

D4.6: Final report on the Unified Air Interface, User-Centric and Beamforming Solutions

Revision: V1.5

Work package	WP 4
Task	Task 4.1 and Task 4.2
Due date	31/03/2026
Submission date	30/03/2026
Deliverable lead	CNIT
Version	V1.5
Authors	Marius Caus, Xavier Artiga, Musbah Shaat, Erislandy Mozo, Cristian J. Vaca-Rubio (CTTC) Alessandro Guidotti, Tommaso Foggi, Elisa Conti (CNIT)
Reviewers	Fanny Parzysz (ORA)
Abstract	This document reports the final outcomes of Tasks 4.1 “Flexible and unified Air Interface” and Task 4.2 “User-Centric and Digital Beamforming Solutions” related to the investigation of the NR-NTN standard, including the waveform, and user-centric beamforming technologies. In particular, the enhancements that are needed in the 3GPP NTN to support the architecture defined in WP3 are addressed. Building upon this air interface, user-centric beamforming designs and Radio Resource Management (RRM) solutions are conceived.
Keywords	3GPP, NTN, 5G, 5G-Advanced, NTN procedures, Architecture, Waveform, OFDM, OTFS, User-centric beamforming, RRM

DOCUMENT REVISION HISTORY

Version	Date	Description of change	List of contributor(s)
V0.1	30/09/2024	Document creation	CTTC
V0.2	01/02/2025	Table of contents	CTTC
V0.3	15/05/2025	Mapping schemes and mobility aspects	CTTC
V0.4	25/07/2025	Beam illumination plan for signalling	CTTC
V0.5	15/09/2025	User-centric beamforming for data transmission	CNIT
V0.6	29/09/2025	Earth-fixed beam footprint and beam layout for signalling	CTTC
V0.7	30/09/2025	Spectrum sharing analysis	CTTC
V0.8	08/10/2025	Intermediate QA review	ORA
V0.9	09/10/2025	Intermediate revision according to the received comments from the QA review	CTTC, CNIT
V1.0	18/03/2026	Numerical assessment of TN-NTN coexistence scenarios	CTTC
V1.1	19/03/2026	Beam-centric beamforming for data transmission and 5G NR radio interface adaptation to fit the OTFS waveform	CNIT
V1.2	20/03/2026	Earth-moving beam footprint and beam layout for signalling	CTTC
V1.3	26/03/2026	Final QA review	ORA
V1.4	26/03/2026	Final revision according to the received comments from the QA review	CTTC,CNIT
V1.5.F	30/03/2026	Approved final version for EC portal submission	DLR

DISCLAIMER



5G-STAR DUST (*Satellite and Terrestrial Access for Distributed, Ubiquitous, and Smart Telecommunications*) project has received funding from the [Smart Networks and Services Joint Undertaking \(SNS JU\)](#) under the European Union's [Horizon Europe research and innovation programme](#) under Grant Agreement No 101096573.

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

COPYRIGHT NOTICE

© 2023 - 2025 5G-STAR DUST

Project co-funded by the European Commission in the Horizon Europe Programme		
Nature of the deliverable:	R	
Dissemination Level		
PU	Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page)	✓
SEN	Sensitive, limited under the conditions of the Grant Agreement	
Classified R-UE/ EU-R	EU RESTRICTED under the Commission Decision No2015/ 444	
Classified C-UE/ EU-C	EU CONFIDENTIAL under the Commission Decision No2015/ 444	
Classified S-UE/ EU-S	EU SECRET under the Commission Decision No2015/ 444	

* R: Document, report (excluding the periodic and final reports)

DEM: Demonstrator, pilot, prototype, plan designs

DEC: Websites, patents filing, press & media actions, videos, etc.

DATA: Data sets, microdata, etc.

DMP: Data management plan

ETHICS: Deliverables related to ethics issues.

SECURITY: Deliverables related to security issues

OTHER: Software, technical diagram, algorithms, models, etc.

EXECUTIVE SUMMARY

In this deliverable, 5G-STAR DUST investigates aspects related to the 3GPP new radio (NR) air interface for non-terrestrial networks (NTN) as outcomes of Tasks 4.1 “Flexible and unified Air Interface” and Task 4.2 “User-Centric and Digital Beamforming Solutions,” respectively. The techniques presented in this deliverable are built upon the preliminary report presented in [1].

In [1], guidelines have been reported to establish the connection from a non-terrestrial network (NTN) user equipment (UE) to an NTN gNB with particular attention to aspects that are left open for implementation. The approach adopted by 5G-STAR DUST distinguishes between two management steps:

- Synthesis of broadcast beams combined with periodic beam hopping techniques to manage system and control information that is common to all users;
- Synthesis of data beams using user-centric beamforming techniques to address use traffic demands.

The application of beam hopping to NTN requires air interface enhancements, which have been covered in [1]. The major contribution of this deliverable is to provide a more detailed design involving mapping schemes between NTN beams and NR cells, scheduling of common signalling and beam footprint and layout designs. The initial step begins with an in-depth analysis of the mapping schemes between NTN beams and NR cells. With respect to the preliminary report, emphasis has been given to intra-satellite mobility aspects. The comparative assessment reveals that the mobility across beams can be supported with lower signalling and latency in the multi-NTN beam scenario, where each NR cell handles multiple NTN beams.

The remarkable features of the multi-beam NTN cell scenario motivate the adoption of the standardized beam management procedure to implement the beam hopping mechanism. However, the efficient management of NTN beams is still an open problem, which becomes fundamental in an EIRP limited system like the one targeted by 5G-STAR DUST, where each satellite covers a large region on Earth with a radius on the order of 1500 km, with a very limited number of simultaneous beams with a beam footprint radius at Nadir of only 41 km. It is important to remark that broadcast beams synthesized for beam hopping are characterized to have the same dwell time and revisit time interval. Accordingly, users are served sequentially. Hence, since the user density is not constant across the coverage area, serving users with periodic beam hopping techniques could make an inefficient use of the resources. Hence, the use of broadcast beams should be minimized, so that resources available in data beams are maximized. This strategy is advantageous to satisfy the traffic demand. Therefore, the figure of merit is to minimize the number of broadcast beams, i.e., the number of time/frequency resources required for the beam hopping mechanism to cover the whole coverage area of a single satellite.

The performance metric that is considered to design beam hopping techniques is the number of time resources required to illuminate 95% of the coverage area with a minimum signal-to-interference-plus noise ratio (SINR) of 3dB. The conventional solution is to deploy a homogenous grid with a crossover level of 3dB, taking as reference the footprint at Nadir. This design results into large number of beams to cover the field of view (FoV). Moreover, large overlapping will be experienced at the edge due to the beam deformation. A more suitable alternative is to reduce the beam cross over level, which significantly lowers the number of

beams and the overlapping at the coverage edge. Building on this approach, a strategy has been derived to select the set of beams illuminated simultaneously.

To evaluate the benefits of the proposed beam hopping design, which relies on increasing the crossover level, frame structures for both UL and DL have been studied. The objective is to define a specific scheduling strategy for the common signalling. Since the 3GPP standard already specified the timing of a beam sweep to transmit SSB burst across a cell in a half frame of 5 ms, the proposed approach for the DL is to co-schedule all common signalling within the slots devoted to SSB transmissions, so a single sweep may allow broadcasting all common signalling across the coverage, reducing the amount of time the beam pattern is designed under common signalling criteria. In the proposed solution, it is possible to serve all the beams in 20 ms. Consequently, the SSB default periodicity can be reused to ensure a coverage ratio close to 100%. In the uplink, FoV coverage is primarily required for RACH signaling. To be consistent with the broadcast signalling, the beam illumination plan in the uplink is designed using the same beam pattern that is conceived for the downlink. Accordingly, for both uplink and downlink, the coverage area is divided into the same number of beams. In order to mimic the scheduling of common downlink signals, the idea is to allocate random access resources associated to multiple beams in the same uplink frame. Based on the beam hopping evaluation, the RACH periodicity specified in the standard is sufficient to guarantee 100% uplink coverage ratio in the uplink.

Although Earth-fixed beams are generally preferred, the beam layout has also been analysed for Earth-moving beams. The main motivation is that the Earth-fixed design becomes inefficient from a coverage perspective. The beam-footprint widening caused by Earth curvature leads to either significant overlap, coverage holes, or reduced beam gains, depending on the chosen beam design strategy. In this context, Earth-moving beams offer two key advantages over the Earth-fixed counterpart. First, the inter-beam distance can be increased as the beams widen, reducing the total number of beams required to cover the service area. Second, the beamforming weights can remain fixed regardless of the region that is illuminated. The main drawback is that Earth-moving beams introduce additional signalling overhead to handle the mobility induced by satellite motion. Regarding the beam illumination pattern, the design is driven by the minimization of the number of hops with a minimum SINR of 3dB for at least 95% of the coverage area. The numerical assessment shows that Earth-moving designs can increase the number of slots available for data transmission by up to 24% within the SSB period compared with Earth-fixed designs.

With respect to user-centric beamforming, this document finalises the assessment that was initiated in the first version of this report by considering both system-level and link-level approaches. As for the former:

- Defining the system-level procedures to support beamforming algorithms based on the availability of either the Channel State Information (CSI) or the location estimates at the transmitter. The considered architecture is the Distributed Beamforming Management (DBFM), extensively detailed in [1], in which all beamforming-related operations (i.e., scheduling, computation of the beamforming coefficients, application of the coefficients to the signals) are implemented in the gNB Distributed Unit (gNB-DU) on-board the satellite. The procedures initially detailed in [1] have been extended to take into account a Distributed Antenna System (DAS) in space, in which a swarm of Non-Terrestrial Network (NTN) nodes collectively implements federated beamforming. This analysis has then been completed by discussing (i) how the CSI or location information can be provided by the User Equipments (UEs) to the on-board gNB and (ii) the expected signalling overhead to support CSI- or location-based solutions. This assessment indicates that location-based solutions introduce a significantly lower overhead compared

to CSI-based ones, but they call for the modification of the 3GPP specifications with respect to the availability of the location of the UEs at RAN level (i.e., at the gNB).

- Designing a methodology to derive databases for non-uniform traffic distributions. This methodology is based on the exploitation of the EC JRC population dataset, publicly available. The proposed methodology allows to easily define traffic ranges (minimum and maximum capacity requests) and the type of area to serve (dense urban, urban, sub-urban, and rural). This procedure is applied for the residential broadband set of scenarios. A second methodology has then been designed to provide capacity requests also for air traffic scenarios, taking into account airport locations and airline traffic. Both solutions can be easily adapted in terms of expected NTN terminals in 2030, number of satellites in visibility, and many other parameters introduced in this report.
- Designing scheduling algorithms with flexible priority metrics. Specifically, two algorithms have been designed which take into account the prioritisation of the traffic based on the residual unmet capacity and the residual visibility time from the serving satellite(s). Both algorithms show a promising performance in terms of flexibility and achievable capacity under various traffic request conditions. The two solutions have also been analysed in terms of the big-o complexity (this analysis is the starting point for the introduction of solutions supported by Artificial Intelligence and Machine Learning, discussed in the companion D4.7). The extensive performance assessment showed that, in the standalone NTN node scenario, the system is capable of providing 25-28 Gbps over a visibility period of approximately 7 minutes for a service area that comprises most of the European land areas. However, in dense traffic scenarios (in particular due to an increased minimum capacity request), a significant amount of unmet capacity is still present. Since the scheduling constraints and algorithms have been numerically optimised, this outcome suggests that the performance is close to its optimum and system-level modifications would be needed to improve it, such as: i) increasing the number of satellites while increasing the minimum elevation angle (which reduces the path losses thus improving the overall SINR), increasing the available on-board power, reducing the number of users (we extensively discussed the impact of the number of allocations per user in the study), or using multiple satellites in a federated approach. Finally, the performance assessment has been completed by providing numerical results under the assumption of non-ideal CSI or location estimates, with the latter resulting in a more robust solution to estimation errors.
- Design and assessing the performance of federated beamforming from multiple satellites. The system-level procedures, beamforming algorithms, and scheduling solutions have been extended to consider a DAS from the sky. The performance assessment showed an improved performance, in particular when all the satellites in the swarm transmit with a power level equal to the total power available on-board.

A performance comparison has also been performed to compare the user-centric PRS designed in Task 4.2 with a beam-centric variant, BC-PRS, which takes into account the beam lattices optimised in Task 4.1. As expected, the performance of the user-centric solution is better in terms of both achieved and unmet capacity; however, the performed analyses highlight a potential direction for future research: the joint optimisation of the beam lattice in terms of signalling overhead and beamforming performance.

Finally, Annex A of this document finalises the performance evaluation initiated in [1] related to random scheduling algorithms with uniform traffic requests.

With respect to link-level, the NR 5G standard waveform based on OFDM was here compared to the novel OTFS waveform from the point of view of link-level performance. An OTFS-based

system was implemented and fitted into the NR frame architecture; some variants were investigated among which one was chosen due to its flexible design. Then, scenarios parameters and data (namely beamforming coefficients, Doppler, delay and attenuation values) were derived by exploiting the system-level simulator, for both a single transmitting satellite and two transmitting satellites. These data were employed as inputs to the specifically developed link-level simulator, whose output are the pragmatic capacity of the system, which only relies on the received signal samples irrespective of the code rate and decoder performance, or directly the block-error rate, thus making use of LDPC decoders and therefore providing some useful information on the actual performance of practical systems. The selection of the modulation orders and code rates were implemented through the NR CSI reporting procedure, also taking into account additional margins as a matter of improving the overall throughput and adapt the mod-code selection to the considered satellite scenario.

