data phase in terms of the percentage of users and time employed by these users to successfully achieve uplink synchronization and throughput of the system.

6.1 NB-IoT random access procedure via NTN: system level performances

This section reports the outcomes of the paper "NB-IoT random access procedure via NTN: system level performances" [103]. The aim of this work is to test specific scenarios defined by the 3GPP in [23] where multiple users access the network at the same time within the same beam footprint. Varying the number of users, the probability of collision changes, enabling us to obtain further results regarding the mean user access time as a function of the user density of the NTN NB-IoT network and the time visibility of the beam. Overall, the contributions of this study can be summarized as follows: we model the random access procedure of a single coverage enhancement level as a multi-channel slotted Aloha system and we characterize the collision probability under the back-off mechanism, the number of sub-carriers reserved for the preamble transmission, and the RA periodicity, based on the satellite beam visibility window and the number of users. Performance analysis of the above-mentioned system is conducted for 3GPP-defined values of the number of repetitions, periodicity, and back-off window size, which are the so-called configurable access parameters. The result of this study is a comparison of the RA success probability and the users' access time for different combinations of the configurable access parameters for a given user density and satellite configurations defined by the 3GPP.

6.1.1 System architecture

Referring to Fig. 6.1, the system architecture consists of the following elements: *i)* the terrestrial segment composed of the terrestrial network infrastructure, which is connected to the NTN segment through the GW. *ii)* A regenerative LEO satellite payload, providing connectivity to the UEs through the user link and generating a single moving beam. *iii)* User segment, represented by a plethora of NB-IoT devices,

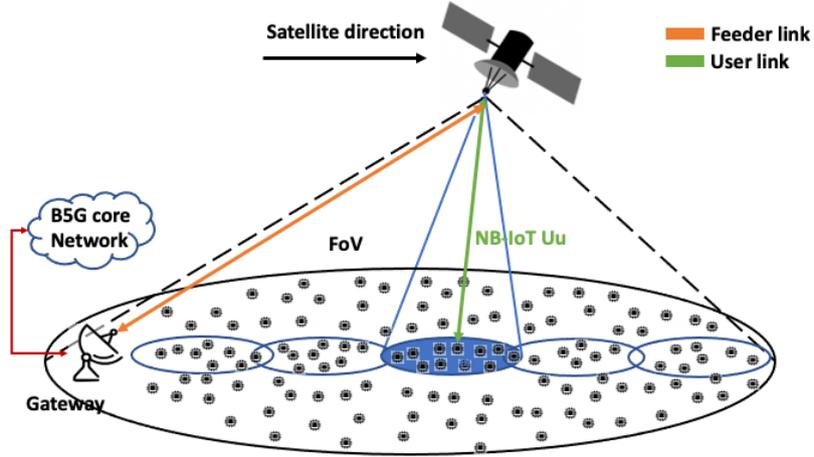


FIGURE 6.1: System architecture for the RA analysis [103].

equipped with the GNSS, spread all over the world. It is worth noting that the devices do not have visibility of the satellites for the entire time. In particular, depending on the satellite's position along its orbit, for a given on-ground coverage area, only a subset of NB-IoT devices is able to communicate with the flying platform.

6.1.2 Analysis of the access parameters

As previously described in Chapter 2, the most critical part of the RA is represented by the preamble transmission, since the devices are not yet synchronized with the network and compete for the same resources. Potentially, every UE can periodically send its preamble in the RAO configured by the network. Within a given interval, T_{max} , the number of RAO is defined as:

$$N_{RAO} = \left\lfloor \frac{T_{max}}{T_{RAO}} \right\rfloor \quad (6.1)$$

where T_{RAO} defines the RAO periodicity. As per 3GPP specification, different values of T_{RAO} are defined to ensure that the device is aware of the procedure's success. Therefore, $T_{RAO} > T_{msg1} + T_{RAR} + T_{CRT}$. Please note that here T_{RAO} is intended as the periodicity of the RA. Therefore, in each RAO, part of the resources are dedicated to the preamble transmission and those remaining to the data on the NPUSCH until the next RAO.

Considering the preamble format 1, in each RAO, the minimum preamble time frame could be repeated to increase the detection probability. Indicated as N_{rep} the number of preamble repetitions, its time duration is computed as follows: $T_{msg1} = 6.4 \cdot N_{rep}$ ms, where 6.4 ms is the length of four SGs, as described in Chapter 2 in Table 2.2. Regarding T_{RAR} and T_{CRT} , they refer to the length of RAR window and the contention resolution timer. Moreover, as mentioned above, the satellite RTD is much higher than the one typical of terrestrial communication systems. However,

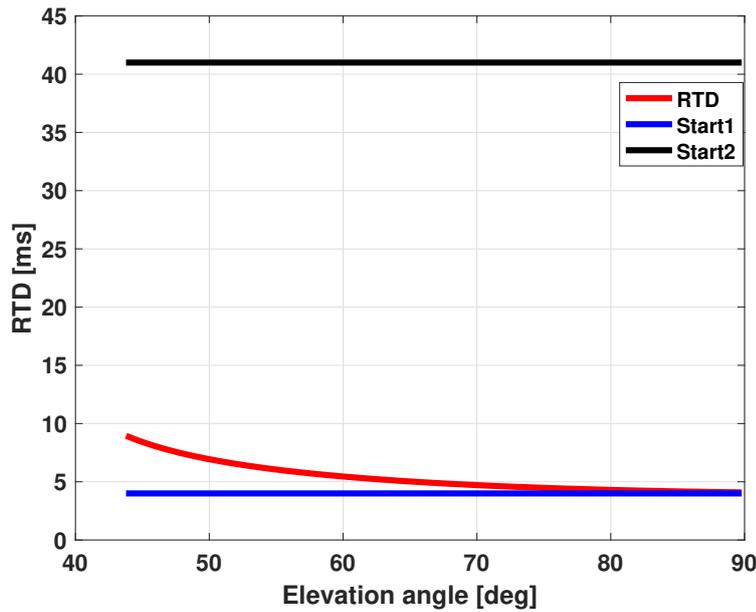


FIGURE 6.2: RTD of LEO satellite varying the beam center elevation angle. Set 3 [103].

since T_{RAR} and T_{CRT} are procedure timers impacted by the single UE RTD, no pre-compensation can be implemented exploiting the GNSS capabilities. As previously described, the NB-IoT standard specifies two starting times for the RAR window, *i.e.*, 4 ms and 41 ms after the last sub-frame containing the preamble [28]. Therefore, it is essential that these two values are long enough to absorb the RTD. Fig 6.2 shows the RTD of a LEO satellite (Set 3) with regenerative payload. It is clear that the first starting time is not long enough to account for the increased delay of the NTN channel. Regarding the CRT, it starts as soon as Msg3 is sent, hence it is necessary to postpone it by a quantity equal to the RTD. Thus, if no modifications of the RA procedure are foreseen, T_{RAO} needs to satisfy the following condition:

$$T_{RAO} > T_{msg1} + T_{RAR} + T_{CRT} + 41 + RTD \quad (6.2)$$

When one of the two timers expires and the device does not receive Msg2 or Msg4, it re-attempts the procedure. In particular, the reasons for a RA failure are: *i)* worst channel condition: the BS is not able to detect the incoming preamble, because the signal power accumulated along the repetitions is not sufficient to exceed the receiver sensibility. Therefore, the BS does not send the RAR to UE, which, after waiting for the expiration of the RAR window, extracts a BO value to be applied before the re-transmission. *ii)* Collision: if two or more devices choose the same preamble, they will receive the same RAR, and, therefore, will send the Msg 3 using the same resources. However, among the colliding users, only one UE will receive the contention resolution message (*i.e.*, Msg4), while the other will restart the procedure after a BO time. The back-off mechanism represents a strategy to reduce the con-

gestion in the RAO, by distributing the number of contending devices over time. Indeed, the UEs which collide in an RAO will randomly choose a back-off value in the interval $[0, BO]$. Thus, the probability that a collided NB-IoT terminal in the $n - th$ RAO will transmit in the $k - th$ one is computed as follow:

$$P_{BO} = \begin{cases} \frac{T_{RAO}}{BO} & \text{if } n + \lfloor \frac{BO}{T_{RAO}} \rfloor \leq k \\ 0 & \text{if } n + \lfloor \frac{BO}{T_{RAO}} \rfloor > k \end{cases} \quad (6.3)$$

The BO value has a great influence on the performance of the RA. Indeed, on the one hand, a small BO allows devices to re-transmit the preamble after a short period of time, increasing the collision probability during burst arrivals. On the other hand, a large BO value may increase the probability of successful access causing high access time. Therefore, it is important to find the optimal BO value which increases the probability of concluding the RA procedure with a reasonable access time. Furthermore, another factor that impacts the performance of the RA is the number of sub-carriers dedicated to the preamble transmission. Indeed, based on the number of coverage enhancement zones and the traffic, the number of sub-carriers could be distributed among these zones with the aim to provide more preambles to the zone with the highest number of users. Defined S as the total available sub-channels and J as the users attempting the access in the same RAO, the probability that each device randomly selects one of the available sub-carrier is $P_s = 1/S$. Therefore, the probability that j users select the same preamble is given by:

$$P_{n,s} = \binom{J}{j} \left(\frac{1}{S}\right)^j \left(\frac{S-1}{S}\right)^{(J-j)} \quad (6.4)$$

It should be noted that the transmission is considered successful if a preamble is chosen only by one user, thus, this probability is equal to:

$$P_{1,s} = \frac{J}{S} e^{-\frac{J}{S}} \quad (6.5)$$

Therefore, the number of UEs that re-attempt the procedure is:

$$J_{coll} = J(1 - e^{-\frac{J}{S}}) \quad (6.6)$$

Then, each of these J_{coll} devices randomly selects one of the available sub-carriers in the $k - th$ BO interval with probability $P_s \cdot P_{BO}$. Therefore, the probability mass function (pmf) obtained by distributing the J_{coll} UEs over the $(S \cdot \frac{BO}{T_{RAO}})$ sub-channels in the $k - th$ BO interval is:

$$P_{j,s,k} = \binom{J_{coll}}{j} (P_s \cdot P_{BO})^j (1 - P_s \cdot P_{BO})^{(J_{coll}-j)} \quad (6.7)$$

where $k \in [1, N_{RAO}]$. Therefore, the collision probability in the k -th RAO is given by:

$$P_{coll,s,k} = 1 - e^{-(J_{coll} \cdot P_s \cdot P_{BO})} \quad (6.8)$$

6.1.3 Simulation Model

In the proposed model, an LEO satellite covers a fixed area defined by FoV, where the stationary NB-UEs, equipped by an omnidirectional antenna, follow a uniform distribution. In order to evaluate the access success rate in the considered area, the satellite motion over the simulation duration, according to the Keplerian orbit parameters of the satellite, is computed. Starting from a reference epoch and given the Keplerian orbital elements, the state vectors, *i.e.*, position, and velocity, for the satellite are computed at each simulation step. Knowing the successive spacecraft positions at each simulation step, *i.e.*, ephemerides, thus its trajectory, it is possible to compute the visibility period of the satellite with respect to the on-ground UE. Thus, based on the satellite - UE geometry, the uplink SNR is computed by means of eq. 4.13. The EUs are able to communicate with the satellite, only if $\Omega(\theta_u^s) > \Omega(\theta_{edge}^s)$, where $\Omega(\theta_{edge}^s)$ is the value of the normalized radiation pattern at the beam edge. Therefore, all the users, whose radiation pattern is bigger than that at the beam edge, have the possibility to perform the RA, by generating their own traffic. This means that the traffic is modeled as pick traffic, simulating a daily uplink report. This choice is due to the fact that, under a massive access scenario, it imposes more challenges on the network side with respect to other arrival process models, such as the beta distribution [104]. Finally, the overall interference I_s^{uplink} in the uplink for the preamble S is

$$I_s^{uplink} = 10 \log_{10} \left(\sum_{j=0}^{J_{coll}-1} 10^{(0.1 \cdot I_{s,j})} \right) \quad (6.9)$$

with $j = 0, \dots, J_{coll} - 1$, being J_{coll} the number of interfering UEs with the same preamble and the SINR is computed by means of Eq. 4.16. In order to solve the contention due to the multiple collisions, SINR of each UE is exploited to estimate the BLER of the received signal, using SNR-BLER curves of the NPUSCH. Indeed, the proposed model is based on the SNR level of the Msg3, since, thanks to this message, the contentions are solved. To this end, the BLER is estimated by considering the chosen MCS, number of Resource Unit, number of NPUSCH repetitions, and the SINR experienced at the satellite during the reception. In case more than one UE has a SINR that exceeds the one providing a BLER of 10^{-1} , the highest one is selected.

6.1.4 Performance evaluation

In this section, the proposed model is evaluated through a system-level study, which highlights how the network parameters and the NTN configurations significantly

TABLE 6.1: Satellite parameters [23]

Satellite orbit	Set 2 LEO 600 km	Set 3 LEO 600 km	Set 4 LEO 600 km
Equivalent antenna aperture	1 m	0.4 m	0.097 m
EIRP density	28 dBW/MHz	28.3 dBW/MHz	21.45 dBW/MHz
Tx/Rx G_{max}	24 dBi	16.2 dBi	11 dBi
ϑ_{3dB}	8.8320 degrees	22.1 degree	104.7 degree
Beam diameter	90 km	234 km	1700 km
G/T	-4.9 dB K^{-1}	-12.8 dB K^{-1}	-18.6 dB K^{-1}
T_{max}	13.4 s	33.86 s	246.9 s

TABLE 6.2: Access configuration parameters [105]

Parameters	Value
N_{rep}	[1,2,4,8,16,32,64,128]
T_{RAO}	[40,80,160,240,320,640,1280,2560]ms
BO	[0, 256 · 2 ^j]ms, $j \in [0, 11]$
BO index	$B \in [1, 13]$

impact the system's performance. The numerical evaluation is provided in terms of the percentage of users who successfully complete the procedure, access time, number of attempts taken by both users who complete and those who do not complete the RA, and remaining beam visibility time for data transmission. To this aim, it is worthwhile highlighting that only one coverage enhancement zone is considered and all the preambles are available for the users, *i.e.*, all the 48 sub-carriers. In the proposed model, T_{max} represents the beam visibility window, which, in turn, defines the upper bound to the number of available RAO and to the number of re-transmissions that a UE can perform to finalize the procedure. Notably, the wider the beam, the greater the number of users served, and the greater the interference. Two user density values have been tested, *i.e.*, $\alpha = 0.1$ [users/km²] and $\alpha = 1$ [users/km²]. It should be noted that these are reasonable values in rural environments; besides only one bearer is considered (*i.e.*, only channel of 180 kHz). Once fixed the value of T_{RAO} for the different NTN configurations, the BO value is chosen based on the collision probability previously computed. The satellite parameters are reported in Table 6.1 and the access parameters in Table 6.2.

TABLE 6.3: Access parameters for satellite configurations. $s = \text{success}$, $f = \text{failure}$, $W = \text{remaining time of beam visibility window}$

Set	Density	Access parameters	Mean access delay	% of completed users	Mean N° of transmissions	Mean W
Set2	0.1 [users/km ²]	$RAO = 160ms, B = 4,$ $N_{rep} = 2$	0.15 [s]	99.56 %, 85.85% (first attempt)	2 (s) - 3 (f)	1.67[s]
		$RAO = 320ms, B = 5,$ $N_{rep} = 8$	0.45 [s]	99.01 %, 71.91% (first attempt)	2(s) - 3 (f)	0.67[s]
		$RAO = 640ms, B = 6,$ $N_{rep} = 16$	1.95 [s]	92.44 %, 34.77% (first attempt)	3(s) - 3(f)	0.4[s]
Set3	1 [users/km ²]	$RAO = 320ms, B = 10,$ $N_{rep} = 16$	2.12 [s]	22.44 %, 14.7% (first attempt)	2(s) - 1(f)	1.18[s]
		$RAO = 640ms, B = 11,$ $N_{rep} = 32$	2.1 [s]	11.25 %, 7.71% (first attempt)	2(s) - 1(f)	1.29[s]
		$RAO = 1280ms, B = 12,$ $N_{rep} = 64$	3.86 [s]	21.38 %, 16.53% (first attempt)	2(s) - 1(f)	2.5[s]
Set4	0.1 [users/km ²]	$RAO = 320ms, B = 11,$ $N_{rep} = 16$	11.49 [s]	8.57%, 2.08% (first attempt)	2(s) - 2 (f)	6.67[s]
		$RAO = 640ms, B = 12,$ $N_{rep} = 32$	10.55 [s]	4.29%, 1.30% (first attempt)	2(s) - 1(f)	6.38[s]
		$RAO = 320ms, B = 12,$ $N_{rep} = 16$	103.38 [s]	9.01%, 0.07% (first attempt)	2(s) - 2 (f)	83.21[s]
		$RAO = 640ms, B = 13,$ $N_{rep} = 32$	100.66 [s]	4.55%, 0.09% (first attempt)	2(s) - 2(f)	80[s]

6.1.4.1 Parameters configuration

All the figures reported in this Section refer to Set 3 with $\alpha = 0.1[\text{users}/\text{km}^2]$. Different configurations, defined by the combination of the RA periodicity, number of repetitions, and the BO index are tested. As shown in Fig. 6.3a, only with a periodicity of 160 ms, more than 90% of the UEs finalize the RA. However, for the same RAO and number of repetitions (*i.e.*, 4), a discriminating factor is represented by the BO value in terms of the percentage of users who complete the procedure. Despite an increased BO value spreads the users over more RAOs, thus reducing the collision in the same opportunity, it increases the probability that a user is not served, being outside the beam visibility window. Indeed, not all users have the same number of RAO, as they have different elevation angles, and thus, different satellite visibility periods. This behavior is reflected in the percentage of devices achieving the connection at their first attempt and in the percentage of users unable to conclude the procedure. Indeed, when the periodicity is 160 ms and the BO index is 8, 58% of the completing UEs have granted the network in their earliest RAO; while, in the case of a smaller BO value, this is reduced to 55%. On the contrary, with a lower BO index, the percentage of devices that are out of the system decreases. For the other two tested values of RAO, it is possible to observe the same behavior. Moreover, the performances worsen as the periodicity increases. This is motivated by the fact that in the given visibility window, higher values of T_{RAO} reduce the number of RAOs. Therefore, being each RAO clogged, even with the use of the BO, the users are spread on few occasions, and, thus, collisions occur. The number of repetitions used for RAO equal to 320 ms and 640 ms is 8 and 16, respectively. Fig. 6.4a reports the CDF of the time taken by the devices to conclude the procedure. At first glance, it seems that the best performance is obtained with an opportunity of 640 ms and a BO index equal to 12, as 87% of the UEs have time to conclude the procedure in the order of ms. However, a good trade-off, in terms of success access rate and access delay, is obtained with a periodicity of 160 ms and a BO index equal to 7. Indeed, its CDF in Fig. 6.4a has the steepest curve, allowing to 92% of the devices to have an access time smaller than 7 s. As a consequence, these users have more time for data transmission. This trend is shown in Fig. 6.4b, where the curves on the right perform better in terms of the remaining time of beam visibility for the data transmission. Fig. 6.3b shows the traffic overload, *i.e.*, the total number of users trying to access the network, including the users re-transmitting after collisions. The worst performance is obtained with a RAO equal to 160 ms and a BO index to 7, which is confirmed also from the CDF plots in 6.5a and in 6.5b. This is due to the fact that in this configuration, most of the users can perform the RA within the satellite visibility. While with other configurations, some of them have the right to transmit their preamble only after they are not anymore in the satellite visibility, due to the long BO values. However, performing an increased number of attempts translates into higher energy consumption of the device battery.

6.1.4.2 Impact of the user density

Referring to Table 6.3, as expected, a greater value of α leads to a higher number of collisions. Indeed, since more devices try to send a preamble in the same RAO, the collision probability increases accordingly. As a consequence, the percentage of users that finalizes the procedure decreases with an increased traffic load. This trend is visible in all the satellite configurations. High density requires high BO values to reduce the collision probability since the colliding users are distributed among different RAOs. However, as previously described, high BO increases the probability that a UE randomly extracts a BO index that prevents him to re-attempt the RA as he is no anymore in the beam. Moreover, high-density values imply the adoption of high repetition to increase the detection probability as explained in the following Section.

6.1.4.3 Impact of the number of repetitions

This parameter needs to guarantee a preamble detection probability at least of 99% (P_{d99}) at the BS [28]. In case of a smaller repetition value than the one providing P_{d99} is adopted, then the miss-detection probability increases, worsening the overall performance of the system in terms of the number of re-transmissions, and therefore of the device's battery consumption. On the contrary, a better performance in terms of detection rate is not guaranteed with a higher number of repetitions and the access time increases due to an increased value of T_{msg1} .

6.1.4.4 Impact of the beam visibility window

The satellite footprint diameter, and therefore the beam visibility window, has a great impact on the RA. As can be deduced from Table 6.3, a small radius reduces the number of terminals accessing the beam, thus reducing the NB-IoT preamble collisions and providing a lower mean access time with respect to the other NTN configurations. Also in terms of SNR it performs better, hence, a lower number of preamble repetitions is necessary to achieve P_{d99} . However, in addition to the low quantity of served users (34000 with $\alpha = 0.1$), the drawback of this configuration is the short satellite visibility window, which in turn, reduces the time for the data sending. In the case of a wider beam, such as the one obtained in Set 3, a greater number of users can be served (89000 with $\alpha = 0.1$). In order to achieve a success rate higher than 90%, greater BO values and a number of repetitions must be taken into account w.r.t the ones considered with a beam of 90 km, at the expense of the access time. The worst performance is obtained with a beam diameter of 1700 km, *i.e.*, Set 4. As expected, the number of devices entering the beam is so high (839000 with $\alpha = 0.1$) that, even with the greatest BO value, the number of collisions is not reduced. In that case, it is necessary to implement other techniques to avoid such an excessive number of terminals transmitting at the same time, *e.g.*, the access barring.

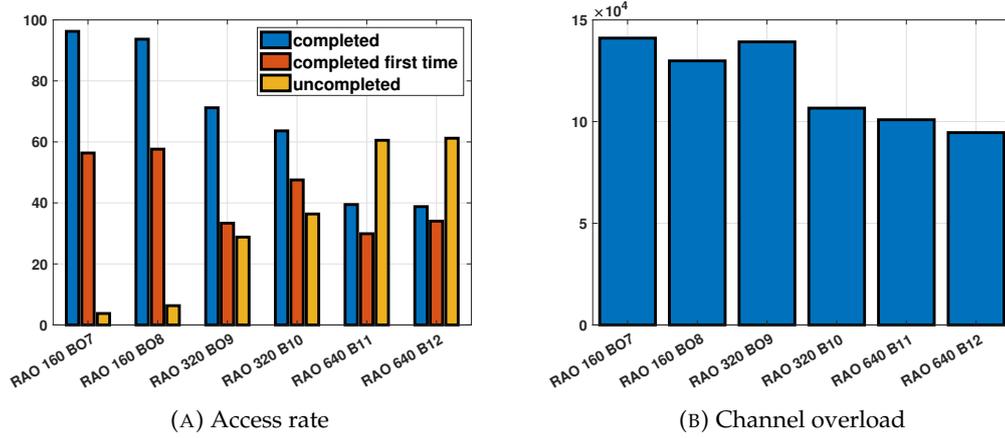


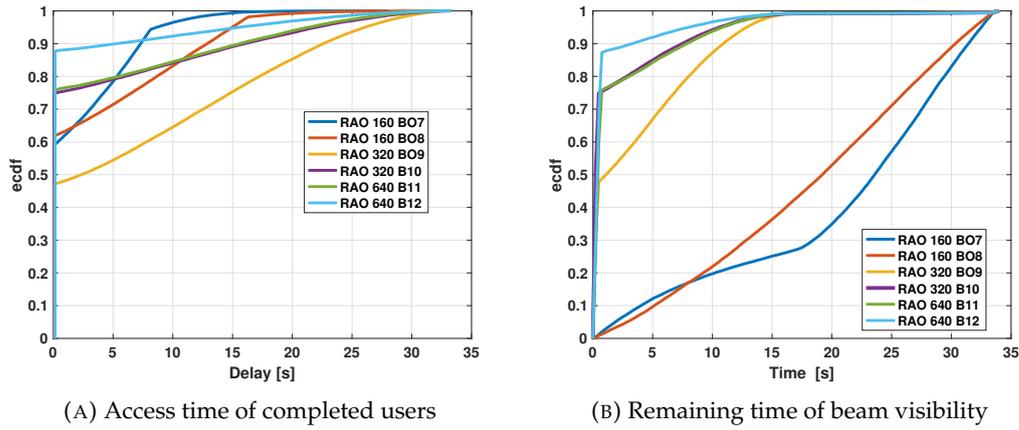
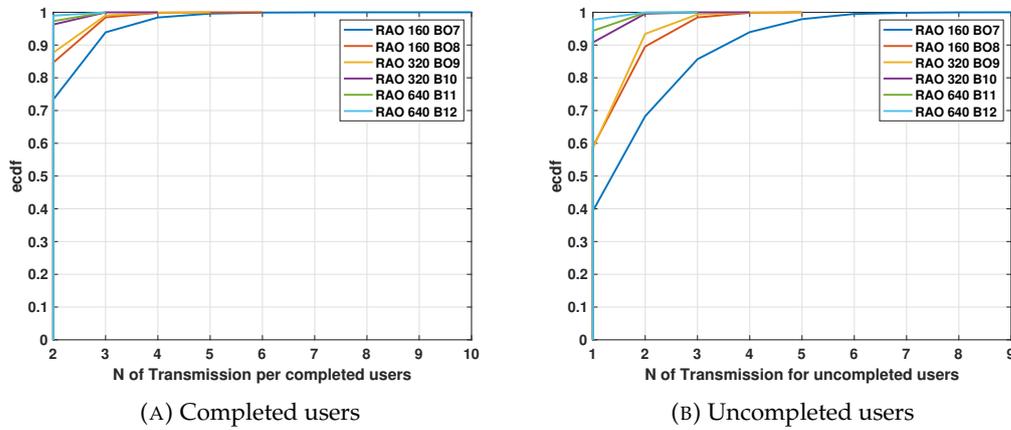
FIGURE 6.3: Performances of the random access procedure. Set 3, $\alpha = 0.1$ [103].

6.1.5 Discussion

In this work, we decided not to modify the NB-IoT standard and, therefore, the RAR start time must be 41 ms, as the former is not as long as the RTD of the beam. Therefore, it is not possible to use short periodicity of 40 ms and 80 ms. In contrast, when a long periodicity is taken into account, such as 1280ms, the number of RA occasions is drastically reduced. Thus, it is possible to observe a significant increase in the number of collisions, as more users compete for the access at the same period, requiring a high number of preamble repetitions and a much longer time to gain random access. The latter has a negative effect both on the battery consumption, due to the high re-transmission factor, and on the remaining time of the beam visibility window for the data transmission. The utilization of medium-length periodicities, *i.e.*, 160ms and 320ms, is recommended, especially if the beam diameters in Set 2 and Set 3 are taken into account. Indeed, these RAOs provide a good balance between the resources dedicated to the preambles and the data with an acceptable number of collisions and a high number of correctly detected signals. In addition, a low number of preamble repetitions (2 and 4) can be sent. As longer periodicities are taken into account, a higher number of repetitions shows better performance. Finally, the simulation results demonstrate that an increase by a factor of 10 of the transmitting users significantly affects the NB-IoT NTN system performance, even adopting different preamble repetition, RAO, and BO values. Hence the number of users is considered a driving factor for choosing the RA parameters and implies the design of a satellite constellation able to serve all the terminals.

6.2 A Joint Access and Data Phases Design for NB-IoT via Non-Terrestrial Networks

This section reports the results of the work pursued during the author's research period abroad that will be collected in the paper "A Joint Access and Data Phases

FIGURE 6.4: Time analysis. Set 3, $\alpha = 0.1$ [103]FIGURE 6.5: Number of transmissions. Set 3, $\alpha = 0.1$ [103].

Design for NB-IoT via Non-Terrestrial Networks" [106].

As seen so far, both in the previous section and from the state of the art, the performance of Random Access greatly affects the data transmission phase. However, so far not enough attention was given to the optimization of both the access and data phases under the limited visibility time of the satellite and high congestion environment. In particular, considering the IoT scenario, non-continuous coverage is the baseline approach, as this type of terminal only provides sporadic uplink data reports. In this framework, all the UEs within the coverage area are competing to access the channel and they need to be served within a relatively short period (2-4 min). This results in a critical and problematic situation, as the number of IoT devices within the beam footprint is very large due to the wide coverage of the satellite. The congestion generated during the signaling phase of message flows of the protocol can lead to the deterioration of the system's performance. At first glance, this problem could be alleviated in two ways:

- By increasing the resources allocated to the NPRACH. However, due to limited uplink resources, the higher the number of resources available in the NPRACH,

the lower the ones available in the NPUSCH for the data sending. Thus, many IoT devices, which have successfully completed their RA attempt, could not find enough transmission resources for the data.

- Or by scheduling the terminals that achieve uplink synchronization in the current RAO, in the NPUSCH resources allocated in subsequent opportunities. On the other hand, the dynamic nature of the NTN system makes this strategy impractical, due to the aging of users' channel coefficients.

Notably, one of the key design parameters for the system performance is the periodicity of the random access, where the uplink resources are shared between the PRACH and NPUSCH. The definition of this parameter is crucial when the time to communicate with the network is limited, such as in the satellite domain. Moreover, the number and the length of RAOs, being limited by the satellite visibility window, influence the choice of other parameters, *e.g.*, the number of uplink repetitions, the BO value, and the scheduling strategy of the resources for the data transmission. Besides, not all the users in the same beam have the same number of RAO, due to the different satellite visibility windows.

In terms of congestion management techniques for the RA, only algorithms naturally designed for terrestrial networks, such as the BO and the Access Barring (AB), have been considered. However, they are not suitable for the integrated IoT-NTN system.

Therefore, the problem is twofold: i) the access management, *i.e.*, how many users can perform the RA procedure at the same time, and ii) the dimensioning of the uplink resources to be shared between PRACH and PUSCH channels based on the channel conditions and the number of active users. Thus, motivated by all these reasons, in this work, a novel and practical algorithm aimed at increasing the spectral efficiency of the NB-IoT NTN system by jointly considering the access and data phase under the peculiarities of the satellite channel is designed. Given the user density, the user demand in the congested zone, and the satellite parameters (*i.e.*, the antenna characteristics and the altitude), a heuristic algorithm that optimally combines the number of satellite passages, the length and the number of the RAO, and the number of uplink repetitions is designed with the aim to maximize the capacity of the system. The algorithm does not foresee scheduling strategies for the data phase. The idea behind the heuristic approach is to reduce the complexity of the system in order to make it applicable in practical situations, such as the design of a satellite constellation to provide IoT services. Moreover, we design a practical user-barring algorithm to alleviate the congestion by redistributing the access requests among the satellite passages and we compute the RA success probability for the IoT devices considering the satellite passages, the number of uplink repetitions, and the probability of being in LoS.

As in [107], the proposed IoT over Satellite framework is based on the stochastic geometry tools, but we consider each element of the link budget that depends on the

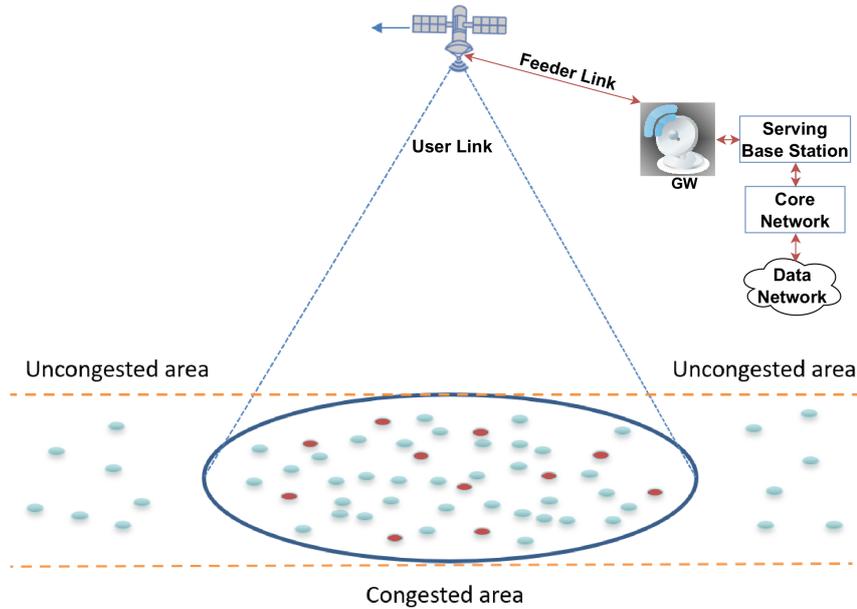


FIGURE 6.6: System architecture [106].

user position.

6.2.1 System architecture & assumptions

Referring to Fig. 6.6, the main elements of the high level system architecture are [39]: *i*) the terrestrial segment composed of the terrestrial network infrastructure, which is connected to the NTN segment through the GW. *ii*) A LEO satellite with a regenerative payload operating in S-band with a moving beam. *iii*) User segment, represented by a plethora of NB-IoT directly connected to the satellite and equipped with the GNSS.

6.2.1.1 NTN peculiarities that compromise the NB-IoT performance

The current NB-IoT protocol is designed in a way that it is able to cope with the terrestrial channel components. When a satellite-based NB-IoT system is considered novel impairments are introduced, or the already existing ones are emphasized. The ones of interest for the purpose of this work are the increased round trip delay and the limited visibility time. Their impact on the NB-IoT performance will be described.

Increased round trip delay

The uplink transmission in the NB-IoT system can be characterized by two main phases, *i*) the access phase where the users will be introduced into the network and achieve uplink synchronization; and *ii*) the data phase where every individual user will send its uplink data in the available resources (assigned by the serving base station). An illustration is given in Figure 6.7. Considering that the access phase of

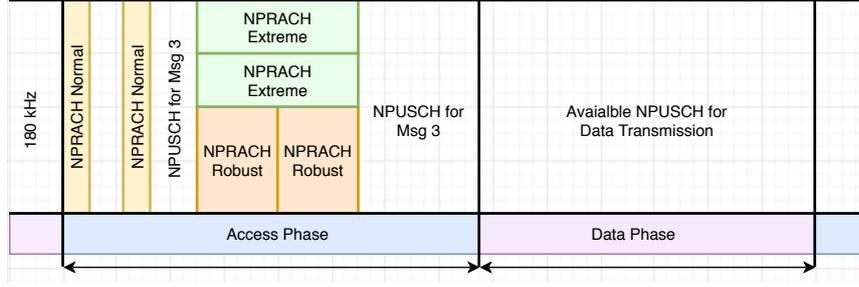


FIGURE 6.7: Access and data phases in NB-IoT [106]

NB-IoT involves a 4-message exchange between the users and the base station, the access phase length T_{access} can be calculated as follows:

$$T_{access}(N_{rep}) = T_{msg1}(N_{rep}) + 2 \cdot RTD + 2 \cdot K^{DL} + K^{UL} + T_{msg2} + T_{msg3}(N_{rep}) + T_{msg4} \quad (6.10)$$

where RTD is the round trip delay (depending on the satellite orbit), T_{msg2} and T_{msg4} are the length of msg2 and msg4 (related to downlink). The equation includes also the time taken by the UE to switch from the reception to the transmission and vice-versa, *i.e.*, K^{DL} and K^{UL} , as for 3GPP specifications. $T_{msg3}(N_{rep})$ represents the length of Msg3 as a function of the number of repetitions. As it can be noted, the increased RTD in the signal propagation would result in a longer access phase. On the other hand, the available resources for data transmission will be smaller for a given RAO length. Obviously, this will impact the performance of the overall NB-IoT system.

Limited visibility time The limited visibility time provided by the satellite has a two-fold impact. First, congestion of the network will occur due to the more limited occasions that the users have to access the network compared to terrestrial infrastructures. Notably, this is a well-known problem also for a terrestrial NB-IoT system, and the existing solutions mainly rely on the access barring principle. The main idea is to distribute the users in several groups so that access to the network is done at different times and some of the congestion is removed. Second, the limited visibility time directly impacts the available resources for data transmission. Hypothetically, even though all the users under the coverage of the satellite can have a successful access phase, this does not guarantee that they will be able to transmit their data. In such a case, they will be forced to repeat the access phase again because once the satellite is not available anymore, access to the network is lost. This is very particular if we compare it to a terrestrial infrastructure where the serving base station would be always available. A balance between the two phases is required in order to maximize the system's performance.

6.2.2 Problem formulation

In order to maximize the system performance of a system it is necessary to maximize its spectral efficiency. In particular, in the case of a satellite-based NB-IoT system, a joint design of the access and data phase is required, taking into account the peculiarities introduced by the NTN channel. The three dimensions that drive the performance are: *i*) the number of uplink repetitions N_{Rep} in the access phase; *ii*) the random access opportunity length T_{RAO} ; and *iii*) the number of satellite passages M over the area of interest.

6.2.2.1 Mathematical Formulation

In the following, we provide the mathematical formulation of the *spectral efficiency* η_s as a function of N_{rep} , T_{RAO} , and M . Given the number of users willing to access the network U , the average capacity per modulation and coding scheme, *i.e.*, $\overline{C_{MCS}}$ [bits/RU], and the average demand per user in the area \overline{D} [bits], the spectral efficiency is:

$$\eta_s = \begin{cases} \frac{\min\left(\frac{U}{M}, N_{sc}^{RA}\right) \cdot P_s(N_{rep}, M) \cdot \overline{D}}{B \cdot T_{RAO}} & \text{if } S_{req}(M, N_{rep}) < S_{av}(T_{RAO}, N_{rep}) \\ \frac{S_{av}(T_{RAO}, N_{rep}) \cdot \overline{C_{MCS}}}{B \cdot T_{RAO}} & \text{o/w} \end{cases} \quad (6.11)$$

It is worth highlighting that $\overline{C_{MCS}}$ depends on the MCS which in turn is a function of the average SNR of the beam, which can be calculated knowing the satellite and the UE antenna parameters. Eq. 6.11 shows that η_s depends on whether the requested resources S_{req} for data transmission are greater or smaller than the available ones, S_{av} . Where S_{req}

$$S_{req} = \frac{\overline{D}}{C_{MCS}} \cdot \min\left(\frac{U}{M}, N_{sc}^{RA}\right) \cdot P_s(N_{rep}, M) \quad (6.12)$$

and S_{av} is:

$$S_{av} = N_{sc} \left[\frac{T_{RAO} - T_{access}(N_{rep})}{T_{RU}} \right] \quad (6.13)$$

where N_{sc} refers to the number of sub-carriers dedicated to the transmission. Please note that N_{sc} and N_{sc}^{RA} are different, as the sub-carrier spacing in the PRACH is 3.75 kHz, while in the NPUSCH is 15 kHz.

When the requested resources are lower than the available ones, the spectral efficiency is driven by the number of bits that are sent by the users with a successful access phase. In this case, resources are being wasted. On the contrary, η_s is led by the number of bits that can be sent in the available resources since this represents the maximum number of bits that can be sent by the users. In Eq. 6.12, $P_s(N_{rep}, M)$ is the success probability of the access phase (explained in the next Section). For the

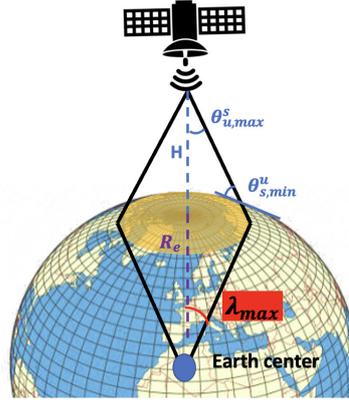


FIGURE 6.8: Definition of the satellite admittance region [106].

sake of clarity, the users winning the access phase are indicated as follows:

$$K = \min\left(\frac{U}{M}, N_{sc}^{RA}\right) \cdot P_s(N_{rep}, M) \quad (6.14)$$

Accordingly, the spectral efficiency η_s is a function of the three design parameters (N_{rep}, T_{RAO}, M) . As a result, to maximize the spectral efficiency of the systems, especially in highly congested areas, it is crucial to find the optimal value of the tuple N_{rep}, T_{RAO}, M , given as inputs U, \bar{D} , and the satellite parameters. Please note that the latter defines the average SNR, and thus the MCS, and the RTD. Then, our problem can be written as follows:

$$\begin{aligned} P_1 : & \max_{T_{RAO}, N_{rep}, M} \eta_s(T_{RAO}, N_{rep}, M) \\ \text{s.t. } C1 : & N_{rep} \in [1, 2, 4, 8, 16, 32, 64, 128] \\ C2 : & T_{RAO} \in [40, 80, 160, 320, 640, 1280, 2560] \text{ms} \\ C3 : & T_{RAO} > T_{access}(N_{rep}) \\ C4 : & \sum_{m=1}^M \overline{N_{RAO}} \cdot B \cdot \eta(m, T_{RAO}, N_{rep}) \cdot T_{RAO} \geq U \cdot \bar{D} \end{aligned} \quad (6.15)$$

In P_1 , constraints $C1$ and $C2$ guarantee that the number of uplink repetitions and the length of the RAO take the values provided by the NB-IoT standard; $C3$ means that the RA periodicity is long enough to absorb the access time; finally, $C4$ ensures that the overall demand is satisfied. $\overline{N_{RAO}}$ represents the average number of RAO.

6.2.2.2 Success Probability

Let us consider an uplink satellite communication network on a circular orbit and at an altitude equal to H , where the on-ground UEs can communicate with the flying platform only if they are located within its coverage area. The devices, with density α , are assumed to be distributed according to a homogeneous Poisson Point process

(PPP) on the Earth's surface. The distance (D_u) from the satellite to the generic u -th user depends on the relative position of the UE within the footprint of the satellite, defined through the Earth-centered angle λ_u , *i.e.*, the angle measured at the center of the Earth from the sub-satellite point to the UE. Thus, D_u is obtained by means of the cosine rule as follows:

$$D_u(\lambda_u) = \sqrt{d_{sat}^2 + R_e^2 - 2R_e d_{sat} \cos(\lambda_u)} \quad (6.16)$$

where $d_{sat} = R_e + H$ and R_e is the Earth radius. The earth central angle is in a geometrical relationship with the elevation angle (ϑ_s^u), measured at the terminal between the spacecraft and the local horizon, and the nadir angle, ϑ_u^s , which is the angle between the boresight of the antenna on the satellite and the u -th UE, as also shown in Chapter 3. The three of them define the relative geometry between the satellite and the user. In the proposed system, shown in Fig. 6.8, the admittance region of the devices for communication purposes is limited by the beamwidth of the satellite, *i.e.*, $2 \cdot \vartheta_{u,max}^s$. Consequently, also the other two angles are bounded, *i.e.*, there will exist a minimum elevation angle $\vartheta_{s,min}^u$ and a maximum Earth central angle $\lambda_{u,max}$ which allows the communication between the user and the satellite. The serving flying platform can be seen from any terminal within the following area:

$$A_{cap} = 2\pi R_e^2 (1 - \cos \lambda_{u,max}) \quad (6.17)$$

Thus, based on the satellite - UE geometry, the uplink power P_r^{UL} is computed as follows:

$$P_{r,u} = EIRP + G^{RX}(\lambda) - PL(\lambda)_s^u \quad (6.18)$$

where the $EIRP = P_t + G^{TX} - L_E$ is the user's EIRP, which depends on the power, the gain, and the equipment losses at the transmitter. While the receiving gain of the satellite depends on the user's position and it is computed according to the Bessel function in [39]. Clearly, in the Bessel function, the nadir angle is required, which is in a geometrical relationship with the earth central angle. Finally, $PL(\lambda)_s^u$ represents the overall path loss, which includes the free space loss (L_{fs}), the *shadow fading* Ψ , and signal loss due to tropospheric or ionospheric scintillations (L_s). Clearly, also the path loss depends on the user's position. The *shadow fading* is modeled as a Gaussian mixture model (GMM) with two different components corresponding to the LoS and the Non-LoS (NLoS) propagation conditions. Therefore, its distribution is modelled as [108]:

$$\Psi(\lambda) = p_{los} \mathcal{N}(-\mu_{los}, \sigma_{los}^2) + p_{nlos} \mathcal{N}(-\mu_{nlos}, \sigma_{nlos}^2) \quad (6.19)$$

In the formula the LoS probability p_{los} represents the component mixture ratio and it depends on the position of the user within the satellite footprint as follows:

$$p_{los} = e^{-\delta \frac{\sin(\lambda_u)}{\cos(\lambda_u) - \frac{R_e}{d_{sat}}}} \quad (6.20)$$

where δ is a parameter related to the propagation environment. It is worth highlighting that the NLoS probability is obtained as $p_{nlos} = 1 - p_{los}$. The mean μ_ξ and standard deviation σ_ξ of ψ are subject to the propagation model $\xi \in \{los, nlos\}$.

Congestion management algorithm

When the users are admitted in the satellite coverage they try to access the network simultaneously, creating high congestion and, thus, decreasing the probability to win the Random Access procedure and to send the data. In order to avoid large delays, traffic overload, extra energy consumption, and packet losses, a simple algorithm to manage the congestion is proposed. Similarly to the terrestrial Access Class Barring scheme, the BS broadcast access barring factor P_{SP} , which is inversely proportional to the number of satellite passages M :

$$P_{SP} = 1/M \quad (6.21)$$

where the values of M are decreased at every satellite passage. Then, at the beginning of the random access procedure, each UE determines its barring status with the information provided by the BS. In order to do that, the UE generates a random number between 0 and 1, $U[0,1)$. If this number is less than or equal to P_{sp} , the user selects and transmits its preamble. Otherwise, it waits for the next satellite passage. The time during which the device is barred from the RA procedure depends on the periodicity of the satellite and the dimension of the constellation.