Moreover, novel detectors specifically tailored for the OTFS user-centric paradigm were designed and investigated. An interference cancellation (IC) detector, especially, was implemented, also adapted for the OFDM waveform, and employed to derive results together with the standard MMSE detector. Results show promising results for both the OTFS waveform over the OFDM one, and for the IC detector over the MMSE, as the former entails comparable performance but with lower complexity.

As concerns the two different satellite scenarios, the latter including two transmitting satellites provides interesting results despite the simple clear-sky non-synchronized channel employed in the simulations, given that the system would mostly benefit from the cell-free user-centric framework with tight synchronization among transmitted signals and in presence of uncorrelated fading channels, with respect to the case of a single transmitting satellite.

The proposed solutions were also evaluated and demonstrated in presence of a narrow-band fading channel, where mainly shadowing is considered. In this scenario OFDM and OTFS, with both the standard MMSE and a novel low-complexity detector based on message-passing and interference cancellation, were compared in terms of pragmatic capacity. When two transmitting satellites are employed, the case of users equipped with antenna arrays was also investigated; this solution enables the coherent combining of the received signals, and further improvements were achieved with the exploitation of the SNR estimates from both satellites in the CSI reporting procedures that set the modulation and coding selection at the transmitting nodes.

The last part of the document is dedicated to analysing the coexistence of TN and NTN on shared frequency resources. The channel arrangements in the Ka-band highlight that interference events primarily affect the spectrum that is allocated to the NTN uplink. In this context, various spectrum-sharing scenarios between satellite and terrestrial networks have been discussed. Link budgets in specific spectrum-sharing scenarios are computed, demonstrating the necessity of interference mitigation. The most detrimental situation occurs when NTN UEs interfere the reception of TN gNBs. Remarkably, no specific protection measures are envisioned when signals transmitted by a large population of TN UEs leak into the main beam of a satellite hosting a gNB. In these scenarios, the magnitude of the interference becomes more prominent at Nadir. However, even under the most adverse conditions, the aggregate interference does not significantly degrade the NTN link. This negligible impact on satellite reception is mainly due to the high propagation losses experienced in the Ka band. Another important aspect that explains this result is that terrestrial UEs synthesize directional beams to connect to the base station. This means that only a small subset of interfering signals leaks through the main lobes, resulting in a limited impact on satellite reception.

TABLE OF CONTENTS

Document Revision History	2
Disclaimer	2
Copyright notice	3
1 INTRODUCTION	22
1.1 Common framework	22
1.2 Organization of the document	23
2 MAPPING NTN BEAMS TO NR CELLS	25
2.1 Single NR cell per NTN beam	25
2.1.1 Intra-satellite mobility aspects	25
2.2 Multi-NTN beam cell	31
2.2.1 Intra-satellite mobility aspects	32
2.3 Concluding remarks	35
3 APPLICATION OF BEAM HOPPING TO NTN	36
3.1 Beam footprint and beam layout	36
3.1.1 Beam widening by phase tapering	41
3.1.2 Power equalization	44
3.1.3 Numerical evaluation	44
3.2 Beam illumination plan	48
3.2.1 Broadcast signalling	49
3.2.2 RACH signalling	58
3.3 Earth-Moving Beams	61
3.3.1 Beam layout	62
3.3.2 Beam hopping/multiplexing design	62
3.3.3 Numerical evaluation	64
3.3.4 Earth-fixed vs Earth-moving	66
4 USER-CENTRIC BEAMFORMING	67
4.1 System definition	67
4.1.1 Architecture and network elements	67
4.1.2 System-level procedures	69
4.1.3 Beamforming algorithms and ancillary information	70
4.2 User location and traffic distribution	73
4.2.1 Residential broadband scenario	74
4.2.2 Air traffic scenario	80
4.3 System-level analyses	81
4.3.1 Standalone beamforming	81

4.3.2	Federated beamforming	134
4.4	Link-level analyses	145
4.4.1	OTFS system description	145
4.4.2	OTFS modulation/demodulation via Discrete Zak transform	148
4.4.3	Novel interference cancellation (IC) detector	154
4.4.4	Numerical results	169
4.4.5	Advanced receiver processing over fading channel.....	175
5	SPECTRUM SHARING	183
5.1	Parameters	184
5.2	INTERFERENCE ANALYSIS.....	187
5.3	NUMERICAL RESULTS.....	190
5.3.1	Scenario 1	191
5.3.2	Scenario 2.....	191
6	CONCLUSIONS	195
7	REFERENCES	200
ANNEX A: EXTENDED PERFORMANCE ASSESSMENT WITH UNIFORM DISTRIBUTIONS		
	203	
7.1	Overview of the system model	203
7.2	Performance Assessment.....	206
7.2.1	Polar constellation	206
7.2.2	Inclined constellation	212
7.3	Lessons learnt	217

LIST OF FIGURES

FIGURE 1: ONE-TO-ONE MAPPING BETWEEN NTN BEAMS AND NR CELLS	25
FIGURE 2: ENTRY CONDITION FOR EVENT D1.	27
FIGURE 3: RACH-LESS HO PROCEDURES	28
FIGURE 4: DAPS HANDOVER.....	30
FIGURE 5: MANY-TO-ONE MAPPING BETWEEN NTN BEAMS AND NR CELLS.....	32
FIGURE 6: BEAM ACQUIREMENT FOR INITIAL ATTACH	33
FIGURE 7: CONNECTED MODE PROCEDURES P-1, P-2, AND P-3	33
FIGURE 8: BEAM MANAGEMENT PROCEDURES FOR NR.....	34
FIGURE 9: -3 DB BEAM FOOTPRINTS FOR AN EARTH-FIXED GRID ACCORDING TO: (LEFT) THE BEAM FOOTPRINT AT NADIR; (RIGHT) BEAM FOOTPRINT AT COVERAGE EDGE. 37	
FIGURE 10: ANTENNA CONFIGURATION (LEFT) AND RADIATION PATTERN CUTS FOR DIFFERENT AZIMUTH ANGLES (Φ) WHEN POINTING AT NADIR.	38
FIGURE 11: -3 DB BEAM CONTOURS OF BEAMS MAPPED TO BEAM INDEX #0 FOR NINDEX=64 (LEFT) AND NINDEX=61 (RIGHT).....	40
FIGURE 12: EXAMPLE OF INDEX MAPPING FOR NINDEX=61.	41
FIGURE 13: COMPARISON OF BEAM WIDENING SOLUTIONS	43
FIGURE 14: BEAM DESIGN FOR ALTERNATIVE A (UNIFORM BEAMS). LEFT: BEAM FOOTPRINTS OVER EARTH; RIGHT: BEAM CONTOURS IN U-V PLANE.	44
FIGURE 15: CDF OF SINR OVER THE WHOLE COVERAGE. LEFT: MAXIMUM EIRP BEAMS (DESIGN B); RIGHT: MAXIMUM EIRP BEAMS WITH BEAM CROSSOVER AT 6° (DESIGN D). 45	
FIGURE 16: CDF OF SINR OVER THE WHOLE COVERAGE FOR THE DESIGNS ACHIEVING 95% COVERAGE OVER 3 DB, SUMMARIZED IN TABLE 5. LEFT: FULL PLOT; RIGHT: ZOOM TO LOW PERCENTILE RANGE.	46
FIGURE 17: NUMBER OF BEAM INDEXES (TIME RESOURCES) REQUIRED TO ILLUMINATE THE WHOLE COVERAGE FOR THE DESIGNS SELECTED IN TABLE 5.....	47
FIGURE 18: SINR DISTRIBUTION ACROSS THE COVERAGE AREA FOR THE TWO MOST EFFICIENT DESIGNS: D3 (LEFT) AND A (RIGHT).	48
FIGURE 19: NR SSB STRUCTURE.....	50
FIGURE 20: CORESET#0 AND SSB SCHEDULING FOR MULTIPLEXING PATTERN 3	52
FIGURE 21: COMMON SIGNALLING ALLOCATION ASSUMING TWO SSB TRANSMISSIONS PER SLOT (LEFT) AND A SINGLE TRANSMISSION PER SLOT (RIGHT)	55
FIGURE 22: CS SCHEDULING TIMING COMPARISON FOR SSB PERIODICITY OF 20 MS AND DESIGNS D3 AND B IN SECTION 3.1.3. GREEN SUBFRAMES ARE DEVOTED TO DATA TRANSMISSION, ORANGES ARE SHARED BETWEEN CS AND DATA TRANSMISSION, AND REDS ARE EXCLUSIVELY USED FOR CS.	57
FIGURE 23: ONE SSB PER RACH OCCASION: (LEFT) ONE PRACH TRANSMISSION IN ONE TIME INSTANCE, (RIGHT) TWO PRACH TRANSMISSIONS IN ONE TIME INSTANT.	58
FIGURE 24: 2 SSBs PER RACH OCCASION.	58
FIGURE 25: ONE SSB IN 2 RACH OCCASIONS: (LEFT) ONE PRACH TRANSMISSION IN ONE TIME INSTANCE, (RIGHT) TWO PRACH TRANSMISSIONS IN ONE TIME INSTANT,	59

FIGURE 26: UL FRAME STRUCTURE FOR RANDOM ACCESS SIGNALLING AND DATA TRANSMISSION	59
FIGURE 27: -3 DB BEAM CONTOURS FOR SCHEMES E (RED) AND F (RED AND BLUE) ..	62
FIGURE 28: PROBABILITY OF SINR<3DB PER BEAM INDEX AS A FUNCTION OF THE NUMBER OF BEAMS SIMULTANEOUSLY ILLUMINATED WITH THE SAME INDEX FOR NINDEX=41	63
FIGURE 29: PERCENTILE 5 OF SINR ACROSS THE COVERAGE AS A FUNCTION OF THE NUMBER OF BEAM INDEXES (GROUPS) NINDEX FOR SCHEME E.....	64
FIGURE 30: PERCENTILE 5 OF SINR ACROSS THE COVERAGE FOR DIFFERENT BEAM LAYOUT SCHEMES.	65
FIGURE 31: DISTRIBUTION OF COVERAGE POINTS WITH SINR<3DB.....	66
FIGURE 32: DISTRIBUTED BEAMFORMING MANAGEMENT (DBFM) ARCHITECTURE WITH ONE OR MORE NTN NODES.	67
FIGURE 33: HIGH-LEVEL RELATIONSHIP BETWEEN SCHEDULING AND REPORTING INTERVALS, [1].	69
FIGURE 34: DBFM SYSTEM-LEVEL PROCEDURES.....	70
FIGURE 35: ANCILLARY INFORMATION ESTIMATION AND REPORTING IN CSI-BASED BEAMFORMING.	71
FIGURE 36: ESTIMATION OF THE OVERHEAD PER USER, PER RADIATING ELEMENT, PER REPORTING PERIOD, TO PROVIDE THE CSI FEEDBACK TO THE NTN SEGMENT AS A FUNCTION OF THE BITS PER COMPLEX CHANNEL COEFFICIENT.....	72
FIGURE 37: ANCILLARY INFORMATION ESTIMATION AND REPORTING IN LOCATION-BASED BEAMFORMING.....	73
FIGURE 38: CONSIDERED AREAS FOR THE ELABORATION OF THE JRC GHSL DATASET.	73
FIGURE 39: POPULATION DATA FOR THE SELECTED AREA OF INTEREST.....	74
FIGURE 40: METHODOLOGY FOR THE GENERATION OF THE NON-UNIFORM TRAFFIC DATASET.....	75
FIGURE 41: CLASSIFICATION OF THE GRID POINTS BASED ON THE POPULATION.....	76
FIGURE 42: POPULATION PER GRID POINT (LEFT) AND CLASSIFICATION (RIGHT) IN THE SERVICE AREA FOR THE POLAR ORBIT CONSTELLATION (1000 KM) AND $\varepsilon S = 30^\circ$.	76
FIGURE 43: AVERAGE TRAFFIC OVER ALL BEAMS ON THE FORWARD AND RETURN LINKS AS A FUNCTION OF THE TIME OF THE DAY, OBTAINED FROM THE HISPASAT DATASET.....	78
FIGURE 44: EXAMPLE OF POPULATION DISTRIBUTION FOR ENERGY SAVING (LEFT) AND DIGITAL DIVIDE (RIGHT) SCENARIOS, POLAR ORBIT CONSTELLATION (1000 KM) AND $\varepsilon S = 30^\circ$	79
FIGURE 45: EXAMPLE OF TRAFFIC REQUEST DISTRIBUTION WITH $C_{min} = 5$ blue, 20 orange MBPS AND $C_{max} = 100$ solid, 500 dashed MBPS, POLAR ORBIT CONSTELLATION (1000 KM) AND $\varepsilon S = 30^\circ$	80
FIGURE 46: EXAMPLE OF TRAFFIC DISTRIBUTION PER AIRPORT IN THE SWITCH AT BOARDING USE CASE.....	81
FIGURE 47: SERVICE AREA AND GROUND TRACK FOR STANDALONE USER-CENTRIC BEAMFORMING WITH $latS = 45^\circ N$, $lonS = 5^\circ E$, $\varepsilon S = 30^\circ$, $hsat = 1000$ KM (5G-STAR DUST POLAR CONSTELLATION, $i = 99^\circ$).....	82