Mean aggregate interference

Since there is no resource scheduling at this stage of the communication, each of the terminals, capable of initiating the RA procedure, randomly selects one of the available preambles among the N_{sc}^{RU} sub-carriers. Then, the collision occurs if the same preamble is chosen by more than one UE. Thus, within A_{cap} , the density of the users transmitting the same preamble can be expressed according to thinning process [109]:

$$\alpha_{sp} = \alpha / N_{sc}^{RA} \quad (6.22)$$

The aggregate interference I_i^{agg} in uplink for the preamble $i \in [0, \dots, N_{sc}^{RA} - 1]$ can be expressed as:

$$I_i^{agg} = \sum_{j=1}^{J_{coll}} P_{r,j}^{UL} \quad (6.23)$$

with $j = 1, \dots, J_{coll}$, being J_{coll} the number of interfering UEs with the same preamble and $P_{r,j}^{UL}$ is the received power from the j -th interferer. Notably, the number of user terminals along with their relative positions with respect to the satellite are

the main contributions affecting the system performance in terms of SINR. Therefore, the aggregate interference I_{agg} is a stochastic process whose distribution depends on the number and location of the interferers as captured by the point process, the statistics of each interfering signal, and the signal propagation model. Defining the number, the location, and the channel statistic of the interferers as J_{coll} , Λ , and Ψ , it is possible to write:

$$\begin{aligned}\mu_{I_{agg}} &= \mathbb{E}_{J_{coll}, \Lambda, \Psi} \left[\sum_{j=1}^{J_{coll}} P_{r,j}^{UL} \right] \\ \mathbb{E}_N &= \left[\left(\sum_{j=1}^{J_{coll}} \mathbb{E}_{\Lambda, \Psi} P_{r,j}^{UL} \right) | J_{coll} = j \right]\end{aligned}\quad (6.24)$$

Since all the devices on-ground have the same transmission power, it is worth emphasizing that each $P_{r,j}^{UL}$ is assumed to be identical and independently distributed (*i.i.d*) random variable with a certain distribution P_r . From the property of PPP, J_{coll} follows a Poisson distribution with mean value $\alpha_{sp} P_{sp} |A_{cap}|$. Therefore, it is possible to write:

$$\begin{aligned}\mu_{I_{agg}} &= \mathbb{E}_N [j \cdot \mathbb{E}_{\Lambda, \Psi} [P_r^{UL}]] \\ &= \alpha_{sp} P_{sp} |A_{cap}| \cdot \mathbb{E}_{\Lambda, \Psi} [P_r^{UL}]\end{aligned}\quad (6.25)$$

In order to compute $\mathbb{E}_{\Lambda, \Psi} [P_r^{UL}]$, the CDF of the earth central angle λ is evaluated. For the sake of clarity, the dependency of the subscript u is removed. Thus, the probability to find UEs within a given region defined by the angle λ is:

$$\mathbb{P}(\Lambda \leq \lambda) = \frac{2\pi \cdot R_e^2 (1 - \cos \lambda)}{|A_{cap}|}\quad (6.26)$$

Therefore, the expected value of P_r can be written as follows:

$$\begin{aligned}\mathbb{E}_{\Lambda, \Psi} [P_r^{UL}] &= \int_0^{\lambda_{max}} \mathbb{E}_{\Psi} [P_r^{UL} | \Lambda = \lambda] f_{\Lambda}(\lambda) d\lambda \\ &= \frac{2\pi \cdot R_e^2}{|A_{cap}|} \int_0^{\lambda_{max}} \mathbb{E}_{\Psi} [P_r^{UL} | \Lambda = \lambda] \sin(\lambda) d\lambda\end{aligned}\quad (6.27)$$

where $f_{\Lambda}(\lambda)$ is the pdf of $\mathbb{P}(\Lambda \leq \lambda)$. The last step is to consider the shadow fading Ψ :

$$\begin{aligned}\mathbb{E}_{\Psi} [P_r | \Lambda = \lambda] &= \mathbb{E}_{\Psi} [P_r | \Lambda = \lambda, LoS] \cdot p_{los} \\ &\quad + \mathbb{E}_{\Psi} [P_r | \Lambda = \lambda, NLoS] \cdot p_{nlos}\end{aligned}\quad (6.28)$$

In order to compute the expectation of P_r given Ψ , it should be noted that:

$$\begin{aligned}\mathbb{E}_{\Psi} [P_r | \Lambda = \lambda] &= \mathbb{E}_{\Psi} \left[\frac{EIRP \cdot G^{RX}(\lambda)}{L_{fs}(\lambda) L_s \Psi(\lambda)_{\xi}} \right] \\ &= \left(\frac{EIRP \cdot G^{RX}(\lambda)}{L_{fs}(\lambda) L_s} \right) \mathbb{E}_{\Psi(\lambda)} \left[\frac{1}{\Psi(\lambda)_{\xi}} \right]\end{aligned}\quad (6.29)$$

Both Ψ_{los} and Ψ_{nlos} follow a log-normal distribution, therefore:

$$\mathbb{E}_{\Psi} \left[\frac{1}{\Psi_{\xi}} \right] = e^{(-v\mu_{\xi} + v^2\sigma_{\xi}^2/2)} = 10^{0.1(-\mu_{\xi} + v\sigma_{\xi}^2/2)} \quad (6.30)$$

where

$$\ln\left(\frac{1}{\Psi_{\xi}}\right) \sim \mathbb{N}(-v\mu_{\xi}, v^2\sigma_{\xi}^2), \quad v = \ln(10)/10, \quad \xi \in Los, NLoS \quad (6.31)$$

Now it is possible to write the *mean aggregate interference* $\mu_{I_{agg}}$ as follow:

$$\begin{aligned} \mu_{I_{agg}} &= \frac{EIRP \cdot 2\pi R_e^2 \alpha_{sp} P_{sp}}{A_f L_s} \Delta(\lambda_{max}) \\ \Delta(\lambda_{max}) &= \int_0^{\lambda_{max}} \frac{\sin(\lambda) G^{RX}(\lambda)}{D(\lambda)^2} \nu(\lambda) d\lambda \end{aligned} \quad (6.32)$$

with $A_f = \left(\frac{4\pi f}{c}\right)^2$ and

$$\begin{aligned} \nu(\lambda) &= e^{(-\delta \frac{\sin(\lambda)}{\cos(\lambda) - Re/d_{sat}})} e^{(-v\mu_{Los} + v^2\sigma_{Los}^2/2)} + \\ &\quad (1 - e^{(-\delta \frac{\sin(\lambda)}{\cos(\lambda) - Re/d_{sat}})}) e^{(-v\mu_{NLoS} + v^2\sigma_{NLoS}^2/2)} \end{aligned} \quad (6.33)$$

RA Success probability

Finally, given a user with an angle λ_c , the success probability, *i.e.*, the probability that the user successfully computes the random access, is evaluated as follows:

$$\begin{aligned} \mathbb{P}_{s|\lambda_c} &= \mathbb{P}[SINR > \gamma] = \\ &\quad \mathbb{P}[SINR > \gamma] \cdot p_{los} + \mathbb{P}[SINR > \gamma] \cdot p_{nlos} \end{aligned} \quad (6.34)$$

where:

$$SINR = \begin{cases} \frac{EIRP + G^{RX}(\lambda_c)}{(\mu_{I_{agg}} + N) L_{fs}(\lambda_c) L_s \Psi_{los}(\lambda_c)} & \text{for LoS} \\ \frac{EIRP + G^{RX}(\lambda_c)}{(\mu_{I_{agg}} + N) L_{fs}(\lambda_c) L_s \Psi_{nlos}(\lambda_c)} & \text{for NLoS} \end{cases} \quad (6.35)$$

and γ is the SINR threshold and N represent the noise power observed at the receiver.

Therefore using Eq. 6.17, 6.28, and 6.30 it is possible to write:

$$\begin{aligned} \mathbb{P}_{s|\lambda_c} &= \mathbb{P} \left[\Psi_{los} < \frac{EIRP \cdot G^{RX}(\lambda_c)}{(\mu_{I_{agg}} + N) L_{fs}(\lambda_c) L_s} \right] p_{los} \\ &\quad + \mathbb{P} \left[\Psi_{nlos} < \frac{EIRP \cdot G^{RX}(\lambda_c)}{(\mu_{I_{agg}} + N) L_{fs}(\theta_s^u(\lambda_c) L_s)} \right] p_{nlos} \\ &= \frac{p_{los}}{2} \left[1 + erf \left(\frac{x - \mu_{los}}{\sigma_{los} \sqrt{2}} \right) \right] \\ &\quad + \frac{p_{nlos}}{2} \left[1 + erf \left(\frac{x - \mu_{nlos}}{\sigma_{nlos} \sqrt{2}} \right) \right] \end{aligned} \quad (6.36)$$

where

$$x = 10 \log 10 \left[\frac{EIRP \cdot G^{RX}(\lambda_c)}{(\mu_{I_{agg}} + N) L_{fs}(\lambda_c) L_s} \right] \quad (6.37)$$

For a device attempting its transmission N_{rep} times, the probability that at least one repetition is successfully received is formulated as:

$$\mathbb{P}_{s|\lambda_c}(N_{rep}) = 1 - [1 - \mathbb{P}_{s|\lambda_c}(1)]^{N_{rep}} \quad (6.38)$$

In order to find the average success probability for a given coverage region given by λ_{max} , Eq. 6.38 is de-conditioned over the distribution of λ as follows:

$$\mathbb{P}_s(N_{rep}) = \int_0^{\lambda_{max}} \mathbb{P}_{s|\lambda_c}(N_{rep}) f_{\Lambda}(\lambda_c) d\lambda \quad (6.39)$$

6.2.3 Solution to spectral efficiency maximization

In this subsection, the solution to the problem P_1 in 6.15 is provided. In P_1 , it is possible to observe that, according to $C1$ and $C2$, the search space of T_{RAO} , N_{rep} , are limited by the NB-IoT standard. However, M can be very high; potentially, it can be equal to $\frac{24 \cdot 3600}{T_{RAO}}$. Testing each value of T_{RAO} , N_{rep} , M to find the tuple maximizing the spectral efficiency of the system would provide near-optimal performance, at the expense of the computation complexity.

Therefore, we propose a heuristic search algorithm, which compromises the performance against the complexity. The key idea behind the heuristic algorithm is to restrict the search space of M to the value providing the maximum spectral efficiency. Indeed, η_s has only one maximum. The function's behavior demonstrated in the following sub-paragraph, forced us to define this heuristic approach instead of exploiting already developed algorithms, such as the bisection method.

The Heuristic approach proposed to solve P_1 is illustrated in Algorithm 1. It starts by initiating the values of T_{RAO} , and by iterating N_{rep} , it computes the time for the data transmission, verifying that $C3$ is satisfied, *i.e.*, $T_{RAO} - T_{access}(1, N_{rep}) > 0$. If this condition is not met, the next RAO is selected. Once the couple $\{(T_{RAO}), (N_{rep})\}$ has been determined, such that condition $C3$ is satisfied, the capacity of the system is computed, for each satellite passage, and compared with the user demand. The algorithm stops as soon as the capacity provided by the satellite passage $m + 1$ is lower than the one provided by the m -th one.

Thus, if $C5$ is satisfied, η_{m-1} and the satellite passage $m-1$ are saved. Considered the same T_{RAO} , these operations are repeated for all the values of N_{rep} that meet $C3$ and for each of them (η_{max}, m_{max}) are saved; then, the algorithm selects the maximum value of the spectral efficiency among the available repetitions and it keeps iterating as long as there are values of RAO. Eventually, after collecting the maximum η_s per random access opportunity, the algorithm selects the tuple $T\{RAO, N_{rep}, m\}$ providing the highest value of the spectral efficiency for a given user density, demand, and satellite parameters.

Proof of the uniqueness of the maximum

The proof of the uniqueness of the maximum relies on the non-linear behavior of the spectral efficiency. This depends on the number of users winning the access phase.

Recalling eq. 6.14, it is possible to identify two functions:

$$\begin{aligned} f(M) &= P_s(N_{rep}, M) \\ g(M) &= \min\left(\frac{U}{M}, N_{sc}^{RA}\right) \end{aligned} \quad (6.40)$$

where $g(M)$ defines the number of users performing the access phase and $f(M)$ represents the success probability. It is worth mentioning that, as long as the number of users in the system is greater than the resources available in the NPRACH, $g(M)$ is a constant equal to N_{sc}^{RA} . Then, when $\frac{U}{M} < N_{sc}^{RA}$, $g(M)$ decreases as M increases. Indeed, its first derivative $g'(M) = -\frac{1}{M^2} < 0$.

Regarding $f(M)$, it is necessary to differentiate between two cases: i) when the requested resources are lower than the available ones and ii) when the requested resources are greater than the available ones.

If $S_{req}(M, N_{rep}) < S_{av}(T_{RAO}, N_{rep})$ and $\frac{U}{N} > N_{sc}^{RA}$, then η_s is an increasing function. For the sake of clarity, let us focus on the case of one repetition and only one of the two components of the GMM.

It is worth emphasizing that the function $f(M)$ is a composite function, given by erf and $\log_{10}\left[\frac{G_1}{\left(\frac{G_2}{M} + N\gamma\right)}\right]$, where G_1 is a constant to express the transmitted power included the losses, G_2 is the mean value of the interference multiplied by the constant threshold γ , and C_{los} refers to the ratio between the mean and the variance of the shadow fading in the LoS, which is also a constant. The first derivative of the logarithmic function and of the erf are always greater than 0 and, since they are multiplied by constants always positive, then $f'(M) > 0$. The same considerations are valid for the second component of the GMM, *i.e.*, the NLoS component; therefore, it is licit to say that $P_s(N_{rep}, M)$ is an increasing function.

When $\frac{U}{M} < N_{sc}^{RA}$, then the objective function is a decreasing function. To prove that, it is of paramount importance to evaluate the first derivative of $\eta_s(M) = g(M)f(M)$:

$$\eta'(M) = f'(M)g(M) + g'(M)f(M) \quad (6.41)$$

Since, $f(M)$, $f'(M)$, and $g(M)$ are positive along all the values of the M , then:

$$f(M)g'(M) > -g(M)f'(M) \quad (6.42)$$

which holds only for positive values of $g(M)$. This proves that η_s is always decreasing when $M \in \left[\frac{U}{N_{sc}^{RA}+1}, \infty\right)$. Moreover, this maximum, given by the point where the function becoming decreasing, is a cusp being different between the upper and the lower limits around this value of M .

Finally, to conclude the analysis, when $S_{req} > S_{av}$, the maximum value of η_s will be only provided by all the values of M for which the users winning the RA (*i.e.*, K) will request more resources than the available one. In all the other cases, the users will request less resources than the available ones, thus η_s will not reach its maximum.

Algorithm 1: Heuristic algorithm to solve P_1

```

Input
 $U, \bar{D}$ ;
Satellite parameters: shown in Table 6.4;
Output
 $\eta_s^*, RAO^*, N_{rep}^*, M^*$ 
Initialize ( $\eta_{s,RAO} = \text{zeros}(RAO, 1)$ )
for  $k=1:RAO$  do
  Initialize ( $\eta_{s,Rep} = \text{zeros}(Rep, 1)$ )
  for  $j=1:N_{rep}$  do
    Initialize ( $\eta_{s,M} = \text{zeros}(M, 1)$ )
    Solve (C3)
    if successful then
      for  $i=1:M$  do
        Compute the spectral efficiency with Eq. 6.11
        Solve (C5)
        if successful then
           $\eta_{s,M}[i] \leftarrow \eta_s$ 
          if  $\eta_{s,M}[i-1] \geq \eta_{s,M}[i]$  then
             $\eta_{s,M}^* \leftarrow \eta_{s,M}[i-1]$ 
             $M^* \leftarrow M[i-1]$ 
            break
          end
        end
      end
    else
      Break
    end
  end
   $\eta_{s,Rep}[j] \leftarrow \eta_{s,M}^*$ 
end
   $\eta_{s,RAO}[k] \leftarrow \text{argmax}_j (\eta_{s,Rep})$ 
   $N_{rep}^* \leftarrow j^*$ 
end
 $\eta_s^* \leftarrow \text{argmax}_k (\eta_{RAO})$ 
 $RAO^* \leftarrow k^*$ 

```

TABLE 6.4: Satellite parameters

Satellite orbit	Set 3 LEO 600 km	Set 4 LEO 600 km
Equivalent satellite antenna aperture	0.4 m	0.097 m
Sat EIRP density	28.3 dBW/MHz	21.45 dBW/MHz
Sat Tx/RX max Gain	16.2 dBi	11 dBi
3dB beamwidth degrees	22.1 degree	104.7 degree
Sat beam diameter	234 km	1700 km
G/T	-12.8 dB K ⁻¹	-18.6 dBK ⁻¹
Beam visibility window	33.86 s	246.9 s

TABLE 6.5: Results of P_1 , for $\bar{D} = 166$ bytes, $\alpha = 0.03$ [users/km²/channel] Set 3

EDT	η^*	N_{rep}^*	RAO^*	M^*	$M_{bisection}^*$
No	2.87 [bit/s/Hz]	4	320 [ms]	20	33
Yes	5.58 [bit/s/Hz]	4	160 [ms]	20	32

6.2.4 Numerical results

In this Section, the numerical results are presented, with the aim to evaluate the performance of the proposed algorithm. It is essential to highlight that the output of the function P_1 is obtained for a congested system, *i.e.*, when there are more users than the available resources in the Random Access phase and, therefore, collisions will occur. The scope is to provide an optimized system for one beam and bearer (*i.e.*, one channel of 180 kHz). Then, the output can be scaled accordingly to the number of beams per satellite, the number of satellites, and the available bearers. In the scenario, only one coverage enhancement zone is considered and $N_{sc}^{RA} = 48$ preambles are available for the users. Regarding the number of sub-carrier for the data transmission, only single-tone is taken into account with 15 kHz spacing, *i.e.*, $N_{sc} = 12$, in order to increase the resources available during the data transmission stage. Regarding the data demand, a Pareto distribution with shaping parameter $\beta = 2.5$ is taken into account. Thus, $\bar{D} = \frac{\beta x}{\beta - 1}$ with $x \in [5, 200]$ bytes. The satellite parameters are summarized in Table 6.4. Finally, the δ parameter for the calculation of the LoS probability is equal to 3.2.

Moreover, the proposed algorithm is applied to the EDT and the results are then compared to the classical 4-step RA followed by the data transmission. In this case, the time to access is given by:

$$T_{access}^{EDT}(N_{rep}) = T_{msg1}(N_{rep}) + RTD + K^{DL} + K^{UL} + T_{msg2} \quad (6.43)$$

6.2.4.1 User Barring performance

The input parameter with the greatest impact on the η_s function is the number of users U . Fig. 6.9 plots the RACH success probabilities with the proposed access

TABLE 6.6: Results of P_1 , for $\bar{D} = 166$ bytes, $\alpha = 0.003[\text{users}/\text{km}^2/\text{channel}]$ Set 4

EDT	η^*	N_{rep}^*	RAO^*	M^*	$M_{bisection}^*$
No	0.96 [bit/s/Hz]	8	640 [ms]	49	70
Yes	2.04 [bit/s/Hz]	16	320 [ms]	14	30

barring scheme. As expected, the RACH success probability increases by decreasing P_{SP} , since the users are distributed along the satellite passes and, in this fashion, the aggregate interference and collision probability are lessened. Indeed, the RACH success probability decreases with the increase of the density of IoT devices as they generate interference leading to lower received SINR at the BS. However, if M is too large then only few users can perform the RA. In the case of lower UEs than the available resources, the spectral efficiency is reduced. Moreover, a large number of satellite passages represents a cost for the provider of the service. Regarding the up-link repetitions, as expected, an increased value of repetitions improves the success probability as shown in Fig. 6.9. However, it is worth emphasizing that if the number of repetitions is overestimated, there will be a waste of potential resources for data transmission, resulting in reduced efficiency of resource utilization.

To verify the goodness of the proposed heuristic method, it is compared with the bisection method. From Tables 6.5 and 6.6, it is possible to see that our algorithm outperforms the bisection method. Indeed, since different numbers of satellite passages provide the same maximum achievable spectral efficiency, due to the fact that $S_{av} < S_{req}$, the bisection method outputs one of them, but not the minimum. Instead, the proposed method provides the minimum number of satellite passages that guarantee the highest spectral efficiency, by stopping at the first value of M that reaches η_s maximum, thus reducing the system costs.

6.2.4.2 Impact of beam dimension

The footprint diameter and the altitude of the satellite greatly impact the RA and data transmission phases, both in terms of visibility and link budget. Fig. 6.10 shows that the mean interference power increases by increasing the beam (considering the same user density), thus by increasing ϑ_u^s and decreasing ϑ_s^u , due to the presence of more interfering terminals within the beam. Indeed, from the PPP assumption with fixed density, the average number of interferers within A_{cap} is equal to $\alpha 2\pi R_E^2 (1 - \cos(\lambda_{max}))$. Therefore, considering Set 4, a greater number of users perform the RA. This is translated into a higher number of satellite passages to achieve the highest η_s . Clearly, this is due to the fact that users performing the procedure linearly scale with the number of satellite passages. On the other hand, a larger beam entails an increment in the average number of RAOs, being the visibility window longer. Indeed, the UEs can be served in more RAOs. Therefore, Set 4, with a wide antenna beam on Earth, allows operators to aggregate the IoT traffic through a lower number of LEO satellites. However, a larger beam diameter implies lower elevation

angles, and, thus, lower SNR, limiting the MCS and the number of bits that can be sent in one transmission. Thus, during the RA, more uplink repetitions are necessary to win the contention phase, as shown in Table 6.6. Moreover, since in this configuration, the average SNR is -3 dB and the lowest MCS achieves a $BLER = 10^{-1}$ dB with a $SNR = -0.51$ dB, it is necessary to consider also repetitions for the data transmission, thus, increasing the transmission time. Besides, it is of paramount importance to consider the higher RTD due to the lower values of the elevation angle. Indeed, to meet C3, a longer periodicity is necessary, as it is shown in Table 6.6. Thus, to better exploit the longer visibility window and to reduce the impact of RTD typical of this satellite configuration, the best choice is to enable the EDT, as shown in Table 6.6. Indeed, the latter is particularly useful when it is not possible to send large data packets and to reduce the signaling overhead. In set 4, with the EDT the number of uplink repetitions is higher than the one obtained with the classical 4-step RA as the available resources are saturated when $K > N_{sc}^{RA}$, leading to a lower SINR in the access phase and lower access probability, which is counteracted by increasing the number of repetitions.

6.2.4.3 Impact of the requested capacity

Finally, the last input impacting the spectral efficiency is the data demand. As previously explained, high SNR allows the transmission of RU with a greater number of bits. Since Set 3 and Set 4 have different average SNR, the maximum number of bits per Resource Unit (RU) changes based on the MCS. Indeed, when Set 4 is considered, a maximum of 42 bytes can be sent. Moreover, the difference between the values of \bar{D} when EDT is enabled or disabled depends on the fact that the maximum number of bits that can be sent in Msg3 of EDT is 10^3 . Fig. 6.11a, 6.12a, 6.13a, and 6.14a show the number of satellite passes and uplink repetitions varying the \bar{D} when EDT is disabled. For Set 3, the value of Random Access Occasion that optimizes the spectral efficiency is 320 ms, while for Set 4 is 320 ms up to $\bar{D} = 25$ bytes, then it moves to 640 ms. As it is possible to notice, these values are not constant with the increase of \bar{D} , since S_{req} increases too. This behavior is motivated by the fact that when the capacity is low, more satellite passes are necessary to achieve the highest spectral efficiency in the system. Indeed, in this case, K , *i.e.*, the users that gain access and send the data, are equal to N_{sc}^{RA} . By increasing the data payload dimension, the algorithm decreases the number of uplink repetitions to increase the resources for data transmission. Therefore, the number of satellite passages provides a number of users performing the RA, such that the success probability is high. Indeed, $K \cdot S_{req}$ needs to be close to the available resources. We observe the same trends in Set 4. The only difference is that with $RAO = 320$ ms and $\bar{D} = 25$ bytes the available resources are saturated, while when the available resources are greater than the requested ones (*e.g.*, with $\bar{D} = 8, 17, 33, 42$ bytes), the algorithm outputs a number of satellite passes and uplink repetitions to waste less available resources as possible, *i.e.*, $K = N_{sc}^{RA}$

When EDT is enabled, for Set 3, the spectral efficiency is maximum with a RAO equal to 160 ms until $\bar{D} = 83$ bytes, then it is double to 320 ms; with Set 4, RAO is 160 ms until $\bar{D} = 25$ bytes, then it becomes 320 ms. The results are shown in Fig. 6.11b, 6.12b, 6.13b, and 6.14b. We observe the exact behavior as for the classical 4-step RA with the main difference that in this case the number of satellite passages is reduced at the expense of the repetitions (which is higher than in the case of EDT disabled) as the RAO length is halved.

6.2.5 Conclusion and future work

In this work, we investigated a LEO satellite-based NB-IoT system with the aim to design a joint access and data phases, considering the satellite characteristics, and thus, the average SNR and RTD, the average amount of data per user, and the number of users in the system. In this context, we formulated a novel problem to maximize the spectral efficiency via a joint definition of the length of the Random Access Opportunity, the number of uplink repetitions, and the number of satellite passages while satisfying the total demand. We solve the problem by defining a heuristic approach, which reduces the search space of the satellite passages, which can be potentially large. The proposed approach outperforms conventional algorithms, such as the bisection method. The extensive numerical analyses show that due to the linear relationship between the number of satellite passages and the users performing the access, when the system is crowded it is necessary to increase M and the N_{rep} to increase the RA success probability. The other factor that drives the choice of the three design parameters is the data demand. Indeed, considering a fixed user density, when the requested capacity is lower than the available ones, the algorithm outputs a value of M and N_{rep} such that K is equal to N_{sc}^{RA} , which is the maximum number of users that can win the RA. Clearly to reach this value, high M and N_{rep} are required. When the available resources are lower than the requested ones, the algorithm returns the first value of M and N_{rep} which will saturate the resources. This implies short periodicity, in order to increase the number of RAOs within the satellite visibility. Moreover, with short periodicities, the effect of the aging parameters is reduced. In terms of RA occasion, medium-length RAOs, *i.e.*, 160ms, 320 ms, 640 ms, are preferred compared to short and long ones. The reason is twofold: i) short Occasions are discarded as the satellite domain is characterized by long propagation time which decreases the resources dedicated to the data phase; ii) longer RAO reduces the number of attempts to perform the RA, and then the data transmission. Finally, we have applied the proposed method to Early Data Transmission. The results show that with this algorithm EDT is preferable in the case of satellites that generate wide beams, as the RTD is high and reduces the time for data. EDT brings benefits also with lower beam dimension, even if the maximum spectral efficiency is gained when $\frac{U}{M} > N_{sc}^{RA}$ meaning there are collisions among the users.

The overall framework proposed in this article to aggregate the IoT traffic through an LEO satellite is novel. Future works foresee the design of a satellite constellation

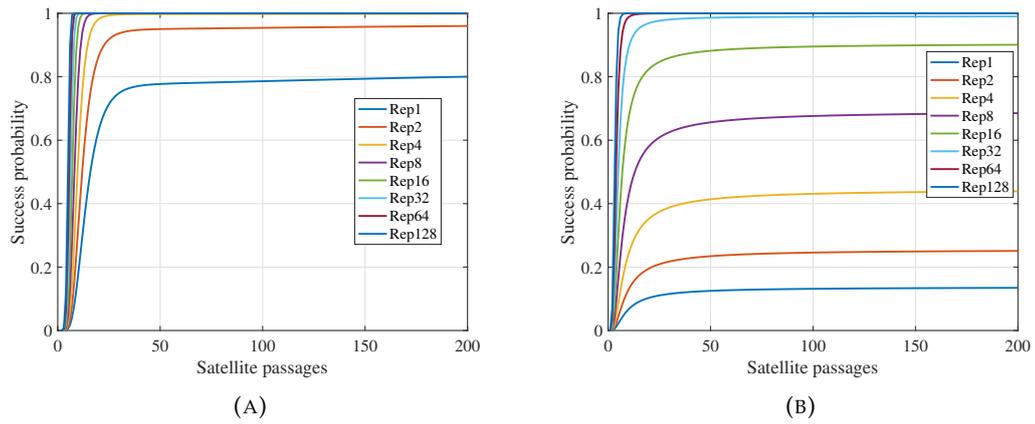


FIGURE 6.9: Success probability in Set 3 and Set 4. $\alpha = 0.002[\text{users}/\text{km}^2/\text{channel}]$ [106].

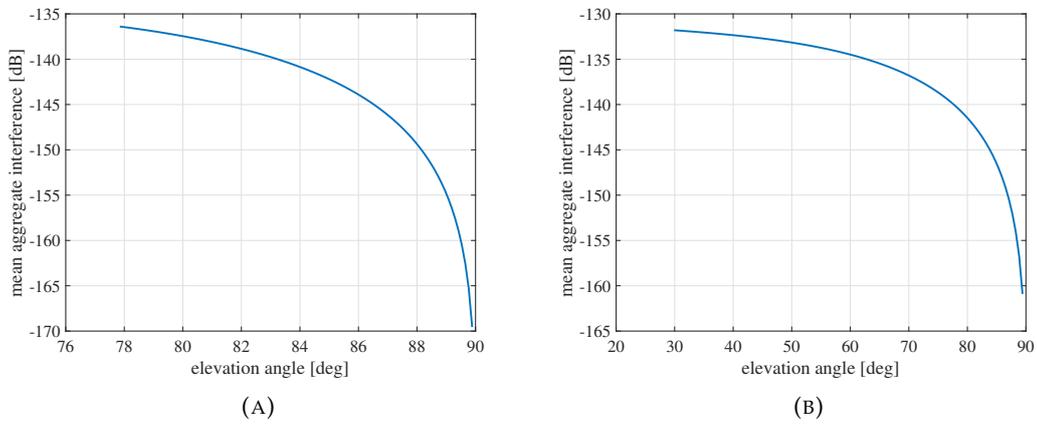


FIGURE 6.10: Average interference power in Set 3 and Set 4 [106].

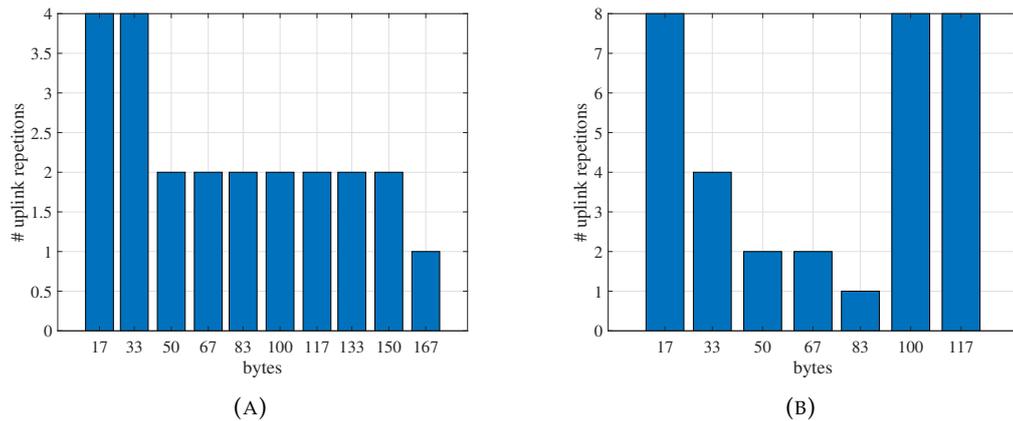


FIGURE 6.11: Number of uplink repetitions varying \bar{D} for Set 3 without (a) and with EDT (b). $\alpha = 0.01$ [106].

for NB-IoT via NTN to take into account not only the number of uplink repetitions and the random access occasion but also other parameters such as the Extended Discontinuous Reception.

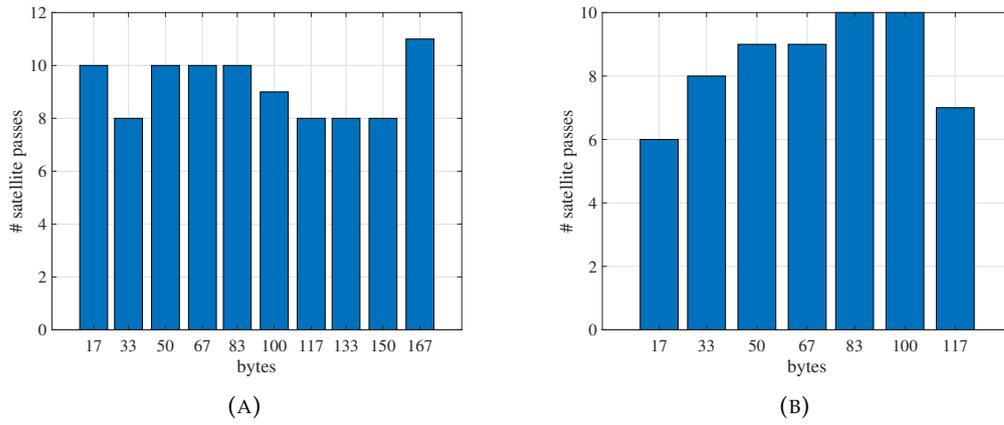


FIGURE 6.12: Number of satellite passes varying \bar{D} for Set 3 without (a) and with EDT (b). $\alpha = 0.01$ [106].

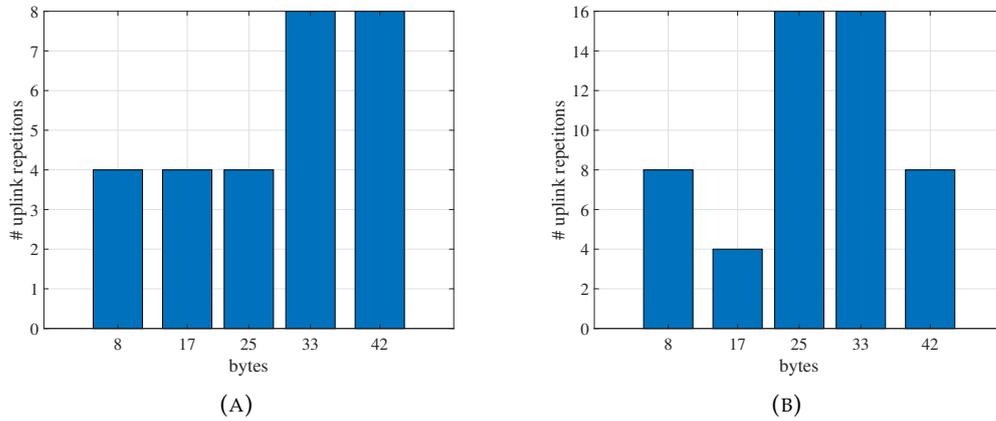


FIGURE 6.13: Number of uplink repetitions varying \bar{D} for Set 4 without (a) and with EDT (b). $\alpha = 0.001$ [106].

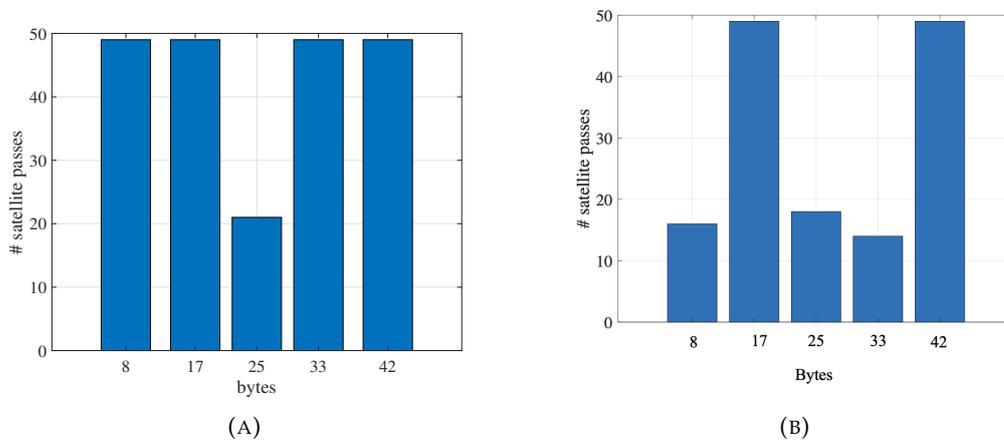


FIGURE 6.14: Number of satellite passes varying \bar{D} for Set 4 without (a) and with EDT (b). $\alpha = 0.001$ [106].

Chapter 7

Conclusions

The main objective of this thesis was to investigate the challenges introduced by the satellite channel impairments in the PHY and MAC layer procedures of the NB-IoT protocol and to develop new algorithms capable of allowing reliable communication by counteracting them. In particular, these algorithms foresee minimal adaptations in the current standard, to foster a fully functional NB-IoT SatCom system. In this framework, a thorough link budget analysis, following the 3GPP specifications, has been performed for an NB-IoT via NTN systems. Fixed IoT devices on Earth, LEO satellites in different constellations, and GEO satellites have been considered with the purpose to compute the multi-beam interference in the downlink and the multi-user interference in the up-link. At the time of writing, the 3GPP SI on NB-IoT via NTN had just been released, new satellite configurations were being defined, and the terminal characteristics had also not yet been specified. Therefore, this study was the basis for all the analyses performed in the thesis, both at the link level (to calculate coefficients) and at the system level (to compute the SNR and SINR).

Then, the work focused on the impact of satellite impairments on the RA procedure. In particular, being the preamble transmission the most critical part, a novel preamble detection method based on wavelet transform has been presented. The proposed algorithm allows the detection of the preamble at -22 dB with an AWGN channel and achieved an estimation accuracy of the ToA of $1 \mu s$. However, an high number of repetitions is required.

The next step was to analyze the performance of RA via NTN at the system level. Indeed, when there are many users who have to transmit in a short time, collisions occur. They have a negative impact on the user's network access time and battery life. In fact, the more times a user retries the RA procedure after a collision, the higher the power consumption. Thus, it is fundamental to find the right combination of the access parameters given the satellite configuration and the number of users you need to serve. The results show that a random access periodicity of medium length provides the best trade-off between the probability of successful access and the time required to complete the procedure. A critical parameter is the number of uplink repetitions, as increasing this value also the probability of detection increases. However, an overestimated number of uplink repetitions does not improve the performance but decreases data transmission resources.

To obtain a fully operative Satcom-based NB-IoT system, a joint access and data phase need to be designed. To this aim, different criteria can be followed, such as the maximization of the throughput or the user fairness, with different methodologies, *e.g.*, the user scheduling based on the size of the data payload and satellite visibility period per user. However, due to the very low period during which the users can communicate with the flying platform and its high speed, which makes obsolete the users' channel coefficients, it is necessary to serve them in the current RAO. To this aim, a novel algorithm that jointly designs the access and data phases has been proposed, which not only considers the high RTD of the SatCom system but also takes into account the dynamicity of the whole system by incorporating the distinct channel conditions and the various data demands coming from the on-ground users. This was an untreated problem in the literature and the methodology proposed in this thesis can act as a baseline for future works in the area of satellite constellation design, particularly for LEO satellite-based NB-IoT systems.

Next, the problem of the collisions has been considered also at the link level, proposing more sophisticated algorithms to detect the collision on the incoming preambles (by means of a Convolutional Neural Network) and to jointly decode the symbols of users that transmit on the same time/frequency resources through a NOMA algorithm, named Sparse Code Multiple Access. The NN-based detector requires an in-depth analysis of the complexity of the network and the type of required inputs. The results show that the CNN is able to correctly classify up to two users choosing the same preamble, while the classification accuracy reaches the 60% in case of four colliding terminals. Regarding the classical NOMA schemes, being the SCMA a short sparse code, it is possible to exploit it in a crowded scenario, such as the mMTC. One of the main advantages of NOMA is the increased throughput, as more than one user can be served on the same time/frequency resources. However, contrary to terrestrial communication systems, SatCom-based systems are characterized by the high speed of the satellite, which makes obsolete the channel coefficients and requires frequent channel estimation, reducing the resources dedicated to the data transmission. Moreover, synchronization among the users is required.

Eventually, an application of NB-IoT communication system has been presented. With the aim to accurately localize a passenger train along the railway, a custom low-power sensor node with GNSS, RTK, IMU, and a NB-IoT module has been designed. The latter is in charge of receiving the RTK corrections to enable centimeter-level positioning accuracy. The accuracy of the GNSS receiver is not only essential for applications that are strictly related to navigation but also for all those applications related to communication. Indeed, given the essential role of GNSS in facilitating the integration of terrestrial and non-terrestrial communication systems, it becomes imperative to evaluate the performance of positioning. Significantly, users can proactively compensate for Doppler shifts and RTD by acquiring information regarding their own location and the ephemeris data of satellites. Evidently, the precision of the position directly influences the accuracy of pre-compensation.

The work pursued in this thesis can be extended in several directions. The main topics left for future works are:

- Absence of the GNSS receiver for pre-compensation of delay and Doppler: in the thesis, the users are always assumed equipped with the GNSS. However, starting from Rel.19 this assumption needs to be relaxed. Thus, new algorithms to compute the position of the device must be designed.
- Neural Network-based NOMA: in order to support the massive access of IoT devices, grant-free access NOMA has received great deal of attention. In grant-free transmission, each device transmits its data without the granting process, thus, the base station needs to identify the active devices among all potential devices. This process called active user detection (AUD), is a challenging problem in NOMA-based systems since it is difficult to identify active devices from the superimposed received signal. Exploiting the patterns of the MA signatures of NOMA schemes, it is possible to use a NN to mitigate this problem.
- System-level design: since most of the IoT applications are based on sporadic uplink data reports, continuous network coverage might not be needed. Thus, an interesting line of future work is to perform a system-level design from the space-segment perspective (*e.g.* the number of satellites, orbital parameters etc.) based on realistic user distribution on Earth, and realistic user demand for various IoT applications.

Appendix A

Appendix Localization systems for railway

This chapter is focused on the performance of the GNSS positioning since part of the work during the third year has been dedicated to the development of techniques and technologies to accurately localize a train. GNSS, which is a navigation system, can be viewed as a communications system, where the transmitters, *i.e.*, satellites, broadcast navigation data where are embedded the precise satellite clock as well as satellite position to allow the receiver to determine both the satellite time and position at the time of transmission. In order to localize itself, the GNSS receiver needs to measure the signals coming from four or more satellites simultaneously or nearly simultaneously without mutual interference. To avoid interference, the flying platforms use the Code Division Multiple Access (CDMA) scheme.

Furthermore, considering that GNSS currently serve as a fundamental requirement for enabling terrestrial communication standards via NTN, it is of paramount importance to assess the performance of positioning. Indeed, by obtaining knowledge of one's own location and the ephemeris data of the satellites, users can preemptively correct for Doppler shifts and RTD. Clearly, the precision of position directly impacts the accuracy of the pre-compensation.

Moreover, in this work, an application of NB-IoT communication is provided. To properly localize the train, a custom sensor board, which includes a GNSS receiver, an IMU and a NB-IoT modules, has been realized. The latter is used to receive Real Time kinematic (RTK) corrections to achieve centimeter-level positioning accuracy.

Where am I on the Earth?

Even if the answer to this question looks very simple, it can sometimes be a matter of life and death, *e.g.*, the crew of vessels in distress looking for assistance. Without being so dramatic, the knowledge of its own position on the Earth is an enabler of a limitless number of applications and scenarios, such as autonomous vehicles, railway signaling, home banking applications on smartphones,

Among the most stunning technological developments in recent years have been the immense advances in the realm of satellite navigation or GNSS technologies, both in terms of space segments and receiver design.

The GNSS describes any constellation of satellites that provides positioning, navigation, and timing (PNT) services on a global or regional basis. The most widely used GNSS is the American Global Positioning System (GPS), but other nations have long been fielding their own systems to provide complementary and independent PNT capabilities. There are currently five other GNSS systems besides GPS:

- BeiDou Navigation Satellite System (BDS), owned by the People's Republic of China. The BDS was formally commissioned in 2020. The operating system consists of 35 satellites.
- Glonass (Globalnaya Navigazionnaya Sputnikovaya Sistema) owned by the Russian Federation and has 24+ satellites.
- Indian Regional Navigation Satellite System (IRNSS)/ Navigation Indian Constellation (NavIC): owned by the government of India. It is an autonomous system designed to cover the Indian region by a margin of 1,500 km. The system consists of seven satellites.
- Quasi-Zenith Satellite System (QZSS): a regional system owned by the Japanese government and operated by QZS System Service Inc. QZS complements the GPS system to improve coverage in East Asia and Oceania. Japan officially declared the start of QZSS services in 2018 with four operational satellites with the goal of expanding the constellation to seven satellites by 2023.
- Galileo: European Union navigation system, which has been designed with the goal of providing navigation services independently of existing GNSS systems. It consists of 24 active satellites plus three spare satellites.

In addition, it is necessary to consider Satellite Based Augmentation Systems (SBAS), which transmit GNSS-like signals mainly dedicated to providing integrity information and wide area corrections and can also be used as extra navigation signals. GNSS, Regional Navigation Satellite Systems (RNSS), and SBAS providers are coordinating their efforts to adopt modernized open signals with compatible frequency planes, common multiple access schemes. With a GNSS system, the accuracy of the position (longitude, latitude, and altitude coordinates) is within 20m to approx. 1cm, while the exact time (Universal Time Coordinated, UTC) reaches an accuracy within 60ns to approx. 5ns. Moreover, speed and direction of travel (course) can be derived from these values, which are obtained from satellites orbiting the Earth. Speed of travel may also be determined directly by means of Doppler shift measurements. Each Satellite Navigation (SatNav) system uses the same basic principles to determine coordinates:

- Satellites with a known position send a signal;
- The position of the receiver is estimated based on the ToA of the radio waves.