FIGURE 48: EXAMPLE OF VISIBLE USERS ACCORDING TO THE NTN NODE POSITION ON ITS ORBIT WITH $latS = 45^\circ N, lonS = 5^\circ E, \varepsilon S = 30^\circ, h_{sat} = 1000 \text{ KM}$ (5G-STARUST POLAR CONSTELLATION, $i = 99^\circ$).....	84
FIGURE 49: EXAMPLE OF SCHEDULED USERS AT 12:07:38 WITH DIFFERENT VALUES OF THE GEOGRAPHICAL SCHEDULING CONSTRAINT (SEE SECTION 4.3.1.2.)	85
FIGURE 50: EXAMPLE OF GEOGRAPHICAL SCHEDULING CONSTRAINT IN EQ. (20) WITH POTENTIAL USERS AND BARRED USERS IN THE POLAR CONSTELLATION (INITIAL NTN NODE POSITION ON ITS ORBIT).	90
FIGURE 51: EXAMPLE OF CSI SCHEDULING CONSTRAINT IN EQ. (21) WITH POTENTIAL USERS AND BARRED USERS IN THE POLAR CONSTELLATION (INITIAL NTN NODE POSITION ON ITS ORBIT).	91
FIGURE 52: HIGH-LEVEL DIAGRAM OF THE PRIORITY QUEUE SCHEDULER.	92
FIGURE 53: HIGH-LEVEL DIAGRAM OF THE PRIORITY RANK SCHEDULER.	96
FIGURE 54: BIG-O COMPUTATIONAL COMPLEXITY OF THE FIXED PHASE IN PRS AND PQS WITH $NPC = 2,4,6,8,10$ AND $N_{sched} = 200$	102
FIGURE 55: ACHIEVED AND UNMET CAPACITY WITH PQS AND GEOGRAPHICAL SCHEDULING CONSTRAINT, STANDALONE NTN NODE. SETUP: $C_{min} = 5 \text{ MBPS}$, $C_{max} = 100 \text{ MBPS}$, $\sigma_{vis} = 10$, $\sigma_{cap} = 2$	105
FIGURE 56: ASYMPTOTIC ACHIEVED CAPACITY AND ALLOCATIONS PER USER WITH PQS AND GEOGRAPHICAL SCHEDULING CONSTRAINT, STANDALONE NTN NODE. SETUP: $C_{min} = 5 \text{ MBPS}$, $C_{max} = 100 \text{ MBPS}$, $\sigma_{vis} = 10$, $\sigma_{cap} = 2$. (LB-)MMSE (SOLID LINE) AND CBF (DASHED LINE).....	106
FIGURE 57: PERCENTAGE OF UNUSED BEAMS WITH RESPECT TO THE MAXIMUM NB WITH PQS AND GEOGRAPHICAL SCHEDULING CONSTRAINT, STANDALONE NTN NODE. SETUP: $C_{min} = 5 \text{ MBPS}$, $C_{max} = 100 \text{ MBPS}$, $\sigma_{vis} = 10$, $\sigma_{cap} = 2$. (LB-)MMSE (SOLID LINE) AND CBF (DASHED LINE).....	107
FIGURE 58: ACHIEVED AND UNMET CAPACITY WITH PQS AND CSI SCHEDULING CONSTRAINT, STANDALONE NTN NODE. SETUP: $C_{min} = 5 \text{ MBPS}$, $C_{max} = 100 \text{ MBPS}$, $\sigma_{vis} = 10$, $\sigma_{cap} = 2$	107
FIGURE 59: ASYMPTOTIC ACHIEVED CAPACITY AND ALLOCATIONS PER USER WITH PQS AND CSI SCHEDULING CONSTRAINT, STANDALONE NTN NODE. SETUP: $C_{min} = 5 \text{ MBPS}$, $C_{max} = 100 \text{ MBPS}$, $\sigma_{vis} = 10$, $\sigma_{cap} = 2$. (LB-)MMSE (SOLID LINE) AND CBF (DASHED LINE).	108
FIGURE 60: PERCENTAGE OF UNUSED BEAMS WITH RESPECT TO THE MAXIMUM NB WITH PQS AND CSI SCHEDULING CONSTRAINT, STANDALONE NTN NODE. SETUP: $C_{min} = 5 \text{ MBPS}$, $C_{max} = 100 \text{ MBPS}$, $\sigma_{vis} = 10$, $\sigma_{cap} = 2$. (LB-)MMSE (SOLID LINE) AND CBF (DASHED LINE).....	109
FIGURE 61: ACHIEVED AND UNMET CAPACITY WITH GEOGRAPHICAL SCHEDULING CONSTRAINT WITH (LB-)MMSE BEAMFORMING IN THE STANDALONE SCENARIO. SETUP: $\sigma_{vis} = 10$, $\sigma_{cap} = 2$	110
FIGURE 62: PERCENTAGE OF UNUSED BEAMS WITH RESPECT TO THE MAXIMUM NB WITH PQS AND GEOGRAPHICAL SCHEDULING CONSTRAINT, STANDALONE NTN NODE. SETUP: $C_{min} = 5 \text{ MBPS}$, $C_{max} = 100,500 \text{ MBPS}$, $\sigma_{vis} = 10$, $\sigma_{cap} = 2$	111
FIGURE 63: ACHIEVED AND UNMET CAPACITY WITH GEOGRAPHICAL SCHEDULING CONSTRAINT WITH CBF BEAMFORMING IN THE STANDALONE SCENARIO. SETUP: $\sigma_{vis} = 10$, $\sigma_{cap} = 2$	112

FIGURE 64: ACHIEVED AND UNMET CAPACITY WITH CSI SCHEDULING CONSTRAINT WITH MMSE BEAMFORMING IN THE STANDALONE SCENARIO. SETUP: $\sigma_{vis} = 10$, $\sigma_{cap} = 2$	112
FIGURE 65: ACHIEVED AND UNMET CAPACITY WITH CSI SCHEDULING CONSTRAINT WITH CBF BEAMFORMING IN THE STANDALONE SCENARIO. SETUP: $\sigma_{vis} = 10$, $\sigma_{cap} = 2$	113
FIGURE 66: ACHIEVED AND UNMET CAPACITY WITH GEOGRAPHICAL SCHEDULING CONSTRAINT WITH (LB-)MMSE BEAMFORMING IN THE STANDALONE SCENARIO. SETUP: $\sigma_{vis} = 50, 150$, $\sigma_{cap} = 2$	115
FIGURE 67: ACHIEVED AND UNMET CAPACITY WITH GEOGRAPHICAL SCHEDULING CONSTRAINT WITH CBF BEAMFORMING IN THE STANDALONE SCENARIO. SETUP: $\sigma_{vis} = 50, 150$, $\sigma_{cap} = 2$	115
FIGURE 68: ACHIEVED AND UNMET CAPACITY WITH CSI SCHEDULING CONSTRAINT WITH (LB-)MMSE BEAMFORMING IN THE STANDALONE SCENARIO. SETUP: $\sigma_{vis} = 50, 150$, $\sigma_{cap} = 2$	116
FIGURE 69: ACHIEVED AND UNMET CAPACITY WITH CSI SCHEDULING CONSTRAINT WITH CBF BEAMFORMING IN THE STANDALONE SCENARIO. SETUP: $\sigma_{vis} = 50, 150$, $\sigma_{cap} = 2$	117
FIGURE 70: ACHIEVED AND UNMET CAPACITY WITH CSI SCHEDULING CONSTRAINT WITH (LB-)MMSE BEAMFORMING IN THE STANDALONE SCENARIO. SETUP: $\sigma_{vis} = 50$, $\sigma_{cap} = 1.2, 1.5$	118
FIGURE 71: ACHIEVED AND UNMET CAPACITY WITH CSI SCHEDULING CONSTRAINT WITH CBF BEAMFORMING IN THE STANDALONE SCENARIO. SETUP: $\sigma_{vis} = 50$, $\sigma_{cap} = 1.2, 1.5$	119
FIGURE 72: ACHIEVED AND UNMET CAPACITY WITH PQS AND GEOGRAPHICAL SCHEDULING CONSTRAINT, STANDALONE NTN NODE. SETUP: $C_{min} = 20$ MBPS, $\sigma_{vis} = 10$, $\rho_{GEO} = 4$, $\alpha = 0.2$	119
FIGURE 73: ACHIEVED AND UNMET CAPACITY WITH PQS AND CSI SCHEDULING CONSTRAINT, STANDALONE NTN NODE. SETUP: $C_{min} = 20$ MBPS, $\sigma_{vis} = 10$, $\rho_{CSI} = 0.04$, $\alpha = 0.2$	120
FIGURE 74: ACHIEVED AND UNMET CAPACITY WITH PRS AND CSI SCHEDULING CONSTRAINT, STANDALONE NTN NODE. SETUP: $C_{min} = 5$ MBPS, $C_{max} = 100$ MBPS, $\psi_C = 0.94$	121
FIGURE 75: ASYMPTOTIC ACHIEVED CAPACITY AND ALLOCATIONS PER USER WITH PRS AND CSI SCHEDULING CONSTRAINT, STANDALONE NTN NODE. SETUP: $C_{min} = 5$ MBPS, $C_{max} = 100$ MBPS, $\psi_C = 0.94$. (LB-)MMSE (SOLID LINE) AND CBF (DASHED LINE).....	122
FIGURE 76: PERCENTAGE OF UNUSED BEAMS WITH RESPECT TO THE MAXIMUM N_B WITH PRS AND CSI SCHEDULING CONSTRAINT, STANDALONE NTN NODE. SETUP: $C_{min} = 5$ MBPS, $C_{max} = 100$ MBPS, $\psi_C = 0.94$	122
FIGURE 77: ACHIEVED AND UNMET CAPACITY WITH CBF AND PRS WITH THE CSI SCHEDULING CONSTRAINT, STANDALONE NTN NODE. SETUP: $C_{min} = 20$ MBPS, $C_{max} = 100$ MBPS, $\rho_{CSI} = 0.05$	124
FIGURE 78: ACHIEVED AND UNMET CAPACITY WITH CSI SCHEDULING CONSTRAINT AND (LB-)MMSE BEAMFORMING IN THE STANDALONE SCENARIO. OPTIMAL PRS CONFIGURATION.	124
FIGURE 79: ACHIEVED AND UNMET CAPACITY WITH CSI SCHEDULING CONSTRAINT AND CBF BEAMFORMING IN THE STANDALONE SCENARIO. OPTIMAL PRS CONFIGURATION.	125