Obviously, if the same type of measurement is performed with three satellites simultaneously, it is possible to locate the user in three dimensions at the intersection of three spheres and obtain the desired real-time navigation. However, in order to achieve that, the user terminal must be synchronized with the transmitters. Assuming that the user clock has an unknown bias error b_u , the user clock reads the signal at the time $t'_u = t_u + b_u$, where t_u is the "true" system time at the user's time of reception. By adding a ranging measurement to a fourth satellite, the solution can be found for both $x(t_u)$, *i.e.*, the user position at the time t_u and true user clock time t_u at the time at which the measurement is taken. The difference between satellite clock time and user clock time when the user clock is not precise is referred as "pseudorange." For the i -th satellite, this range difference is denoted as ρ_{iT} , where subscript T denotes true pseudorange. The latter, in the idealized error-free condition, is the true range plus the user clock bias correction:

$$\rho_{iT} = c(t_u - t_{si}) + c \cdot b_u \text{ [s]} \quad (\text{A.1})$$

where t_{si} is the clock reading of the i -th satellite. However, in the real measurement, there are random noise effects, various other bias errors, and propagation errors (and relativistic effects); thus, the measured pseudorange is:

$$\rho = \rho_{iT} + \Delta D_i - c\Delta b_i + c(\Delta T_i + \Delta I_i + \nu_i + I_{m_i}) \quad (\text{A.2})$$

where Δb_i is the satellite bias clock error (s); ΔD_i is the satellite position error effect on range; ν_i is the receiver measurement noise error for the i -th satellite; ΔI_i is the ionospheric delay error (s); ΔT_i is the tropospheric delay error (s); and I_m is the error due to the multipath.

The Ionospheric delay is typically the largest contributor to the computation of the pseudorange, especially at higher latitudes and during periods of high solar activity. The ionosphere can cause a delay of up to several tens of meters in the GNSS signal's arrival time, which can result in a significant error in the position solution as described by equation A.3:

$$I = \frac{40.3}{f_c^2} \text{TEC} \quad (\text{A.3})$$

Where I is the error in meters produced by the ionosphere; f_c is the carrier frequency of the GNSS signal; TEC is the number of electrons per unit area along its propagation path. This delay appears as the ionosphere is a dispersive medium refracting the GNSS signal while passing through it. This effect depends on the level of ionization and the frequency of the signal. It can be removed up to 95% by using multifrequency receivers, in the case of a single frequency, the receivers should apply a prediction model such as Klobuchar [110] or Nequick [111] to reduce the effect in the measurements.

Regarding the Tropospheric delay, the main feature of the troposphere is that it is a non-dispersive media with respect to electromagnetic waves up to 15GHz, *i.e.*, the

tropospheric effects are not frequency dependent for the GNSS signals. Therefore, the only way to mitigate the tropospheric effect is to use models and/or to estimate it from observational data. This is typically a smaller contributor to the pseudorange, causing a delay in the GNSS signal's arrival time which leads to an error in the position up to several meters. This delay depends on the temperature, pressure, and humidity as well as the transmitter and receiver antennas' location.

The multipath effect applies when multiple copies of the signal arrive at the receiver at different times. These copies appear as reflections or diffractions in the environmental features of the main signal, causing error in the estimation of the signal's arrival time and resulting in an error in the position solution. This can cause an error of several meters in the position solution.

The receiver noise can cause uncertainty in the signal strength measurement and thus the distance calculation. This can cause an error of several centimetres to several meters in the position solution, depending on the type of receiver and the environment in which it is being used.

The ephemeris errors (ΔD_i) can cause errors in the satellite position calculation, which affect the distance calculation from the receiver to the satellite. This causes an error of several meters in the position solution.

Satellite clock errors can cause an error of several meters in the position solution. The clock error in the satellite can cause uncertainty in the signal transmission time, and thus the distance calculation from the receiver to the satellite. It can be estimated by splitting its error into two components, one broadcasted in the navigation message and the second term due to the orbital eccentricity called Relativistic Clock Correction.

These delays cause errors in the position and time solutions calculated by the receiver, and can also affect the receiver's ability to track the GNSS signal. Some of these errors can be corrected using techniques such as differential GNSS (DGNSS), *e.g.*, RTK positioning. It is worth noting that these errors are statistical and can vary depending on the location, time, and the specific receiver. Moreover, different GNSS systems have different designs and characteristics that can affect the computation of the pseudorange.

A.1 European Rail Traffic Management System

The European Rail Traffic Management System (ERTMS) is a single European signalling and speed control system that ensures interoperability of the national railway systems, reducing the purchasing and maintenance costs of the signalling systems as well as increasing the speed of trains, the capacity of infrastructure and the level of safety in rail transport. ERTMS comprises of:

- the European Train Control System (ETCS), a train control standard, based on

in-cab equipment that can supervise train movements and stop the train according to the speed allowed in each section of the line, along with the calculation and supervision of the train's maximum speed at all times. Information is received by ETCS equipment next to the track (balise or radio) depending on the operating level. The response of the Conducting Staff is constantly monitored and, if necessary, the system imposes emergency braking.

- the Global System for Mobile communications for Railways (GSM-R) is the European radio communications standard for rail operations. Based on GSM radio technology, GSM-R uses unique frequency bands to enable communication between the train and the infrastructure

Within the ERTMS system, it is possible to identify three different levels depending on the type of train movement monitoring, communication, and the entity performing the train detection:

- Level 1 provides continuous supervision of train movement while non-continuous communication between the train and the track (normally through euro-balises) signals are required along the line and train detection is performed by the trackside equipment outside of the scope of ERTMS.
- Level 2 provides continuous supervision of train movement with continuous communication, provided by GSM-R, between the train and the track. Signals along the line are optional in this case, and train detection is performed by the trackside equipment outside the scope of ERTMS.
- Level 3 is also a signalling system that provides continuous train supervision with continuous communication between the train and the track. The main difference with Level 2 is that train position and integrity are managed within ERTMS, *i.e.*, no trackside signals or trackside train detection systems other than euro-balises are required. Train integrity is controlled by the train, *i.e.*, the train checks that it is complete and has not been accidentally split.

In order to implement the third level, an autonomous positioning system is required. Currently, the GNSS is the principal way of providing globally referenced positioning in integrated navigation systems. Although stand-alone GNSS would minimize costs, its use in the railway sector is still limited [112]. Due to the errors described above, the achieved accuracy is around 1-5 m, which is not sufficient for safety-critical applications. Indeed, GNSS adoption in this field faces restrictions due to the high accuracy and availability requirement due to potential Signal-in-Space (SiS) obstructions (*e.g.*, tunnels, trees) and excessive position errors caused by local effects (*e.g.*, multipath, spoofing). For example, train positioning on parallel rail requires a Horizontal Protection Level (HPL) less than 3 m [112]. Therefore, more complex positioning methods have been developed to fulfil the needs of these critical applications [113]. By requiring more accurate input data, these algorithms are able

to improve and enhance the positioning solutions obtained with stand-alone GNSS. Some examples of these algorithms are Precise Point Positioning (PPP) and RTK. Despite the development of these technologies, it is not always possible to guarantee an acceptable navigation accuracy, due to the lack of GNSS signal under certain operating conditions. Therefore, in order to improve the robustness and effectiveness of GNSS-based train localization systems, different integrated positioning solutions have been designed. These aim to compensate for the GNSS signal unavailability, especially during long signal outages, with the aid of extra sensors, such as inertial sensors. In this way, the integration of GNSS and INS (Inertial Navigation Systems) should detect a GNSS signal fault and provide redundant measurement data from the sensors, increasing the reliability of the obtained position. Currently, a plethora of different GNSS receiver systems is present in the transport sector[114], especially with the adoption of high-volume devices in the mass market. Most of them are receivers with built-in GNSS and IMU sensors and implement their own sensor fusion algorithms to provide an accurate and reliable position. On the other hand, a stand-alone wireless sensor node that can satisfy the stringent positioning requirements in the severe propagation environment typical in railway applications still presents a challenge and such devices are not yet available on the market.

A.2 Towards the Future Generation of Railway Localization and Signaling Exploiting sub-meter RTK GNSS

Motivated by all these reasons, in the paper "Towards the Future Generation of Railway Localization and Signaling Exploiting sub-meter RTK GNSS" [115] the design and the implementation of a multi-sensor wireless node that aims at establishing a proof of concept for localizing trains with centimetre accuracy fusing GNSS with RTK and inertial sensors has been presented. The sensor node has been developed and evaluated both in the laboratory and in the city of Zürich both with static and dynamic test.

A.2.1 State of the Art and Background

A.2.1.1 Related work

In order to meet the stringent requirements in terms of integrity, availability, and reliability of the GNSS solution for railway applications, some important requirements must be satisfied, such as the simultaneous reception of signals from different constellations on different frequencies [114], multi-band RTK with fast convergence times, and a high update rate for the dynamic scenario [116]. Indeed, with RTK and Network-based RTK (N-RTK), GNSS can provide localization results with centimetre-level accuracy in clear sky conditions [117]. However, as previously mentioned, the performance of most GNSS receivers degrades as the hostility of the environment increases, particularly in terms of LoS and tracked satellites' geometric

availability [118]. Recent literature has presented numerous solutions to address various aspects of the degradation caused by local phenomena, by detection and mitigation techniques [119]–[121]. In particular, the performance of multipath and/or NLoS mitigation techniques is largely addressed [118], [122], [123]. An example is provided in [124], where the authors analyzed raw GNSS data, collected from a large number of tests along 2,000 km of several railway lines in the area of the West Japan Railway Company, in Japan, to investigate the large errors, such as the multipath and find countermeasures to reduce the horizontal error on the positioning outcomes.

A recent subject of study has been Signal in Space (SiS) integrity monitoring [125], [126]. In [127], the authors analyze the deviation of the expected and observed pseudorange of the reference station in the RTK and N-RTK algorithms, to increase the reliability of these techniques in railway applications. In addition, error reduction is also considered in the context of a hybridized solution with other positioning sensors, in particular with IMUs [128], [129]. Thanks to their complementary properties, GNSS and INS can be integrated so to improve the accuracy and robustness of localization [129], [130]. The fusion of GNSS and INS can achieve measurement errors of less than 0.6 cm and 1.1 cm in the horizontal and vertical directions respectively, when GNSS reception is uninterrupted along the railway line [129]. Nevertheless, establishing the accuracy of the localization solution provided by an on-board GNSS-based system remains a challenging issue due to the environment in which trains operate. Therefore, the positioning performance in the railway sector must be verified in different operating conditions exploiting commercial of the shelf components.

The suitability of GNSS systems for railway applications has also been investigated and evaluated through field tests and proofs of concept as part of numerous funded projects. SATLOC-2¹, ERSAT-EAV², 3InSat³, and GaLoROI⁴ developed and validated the use of GNSS in train localization at signalling system level and showed that the use of complementary positioning techniques is essential. The objective of X2RAIL-4⁵ and X2RAIL-5⁶ is to enhance performance at a railway system level by introducing new functionalities that aim to evolve future signaling and automation concepts by relying on the use of GNSS. The primary goal of RHINOS⁷ was to increase the use of European GNSS to support the train localization functions for train control. Thus, inspired by previous works on the combination of GNSS with RTK and the use of INS, this work presents the design and implementation of a wireless sensor node for high-precision localization of trains on the railway. In particular, the experimental evaluations are presented in terms of functionality and accuracy

¹<https://uic.org/projects/article/satloc-2>

²www.ersat-eav.eu

³<https://business.esa.int/projects/3insat>

⁴www.galoroi.eu

⁵https://projects.shift2rail.org/s2r_ip2_n.aspx?p=X2RAIL-4

⁶https://projects.shift2rail.org/s2r_ip2_n.aspx?p=X2RAIL-5

⁷<https://cordis.europa.eu/project/id/687399>

in the city of Zürich exploiting two different commercial GNSS modules, while the wireless sensor node is ready for future experiments on trains.

A.2.1.2 Background: N-RTK

RTK aims at providing Positioning, velocity, and Timing (PVT) solutions with centimetre-level accuracy by using correction data from a terrestrial base station located at a known position. The base station periodically transmits to the users its location, the pseudorange, and the phase measurements collected from all the satellite signals that are visible to the base station itself. This technique exploits the high correlation in space and in time of the above-mentioned GNSS error sources. In particular, RTK uses the correction data to cancel out or at least mitigate the clock, the satellite orbit, and the atmosphere-related errors by differentiating the measurements made for the different satellites between the user and the base station. Therefore, since the base station estimates the errors by leveraging its known position and provides this information in the form of corrections to the users, it is expected that the user's position accuracy will be improved as a result. However, the accuracy improvement is heavily dependent on the distance between the receiver and the base station. In order to achieve centimetre-level performance, the maximum distance is around 10-15 kilometers [131].

To overcome this limitation, the concept of N-RTK using Networked Transport of RTCM (Radio Technical Commission for Maritime Services) via Internet Protocol (NTRIP) has been developed based on the use of multiple reference base stations. Here, the correction information is sent as a RTCM message via NTRIP. The N-RTK solutions are of three types, named *Stand-alone* mode, *Fix* mode, and *Float* mode [132]. *Stand-alone* mode works without corrections, *i.e.*, it delivers the positioning solution with the same accuracy as the stand-alone GNSS. *Fix* mode reports the ambiguity of the carrier phase as an integer value, meaning that the position has been determined, while *Float* mode provides the carrier phase as a floating number. Therefore, in *Float* mode, the RTK is active but is unable to determine the exact position. Basically, the N-RTK system continuously receives the RTCM message sources (*i.e.*, NTRIP Source) from the different Base Stations, which are sent to the Ntrip-Caster through the NtripServer. Finally, the user (NTRIP client) receives the RTCM messages via Internet in the form of reference receiver measurements. N-RTK relies on different techniques to calculate and transmit this information, including the Virtual Reference Station (VRS) method [133]. This method aims to further reduce the baseline distance between the receiver and the station by generating a VRS close to the user position, in order to efficiently eliminate the spatially and temporally correlated errors.

To calculate the position of the VRS, the user must compute its raw position and transmits this information to the computation server using a standard National Marine Electronics Association (NMEA) format. The computation center generates in real time a virtual reference station at or close to the initial user position. This is

done by geometrically translating the pseudorange and carrier phase data from the nearest reference station to the virtual location and then adding the interpolated errors from the network error models. Finally, this generated VRS data is then sent to the user through a wireless connection, via NTRIP. Therefore, N-RTK requires a bi-directional link between the user and the server, which increases the required bandwidth and limits the number of simultaneously connected users (w.r.t the RTK). In addition, as the user moves away from the VRS location, the resulting solution may have the same accuracy as the baseline RTK, thus no longer at the centimetre level. Therefore, a new VRS needs to be generated, whose location is not updated in real-time.

A.2.1.3 Background: Dead reckoning

The Dead Reckoning (DRK) localization technique relies on inertial measurements and computes the current position of the vehicle based on the vehicle's latest location estimation. Given the 2-dimensional position of the user at the previous step (x_k, y_k) , its position at the current step can be computed as:

$$\begin{aligned} x_{k+1} &= x_k + \Delta a_k \cos(\epsilon_k) + n_k \\ y_{k+1} &= y_k + \Delta a_k \sin(\epsilon_k) + n_k \end{aligned} \tag{A.4}$$

where ϵ_k is the real-time heading data collected from the sensor, n_k is the zero mean Gaussian noise, and a_k is the distance travelled by the vehicle, obtained by multiplying the speed by the time interval. It is worth observing that the DRK position is affected by the initial position error, errors in the previous time step, as well as mounting and attitude errors.

A.2.2 HW and NTRIP network selection

A.2.2.1 Wireless sensor node

Fig. A.2d shows the designed custom board. It consists of: i) an STM32L4 Microcontroller Unit (MCU), ii) an ASM330LHH IMU, a microSD Card slot, iii) a GNSS module, and a iv) NB-IoT SARA-R410M module. The board collects data from the IMU, which generates two independent data flows from the accelerometer and gyroscope, with a sampling rate up to 3.3 kHz, and from the GNSS module. The board is able to receive and process both common and standard proprietary GNSS messages (NMEA and UBX). The data is stored on the SD card via an SPI interface and appended to a csv file along with an individual time stamp. The maximum file size is set to 1 MB. When the latter is reached, a new file is created and the data logging continues, ensuring the minimum data loss in case of power loss or an unexpected software crash. The NB-IoT module is in charge of sending the positioning information to the NTRIP server and receiving the RTCM corrections. This design offers great flexibility, as it allows each individual module to be easily changed for testing

purposes. During the data logging, the sensor node without the SARA-R4 module has an average power consumption equal to 673 mW. On the other hand, with the cellular interface enabled, the average power is up to 1 W.

A.2.2.2 GNSS module selection

As previously described, a multi-frequency receiver is necessary to correct the source of GNSS errors, while a multi-constellation receiver makes more robust the PVT solution against local effects, such as spoofing. Therefore, in order to achieve sub-meter accuracy, the ZED platform from U-blox has been selected, since it integrates multi-band GNSS, RTK, and DRK technologies. Indeed, ZED modules concurrently use GNSS signals from all four GNSS constellations (GPS, GLONASS, Galileo, and BeiDou) from multiple frequency bands. In addition, it supports the RTCM corrections from virtual reference stations and offers a range of correction services. More in detail, the ZED-F9P and ZED-F9R modules have been selected. The P variant can be used either as a rover or base station, meaning it is possible to build an NTRIP network for testing purposes in a region where one is not already present. The R variant cannot be used as a base station, therefore it can only receive RTCM corrections through the NTRIP network. To achieve centimetre-level accuracy, this version uses high-precision sensor fusion (HPS) based on a 3D IMU included within the module. The algorithm processes both the sensor data and the GNSS signals to provide highly accurate and continuous positioning solutions. Thanks to DRK, it is possible to localize the train also in case of complete or partial GNSS signal obstruction.

A.2.2.3 NTRIP network

The Swipos-GIS/GEO from Swiss Positioning Service swipos has been chosen as the RTK network. It provides RTK correction data from 31 AGNES stations (i.e., the stations of the Automated GNSS Network for Switzerland), which receive GNSS signals from all four satellite systems. The Swipos-GIS/GEO uses the VRS method to calculate the corrections data. To receive this information from that network, the receiver must be able to output its raw position in NMEA format, receive correction data in the RTCM 3.2 Multi Signal Message 4 (MSM4), and support the NTRIP protocol.

A.2.3 Experimental Results

In the following, the main results from the evaluation of the position accuracy of the sensor board with the two GNSS modules are described. In particular, for the static scenario, the mean (μ) and standard deviation (σ) of the positioning error with respect to the location of the reference point are computed. For the dynamic scenario, both quantitative and qualitative results are reported. The mean and the standard deviation of the error between the vehicle position and center of the lane are computed and a trace of the vehicle on the map of the zone under test is provided.

TABLE A.1: Localization accuracy for static tests w and w/o RTK

Metrics		Rural	Suburban	Alley	Urban
μ (m)	RTK off	2.465	1.0675	2.475	4.897
	RTK on	0.019	0.019	5.0349	8.1719
σ (m)	RTK off	0.237	0.354	0.595	1.353
	RTK on	0.008	0.005	0.033	0.022

During the dynamic test, the GNSS antenna was installed on the roof of the car. Both static and dynamic tests were performed in different propagation environments, *i.e.*, rural, suburban, and urban, to quantify the accuracy under LoS and NLoS conditions both when RTK is enabled and when it is not. Finally, we evaluate the high-precision sensor fusion of the ZED-F9R module by performing a car test in environments where the GNSS signal is completely or partially obstructed. Both GNSS modules operate as RTK receivers and are connected via USB to a laptop. The built-in NTRIP client, within the U-center application (on the laptop), receives the RTK correction messages through a cellular connection and sends them to the GNSS modules via USB. The raw observations in the different environments and scenarios are recorded at 1 Hz, starting 5 minutes after the GNSS module is switched on.

A.2.3.1 Static Tests

In this scenario, the Geodetic points from the Federal office of topography swisstopo and the Stadt Zürich - Tiefbau - und Entsorgungsdepartement - Geomatik + Vermessung have been used as ground truth. These points are bolts on rocks in the terrain which have an average accuracy of ± 3 cm (shown in Fig. A.1). In this phase, only the measurements done with the ZED-F9P module, whose performance in terms of mean and variance of the position error is described in Table A.1, have been considered. The RTK corrections reduce the variance of the position error at least by a factor of 10 compared to the measurements without RTK, especially in the rural and suburban environments. For the latter, the mean of the error is drastically reduced, since most of the GNSS source errors are spatially and temporarily correlated, and therefore easy to remove with the aid of RTK. The opposite behavior is visible in the narrow alley and in the urban environment, where the multipath is the predominant component of the error. Looking at Table A.1, it is clear that the largest error is obtained when RTK is enabled. This is due to the fact that multipath distorts the phase of the carrier and therefore degrades the accuracy of the RTK measurements. Furthermore, since the multipath errors are uncorrelated from base station to base station, the difference in measurement error caused by multipath between two of them has a variance given by the sum of the multipath error variance of each base station alone [134].

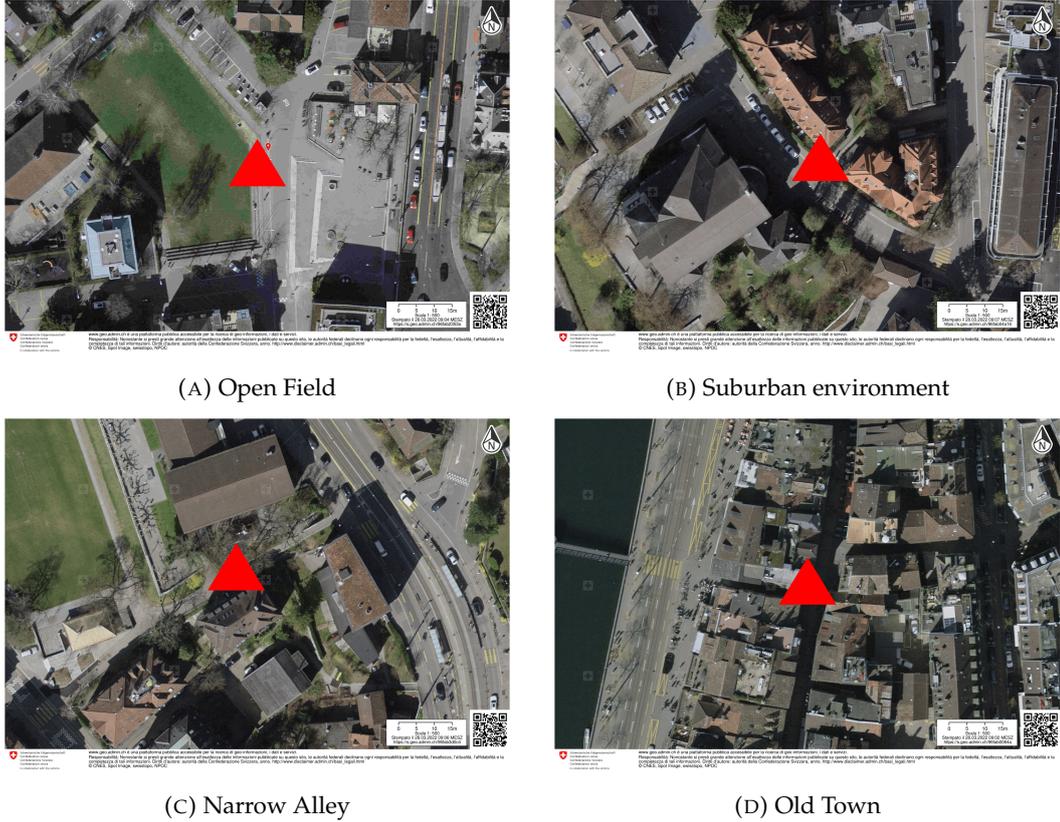


FIGURE A.1: Reference locations for GNSS static tests rural, suburban and urban environments. [115].