FIGURE 80: ACHIEVED AND UNMET CAPACITY WITH CSI SCHEDULING CONSTRAINT AND CBF BEAMFORMING IN THE STANDALONE SCENARIO. OPTIMAL PQS CONFIGURATION WITH $T_{sched} = 10$ MS AND $C_{min} = 5$ MBPS.....	125
FIGURE 81: PERFORMANCE WITH NON-IDEAL LOCATION ESTIMATION. SETUP: $C_{min} = 5$ MBPS, $C_{max} = 100$ MBPS, PRS WITH $\psi_{cap} = 1$ AND $\rho_{CSI} = 0.05$	128
FIGURE 82: PERFORMANCE WITH NON-IDEAL LOCATION ESTIMATION. SETUP: $C_{min} = 5$ MBPS, $C_{max} = 500$ MBPS, PRS WITH $\psi_{cap} = 0.98$ AND $\rho_{CSI} = 0.05$	128
FIGURE 83: PERFORMANCE WITH NON-IDEAL LOCATION ESTIMATION. SETUP: $C_{min} = 20$ MBPS, $C_{max} = 100$ MBPS, PRS WITH $\psi_{cap} = 1$ AND $\rho_{CSI} = 0.03$	129
FIGURE 84: PERFORMANCE WITH NON-IDEAL LOCATION ESTIMATION. SETUP: $C_{min} = 20$ MBPS, $C_{max} = 500$ MBPS, PRS WITH $\psi_{cap} = 0.94$ AND $\rho_{CSI} = 0.04$	129
FIGURE 85: PERFORMANCE WITH NON-IDEAL CSI ESTIMATION. SETUP: $C_{min} = 5$ MBPS, $C_{max} = 100$ MBPS, PRS WITH $\psi_{cap} = 1$ AND $\rho_{CSI} = 0.05$	130
FIGURE 86: PERFORMANCE WITH NON-IDEAL CSI ESTIMATION. SETUP: $C_{min} = 5$ MBPS, $C_{max} = 500$ MBPS, PRS WITH $\psi_{cap} = 0.98$ AND $\rho_{CSI} = 0.05$	130
FIGURE 87: PERFORMANCE WITH NON-IDEAL CSI ESTIMATION. SETUP: $C_{min} = 20$ MBPS, $C_{max} = 100$ MBPS, PRS WITH $\psi_{cap} = 1$ AND $\rho_{CSI} = 0.03$	131
FIGURE 88: PERFORMANCE WITH NON-IDEAL CSI ESTIMATION. SETUP: $C_{min} = 20$ MBPS, $C_{max} = 500$ MBPS, PRS WITH $\psi_{cap} = 0.94$ AND $\rho_{CSI} = 0.04$	131
FIGURE 89: BEAM LATTICES CONSIDERED IN THE BEAM-CENTRIC ANALYSIS.	132
FIGURE 90: HIGH-LEVEL DIAGRAM OF THE BEAM-CENTRIC PRS (BC-PRS).....	132
FIGURE 91: PERFORMANCE COMPARISON BETWEEN PRS (USER-CENTRIC) AND BC-PRS (BEAM-CENTRIC).....	133
FIGURE 92: SCENARIO FOR CONSTELLATION VISIBILITY ASSESSMENT.	135
FIGURE 93: GROUND TRACKS OF THE TWO 5G-STAR DUST SUB-CONSTELLATIONS. .	135
FIGURE 94: VISIBILITY ANALYSIS OF THE 5G-STAR DUST INCLINED SUB-CONSTELLATION.	136
FIGURE 95: VISIBILITY ANALYSIS OF THE 5G-STAR DUST POLAR SUB-CONSTELLATION.	137
FIGURE 96: SERVICE AREA AND GROUND TRACK FOR FEDERATED BEAMFORMING WITH $lat_S = 48.42^\circ N, lon_S = 3.68^\circ E, \varepsilon_S = 30^\circ, h_{sat} = 1300$ KM (5G-STAR DUST INCLINED CONSTELLATION, $i = 55^\circ$), $N_{sat} = 2$	138
FIGURE 97: EXAMPLE OF VISIBLE USERS ACCORDING TO THE NTN NODES POSITION ON THEIR ORBIT WITH $lat_S = 48.42^\circ N, lon_S = 3.68^\circ E, \varepsilon_S = 30^\circ, h_{sat} = 1300$ KM (INCLINED CONSTELLATION, $i = 55^\circ$), $N_{sat} = 2$	139
FIGURE 98: ACHIEVED AND UNMET CAPACITY WITH FEDERATED (LB-)MMSE BEAMFORMING, PRS WITH $\psi_{cap} = 1, C_{min} = 20$ MBPS.	142
FIGURE 99: ACHIEVED AND UNMET CAPACITY WITH FEDERATED CBF BEAMFORMING, PRS WITH $\psi_{cap} = 1, C_{min} = 20$ MBPS.	142
FIGURE 100: ACHIEVED AND UNMET CAPACITY WITH FEDERATED (LB-)MMSE BEAMFORMING, PRS WITH $\psi_{cap} = 0.98$	143
FIGURE 101: ACHIEVED AND UNMET CAPACITY WITH FEDERATED CBF BEAMFORMING, PRS WITH $\psi_{cap} = 0.98, C_{min} = 20$ MBPS.....	144
FIGURE 102: ACHIEVED AND UNMET CAPACITY WITH FEDERATED (LB-)MMSE BEAMFORMING, PRS WITH $\psi_{cap} = 1, C_{min} = 5$ MBPS.	144

FIGURE 103: ACHIEVED AND UNMET CAPACITY WITH FEDERATED CBF BEAMFORMING, PRS WITH $\psi_{cap} = 1$, $C_{min} = 5$ MBPS.	144
FIGURE 104: OFDM TIME FREQUENCY GRID OF ONE NR FRAME.	146
FIGURE 105: OTFS TIME FREQUENCY GRID OF ONE NR FRAME.	147
FIGURE 106: OTFS WAVEFORM FIR FILTER MAGNITUDE AND PHASE.	147
FIGURE 107: OTFS TRANSMITTED SYMBOL, WITH PILOT FOR CHANNEL ESTIMATION.	148
FIGURE 108: SCHEMATIC OF THE SFFT IMPLEMENTATION OF THE OTFS SYSTEM.....	149
FIGURE 109: SCHEMATIC OF THE OTFS TRANSMITTER.....	150
FIGURE 110: SCHEMATIC OF THE ZAK IMPLEMENTATION OF THE OTFS SYSTEM.	152
FIGURE 111: EXAMPLE OF THE CHANNEL EFFECT ON A GENERIC TRANSMITTED SYMBOL. IN THE LEFT SUBfigure, THE	154
FIGURE 112: COMPARISON OF THE ENERGY DISTRIBUTION FOR THE MATRIX Ψ (ON THE LEFT) AND G (ON THE RIGHT).....	156
FIGURE 113: STRUCTURE OF THE FG FOR THE Ψ -BASED ALGORITHMS.	158
FIGURE 114: STRUCTURE OF THE FG FOR THE MP_G ALGORITHM.	162
FIGURE 115: STRUCTURE OF THE HIERARCHICAL MP ON $HBCJRG$	163
FIGURE 116: CUMULATIVE DISTRIBUTION FUNCTION OF THE PRAGMATIC-CAPACITY-BASED THROUGHPUT FOR OFDM AND OTFS WITH TWO TRANSMITTING SATELLITES, FADING CHANNEL AND ANTENNA ARRAYS AT THE RECEIVE SIDE (MRC).....	179

LIST OF TABLES

TABLE 1: 5G-STAR DUST SUB-CONSTELLATIONS AS DEFINED IN D3.2.....	22
TABLE 2: MEASUREMENT EVENTS.	26
TABLE 3: PROS AND CONS OF THE DIFFERENT HANDOVER MECHANISMS.....	30
TABLE 4: SYSTEM PARAMETERS.....	38
TABLE 5: SUMMARY OF SIMULATION RESULTS.....	47
TABLE 6: SET OF COMMON SEARCH SPACES.....	51
TABLE 7: SET OF PHYSICAL RESOURCES ASSIGNED TO CORESET#0.....	52
TABLE 8: REQUIRED SNR TO MEET THE TARGET BLER.....	54
TABLE 9: REQUIRED TIME/FREQUENCY RESOURCES.....	55
TABLE 10: BEAM HOPPING SPECIFICATIONS	56
TABLE 11: BEAM HOPPING SYSTEM LEVEL EVALUATION FOR UPLINK	60
TABLE 12: SUMMARY OF SIMULATION RESULTS.....	65
TABLE 13: COMPARISON OF CS EFFICIENCY.....	66
TABLE 14: SUMMARY OF THE PARAMETERS TO COMPUTE THE ACTIVITY PROBABILITY PER GRID POINT.	79
TABLE 15: SUMMARY OF THE CONSIDERED BEAMFORMING ALGORITHMS AND POWER DISTRIBUTION STRATEGY.	85
TABLE 16: SIMULATION PARAMETERS FOR THE STANDALONE CASE WITH IDEAL ANCILLARY INFORMATION.	104
TABLE 17: CONFIGURATION PARAMETERS FOR THE PQS ALGORITHM.	104
TABLE 18: OPTIMAL VALUE OF THE GEOGRAPHICAL SCHEDULING CONSTRAINT WITH (LB-)MMSE BEAMFORMING IN THE STANDALONE SCENARIO. SETUP: $\sigma_{vis} = 10$, $\sigma_{cap} = 2$.....	109
TABLE 19: OPTIMAL VALUE OF THE GEOGRAPHICAL SCHEDULING CONSTRAINT WITH CBF BEAMFORMING IN THE STANDALONE SCENARIO. SETUP: $\sigma_{vis} = 10$, $\sigma_{cap} = 2$. 111	111
TABLE 20: OPTIMAL VALUE OF THE CSI SCHEDULING CONSTRAINT WITH (LB-)MMSE BEAMFORMING IN THE STANDALONE SCENARIO. SETUP: $\sigma_{vis} = 10$, $\sigma_{cap} = 2$.....	112
TABLE 21: OPTIMAL VALUE OF THE CSI SCHEDULING CONSTRAINT WITH CBF BEAMFORMING IN THE STANDALONE SCENARIO. SETUP: $\sigma_{vis} = 10$, $\sigma_{cap} = 2$. 113	113
TABLE 22: OPTIMAL VALUE OF THE GEOGRAPHICAL SCHEDULING CONSTRAINT WITH (LB-)MMSE AND CBF BEAMFORMING IN THE STANDALONE SCENARIO. SETUP: $\sigma_{vis} = 10, 50, 150$, $\sigma_{cap} = 2$.....	113
TABLE 23: OPTIMAL VALUE OF THE CSI SCHEDULING CONSTRAINT WITH (LB-)MMSE AND CBF BEAMFORMING IN THE STANDALONE SCENARIO. SETUP: $\sigma_{vis} = 10, 50, 150$, $\sigma_{cap} = 2$.....	114
TABLE 24: OPTIMAL VALUE OF THE CSI SCHEDULING CONSTRAINT WITH (LB-)MMSE AND CBF BEAMFORMING IN THE STANDALONE SCENARIO. SETUP: $\sigma_{vis} = 50$, $\sigma_{cap} = 1.2, 1.5, 2$.	117
TABLE 25: CONFIGURATION PARAMETERS FOR THE PRS ALGORITHM.	121

TABLE 26: OPTIMAL VALUE OF THE CSI SCHEDULING CONSTRAINT AND SCHEDULER SETUP ψ_C WITH (LB-)MMSE AND CBF BEAMFORMING IN THE STANDALONE SCENARIO WITH PRS.	123
TABLE 27: CONFIGURATION PARAMETERS FOR THE ANALYSIS WITH NON-IDEAL ANCILLARY INFORMATION.	127
TABLE 28: ORBITAL PARAMETERS OF THE WALKER DELTA CONFIGURATION OF THE 5G-STAR DUST SUB-CONSTELLATIONS.	134
TABLE 29: SIMULATION PARAMETERS FOR THE STANDALONE CASE WITH IDEAL ANCILLARY INFORMATION.	141
TABLE 30: SYMBOLS, SLOTS AND SUBFRAMES PER FRAME.	145
TABLE 31: COMPLEXITY COMPARISON PER SYMBOL PER ITERATION FOR THE TESTED DETECTORS.	168
TABLE 32: SIMULATION PARAMETERS.	169
TABLE 33: BLOCK ERROR RATE PERFORMANCE AS A FUNCTION OF CSI MARGIN.	172
TABLE 34: ITU-R P.681-11 CHANNEL PARAMETERS FOR HIGHWAY ENVIRONMENT	176
TABLE 35: BLOCK ERROR RATE PERFORMANCE AS A FUNCTION OF CSI MARGIN	182
TABLE 36: OPERATING BANDS IN FR2-NTN.	183
TABLE 37: OPERATING BANDS IN FR2-TN	183
TABLE 38: COEXISTENCE SCENARIOS.	184
TABLE 39: SATELLITE PARAMETERS	184
TABLE 40: NTN UE PARAMETERS	185
TABLE 41: BASE STATION ELEMENT PATTERN.	185
TABLE 42: TN GNB PARAMETERS	186
TABLE 43: UE ELEMENT PATTERN.	186
TABLE 44: TN UE PARAMETERS	187
TABLE 45: INR IN COEXISTENCE SCENARIOS 1 FOR DIFFERENT SATELLITE ELEVATION ANGLES	191
TABLE 46: SIMULATION PARAMETERS	191
TABLE 47: PARAMETERS FOR SYSTEM-LEVEL SIMULATIONS WITH A SINGLE NTN NODE AND RANDOM SCHEDULING.	205
TABLE 48: PERFORMANCE COMPARISON WITH FR3 AND FR4 (RELATIVE GAINS/LOSSES IN SQUARE BRACKETS).	212
TABLE 49: PERFORMANCE COMPARISON WITH FR3 AND FR4 (RELATIVE GAINS/LOSSES IN SQUARE BRACKETS) FOR THE INCLINED CONSTELLATION.	217

ABBREVIATIONS

5GC	5G Core network
ACLR	Adjacent Channel Leakage Ratio
ACS	Adjacent Channel Sensitivity
AF	Array Factor
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BFN	Beamforming Network
C-RNTI	Cell RNTI
CBF	Conventional Beamforming
CDF	Cumulative Distribution Function
CHO	Conditional Handover
CORESET	COntrol Resource SET
CP	Cyclic Prefix
CQI	Channel Quality Indicator
CRB	Cramer-Rao Bound
CSI	Channel State Information
CSI-RS	CSI Reference Signal
DAS	Distributed Antenna System
DAPS	Dual Active Protocol Stack
DBFM	Distributed Beamforming Management
DCI	Downlink Control Information
DD	Delay-Doppler
DL	Downlink
DM-RS	Demodulation Reference Signal
DRA	Direct Radiating Array

DZT	Discrete Zak Transform
EC	European Commission
ECEF	Earth-Centered Earth-Fixed
EIRP	Effective Isotropically Radiated Power
FFT	Fast Fourier Transform
FFR	Full Frequency Reuse
FG	Factor-Graph
FoV	Field of View
FSL	Free Space Loss
GEO	Geostationary Earth Orbit
GHSL	Global Human Settlement Layer
GNSS	Global Navigation Satellite System
GSO	Geosynchronous Orbit
GW	Gateway
HO	Handover
IC	Interference Cancellation
INR	Interference-to-Noise Ratio
ISL	Inter-Satellite Link
JRC	Joint Research Center
KPI	Key Performance Indicator
LB-MMSE	Location-Based MMSE
LEO	Low Earth Orbit
LI	Layer Indicator
LMF	Location Management Function
LOS	Line-of-Sight
MAC	Medium Access Control
MCS	Modulation and Coding Scheme

MMSE	Minimum Mean Square Error
MP	Message Passing
MPC	Maximum Power Constraint
MR	Measurement Report
NE	Network Element
NGC	Next Generation Core
NGSO	Non-Geostationary Satellite Orbit
NLOS	Non-LOS
NR	New Radio
NTN	Non-Terrestrial Network
OBBS	On-Board Beamforming and Scheduling Computation
OFDM	Orthogonal Frequency Division Multiplexing
OGBS	On-Ground Beamforming and Scheduling Computation
OSS	Operation Support Systems
OTFS	Orthogonal Time Frequency Space
PAC	Per Antenna Constraint
PBCH	Physical Broadcast Channel
PDCCH	Physical Downlink Control Channel
PDCCP	Packet Data Convergence Protocol
PHY	Physical Layer
PMF	Probability Mass Function
PMI	Precoding Matrix Indicator
PQS	Priority Queue Scheduler
PRACH	Physical Random Access Channel
PRC	Priority Rank Scheduler
PSS	Primary Synchronisation Signal
RACH	Random Access Channel

RAN	Radio Access Network
RI	Rank Indicator
RLC	Radio Link Control
RNTI	Radio Network Temporary Identifier
RRC	Radio Resource Control
RRM	Radio Resource Management
RS	Random Scheduling
RSRP	Reference Signal Received Power
RTT	Round Trip Time
RU	Radio Unit
SCS	Sub-Carrier Spacing
SFFT	Symplectic Finite Fourier Transform
SIB	System Information Block
SINR	Signal-to-Interference-plus-Noise Ratio
SLL	Side Lobe Level
SNO	Satellite Network Operator
SNP	Sub-Node Point
SNR	Signal-to-Noise Ratio
SPC	Sum Power Constraint
SSB	Synchronisation Signal Block
SSS	Secondary Synchronisation Signal
TA	Timing Advance
TN	Terrestrial Network
TTT	Time-To-Trigger
UE	User Equipment
UL	Uplink
VSAT	Very Small Aperture Terminal

1 INTRODUCTION

This deliverable D4.6 reports the final outcomes of Task 4.1 “Flexible and unified Air-Interface” and Task 4.2 “User-Centric and Digital Beamforming Solutions,”. This document is based on the inputs provided in D3.2, [2] and the preliminary report in D4.3, [1]. More specifically, in this document, 5G-STARDUST investigates aspects related to the 3GPP NR Air Interface for NTN, including

- Mapping schemes between NTN beams and NR cells;
- Application of beam hopping to NTN with the focus on the scheduling of the signalling, the beam footprint and the beam layout;
- Assessment of user-centric beamforming at system-level;
- Comparative assessment between OTFS and OFDM at link-level.

1.1 COMMON FRAMEWORK

The first part of this section is devoted to defining a common framework that is used across the deliverable to conduct various analyses.

Table 1: 5G-STARDUST sub-constellations as defined in D3.2.

Inclined orbit sub-constellation	Polar orbit sub-constellation
Altitude: ~1300 km Inclination: 50° Number of planes: 20 Number of satellites per plane: 11	Altitude: ~1000 km Inclination: 99° Number of planes: 6 Number of satellites per plane: 13
	
<u>Inclined orbit - orbital view</u>	<u>Inclined orbit – orbital view</u>

The reference constellations derived in [2], are assumed. The number of satellites is optimized to cover the target service areas. This led to following a typical design with 2 sub-constellations, tailored for the purpose of: i) ensuring worldwide coverage; and ii) satisfying higher demand over high user traffic density areas, typically between +/- 60° latitude.

In particular, unless otherwise stated, the evaluations will focus on the principal sub-constellation featuring 20 planes and 11 satellites per plane at 1300 km and an inclination of 50°. Moreover, it is assumed that each satellite covers a region corresponding to UEs seeing the satellite with a minimum elevation angle of 30°, which with the satellite at 1300 km translates to a FoV radius around 1500 km.

With respect to the satellite, this deliverable assumes the payload antenna model introduced in [2], featuring four apertures (TX-LHCP, TX-RHCP, RX-LHCP, RX-RHCP), each consisting of 512 radiating elements in an approximated circle shape with a Hexa-Deca structure, and inter-element spacing of 0.65λ . The resulting beam footprint at Nadir has a radius of approximately 41 km. The gain of each element is 4.7 dBi, their pattern can be modelled by a $\cos^2(\cdot)$ function and each have an associated power amplifier of 65 mW. Moreover, we assume digital beamforming.

In FR2, as recommended by 3GPP, the normalized antenna pattern of a Very Small Aperture Terminal (VSAT) UE is:

$$G_{UE}(\theta) = \begin{cases} 4 \left| \frac{J_1(2\pi f_c d_A \sin\theta/c)}{2\pi f_c d_A \sin\theta/c} \right|^2 & \theta \neq 0 \\ 1 & \theta = 0 \end{cases}$$

where $J_1(\cdot)$ is the Bessel function of the first kind and first order with the argument, d_A is the antenna aperture radius, θ is the angle measured from the antenna boresight and c is the light propagation speed.

1.2 ORGANIZATION OF THE DOCUMENT

This document is organised as follows:

- Section 2 compares two different mapping schemes between NTN beams and cells. i) multiple NTN beams managed by a single cell; ii) deploying as many cells as NTN beams, with handover identified as the main drawback. The comparison focuses on intra-satellite mobility aspects, with particular emphasis on signalling requirements and the latency incurred when moving across beams within a single satellite.
- Section 3 proposes a methodology to implement periodic beam hopping in NTN. A step-by-step process is provided for calculating the beamformer for all widened beams and for deriving the corresponding beam lattice. Using a predefined set of beams, this section also examines power allocation strategies to ensure that all beams transmitted simultaneously deliver comparable SINR levels to users. The procedure for conducting a comparative assessment between different designs is described in detail. The second part of the section focuses on scheduling signaling in both downlink (DL) and uplink (UL) frames, with the objective of minimizing resource consumption. In addition, methodologies are presented for computing the coverage ratio in both UL and DL.
- Section 4 is dedicated to user-centric beamforming. In particular:
 - Section 4.1 provides an overview of the main system and architecture assumptions based on previous activities within WP3 and WP4 of 5G-STARLUST.
 - Section 4.2 introduces the two methodologies for the generation of non-uniform traffic distributions, covering residential broadband and air traffic scenarios. These methodologies are extremely flexible and can be adapted to consider different traffic

requests (in terms of the minimum and maximum capacity) and system-level configurations (e.g., number of simultaneously visible satellites per user, used polarisations, and expected percentage of NTN terminals, among many others).

- Section 4.3 is dedicated to the assessment of user-centric beamforming at system-level, for both standalone and federated solutions. The performance assessment is provided under various traffic scenarios, with ideal and non-ideal CSI/location estimates at the gNB, and also observing the Big-O complexity of the scheduling algorithms. As indicated in the Executive Summary above, two scheduling algorithms have been designed, which take into account the residual unmet capacity and the residual visibility time of the serving satellite(s) per user to prioritise traffic.
- Section 4.4 reports a description of the OTFS transmitter architecture, of the two possible implementation solutions (based on SFFT or on the Zak transform, where the latter was implemented), and of the novel designed detectors; numerical results on the comparison between OFDM and OTFS in terms of pragmatic capacity, for MMSE and the novel interference cancellation detector, and ,briefly in terms of block-error rates, for both one-transmitting satellite and two transmitting satellites (for the latter only pragmatic capacity).
- Section 5 presents the spectrum landscape in the Ka-band and characterizes four coexistence scenarios between TN and NTN. These scenarios consider both directions: TN as the source of interference with NTN as the victim, and vice versa. To assess the severity of interference, link budgets are computed. The transmitter and receiver parameters for both the terrestrial and space segments are derived from 3GPP technical reports and the specifications provided in [2].
- Annex A reports an extension of the analyses initiated in the previous report, [1], focusing on uniform traffic requests and random scheduling and comparing the performance with Frequency Reuse 3 (FR3) and FR 4 (FR4) coverages. The performance is also evaluated with a variable extension of the service area extension.

2 MAPPING NTN BEAMS TO NR CELLS

In the introduction of NTN into the 3GPP ecosystem, certain aspects are left open to implementation. One key example is the synthesis of broadcast beams, which must efficiently manage system and common control signals for all users. The first challenge concerns the most suitable mapping between NTN beams and NR cells. The standard defines two possible approaches:

- One-to-one mapping, where each NTN beam corresponds directly to a single NR cell;
- Many-to-one mapping, where multiple NTN beams are logically grouped under a single NR cell.

This section examines both options, with particular emphasis on their impact on mobility.

2.1 SINGLE NR CELL PER NTN BEAM

In this mapping scheme, as shown in Figure 1, each NR cell is associated with a single NTN beam.

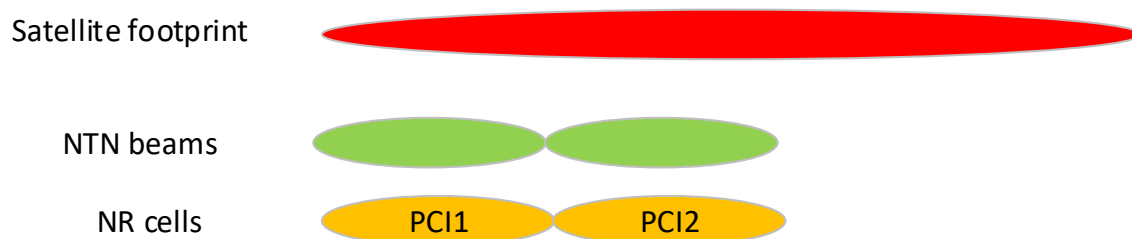


Figure 1: One-to-one mapping between NTN beams and NR cells

2.1.1 Intra-satellite mobility aspects

Mobility across NTN beams is handled as a handover, which requires rapid execution to ensure seamless communication. Intra-satellite handover refers to a specific type of handover in satellite networks, where the handover occurs between different NTN beams of the same satellite.

Conditional Handover (CHO) has been introduced by 3GPP to enhance the mobility robustness of baseline handover (BHO) [3]. The CHO configuration contains the configuration of candidate cells generated by the candidate target gNBs, while the execution conditions are generated by the source gNB. For terrestrial networks, an execution condition may consist of up to two trigger conditions. In terrestrial deployments, the users are characterized by a high SNR imbalance between the cell centre and the cell edge. In such a case, it is expected that A3 and A5 measurement-based events are enough for an efficient CHO. Other relevant measurement events are gathered in Table 2.

Table 2: Measurement events.

Measurements events	
A1	Serving cell becomes better than threshold
A2	Serving cell becomes worse than threshold
A3	Neighbour becomes offset better than serving
A4	Neighbour becomes better than threshold
A5	Serving becomes worse than threshold1 and neighbour becomes better than threshold2
A6	Neighbour becomes offset better secondary cell
B1	Inter-RAT neighbour becomes better than threshold
B2	PCell becomes worse than threshold1 and inter-RAT neighbour becomes better than threshold2
D1	Distance between UE and a reference location referencelocation1 becomes larger than configured threshold1 and distance between UE and a reference location referencelocation2 becomes shorter than configured threshold2
D2	Distance between UE and the serving cell moving reference location is above threshold1 and distance between UE and a moving reference location is below threshold2
T1	Time measured at UE is within a duration from threshold

However, in NTN, triggering events based only on signal strength might not be efficient. The main reason is because the signal strength imbalance between the cell edge and the centre is usually 3dB. Hence, the signal strength is not an informative metric to distinguish which cell shall be selected to serve the UE.

To consider the specific characteristics of the satellite links, new triggering conditions are defined for NTN. In this respect, time-based and location-based condition are introduced, as well as the A4 event.

Location-based CHO is defined by the distance between UE and a reference location. For instance, Event D1 is a new measurement triggering event based on location. For this measurement event, the UE is configured with two reference locations and two threshold distances. The reference location1 and threshold1 can be associated with the serving cell and reference location2 and threshold2 can be associated with a neighbour cell. The event D1 is triggered when the distance between UE and referenceLocation1 is larger than threshold1 and distance between UE and referencelocation2 is shorter than threshold2. The event is triggered

when an entry condition for the event is true for a certain amount of time, i.e. time-to-trigger (TTT), which is a network configured parameters.

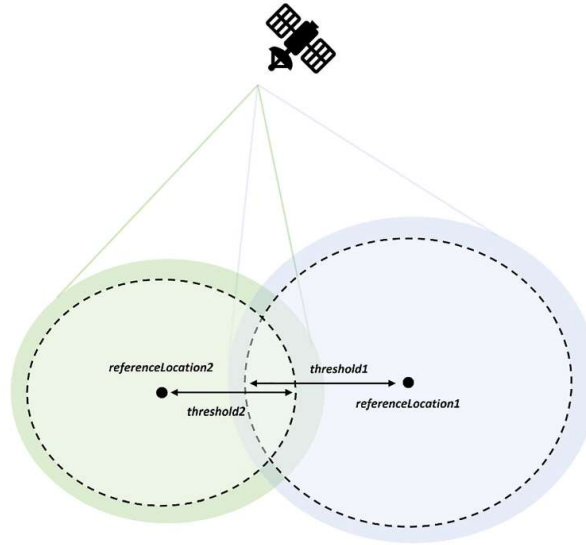


Figure 2: Entry condition for event D1.