TABLE A.2: Localization accuracy for dynamic tests w and w/o RTK.

Metrics		Highway	Urban	Suburban
μ (m)	RTK off	0.97	0.93	1.87
	RTK on	0.04	0.47	1.21
σ (m)	RTK off	0.56	0.68	1.07
	RTK on	0.18	0.12	0.97

A.2.3.2 Dynamic Tests

The dynamic tests were performed along urban and suburban roads, and on a highway, with the aim to assess the behavior of the ZED-F9P module with and without the RTK. Table A.3 shows a comparison of the error performance obtained in both cases, with an average number of satellites in view equal to 26. Along the highway, when the RTK is enabled, it is possible to observe that a centimeter position accuracy is achieved (0.04 m) and remains constant for the duration of the test, confirmed by the low variance. This behavior is verified also in Fig. A.2a, where the position of the vehicle is always inside the right lane of the highway (red line). Therefore, the RTK is able to fix the phase ambiguity for the whole duration of the test. In contrast, the use of GNSS stand-alone, without RTK provides an average error equal to 0.97 m, confirming that in open environments, such as the highway, the error budget is dominated by spatially correlated errors that are easy to mitigate by leveraging the

TABLE A.3: Localization accuracy for dynamic tests w and w/o RTK and DRK.

Metrics		Tunnel long (2176 m)	Tunnel short (1347 m)	Suburban	Highway
μ (m)	RTK + DRK	-	-	0.17	0.17
	DRK	13.0	2.35	0.13	-
σ (m)	RTK + DRK	-	-	0.4	0.32
	DRK	14.6	1.07	0.3	-



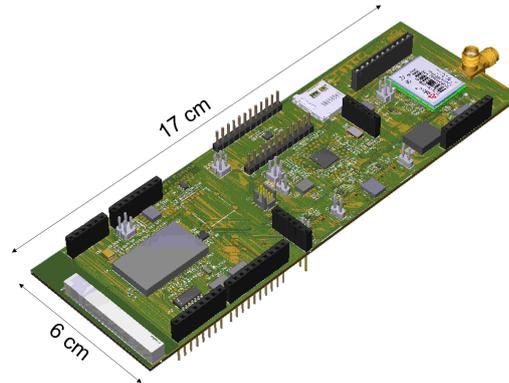
(A) Highway



(B) Urban road

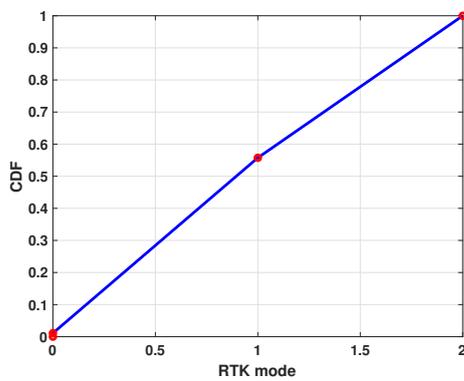


(C) Suburban

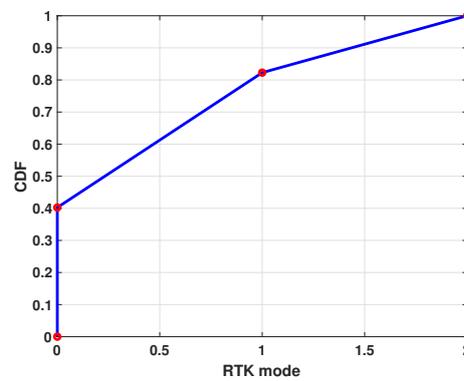


(D) Custom sensor node

FIGURE A.2: References to dynamic test environments including the satellite map and the tracked path (in red). (d) The custom sensor node used for reported tests [115].



(A) Urban scenario



(B) Suburban scenario

FIGURE A.3: CDF of RTK modes (0 = Stand-alone, 1 = Float, 2 = Fix) in dynamic scenario [115].

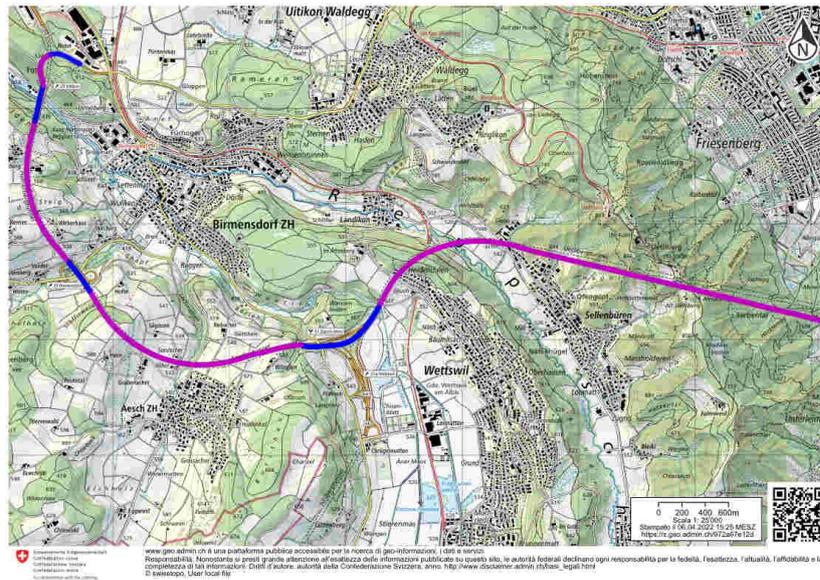


FIGURE A.4: References to dynamic test environments including the satellite map and the tracked path with RTK + DRK (blue line) and only DRK (purple line) [115].

RTK.

The same behavior is shown in an urban environment, where the RTK is active all the time. This results in an accurate vehicle position tracking, as it appears in Fig. A.2b, where the paths taken in both directions are clearly distinguishable. Since in this scenario, the measurement accuracy depends on the RTK, the rates of RTK carrier-phase fixing should be analyzed carefully. Fig. A.3a reports the CDF of the RTK modes. As shown, the RTK operates in *Float* (1) and *Fix* (2) modes for the entire duration of the test. The *Float* mode is due to the presence of the buildings, which increases the NLoS probability. Therefore, the ambiguity term cannot be properly solved due to the poor quality of the GNSS observations, where multipath is assumed to have the largest impact. Despite the *Float* mode, it is still possible to achieve an accuracy of less than one meter (with an average error equal to 0.47 m). The speed of the vehicle (100 km/h on the highway, 50 km/h on the urban road) does not appear to be an impediment in the VRS creation and update. The opposite situation occurs along the suburban road (Fig. A.2c). In that case, the tree foliage often overshadows the user, meaning the received signal is composed of the attenuated LoS signal component and a signal generated by the scattering of the electromagnetic waves by the leaves. As consequence, the module works in *stand-alone mode* (indicated as 0 in Fig. A.3b) for 40% of the test duration. This is reflected in the quality of the measurements done with and without RTK, since in both cases the mean error is higher than one meter, as reported in Table A.2.

A.2.3.3 Dynamic Tests with RTK and IMU

The last tests were performed on the highway with a constant speed of 100 km/h and on the suburban road (with a maximum speed of 50 km/h) with the aim to evaluate the inertial sensor fusion of the ZED-F9R module. As observed in the previous test, the RTK alone was unable to deliver a centimetre-level accuracy along the suburban road. Thus, a more accurate position is provided by the high-precision sensor fusion algorithm, which combines the GNSS measurements with the internal 6-axis IMU of the module. The IMU is capable of delivering a meaningful improvement to the fixing of the ambiguity terms, by supplying a significantly smaller average error compared to the error obtained using only RTK, *i.e.*, 0.17 m and 1.21 m, respectively. Looking at the performance of the stand-alone DRK (Table A.3) along the same path (986 meters long), it seems that centimetre-level accuracy is achieved. However, INS cannot provide long-term navigation by itself due to the fact that the error of IMU measurements will accumulate. This is visible in Fig. A.4, where the trajectory of the vehicle through multiple tunnels is depicted. Along the highway, and therefore, before and after each tunnel, the receiver's position solution relies on the RTK+IMU measurements (blue lines), providing an average position error equal to 17 cm. At the end of each tunnel, when the GNSS signal is available, it is able to deliver a significant improvement to the position performance obtained with the DRK algorithm. Inside the tunnel, the module is only able to compute the position based on the IMU sensor measurements (purple line in Fig. A.4), due to the long GNSS signal outage. In the long tunnel (2.176 km) the DRK produces an error that will accumulate until the module is able to receive the GNSS signal, leading to an average estimation error equal to 13 m. Thus, the IMU embedded in the ZED-F9R module provides centimetre-level accuracy along relatively short path and with a speed around 50 km/h (sub-urban environment). Whereas, by doubling the speed, performance decreases as the distance travelled increases. This behavior could be explained by the fact that the IMU sensor is not rigidly connected to the car frame and, therefore, cannot capture all of the vehicle's movements.

A.2.4 Discussion

As expected, under clear sky conditions, centimetre accuracy is achievable, whereas, in urban environments, multipath has a notable impact on positioning accuracy, limiting the precision achievable with the RTK. The dynamic scenarios follow the same trend. It is worth emphasizing that the high speed (up to 100 km/h) poses no problem in generating and updating the VRS, allowing an accuracy of less than one meter. As revealed by the analysis of the success rates of RTK carrier-phase fixing, *Fix* mode is hard to maintain, but sub-meter accuracy is observed in *Float* mode too. In tunnels and along roads where the GNSS signal is obstructed, the DRK is a necessity to localize the vehicle. Therefore, to enable safety-critical operations, additional sensors and a sensor-fusion algorithm, which receives as input the measurements

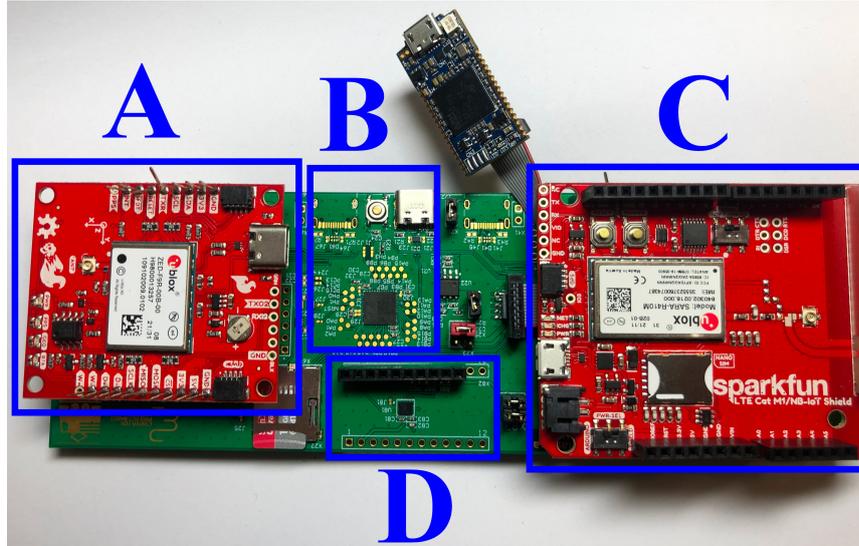


FIGURE A.5: Overview of the custom localization sensor node where: (A) ZED-F9P GNSS module, (B) STM32L452CEU MCU, (C) SARA-R410M cellular module, (D) ASM330LHH IMU [135].

of all the sensors and the GNSS signal, will be required to compensate for the lack of GNSS availability and reliability in tunnels and in poor sky view. These results indicate that it is possible to satisfy the accuracy requirement for some railway applications.

A.3 Towards the Future Generation of Railway Localization and Signaling exploiting sub-meter RTK GNSS: extension

An improvement of the previous work is reported in [135], where, the sensor node has been improved and its GNSS and RTK performance have been tested on-board a passenger train on a real railway line in Emilia-Romagna, Italy.

A.3.1 Sensor node

The custom low-power sensor node consists of an STM32L4 (STM32L452CEU) Microcontroller Unit with an ARM-Cortex-M4F core, an ASM330LHH IMU, a microSD card slot for local data storage, a GNSS module, and a NB-IoT SARA-R410M module. The node is depicted in Fig. A.5. As before, the MCU collects data from the IMU, which in turn generates two data streams from the accelerometer and gyroscope at 1.667 kHz, synchronized with sub-millisecond precision. In parallel, UBX or NMEA messages from the GNSS module and RTK corrections from a remote reference station are collected and distributed at 1 Hz. The board is able to parse the NMEA and UBX and report detailed information in real-time via a serial interface. The collected data is then stored on a SD card using the CSV file format, with each line containing

a timestamp obtained from the internal Real-Time Clock (RTC), a tag indicating the data type (U for UBX, N for NMEA and I for IMU data), followed by the payload of UBX, NMEA, or IMU data.

The NB-IoT module is used to receive the RTK correction packets from two, easily configurable remote sources. The sensor node can establish a connection to a commercial public NTRIP server, usually managed through a user-password authentication procedure. Alternatively, a direct connection to a raw TCP socket can be used instead, to receive corrections from a custom base station consisting of another GNSS module. Support for VRS in the NTRIP case was not implemented, as it was not needed for the tests, but can be easily added if required. The raw RTK packets from the SARA module are intercepted by the MCU for validation purposes and then forwarded to the GNSS module. The validation consists of a watchdog timer which restarts the SARA module whenever the last RTK packet was received more than 10 s ago, thus resetting the cellular connection in case of connection loss. Using a standard USB Type-C connector a simple user interface is exposed, allowing the sensor operator to receive real-time information about the status of the sensors, *e.g.*, number of connected satellites, status on the RTK connection and other diagnostics. In addition, the same interface can be used to manually incorporate timestamped and synchronized comments and labels in the final CSV file together with the stored sensor data, as can be seen in the Section dedicated to the dynamic test, where the plots are labelled with the arrival and departure time at each railway station. Additional labels are also stored automatically whenever the SARA watchdog triggers a reconnect of the cellular connection and on the first valid RTCM packet after each loss of connection.

Regarding the GNSS modules, the ZED platforms from u-blox were chosen. The ZED-F9P and ZED-F9R are the two variants selected from the ZED family.

A.3.2 GNSS with a Shielded Antenna

In order to quantify the possible effect of the train cab on the sky view of the GNSS, a static test was performed using a metal plate to shield a specific portion of the sky view.

Method

The test was performed using a ZED-F9P module and a TOPGNSS TOP106 L1/L2 multi-band antenna, shielded with a grounded metal plate. This plate can rotate along its lower edge, positioned near the antenna mounting point. In this way, it is possible to fix the shield at selected angles with respect to the ground. The shield can be set from a vertical position (90°), with minimum effect on the antenna, up to a horizontal position (0°), with maximum shielding effect. The setup is shown in Fig. A.6a. By diminishing the angle, the shield effectively provides coverage over the antenna, resulting in a decrease in the quantity of satellites that in LoS with the

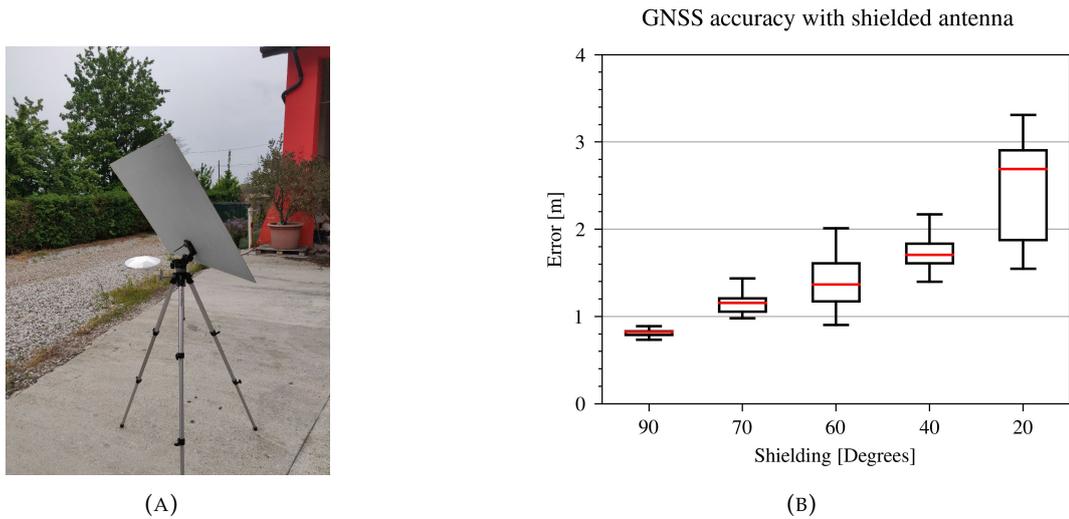


FIGURE A.6: Static antenna shielding test setup (a) and results as box plot showing static measurement accuracy without RTK at different angles (b) [135].

antenna. Consequently, this reduction in visibility negatively impacts the overall performance. The test was performed with 90, 70, 60, 40, and 20 degrees of shielding angle without RTK.

Results

The positioning accuracy is illustrated in Fig. A.6b, where the results show that decreasing the angle of shielding decreases the positioning accuracy and increases the variance of the measurements. For example, it is possible to consider the two extreme cases of (90°) and (20°). In the former case, the median error is 0.8 m with a standard deviation of 4 cm, while in the latter, the median error is 3.7 m with a standard deviation of 0.55 m. Indeed, progressively shielding the antenna decreases the number of GNSS satellites in direct visibility, making the multi-path contribution predominant. The GNSS module receives signals from a constant number of approximately 30 satellites throughout the measurements, but shielding the antenna from the top allows the signal to reach the GNSS receiver only through ground reflections, thus decreasing the measurement accuracy.

A.3.3 Dynamic Tests on a Train

The on-train field tests were performed on the branch line between Formigine and Modena in Emilia-Romagna, Italy, over two days in October 2022. On this line, ETR103 trains manufactured by Alstom are operated by Trenitalia TPER. They have a maximum speed of 160 km/h, although line speed on the line tested does not exceed 80 km/h and is generally no more than 60 km/h. The line is electrified using 3 kV DC overhead line. It passes through a variety of GNSS and cellular environments, with sections of open sky view and sections with many buildings directly

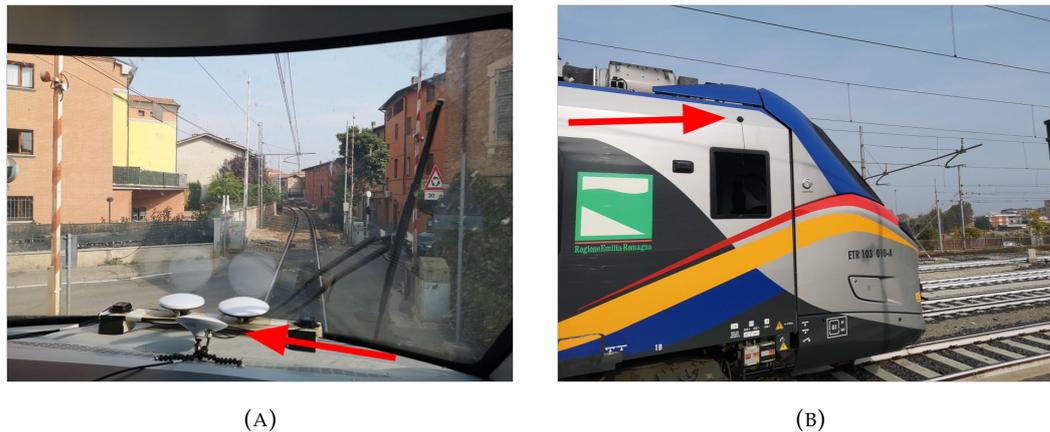


FIGURE A.7: Dynamic tests: Sensor node antenna measurement setup for Sensor Node 1 (a) and Sensor Node 2 (b) [135].

on both sides of the line which heavily limit the sky view. Cellular coverage is intermittent throughout. The line is mostly single-track, with double-track passing loops at Modena Piazza Manzoni and Formigine stations, as well as multiple tracks in Modena station itself.

Two identical sensor nodes were used, with two different antennas. The antennas were placed in the rear cab of the train and moved to the opposite cab at the end of each trip. The first sensor node was connected to a TOPGNSS TOP106 antenna L1/L2 multi-band antenna placed underneath the rear windscreen of the train, on a flexible tripod fixed using adhesive putty and facing upwards. The setup can be seen in Fig. A.7a (note that the other antennas in the background are from an unrelated experiment). The second sensor node used a smaller magnetic mount ANN-MB-00 L1/L2 dual band antenna which was affixed to the outside of the train above the window, approximately 2 m above the floor height of the train and approximately 1.4 m to the left of the left running rail, using adhesive putty, since the aluminum body shell of the train is not magnetic. This is shown in Fig. A.7b. For the RTK corrections, a custom base station was set up by Sadel in Modena, Italy exposing the RTCM data produced by a ZED F9P module configured as an RTK base station via a raw TCP socket. The NTRIP protocol was not used for these tests.

A useful metric to evaluate the performance of the GNSS localization is to calculate the error between the GNSS point recorded by the ZED-F9P module and the actual location of the train and to compare this to the accuracy reported in the $hAcc$ field of the `UBX-NAV-PVT` message by the GNSS module. However, on a moving train, this is challenging, particularly since the maps of the line available are of limited precision. Nevertheless, this was achieved as follows. A ground truth of the line was extracted from OpenStreetMap in GeoJSON format. Both the GeoJSON file and the data extracted from the ZED module are projected from the global WGS84 (EPSG: 4326) Coordinate Reference System to the locally flat UTM 32 projection (EPSG:25832) using the `PyProj` library [136]. The minimum distance between each GNSS point and the ground truth is then calculated using the `shapely` library [137].

While useful, this approach still has several limitations. The sensor node is not located precisely above the track axis, but rather has a certain offset, which is reported as an error. Moreover, this metric calculates the distances from the position reported by the GNSS module to the track axis, but not to the actual position of the train, particularly in the longitudinal direction along the track. Nevertheless, this metric is still useful for assessing localization performance.

A.3.3.1 Results

Fig. A.8 shows the number of satellites detected by each sensor node during the same trip between Formigine and Modena. Arrival and departure times at each station, are indicated in green and red respectively, while the connectivity of the cellular module and RTCM correction messages is indicated in blue. The approximate location of the tunnel between Modena Piazza Manzoni and Modena stations is indicated in grey.

It is clear that both sensor nodes receive signals from a large number of satellites. However, while the number of satellites is relatively consistent throughout the line for Sensor Node 2 (with the small antenna outside), ranging from 26 to 32, the number of satellites received by Sensor Node 1 is much more variable, with a minimum of 18 and a maximum of 32. The exception to this is inside the tunnel, where no more valid GNSS points are received by both sensor nodes shortly after the train enters.

The effect of the tunnel can also be seen clearly in Fig. A.9. The accuracy reported in the `hAcc` field of the `UBX-NAV-PVT` message is shown in blue, while the distance to the ground truth is shown in orange. Note that while accuracy is reported by the GNSS module, the `invalidLlh` flag of the `UBX-NAV-PVT` message is set throughout the tunnel, meaning the reported data is not considered to be valid.