Time-Based CHO is defined by the time between T1 and T2, where T1 is an absolute time value and T2 is a duration started at T1. The time-based CHO uses a time-trigger. Time-trigger shows how much of the serving cell availability duration is left. In other words, this time-based CHO takes into account the remaining serving time of current cell as a metric to trigger CHO.

Remarkably, in the NTN context, the gNB can still configure up to two trigger conditions per candidate target cell and one of these trigger conditions needs to be signal strength measurement-related event.

The second concern addressed in NTN is the signalling storm in case all UEs in NTN cells are in need of a CHO at a given time instant. To prevent signalling storms in earth-fixed cells, the idea is to use group HO instead of individual ones.

For the CHO mechanism, the handover interruption starts when the CHO execution condition is satisfied. After that, the UE releases the connection from the serving cell and attempts to establish a connection to the target cell.

Random access channel (RACH)-less HO has been introduced by 3GPP in [4]. Interestingly, this mechanism can be used to address challenges that are inherent to large coverage areas, such as NTN deployments. RACH-less HO reduces the overhead and the mobility interruption time with respect to BHO by omitting the RACH procedure. The key aspect of intra-satellite beam-to-beam HO within a single LEO satellite is that timing advance (TA) value is kept unchanged. This means that the source cell, the target cell and the UE are synchronized. Thus, the boundaries of the slots and the subframes between the source cell and target cell are assumed to be aligned. The synchronization of target and source cells can be guaranteed as they are managed by the same satellite. In this regard, neither the feeder link nor the gNB are switched. Hence, the source and target cells experience the same RTT delay, which allows the UE to straightforwardly derive the target cell TA from the source cell TA.

In RACH-less HO the connection is transferred from the source to the target cell via RRC signalling. Based on the fact that all the network elements are synchronized, the procedure is as follows:

- When the A3 condition (or other configured event) is met during the TTT period, the UE sends the measurement report (MR) to the source cell. The A3 event is triggered when the reference signal received power (RSRP) of the target cell exceeds that of the serving cell by a predefined offset;
- The source cell sends handover request messages to the potential target cells. The potential cells transmit the request acknowledgement (ACK);
- The source cell stops DL transmission to the UE and sends the HO command of the potential target cells;
- The UE completes the radio resource control configuration with the target cells using the information that is contained in the HO command;
- The target cell provides an uplink grant to the UE. The HO command sent by the source cell shall indicate the pre-allocated grant or the resources for physical downlink shared control channel (PDCCH) monitoring the dynamic grant;
- The UE acquires the desired target cell without an explicit TA command, by send initial UL transmission using the available UL grant;
- Consider RACH-less HO is completed upon receiving the network confirmation.

The procedures of the RACH-less HO are depicted in Figure 3.

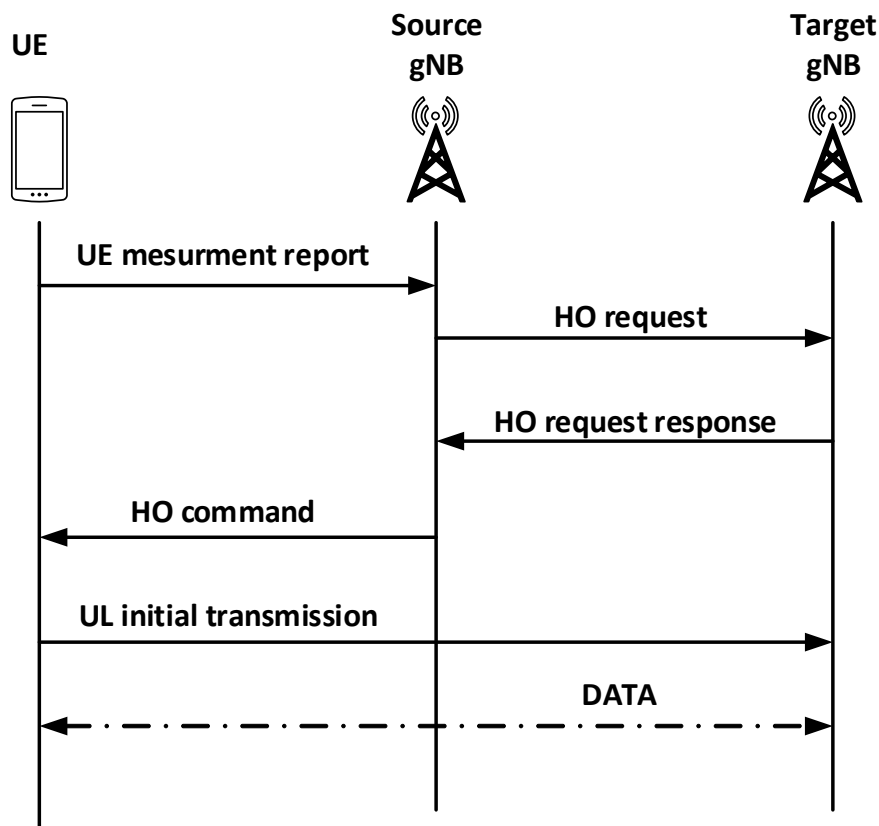


Figure 3: RACH-less HO procedures

An important aspect is that the validity of pre-allocated or dynamic grant. To make an efficient use of reserved resources, the target gNB should ensure the grant is available to the UE within the handover window. This could cause resource waste if the network adopts an aggressive approach to pre-allocate or schedule grants. By contrast, if the network adopts a too conservative approach, the HO failure rate could increase. This trade-off should be considered in the implementation.

Dual active protocol stack (DAPS) handover: In the handover strategies presented so far, the UE releases the connection from the source cell before establishing the connection with the target cell. This interruption time may be critical for time-sensitive services. According to 3GPP TS 38.133, the interruption time in LEO systems at 1300km altitude can be up to 250 ms, considering the largest PRACH configuration period. As a solution, 3GPP has proposed the dual active protocol stack (DAPS), where the UE is able to maintain the connection with the source cell, until the UE completes the random access procedure with the target cell [3]. Hence, for a short period, the UE should be able to receive data from source and target cells. During the registration procedure, the UE indicates its support for DAPS handover in the UE capability message. If the UE does not support DAPS, the network will not configure DAPS handover for the UE.

As shown in Figure 4, to implement DAPS handover is essential to keep two stacks containing physical (PHY), medium access control (MAC) and radio link control (RLC) layers with a common packet data convergence protocol (PDCP) entity.

Remarkably, when the DAPS handover fails, the UE falls back to the source cell configuration, resumes the connection with the source cell, and reports DAPS handover failure without triggering RRC connection re-establishment if the source link has not been released. This holds true if the source cell is in the field of view during the handover procedure. Otherwise, cell reselection needs to be performed after a handover attempt fails.

Another important aspect to comment is that DAPS handover is not supported for NTN.

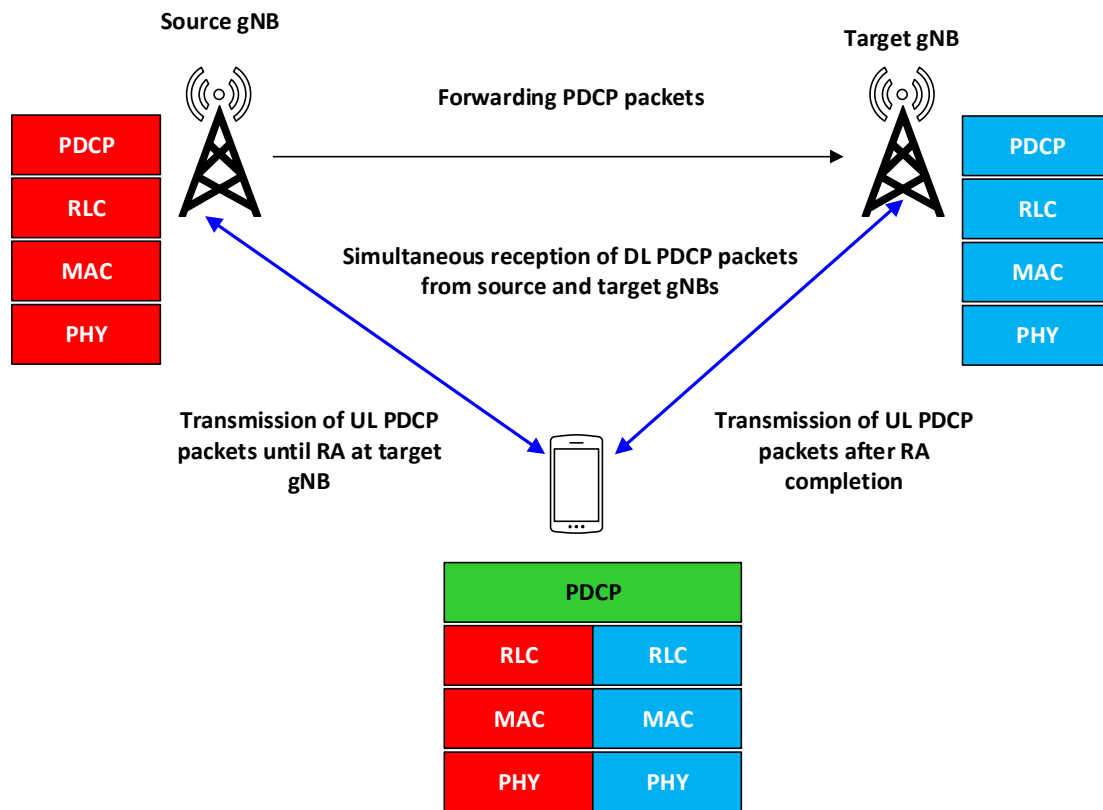


Figure 4: DAPS handover

Summary

Advantages and disadvantages of different intra-satellite handover mechanisms are provided in Table 3.

Table 3: Pros and Cons of the different handover mechanisms

Handover mechanism	Advantages	Disadvantages
Baseline handover	Supported in the standard	Small RSRP variation between the edge and the centre of the cell, which affects the handover reliability The RACH procedure leads to handover interruption time
Conditional handover	Supported in the standard UE independently selects the optimal cell Useful to overcome issue of small RSRP/RSRQ variation in regions of cell overlap The network can easily predict the triggering condition	The UE may need to perform additional measurements to execute the handover The RACH procedure leads to handover interruption time The network needs up-to-date UE coordinates

RACH-less handover	Handover interruption time is reduced by omitting the RACH procedure.	Same handover reliability as the baseline handover Signalling overhead may significantly increase if the UL grant periodicity is not properly adjusted
DAPS handover	UE makes the connection with target cell before breaking connection with the source cell, reducing the handover failure probability and eliminating the handover interruption time	Not specified for NTN Two radio frequency links shall be maintained, which increase the power consumption and the complexity in the UE

The comparative assessment reveals that the baseline handover is not well suited to handle intra-satellite mobility aspects. The main limitations are:

- Signal strength as a handover trigger is ineffective in NTN due to minimal signal strength variation within satellite coverage;
- High latency in ground-space communication links. This increases the handover interruption time.

The conditional handover offers the possibility of defining new triggering conditions, such as time- and location-based conditions. This prevents the UE from triggering the handover based on the signal strength, thus reducing the handover failure probability.

It is important to remark that the conditional handover does not reduce the interruption time with respect to the baseline solution. To this end, the RACH-less handover shall be employed. By omitting the RACH procedure, the connection can be transferred with a substantial reduction of the interruption time. However, special attention shall be paid on the UL grant periodicity to make an efficient use of the resources. For what concerns the reliability, the RACH-less handover mechanisms offers comparable performance to the baseline handover.

To improve the reliability and eliminate the handover interruption time, the most suitable option is to resort to the DAPS handover. It shall be noted that this scheme relies on the multi-connectivity concept to implement the make-before-break approach. Hence, two RF links shall be maintained during the execution of the handover. Consequently, the complexity and the power consumption will be increased. Another aspect that shall be commented is that this option has not been specified yet for NTN.

2.2 MULTI-NTN BEAM CELL

In the multi-NTN beam cell scenario, each NR cell controls multiple NTN beams, as shown in Figure 5. Hence, the legacy beam management can be used as baseline to handle the beams.

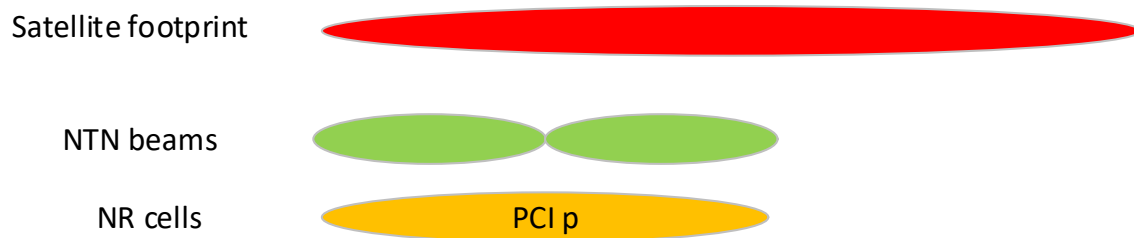


Figure 5: Many-to-one mapping between NTN beams and NR cells

2.2.1 Intra-satellite mobility aspects

Intra-cell mobility lies within the category of L1 and L2 beam management procedures to determine and update the NTN beams for each UE within a serving cell. Remarkably, beam management has been defined in 3GPP Release 15 [5]. Therefore, mobility across beams is handled as a beam switch.

Beam management has been originally conceived to achieve fine alignment of the beam pair of the transmitter and receiver at frequencies above 6 GHz. In NR, the process of identifying the most suitable beam pair is achieved through a set of operations [6]. Beam management relies on three fundamental processes: beam sweeping, beam measurement and reporting, and beam determination. Depending on the UE's state, the execution of these processes differs significantly between idle mode and connected mode.

2.2.1.1 Idle mode beam acquirement for initial attach (Initial Access)

In idle mode, the UE does not have a dedicated RRC connection, and beam management is limited to the discovery of an initial access beam for cell selection and random access.

Beam management in idle mode primarily consists of beam sweeping by the gNB and beam selection by the UE. The gNB periodically transmits synchronization signal blocks (SSBs) across multiple spatial directions to illuminate its coverage area. This sweeping allows the UE to detect and measure the received SSBs using metrics such as RSRP or SINR. The UE then selects the strongest or most suitable beam for initial access, which dictates the uplink direction for the random access channel (RACH) preamble transmission.

Beam reporting in idle mode is implicit, as the selected beam is indicated by the PRACH preamble sent on the corresponding uplink direction. This allows the gNB to align its reception beam without explicit feedback. Idle mode beam management is therefore a coarse process, limited to initial link discovery and basic alignment, with beam refinement deferred until the UE transitions to the connected state. The basic process is summarized in Figure 6.

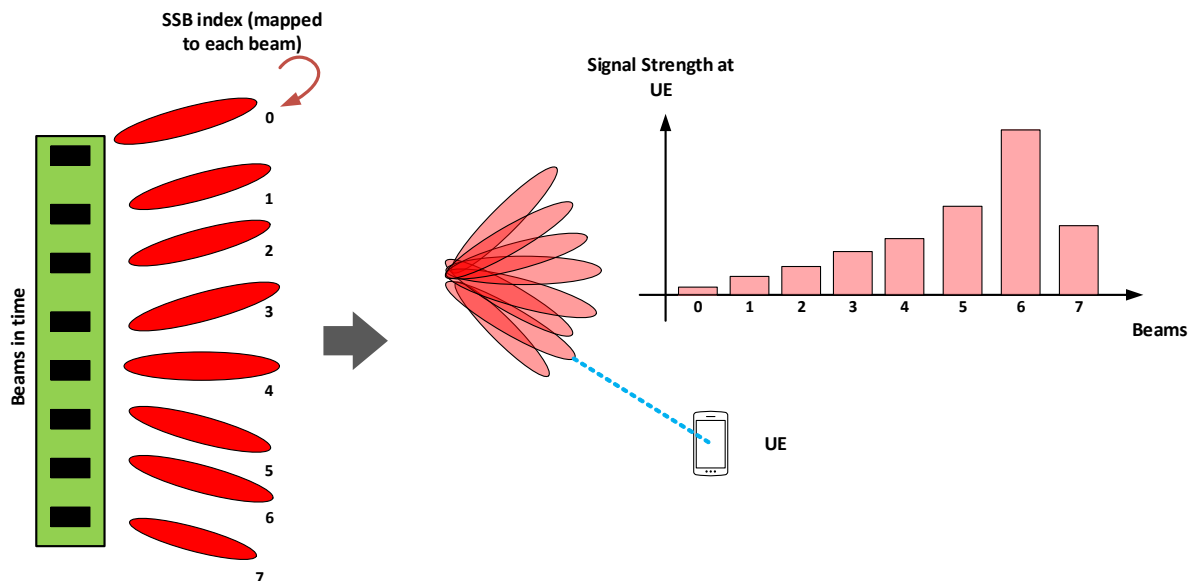


Figure 6: Beam acquisition for initial attach

2.2.1.2 Connected Mode Beam Management (RRC_CONNECTED)

Once the UE is in the RRC_CONNECTED state, beam management becomes continuous and dynamic, aiming to maintain robust connectivity, optimize link quality, and support mobility. In this mode, the UE can utilize dedicated reference signals, such as channel state information reference signals (CSI-RS), and provide explicit feedback to the gNB. Beam management follows the P-1/P-2/P-3 framework defined by 3GPP as depicted in Figure 7.

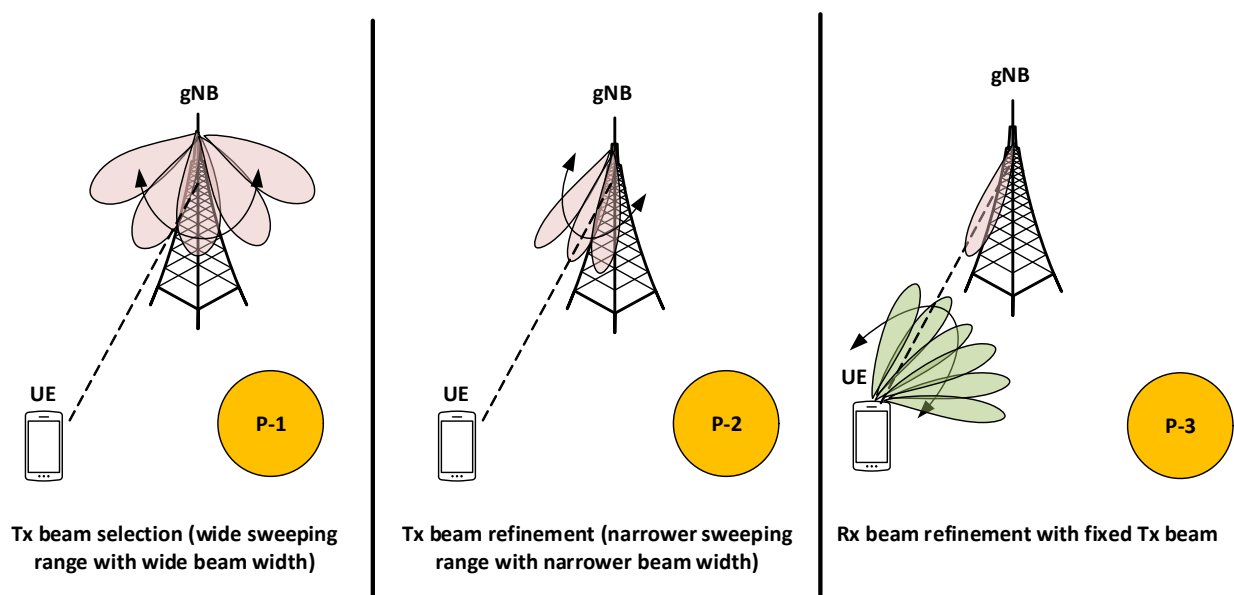


Figure 7: Connected mode procedures P-1, P-2, and P-3

The P-1/P-2/P-3 framework can be detailed as follows:

- P-1 (Initial Beam Establishment) ensures that an initial transmit–receive beam pair is selected and maintained. This P1 step contains several procedures, as depicted in Figure 7, (i) Beam sweeping is performed using both SSBs and CSI-RS. The gNB transmits

reference signals to facilitate initial acquisition and beam switch. To this end, the gNB sweeps the area of interest by sequentially transmitting multiple SSBs in different beams. Each SSB is identified with the corresponding index. For beam refinement purposes, the gNB sends multiple CSI-RSs associated to a single SSB. Remarkably, the beams associated with each CSI-RS can be narrower than those synthesized to transmit the SSB. Hence, the gNB uses dedicated signalling to configure the CSI-RSs for each UE; afterwards, (ii) Beam measurement and reporting is performed where UE tries to detect reference signals, either SSB or CSI-RS, that are transmitted in different beams. The target beams that are checked are usually adjacent to the established beam. For each detected reference signal, the UE measures the link quality of the corresponding beam pairs using RSRP. Then, the UE sends the RSRP and the index that is associated with the beamformed reference signal; (iii) Beam determination is performed in the last stage where the gNB indicates the best beam through a transmission configuration indicator (TCI). In general, the best beam is the one with the highest SNR, although the gNB could change the criterion based on potential interference. If the received data is not received by the UE or BLER is less than the predefined threshold, the UE determines that the beam selected by the gNB is not suitable (beam failure) and starts a beam recovery procedure and requests the gNB to change the beam. Beam correspondence is supported by which the UE is allowed to reuse the beam acquired in downlink for uplink transmission.

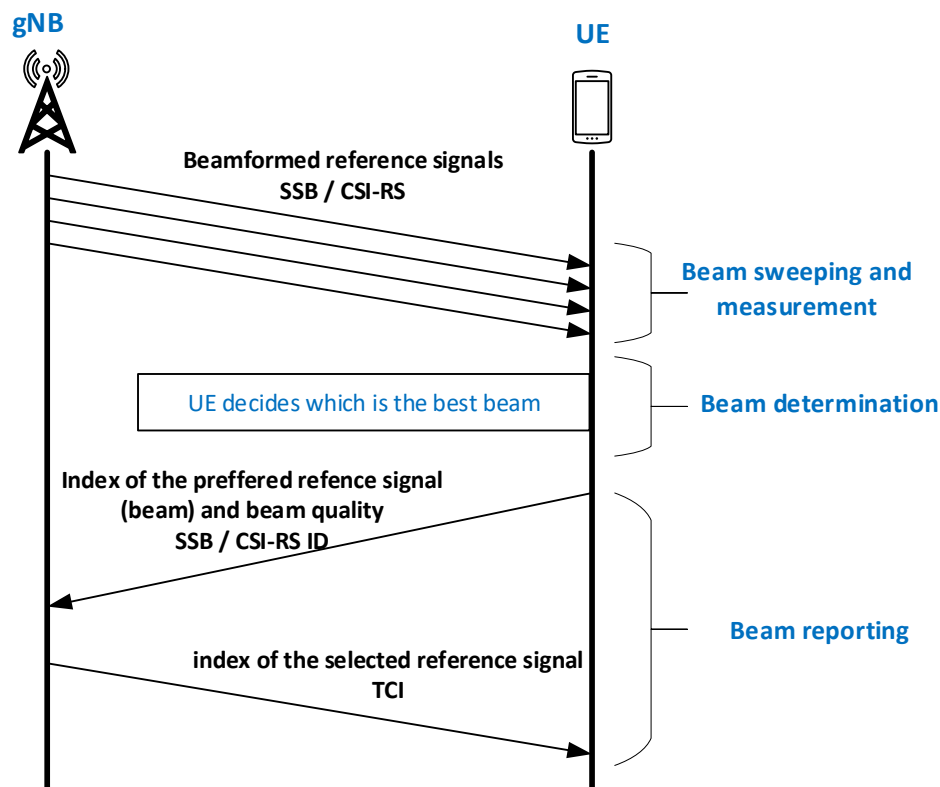


Figure 8: Beam management procedures for NR

- P-2 (Transmit Beam Refinement) focuses on optimizing the UE's uplink transmission direction. Using explicit feedback from CSI measurements, the gNB can command the UE to adjust its transmit beam to maximize uplink signal quality and ensure reliable communication, particularly under mobility or blockage conditions.

- P-3 (Receive Beam Refinement) is responsible for optimizing the gNB's downlink transmission beams. The UE performs detailed measurements on CSI-RS and SSBs and reports metrics such as RSRP, SINR, CQI, and full CSI reports. This information allows the gNB to switch, refine, or recover beams dynamically. If the active beam degrades, the UE can send a Beam Failure Indication (BFI) to trigger Beam Failure Recovery (BFR), ensuring seamless connectivity.

In connected mode, the procedure shall be periodically applied, so that each UE can update its beam pair. The periodicity is determined by the SSB period, which is 20 ms by default.

2.3 CONCLUDING REMARKS

Handover and beam management are two different mechanisms for supporting mobility in NTN. Handover operates at the cell level, ensuring service continuity as the UE transitions between NR cells, albeit with higher signalling complexity and latency. Conversely, beam management functions within a cell, providing fine-grained adaptation of the radio link through beam switching and refinement with lower signalling overhead. Considering this, the preferred option for LEO satellite systems is to rely on beam management to handle multiple NTN beams. Hence, multi-NTN beam cell scenario is considered the most suitable mapping scheme for LEO satellite systems, such as the 5G-STARDUST constellation described in Section 1.1.

3 APPLICATION OF BEAM HOPPING TO NTN

The remarkable features of the multi-beam NTN cell scenario, as outlined in Section 2, motivate the adoption of the standardized beam management procedure to design the beam illumination plan. Beam management is the term used by 3GPP to define the procedures that permit aligning transmit and receive antenna beams when directional antennas are used. In NTN, beam management at the UE side translates to traditional satellite acquisition and tracking operations. At the satellite side, beam management procedures go far beyond identifying the beam in which a given user is located. It should address the mapping of satellite beams, frequencies, and polarizations to 5G cells for an efficient use of resources. These topics have been discussed in Release 17, but no agreement was reached, so they were left open for implementations of each manufacturer/operator. Therefore, the efficient management of satellite beams is still an open problem, which becomes fundamental in an EIRP limited system like the one targeted by 5G-STARUST. The details are provided in Section 1.1. Bearing this in mind, an efficient beam hopping mechanism is required to ensure efficient full coverage while maximizing satellite capacity. However, there is a clear trade-off between the two requirements. To maximize satellite capacity, 5G-STARUST targets user centric beamforming. In contrast, efficient full coverage requires reducing the number of beams to be able to cover the whole field of view with minimum resources, especially for broadcast signalling. As a consequence, two different beam management stages are envisaged:

- Synthesis of broadcast beams to manage user-specific information as well as system and control information that is common to all users;
- Synthesis of data beams based on user centric solutions for user-specific data and control information.