The relationship between the reported accuracy and the actual error to the ground truth is shown in more detail for both sensor nodes between Formigine and Modena Piazza Manzoni (before the tunnel) in Fig. A.10. Moreover, statistics are given in Table A.4. It is clear that there is not necessarily a direct relationship between the accuracy reported by the GNSS module (which is 1 cm almost throughout), and the actual accuracy as measured relative to the ground truth, which is significantly worse than reported by the GNSS module. Nevertheless, the actual accuracy is within 2 m for most of the line, apart from shortly after departure from Modena Fornaci station. Although the raw numbers suggest otherwise, Sensor Node 2 appears to provide a more accurate position, once its offset to the track center-line (of approximately 2 m) is taken into account. Sensor Node 1 is located almost directly above the track center-line and therefore is significantly less accurate in comparison. This is confirmed by the standard deviation σ which is noticeably smaller for Sensor Node 2.

On both error plots in Fig. A.10, there are several peaks shortly after departing Modena Fornaci station. Two satellite map extracts of this section can be seen in Fig. A.11, where the ground truth is indicated in yellow, Sensor Node 1 in blue and Sensor Node 2 in orange. These peaks appear to have two distinct causes. First,

TABLE A.4: Position Error Summary Statistics

	Metrics	Reported Error	Actual Error
μ (m)	Sensor Node 1	0.163983	1.075075
	Sensor Node 2	0.128677	2.065174
σ (m)	Sensor Node 1	0.352838	1.324148
	Sensor Node 2	0.380484	1.123460

directly after leaving Modena Fornaci, the line passes through an underpass under a highway, as shown in Fig. A.11a. This causes the reported accuracy to briefly decrease. This is also reflected in Fig. A.8, where there is a decrease in the number of visible satellites, particularly for Sensor Node 1. It can be seen that the reported positions of the sensor nodes diverge from the path of the line, especially for Sensor Node 1. For Sensor Node 1, this reported error peak is synchronous with a peak in the actual error relative to the ground truth, as would be expected. In contrast, for Sensor Node 2, no such actual error peak is visible. The map in Fig. A.11a also does not indicate a deviation for Sensor Node 2 at this stage.

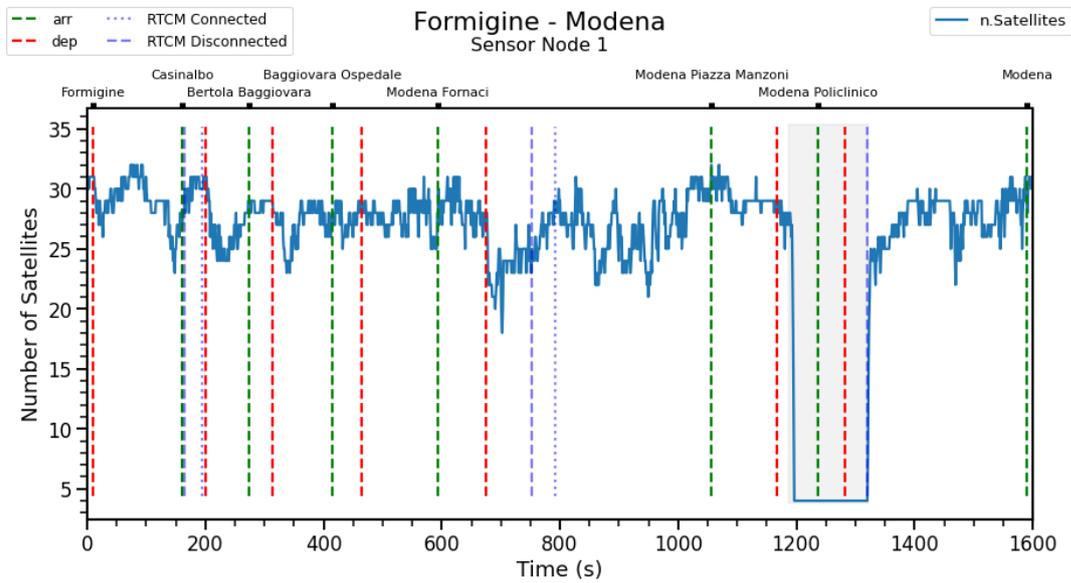
A significantly larger peak in the actual error relative to the ground truth occurs shortly afterwards for both sensor nodes but does not have a corresponding reported error peak. A closer look at the map, as shown in Fig. A.11b suggests that rather than being caused by an incorrectly determined position, an error in the ground truth itself appears more likely. The orthophoto background indicates that the ground truth passes through several trees, while the sensor node measurements pass through an adjacent clearer area while following a far smoother curved path. Due to the relatively poor quality of the satellite orthophoto, it is difficult to fully identify the cause of the error, but given that neither the ground truth nor the satellite image offers sufficient accuracy guarantees. and that the exact origin of this data is unknown, it is not unreasonable to conclude that the GNSS measurements from the sensor nodes may in fact be significantly more accurate than the ground truth and error plot would suggest. A combination of factors is also likely.

For Sensor Node 2, another, flatter peak in the reported error occurs at around $t = 750$ s. This can likely be explained by this sensor node's intermittent connection to the RTCM server at this stage, as can be seen in Fig. A.8b. No such peak is visible for Sensor Node 1, which also does not lose connectivity during this time period.

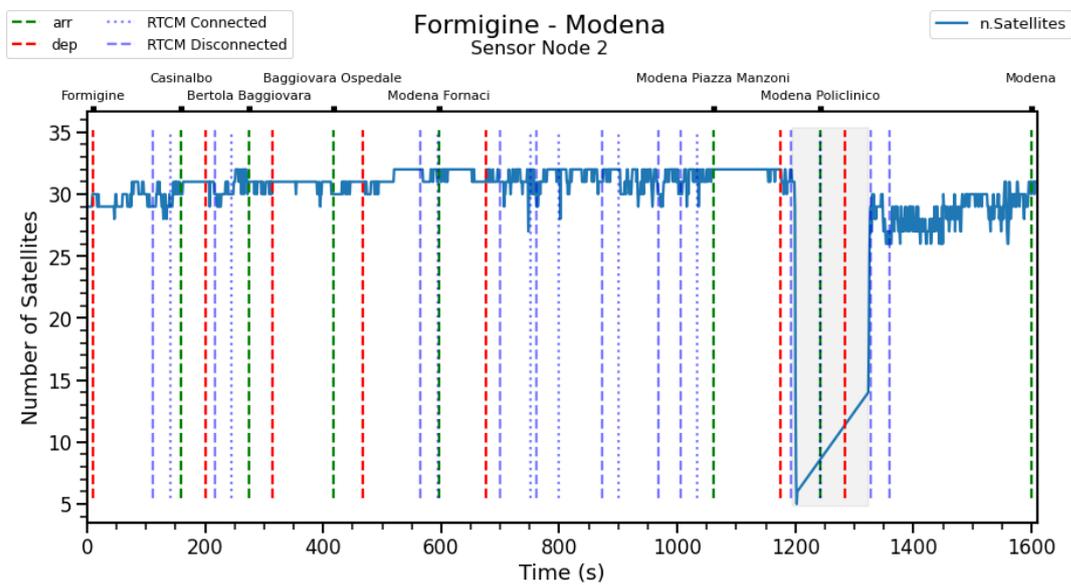
A.3.4 Discussion and Conclusion

Dynamic tests on a railway line between Formigine and Modena in Emilia-Romagna, Italy in which GNSS, RTK, and IMU data was collected in several measurement runs have been performed. In this work, the horizontal accuracy and satellite coverage of the GNSS data with RTK is evaluated, and compared to a ground truth derived from publicly available data. The self-reported accuracy of the GNSS module with RTK enabled is within 16 cm on average, when not obstructed by long over-bridges or tunnels. The actual error, when compared to the track center-line, is much larger,

but still within 1.3 m on average when antenna position is taken into account. However, some of this error is likely to be caused by the limited accuracy of the available ground truth track center-line. The data sets for the dynamic tests will be open and freely available on GitHub for future work by the scientific community. The tests confirm the need for a sensor fusion approach when GNSS is not available or has poor accuracy, but show that RTK is a promising approach for the railway environment. The development of a novel sensor fusion algorithm to integrate the collected IMU data with GNSS and RTK will form the basis for our future work. Further investigation is also needed into how a reliable ground truth can be collected for the railway line so that real-time localization performance can be quantified more easily. This would potentially also enable track-matching approaches to be used.



(A)



(B)

FIGURE A.8: Number of Satellites received by Sensor Node 1 (a) and Sensor Node 2 (b) on a measurement run between Formigine and Modena [135].

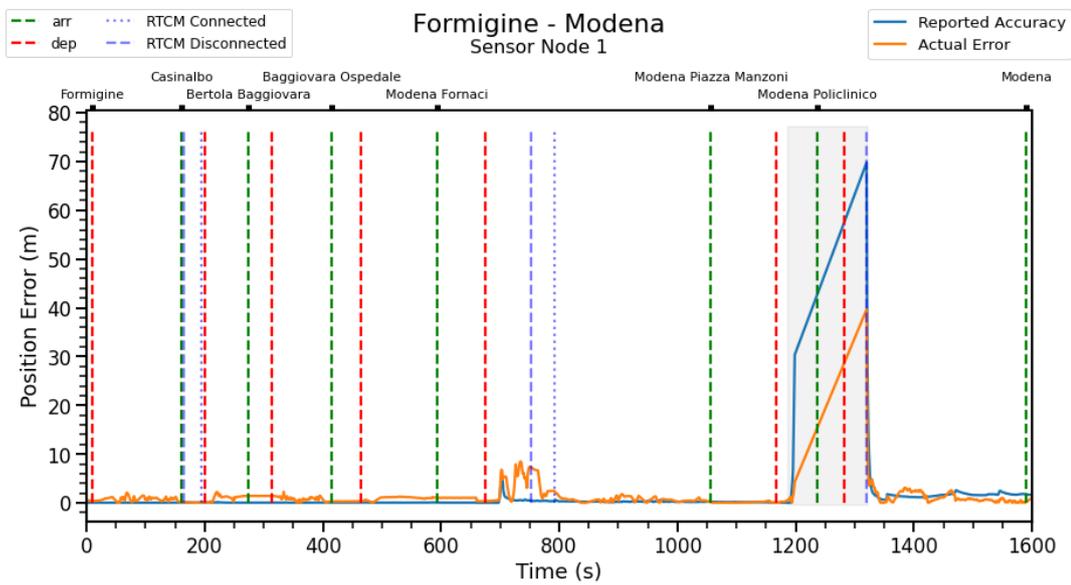
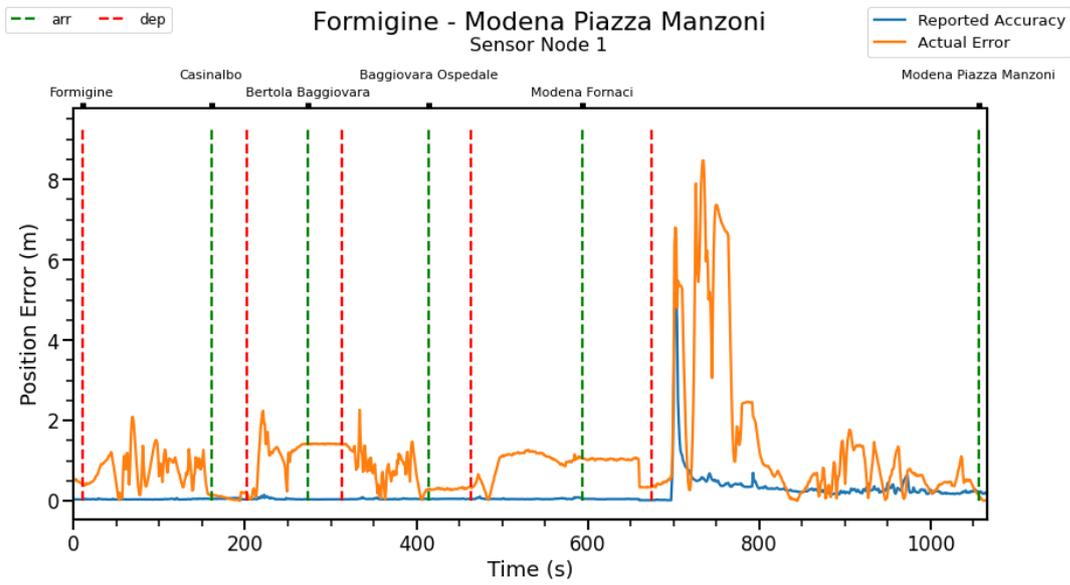
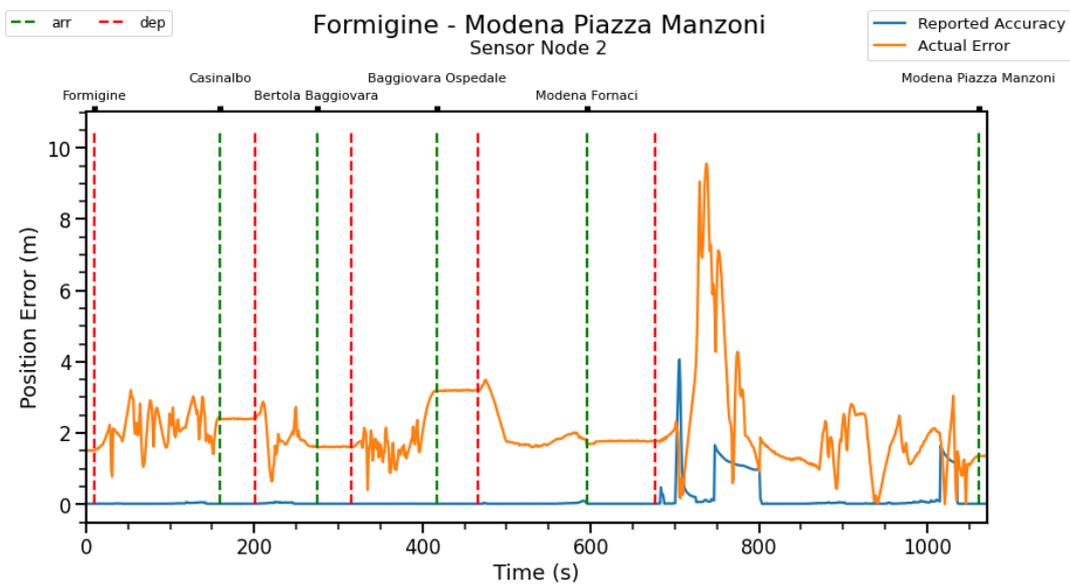


FIGURE A.9: Position Error reported by the GNSS module and the actual error relative to the ground truth for the entire measurement run of Sensor Node 1 between Formigine and Modena, including cellular connectivity status [135].



(A)



(B)

FIGURE A.10: Position Error reported by the GNSS module and the actual error relative to the ground truth, without the tunnel section, for Sensor Node 1 (a) and Sensor Node 2 (b) [135].



(A)



(B)

FIGURE A.11: Maps showing position error near Modena Fornaci in areas of interest under a highway overpass (a) and a region with potentially inaccurate ground truth (b). The ground truth is indicated in yellow, Sensor Node 1 in blue and Sensor Node 2 in orange [135].

Bibliography

- [1] *NB-IoT statistics*. [Online]. Available: <https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/>.
- [2] M. R. Palattella, M. Dohler, A. Grieco, *et al.*, "Internet of things in the 5G era: Enablers, architecture, and business models," *IEEE journal on selected areas in communications*, vol. 34, no. 3, pp. 510–527, 2016, Publisher: IEEE.
- [3] Y.-P. E. Wang, X. Lin, A. Adhikary, *et al.*, "A primer on 3GPP narrowband Internet of Things," *IEEE communications magazine*, vol. 55, no. 3, pp. 117–123, 2017.
- [4] L. Vangelista, A. Zanella, and M. Zorzi, "Long-range IoT technologies: The dawn of LoRa™," in *Future access enablers of ubiquitous and intelligent infrastructures*, tex.organization: Springer, 2015, pp. 51–58.
- [5] *Sigfox technical overview*. [Online]. Available: <https://www.huawei.com/en/news/2020/7/3gpp-itu-imt-2020-5g-standard>.
- [6] E. M. Migabo, K. D. Djouani, and A. M. Kurien, "The narrowband Internet of Things (NB-IoT) resources management performance state of art, challenges, and opportunities," *IEEE access : practical innovations, open solutions*, vol. 8, pp. 97 658–97 675, 2020, Publisher: IEEE.
- [7] 3GPP, "LTE; evolved universal terrestrial radio access (E-UTRA); physical channels and modulation," *3GPP TS 36.211 version 15.2.0 Release 15*, Oct. 2018.
- [8] *NB-IoT deployment guide to basic feature set requirements*. [Online]. Available: <https://www.gsma.com/iot/wp-content/uploads/2019/07/201906-GSMA-NB-IoT-Deployment-Guide-v3.pdf>.
- [9] M. O. Ojo, S. Giordano, G. Procissi, and I. N. Seitanidis, "A review of low-end, middle-end, and high-end iot devices," *IEEE Access*, vol. 6, pp. 70 528–70 554, 2018.
- [10] ITU, "ITU-R P.618 : Propagation data and prediction methods required for the design of Earth-Space telecommunication systems," Tech. Rep., Dec. 2017.
- [11] B. Evans, "6G satellite communications," in *2022 27th asia pacific conference on communications (APCC)*, 2022.
- [12] "ETSI EN 301 790, V1.5.1, Digital Video Broadcasting (DVB); Interaction channel for satellite distribution systems," Tech. Rep., May 2009.

- [13] *OQTEC Technology*. [Online]. Available: <https://www.oqtec.space/>, .
- [14] *Successful launch of tiger-2 satellite aboard spacex transporter-2 mission*. [Online]. Available: <https://www.oqtec.space/news/successful-launch-of-tiger-2-satellite-aboard-spacex-transporter-2-mission>.
- [15] *Sateliot*. [Online]. Available: <https://sateliot.space/>.
- [16] *Sateliot & spanish mod promotes rst nanosatellite constellation for secure iot and data exchange*. [Online]. Available: <https://sateliot.space/sateliot-spanish-mod-promotes-â€¢rst-nanosatellite-constellation-for-secure-iot-and-data-exchange-2/>.
- [17] *GateHouse*. [Online]. Available: <https://gatehouse.com/satcom/satellite-iot-and-m2m-connectivity/5g-satellite-nb-iot/>.
- [18] *Mediatek*. [Online]. Available: <https://www.mediatek.com/>.
- [19] *Mediatek conduct world's rst public test of 5g satellite iot data connection with Inmarsat*. [Online]. Available: <https://corp.mediatek.com/news-events/press-releases/mediatek-conduct-worlds-â€¢rst-public-test-of-5g-satellite-iot-data-connection-with-inmarsat>.
- [20] M. Centenaro, C. E. Costa, F. Granelli, C. Sacchi, and L. Vangelista, "A survey on technologies, standards and open challenges in satellite Iot," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1693–1720, 2021, Publisher: IEEE.
- [21] S.Cioni and others, "On the Satellite Role in the Era of 5G Massive Machine Type Communications," *IEEE Network*, Sep. 2018.
- [22] G. A. Akpakwu, B. J. Silva, G. P. Hancke, and A. M. Abu-Mahfouz, "A survey on 5G networks for the Internet of Things: Communication technologies and challenges," *IEEE access*, vol. 6, pp. 3619–3647, 2017.
- [23] 3GPP, *36.763 - Study on Narrow-Band Internet of Things (NB-IoT) / enhanced Machine Type Communication (eMTC) support for Non-Terrestrial Networks (NTN)*, Jun. 2021.
- [24] M. Kanj, V. Savaux, and M. Le Guen, "A tutorial on NB-IoT physical layer design," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2408–2446, 2020, Publisher: IEEE.
- [25] O. Liberg, M. Sundberg, E. Wang, J. Bergman, and J. Sachs, *Cellular Internet of things: technologies, standards, and performance*. Academic Press, 2017.
- [26] H. Fattah, *5G LTE narrowband internet of things (NB-IoT)*. CRC Press, 2018.
- [27] 3GPP, "LTE; evolved universal terrestrial radio access (E-UTRA); radio resource control (RRC); protocol specification," *3GPP TS 36.331 version 13.8.1 Release 13*, Jan. 2018.

- [28] 3GPP, "Evolved universal terrestrial radio access (E-UTRA); physical layer procedures," *3GPP TS 36.213*, Sep. 2022.
- [29] Y.-J. Yu, "NPDCCH period adaptation and downlink scheduling for NB-IoT networks," *IEEE Internet of Things Journal*, vol. 8, no. 2, pp. 962–975, 2020, Publisher: IEEE.
- [30] A. Hoglund, D. P. Van, T. Tirronen, O. Liberg, Y. Sui, and E. A. Yavuz, "3GPP Release 15 Early Data Transmission," *IEEE Communications Standards Magazine*, vol. 2, no. 2, pp. 90–96, 2018. DOI: 10.1109/MCOMSTD.2018.1800002.
- [31] J. Kim, G. Lee, S. Kim, T. Taleb, S. Choi, and S. Bahk, "Two-step random access for 5G system: Latest trends and challenges," *IEEE Network*, vol. 35, no. 1, pp. 273–279, 2020, Publisher: IEEE.
- [32] 3GPP, *36.211 - Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation*, Jun. 2021.
- [33] G. Maral, M. Bousquet, and Z. Sun, *Satellite communications systems: Systems, techniques and technology*, Sixth edition. Hoboken, N.J: John Wiley & Sons, 2020, ISBN: 978-1-119-38212-6 978-1-119-38207-2.
- [34] A. Guidotti, A. Vanelli-Coralli, V. Schena, *et al.*, "The path to 5G-Advanced and 6G Non-Terrestrial Network systems," in *2022 11th Advanced Satellite Multimedia Systems Conference and the 17th Signal Processing for Space Communications Workshop (ASMS/SPSC)*, IEEE, 2022, pp. 1–8.
- [35] 3GPP, "TR38.811, Technical Specification Group Radio Access Network; Study on New Radio (NR) to support non-terrestrial networks (Release 15), V15.4.1," Technical report, Oct. 2020.
- [36] 3GPP, "TR 22.822: Study on using satellite access in 5G," Technical report, Jul. 2018.
- [37] 3GPP, "TS 22.261: Service requirements for the 5G system," Technical specification, Jul. 2018.
- [38] 3GPP, "TR 28.808: Study on management and orchestration aspects of integrated satellite components in a 5G network," Technical report, Jul. 2018.
- [39] 3GPP, *TR38.821 - Solutions for NR to support Non-Terrestrial Networks (NTN)*, Jun. 2021.
- [40] 3GPP, "TR 23.737: Study on architecture aspects for using satellite access in 5G," Technical report, Jul. 2018.
- [41] 3GPP, "TR 24.821: Study on PLMN selection for satellite access," Technical report, Sep. 2021.
- [42] 3GPP, "TR38.801, Technical Specification Group Radio Access Network; Study on new radio access technology: Radio access architecture and interfaces (Release 14), V14.0.0," Technical report, Mar. 2017.

- [43] 3GPP, "TS 38.470, Technical Specification Group Radio Access Network; NG-RAN; F1 general aspects and principles (Release 16), V16.3.0," Technical report, Sep. 2020.
- [44] A. Guidotti, S. Cioni, G. Colavolpe, *et al.*, "Architectures, standardisation, and procedures for 5G Satellite Communications: A survey," *Computer Networks*, vol. 183, p. 107588, 2020.
- [45] 3GPP, "TS 38.410 Technical Specification Group Radio Access Network; NG-RAN; NG general aspects and principles, (Release 16), V16.3.0," Technical report, Sep. 2020.
- [46] "ETSI EN 302 307-1, V1.1.1, Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications; Part 2: DVB-S2X," Tech. Rep., Oct. 2014.
- [47] C. Amatetti, A. Guidotti, and A. Vanelli-Coralli, "NTN support for mMTC: Architectural and channel model considerations."
- [48] I. Ali, N. Al-Dhahir, and J. E. Hershey, "Doppler characterization for leo satellites," *IEEE Transactions on Communications*, vol. 46, no. 3, pp. 309–313, 1998. DOI: 10.1109/26.662636.
- [49] A. Guidotti, A. Vanelli-Coralli, M. Caus, *et al.*, "Satellite-enabled LTE systems in LEO constellations," in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2017, pp. 876–881. DOI: 10.1109/ICCW.2017.7962769.
- [50] ITU, "ITU-R P.676 : Attenuation by atmospheric gases and related effects," Tech. Rep., Aug. 2019.
- [51] ITU, "ITU-R P.531 : Ionospheric propagation data and prediction methods required for the design of satellite networks and systems," Tech. Rep., Aug. 2019.
- [52] O. Kodheli, N. Maturo, S. Andrenacci, S. Chatzinotas, and F. Zimmer, "Link budget analysis for satellite-based narrowband IoT systems," in *International conference on ad-hoc networks and wireless*, 2019.
- [53] V. Mannoni, V. Berg, S. Cazalens, and P. Raveneau, "NB-IoT for satellite communications: Physical layer analysis and performance," in *2021 IEEE 32nd annual international symposium on personal, indoor and mobile radio communications (PIMRC)*, 2021.
- [54] R Barbau, V Deslandes, G Jakllar, and A.-L. Beylot, "Performance evaluation of different 5G NB-IoT satellite systems," 2021, Publisher: IET.
- [55] G. Charbit, A. Medles, P. Jose, D. Lin, X. Zhu, and I.-K. Fu, "Satellite and cellular networks integration-A system overview," in *2021 joint european conference on networks and communications & 6G summit (EuCNC/6G summit)*, 2021.

- [56] A. Guidotti, A. Vanelli-Coralli, M. Conti, *et al.*, "Architectures and Key Technical Challenges for 5G Systems Incorporating Satellites," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, Mar. 2019, ISSN: 0018-9545, 1939-9359. DOI: 10.1109/TVT.2019.2895263.
- [57] M. Conti, S. Andrenacci, N. Maturo, S. Chatzinotas, and A. Vanelli-Coralli, "Doppler Impact Analysis for NB-IoT and Satellite Systems Integration," in *IEEE ICC'20 - SAC-08 SSC*.
- [58] G. Charbit, D. Lin, K. Medles, L. Li, and I.-K. Fu, "Space-Terrestrial Radio Network Integration for IoT," in *2020 2nd 6G Wireless Summit (6G SUMMIT)*, Mar. 2020. DOI: 10.1109/6GSUMMIT49458.2020.9083854.
- [59] H. Chougrani and others, "NB-IoT Random Access for Non-Terrestrial Networks," *arXiv preprint arXiv:2101.08079*, 2021.
- [60] O. Liberg, S. E. Löwenmark, S. Euler, *et al.*, "Narrowband Internet of Things for Non-Terrestrial Networks," Oct. 2020, arXiv: 2010.04906 [cs, eess]. [Online]. Available: <http://arxiv.org/abs/2010.04906> (visited on 12/22/2020).
- [61] R. Barbau, V. Deslandes, G. Jakllari, J. Tronc, J.-F. Chouteau, and A.-L. Beylot, "NB-IoT over GEO satellite: Performance analysis," in *2020 10th Advanced Satellite Multimedia Systems Conference and the 16th Signal Processing for Space Communications Workshop (ASMS/SPSC)*, IEEE, 2020, pp. 1–8.
- [62] O. Kodheli, A. Astro, J. Querol, *et al.*, "Random access procedure over non-terrestrial networks: From theory to practice," *IEEE access : practical innovations, open solutions*, vol. 9, pp. 109 130–109 143, 2021, Publisher: IEEE.
- [63] G. Sciddurlo, A. Petrosino, M. Quadrini, *et al.*, "Looking at NB-IoT over LEO satellite systems: Design and evaluation of a service-oriented solution," *IEEE Internet of Things Journal*, 2021, Publisher: IEEE.
- [64] A. Petrosino, G. Sciddurlo, S. Martiradonna, D. Striccoli, G. Piro, and G. Boggia, "WIP: An open-source tool for evaluating system-level performance of NB-IoT non-terrestrial networks," in *2021 IEEE 22nd international symposium on a world of wireless, mobile and multimedia networks (WoWMoM)*, 2021.
- [65] J. Sedin, S. Gopinath, X. Lin, *et al.*, "5G massive machine type communication performance in non-terrestrial networks with LEO satellites," in *2021 IEEE global communications conference (GLOBECOM)*, 2021.
- [66] O. Kodheli, N. Maturo, S. Chatzinotas, S. Andrenacci, and F. Zimmer, "NB-IoT via LEO satellites: An efficient resource allocation strategy for uplink data transmission," *IEEE Internet of Things Journal*, vol. 9, no. 7, pp. 5094–5107, 2021, Publisher: IEEE.

- [67] O. Kodheli, S. Andrenacci, N. Maturo, S. Chatzinotas, and F. Zimmer, "Resource Allocation Approach for Differential Doppler Reduction in NB-IoT over LEO Satellite," in *2018 9th ASMS/SPSC*, Sep. 2018. DOI: 10.1109/ASMS-SPSC.2018.8510724.
- [68] O. Kodheli, N. Maturo, S. Andrenacci, S. Chatzinotas, and F. Zimmer, "Link budget analysis for satellite-based narrowband iot systems," in *Ad-Hoc, Mobile, and Wireless Networks: 18th International Conference on Ad-Hoc Networks and Wireless, ADHOC-NOW 2019, Luxembourg, Luxembourg, October 1-3, 2019, Proceedings 18*, Springer, 2019, pp. 259-271.
- [69] X. Ma, J. Zhang, Z. Yu, B. Bai, and Q. Mei, "A PAPR-reduction SCMA codebook design for satellite communication systems," in *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, IEEE, 2019, pp. 1-6.
- [70] J. Jiao, J. Zhou, S. Wu, and Q. Zhang, "Superimposed pilot code-domain noma scheme for satellite-based internet of things," *IEEE Systems Journal*, vol. 15, no. 2, pp. 2732-2743, 2020.
- [71] *NB-IoT4Space - 3GPP Narrow-Band Internet-of-Things (NB-IoT) User Sensor Integration into Satellite | ESA's ARTES Programmes*. [Online]. Available: <https://artes.esa.int/projects/nbiot4space>.
- [72] *Narrowband IoT for SmallSat Networks*. [Online]. Available: <https://artes.esa.int/projects/narrowband-iot-smallsat-networks>.
- [73] *Satellite Backhauling Prototype for Future Narrow Band Internet of Things (NB-IoT) Networks | ESA's ARTES Programmes*. [Online]. Available: <https://artes.esa.int/projects/iot-satback>.
- [74] M. Conti, C. Amatetti, A. Guidotti, and A. Vanelli-Coralli, "NB-IoT over Non-Terrestrial Networks: Link Budget Analysis," in *2020 IEEE Global Communications Conference (GLOBECOM2020)*.
- [75] A. Guidotti, A. Vanelli-Coralli, A. Mengali, and S. Cioni, "Non-terrestrial networks: Link budget analysis," in *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, IEEE, 2020, pp. 1-7.
- [76] H. Huawei, "R1-1911858, Discussion on performance evaluation for NTN, RAN1#99," Technical report, Nov. 2019.
- [77] H. Huawei, "R1-1903998: Discussion on link budget for NTN," Technical report, Apr. 2019.
- [78] 3GPP, "TS36.211, Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Physical Channels and Modulation, V16.1.0," Technical report, Mar. 2020.

- [79] 3GPP, "TR36.888 Technical Specification Group Radio Access Network; Study on provision of low-cost Machine-Type Communications (MTC) User Equipments (UEs) based on LTE (Release 12), V12.0.0," Technical report, Jun. 2013.
- [80] C. Amatetti, M. Conti, A. Guidotti, and A. Vanelli-Coralli, "Preamble detection in NB-IoT via Satellite: A Wavelet based approach," in *2021 IEEE global communications conference (GLOBECOM)*, 2021.
- [81] X. Lin, A. Adhikary, and Y.-P. E. Wang, "Random access preamble design and detection for 3GPP narrowband IoT systems," *IEEE Wireless Communications Letters*, vol. 5, no. 6, pp. 640–643, 2016.
- [82] J.-K. Hwang, C.-F. Li, and C. Ma, "Efficient detection and synchronization of superimposed NB-IoT NPRACH preambles," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 1173–1182, 2018.
- [83] A.Chakrapani, "NB-IoT uplink receiver design and performance study," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 2469–2482, 2019, Publisher: IEEE.
- [84] Q.Wu and others, "An efficient NPRACH receiver design for NB-IoT systems," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 10 418–10 426, 2020, Publisher: IEEE.
- [85] H.Chougrani and others, "Efficient Preamble Detection and Time-of-Arrival Estimation for Single-Tone Frequency Hopping Random Access in NB-IoT," *IEEE Internet of Things Journal*, vol. 8, no. 9, pp. 7437–7449, 2020, Publisher: IEEE.
- [86] 3GPP, "RP-193235, Study on NB-Io/eMTC support for NTN, RAN Meeting #86," Technical report, Dec. 2019.
- [87] 3GPP, "R1-2009095, Discussion on RAN1 aspects of IoT NTN, RAN1 meeting 103e," Technical report, Nov. 2020.
- [88] 3GPP, "TR 36.104 evolved universal terrestrial radio access (E-UTRA); base station (BS) radio transmission and reception v16.7.0," Technical report, Oct. 2020.
- [89] R.Barsanti and others, "Wavelet-based time delay estimates for transient signals," in *The thirty-seventh asilomar conference on signals, systems & computers, 2003*, vol. 1, 2003.
- [90] K.Kuzume and others, "Dyadic lifting wavelet based signal detection," in *2016 3th AIPR*, 2016.
- [91] A.Jain and others, "Algorithms for change detection with sparse signals," *IEEE Transactions on Signal Processing*, vol. 68, pp. 1331–1345, 2020, Publisher: IEEE.
- [92] C. Amatetti, R. Campana, A. Georganaki, and A. Vanelli-Coralli, "Neural network based non orthogonal random access for 6g ntn-iot," in *2022 IEEE Globecom Workshops (GC Wkshps)*, IEEE, 2022, pp. 1389–1394.

- [93] 3GPP, "R1-211658, RAN vice-chair, Moderators summary for discussion NTN evolution, RP-93-e meeting," Technical report, Sep. 2021.
- [94] Y. Piao, Y. Kim, and T.-J. Lee, "Random Power Back-Off for Random Access in 5G Networks," *IEEE Access*, vol. 9, pp. 121 561–121 569, 2021.
- [95] X. Liu, K.-Y. Lam, F. Li, J. Zhao, L. Wang, and T. S. Durrani, "Spectrum sharing for 6g integrated satellite-terrestrial communication networks based on noma and cr," *IEEE Network*, vol. 35, no. 4, pp. 28–34, 2021.
- [96] L. Dai, B. Wang, Y. Yuan, S. Han, I Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 74–81, 2015.
- [97] Nokia and Alcatel-Lucent Shanghai and Bell, "R1-165021, Performance of interleaved division multiple access (IDMA) in combination with OFDM family waveforms," Technical report, Apr. 2016.
- [98] S. YANG, P. CHEN, L. LIANG, J. ZHU, and X. SHE, "Uplink multiple access schemes for 5g: A survey," *ZTE Communications*, vol. 15, no. S1, pp. 31–40, 2020.
- [99] Intel, "R1-166552, On UL non-orthogonal multiple access schemes," Technical report, 2016.
- [100] K. Lai, J. Lei, Y. Deng, L. Wen, G. Chen, and W. Liu, "Analyzing uplink grant-free sparse code multiple access system in massive iot networks," *IEEE Internet of Things Journal*, vol. 9, no. 7, pp. 5561–5577, 2021.
- [101] S. Chaturvedi, Z. Liu, V. A. Bohara, A. Srivastava, and P. Xiao, "A tutorial to sparse code multiple access," *arXiv preprint arXiv:2105.06860*, 2021.
- [102] S. Chaturvedi, Z. Liu, V. A. Bohara, A. Srivastava, and P. Xiao, "A tutorial on decoding techniques of sparse code multiple access," *IEEE Access*, 2022.
- [103] C. Amatetti and others, "NB-IoT random access procedure via NTN: System level performances," in *ICC 2022-IEEE international conference on communications*, 2022.
- [104] S. Narayanan, "ADAM: An Adaptive Access Mechanism for NB-IoT Systems in the 5G Era," *IEEE access : practical innovations, open solutions*, vol. 9, pp. 109 915–109 931, 2021, Publisher: IEEE.
- [105] 3GPP, 36.331 - *Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification*, Jul. 2021.
- [106] C. Amatetti, O. Kodheli, V. Thang, S. Chatzinotas, and A. Vanelli-Coralli, "A Joint Access and Data Phases Design for NB-IoT via Non-Terrestrial Networks."
- [107] B. Manzoor, A. Al-Hourani, and B. Al Homssi, "Improving IoT-over-satellite connectivity using frame repetition technique," *IEEE Wireless Communications Letters*, vol. 11, no. 4, pp. 736–740, 2022.

- [108] A. Al-Hourani and I. Guvenc, "On modeling satellite-to-ground path-loss in urban environments," *IEEE Communications Letters*, vol. 25, no. 3, pp. 696–700, 2020.
- [109] Y. Liu, Y. Deng, N. Jiang, M. ElKashlan, and A. Nallanathan, "Analysis of random access in NB-IoT networks with three coverage enhancement groups: A stochastic geometry approach," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 549–564, 2020.
- [110] J. A. Klobuchar, "Ionospheric Time-Delay Algorithm for Single-Frequency GPS Users," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-23, no. 3, pp. 325–331, 1987. DOI: 10.1109/TAES.1987.310829.
- [111] G. Di Giovanni and S. Radicella, "An analytical model of the electron density profile in the ionosphere," *Advances in Space Research*, vol. 10, no. 11, pp. 27–30, 1990, ISSN: 0273-1177. DOI: [https://doi.org/10.1016/0273-1177\(90\)90301-F](https://doi.org/10.1016/0273-1177(90)90301-F). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/027311779090301F>.
- [112] EUSPA, "Report on rail user needs and requirementsl," Tech. Rep. [Online]. Available: https://www.gsc-europa.eu/sites/default/files/sites/all/files/Report_on_User_Needs_and_Requirements_Rail.pdf.
- [113] EUSPA, "PPP-RTK market and technology report," Tech. Rep. [Online]. Available: https://www.euspa.europa.eu/simplecount_pdf/tracker?file=calls_for_proposals/rd.03_-_ppp-rtk_market_and_technology_report.pdf.
- [114] EUSPA, "GNSS User Technology Report," Tech. Rep. [Online]. Available: https://www.euspa.europa.eu/simplecount_pdf/tracker?file=uploads/technology_report_2020.pdf.
- [115] C. Amatetti, T. Polonelli, E. Masina, *et al.*, "Towards the future generation of railway localization and signaling exploiting sub-meter RTK GNSS," in *2022 IEEE sensors applications symposium (SAS)*, 2022.
- [116] H. Somogyi and A. Soumelidis, "Comparison of High-Precision GNSS systems for development of an autonomous localization system," in *2020 23rd international symposium on measurement and control in robotics (ISMCR)*, IEEE, 2020, pp. 1–6.
- [117] H.-G. Shin, D.-K. Lee, S.-H. Lee, and S.-I. Park, "RTK correction data transmission service for autonomous-driving via ATSC 3.0 in South Korea," in *2021 IEEE international symposium on broadband multimedia systems and broadcasting (BMSB)*, 2021.
- [118] Y. Zhang, D. Lu, B. Cai, and J. Liu, "3D digital track map-based GNSS NLOS signal analytical identification method," in *2021 international conference on electromagnetics in advanced applications (ICEAA)*, IEEE, 2021, pp. 313–318.

- [119] K.-W. Chiang, H.-W. Chang, Y.-H. Li, *et al.*, "Assessment for INS/ GNSS/ odometer/ barometer integration in loosely-coupled and tightly-coupled scheme in a GNSS - degraded environment," *IEEE Sensors Journal*, vol. 20, no. 6, pp. 3057–3069, 2019.
- [120] S. A. Kazim, N. A. Tmazirte, and J. Marais, "Realistic position error models for GNSS simulation in railway environments," in *2020 european navigation conference (ENC)*, 2020.
- [121] M. Caamano, O. G. Crespillo, D. Gerbeth, and A. Grosch, "Detection of GNSS multipath with time-differenced code-minus-carrier for land-based applications," in *2020 european navigation conference (ENC)*, 2020.
- [122] P. Zabalegui, G. De Miguel, A. Pérez, J. Mendizabal, J. Goya, and I. Adin, "A review of the evolution of the integrity methods applied in GNSS," *IEEE access : practical innovations, open solutions*, vol. 8, pp. 45 813–45 824, 2020, Publisher: IEEE.
- [123] S. Jiang, D. Lu, and B. Cai, "GNSS NLOS signal modeling and quantification method in railway urban canyon environment," in *2019 IEEE intelligent vehicles symposium (IV)*, 2019.
- [124] N. Kubo, M. Higuchi, T. Takasu, and H. Yamamoto, "Performance evaluation of GNSS-based railway applications," in *2015 international association of institutes of navigation world congress (IAIN)*, 2015.
- [125] V. Quiñones, A. Águila, and I. Cordero, "GNSS4Rail simulation tool. Supporting Virtual Balise (VB) location for GNSS based railway operations," in *2021 international conference on electrical, computer, communications and mechatronics engineering (ICECCME)*, 2021.
- [126] S. Y. Cho, M. S. Chae, and K. H. Shin, "Reliability analysis of the integrated navigation system based on real trajectory and calculation of safety margin between trains," *IEEE access : practical innovations, open solutions*, vol. 9, pp. 32 986–32 996, 2021, Publisher: IEEE.
- [127] C. Stallo, A. Neri, P. Salvatori, R. Capua, and F. Rispoli, "GNSS integrity monitoring for rail applications: Two-tiers method," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 4, pp. 1850–1863, 2018, Publisher: IEEE.
- [128] M. Bochkati, H. Sharma, C. A. Lichtenberger, and T. Pany, "Demonstration of fused RTK (fixed)+ inertial positioning using Android smartphone sensors only," in *2020 IEEE/ION position, location and navigation symposium (PLANS)*, 2020.
- [129] Y. Zhou, Q. Chen, and X. Niu, "Kinematic measurement of the railway track centerline position by GNSS/INS/odometer integration," *IEEE access : practical innovations, open solutions*, vol. 7, pp. 157 241–157 253, 2019, Publisher: IEEE.

- [130] A. Schütz, D. E. Sánchez-Morales, and T. Pany, "Precise positioning through a loosely-coupled sensor fusion of GNSS-RTK, INS and LiDAR for autonomous driving," in *2020 IEEE/ION position, location and navigation symposium (PLANS)*, 2020, pp. 219–225.
- [131] S. Bisnath, "Relative positioning and real-time kinematic (RTK)," *Position, Navigation, and Timing Technologies in the 21st Century: Integrated Satellite Navigation, Sensor Systems, and Civil Applications*, vol. 1, pp. 481–502, 2020, Publisher: Wiley Online Library.
- [132] S. Mahato, A. Santra, S. Dan, P. Rakshit, P. Banerjee, and A. Bose, "Preliminary results on the performance of cost-effective GNSS receivers for RTK," in *2019 URSI asia-pacific radio science conference (AP-RASC)*, 2019.
- [133] I. Um, S. Park, S. Oh, and H. Kim, "Analyzing location accuracy of unmanned vehicle according to RTCM message frequency of RTK-GPS," in *2019 25th asia-pacific conference on communications (APCC)*, 2019.
- [134] E. Kaplan and C. Hegarty, *Understanding GPS: Principles and applications* (Artech house mobile communications). Artech House, 2006, ISBN: 978-1-58053-894-7.
- [135] D. Mikhaylov, C. Amatetti, T. Polonelli, *et al.*, "Towards the future generation of railway localization exploiting rtk and gnss," *IEEE Transactions on Instrumentation and Measurement*, 2023.
- [136] A. D. Snow, J. Whitaker, M. Cochran, and others, *Pyproj4/pyproj: 3.4.0 Release*, Sep. 2022. DOI: 10.5281/zenodo.7065964. [Online]. Available: <https://doi.org/10.5281/zenodo.7065964>.
- [137] S. Gillies, C. van der Wel, J. Van den Bossche, *et al.*, *Shapely*. [Online]. Available: <https://github.com/shapely/shapely>.

Acknowledgements

This thesis is the outcome of three years of work. I would like to thank the many people who have been close to me during this journey and helped me grow and face difficult times. Indeed the PhD is the longest and shortest time of our life. Moments of frustration, fatigue, and tears. Also excitement, astonishment, joy, friendship, wonder, and accomplishment.

I would like to thank Prof. Alessandro Vanelli who believed in me and gave me the opportunity to embark on this path. He has been a great mentor, teacher, and reference point. I am grateful to him for what he taught me, especially because he encouraged me to work hard and dream big, always aiming for perfection. I learned a lot from him and hope to continue doing so in the future.

I would like to thank prof. Alessandro Guidotti for his continuous and constant support and for all that he has taught me.

I thank my colleagues, actually more friends than colleagues. Matteo, is always a goad to keep improving and present with his advice. Riccardo, for his support, his patience, his kindness, and always ready to help. A sure presence. Bruno, the new "purchase" but already part of the group, is also always available, present and ready to help in any situation. Daniel, Bilal and Rabih with whom I shared these years.

I would like to thank Prof. Benini's ETH team, and especially Tommaso. With them, I shared the last year through a project, which allowed me to meet people who taught me a lot. I hope there will be opportunities to collaborate again in the future.

I thank Prof. Symeon Chatzinotas for welcoming me into SnT, it was a wonderful experience and a great opportunity for growth.

I would like to thank the people I worked with while I was in SnT, for their incredible patience and willingness to teach me. I thank Otjon, not only for his great patience during these seven months, for all the teaching, but also for welcoming me as a true friend from day one and helping me every step of the way. Thank you.

Thanks to Mario and Clement, and all the friends, who make Lux a second home.

Thanks to Kikka, Isa, Fede, Alba and all the friends from Pescara and Bologna for the many good times spent together and for the support in the most difficult moments.

Thanks to Ale, more precious than the four fundamental elements.

Thanks to my family. It is obvious to say that, without them, I would not have arrived here.

Thanks to Chiaretta for her kindness.

Thanks to my aunt Angela not only an aunt but also a mentor.

Thanks to my Dad for his valuable advices and for always being ready to share with me his many passions, which enrich me.

To my mom, with whom a glance is enough to understand each other.