Remarkably, a periodic beam hopping pattern can be used for broadcast beams. The periodic pattern for control information can be adjusted if the network has prior knowledge of the user distribution. For instance, placing coverage holes in areas with no traffic demands. In contrast, the beam hopping pattern of data beams is driven by the traffic demand and the scheduler. Consequently, the pattern of data beams is non-periodic. Data transmission will be discussed in more detail in Section 4.

It is important to remark that broadcast beams are characterized to have the same dwell time and revisit time interval and accordingly, users are served sequentially. Nevertheless, since the user density is not constant across the coverage area, serving users with broadcast beam could make an inefficient use of the resources. Hence, the use of broadcast beams should be minimized, so that resources available in data beams are maximized. This strategy is advantageous to satisfy the traffic demand. Therefore, the figure of merit is to minimize the number of broadcast beams, *i.e.*, the number of time/frequency resources required to cover the whole coverage area of a single satellite. This section focuses on the design of a beam illumination plan, starting with the optimization of the beam layout, followed by an overview description of broadcast signalling and multiple access and ending with a beam hopping scheme proposal.

3.1 BEAM FOOTPRINT AND BEAM LAYOUT

The design of an efficient beam layout for broadcast signalling was already introduced and preliminary evaluated in [1]. Here we extend that evaluation considering phase tapering for beam widening (note that in [1] a kind of upper bound considering array truncation but keeping the transmit power unaltered, so only taking into account the directivity reduction, was used);

deriving the SNR taking into account the specific beams illuminated in each time instant and a given power equalization strategy instead of a mean value over the coverage area; and introducing a new solution not based on beam widening. In any case, for the sake of completeness, we first recall the objectives, challenges, assumptions and proposed solutions. Then we discuss on beam widening strategies, and we end the section with the numerical evaluation of proposed solutions.

OBJECTIVE

To design a beam layout and identify the groups of beams to be illuminated simultaneously to minimize the number of time resources required to serve the whole coverage area.

PROBLEM STATEMENT

The challenges arise from the combination of two system parameter choices: the use of Earth-fixed beams and a rather limited available power at the satellite payload. Using a traditional phased array beamforming scheme, with uniform amplitude and phases according to the steering vector to the beam centre position, it comes out that beam footprints at coverage edges are largely widened with respect to those at Nadir, due to the Earth curvature. Therefore, to synthesize a uniform grid of beams on Earth without coverage holes, the grid needs to be selected according to the beam footprint at Nadir. As shown in Figure 9 left, this results in a very large number of beams, with high beam overlapping at coverage edges. Since the number of beams that can be simultaneously illuminated is limited by the available power, this solution may require a lot of time resources to sweep around the whole coverage area. The alternative is to design the grid according to the beam footprint at coverage edge, as depicted in Figure 9 right, which yields to a reduced number of uniform beam footprints but is affected by an EIRP reduction due to beam widening. Compromise solutions trading off the number of beams and the EIRP per beam can be realized by defining the grid of beams according to the beam footprint size anywhere from Nadir to coverage edges. In these cases, we may use beam widening when required or stick to uniform (and maximum amplitude) beams providing maximum EIRP but decreasing beam crossover levels wherever needed.

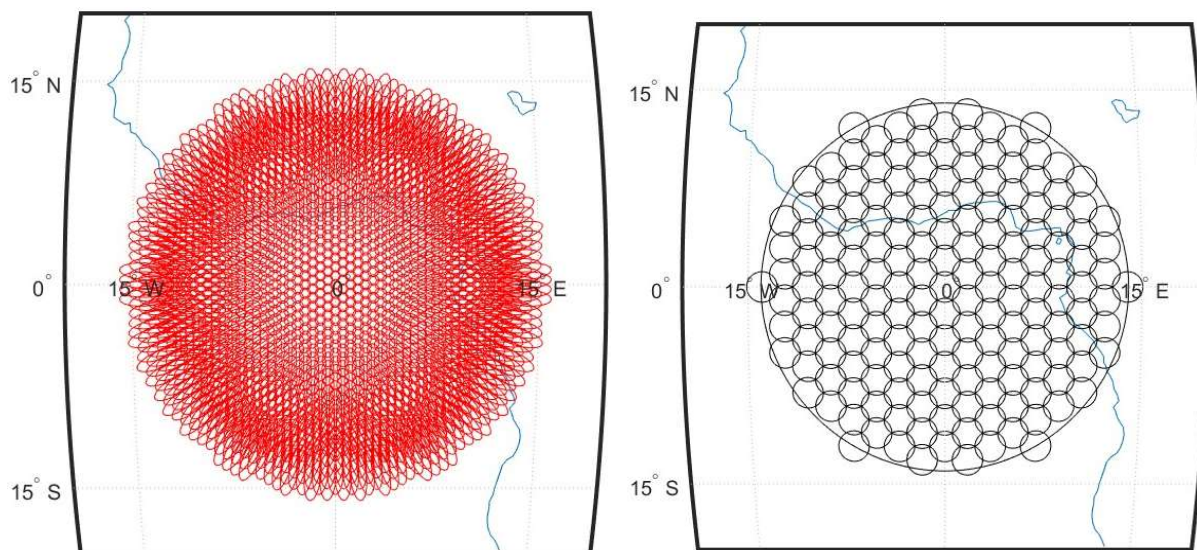


Figure 9: -3 dB beam footprints for an Earth-fixed grid according to: (left) the beam footprint at Nadir; (right) beam footprint at coverage edge.

SYSTEM CONFIGURATION

As derived in [2], the satellites orbit at 1300 km and its targeted coverage area correspond to the Earth region in which the users see the satellite with an elevation angle of 30° . It corresponds to a satellite tilt angle from Nadir of 46° . The downlink antenna, as depicted in Figure 10 left, consists of 512 radiating elements with an available power per element of 65 mW. Figure 10 right shows the resulting radiation pattern cuts at Nadir when the beamforming is a phased array pointing to Nadir. Due to the asymmetry of the antenna, the half-power beamwidth varies with the azimuth angle (φ).

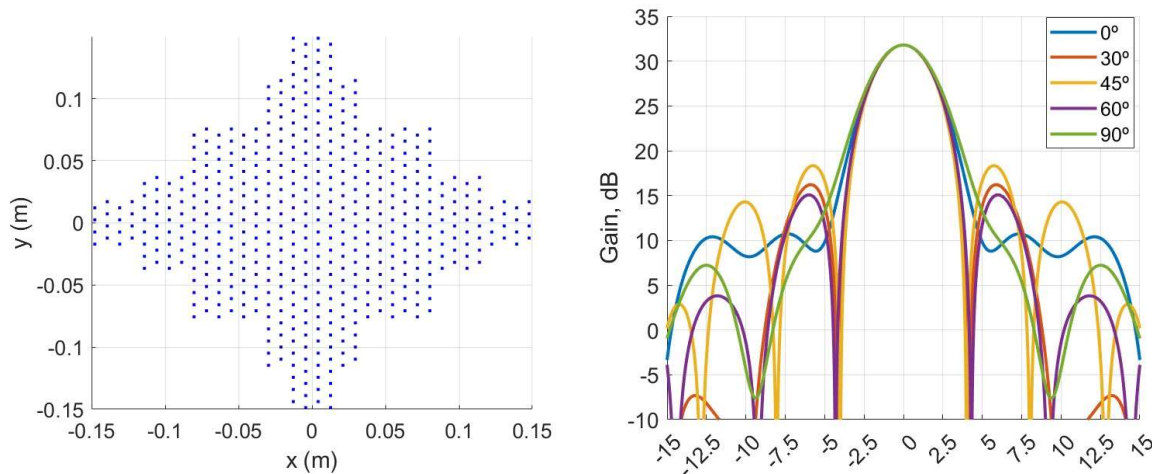


Figure 10: Antenna configuration (left) and radiation pattern cuts for different azimuth angles (φ) when pointing at Nadir.

Other system parameters used for the evaluation of the designed layouts are summarized in Table 4.

Table 4: System parameters

Parameter	Value	Comments
Frequency	20 GHz	
Bandwidth	200 MHz	
Available power	65 mW	Per antenna element
Channel	LOS	
Atmospheric loss	0.5dB @elev=90°, 0.71dB @elev=45°, 1.1dB @elev=30°	Linear interpolation used for angles in between
Shadow fading	Not considered	
Receiver G/T	11dB @elev=90°, 10dB @elev=45°, 8dB @elev=30°	Linear interpolation used for angles in between
Implementation loss	2 dB	

PERFORMANCE METRIC

The performance metric will be the number of time resources required to illuminate 95% of the coverage area with a minimum SINR of 3dB. Time resources here correspond to the number of groups of beams simultaneously illuminated. Targeting 95% is a reasonable choice though the design and evaluation method used here can be used for 99% or more challenging targets. The minimum SINR value is already a challenging choice considering that the SINR required for detection of the different broadcast signalling types varies from -4 dB to -11 dB, as summarized in Table 8. Therefore, a rather large link budget margin is considered to tackle atmospheric and shadowing effects.

BEAM DESIGNS

Four beam design strategies are considered:

- A. Uniform beams: Earth-fixed grid according to the beam footprint radius at coverage edge. Beams widened providing uniform beam footprint radius across the coverage area with minimum beam overlapping (see Figure 9 right).
- B. Maximum EIRP beams with -3dB beam crossover: Earth-fixed grid according to the beam footprint radius at Nadir. Phased-array beamformers with all antennas transmitting at uniform and maximum power, synthesizing minimum beam footprints. High beam overlapping at coverage edges. (see Figure 9 left).
- C. Non-uniform widened beams: Earth-fixed grid according to the beam footprint radius at 50% (i.e. 23°) or 75% (i.e. 32.25°) of the angular distance between Nadir and coverage edge. Beams with a tilt angle below 23°/32.25° are widened whereas above them they are not widened and present overlapping.
- D. Maximum EIRP beams with reduced beam crossover levels: Earth-fixed grid according to the beam footprint radius corresponding to half beam sizes increasing from 4.8° to 9°. This results in maximum EIRP beams presenting coverage holes close to Nadir due to large grid and reducing overlapping at coverage edge. It follows the proposal from Ericsson in [7].

MULTIPLEXING SCHEME

To evaluate the SINR perceived by the users and the number of resources required to illuminate the whole coverage, we need to define which beams are transmitted simultaneously. For this, we assign a beam index to each beam so all beams with the same beam index are transmitted together. Three conditions are imposed:

- 1) Active beam footprints must not overlap to reduce inter-beam interference.
- 2) One beam is only active in one hop within the coverage sweep, so all beams are fairly illuminated.
- 3) The multiplexing solution must not depend on the satellite position to avoid complex calculations as the satellite move over the Earth grid.

The proposed solution consists of assigning a hop index ih to each point in the Earth-fixed grid. All points within the coverage area sharing the same hop index are served simultaneously in the same hop, whereas points with different hop index are time multiplexed in different hops. The resulting number of hops required to sweep over the whole coverage $nIndex$ corresponds to the total number of hop indexes used. To fulfill conditions 2) and 3), a regular hop index assignment guaranteeing maximum distance in both latitude and longitude dimensions between beam footprints within the same hop is proposed. It maps consecutive hop indexes from $ih = 0$ to $ih = nIndex - 1$ to rectangular blocks with a number of columns ($nCols$) and a number of rows ($nRows$) equal to

$$nCols = \lceil \sqrt{nIndex} \rceil$$

$$nRows = \left\lceil \frac{nIndex}{nCols} \right\rceil.$$

where $\lceil \cdot \rceil$ is the ceil function. Rows here correspond to the longitude dimension of the Earth-grid, whereas columns correspond to latitude. Despite the Earth-fixed grid is hexagonal, square blocks still can be applied by decomposing the hexagonal grid in two slightly shifted square grids.

Let us numerically exemplify this hop index assignment process. For $nIndex=64$, blocks of 8×8 are formed, meaning that each beam index is repeated every 9 beams in both latitude and longitude dimensions. For $nIndex=56$, the blocks would be 8×7 . Between 56 and 64, blocks of 8×8 are used but we cannot fill it with unique beam indexes. For the first block we reuse the indexes starting from 0. Then next block starts with the last index used in the previous one, and we repeat this operation until a block is completed and the last index is $nIndex - 1$. For instance, for $nIndex=61$, we would need 8 blocks of 8×8 . The resulting 64×8 block is then regularly repeated along the Earth grid.

Figure 11 compares the beams mapped to SSB index #0 when $nIndex$ is 64 and 61. It can be observed that $nIndex=64$ produces a regular lattice whereas for $nIndex=61$, it is not as regular but still provides the required beam isolation. Figure 12 exemplifies the index mapping procedure for $nIndex=61$.

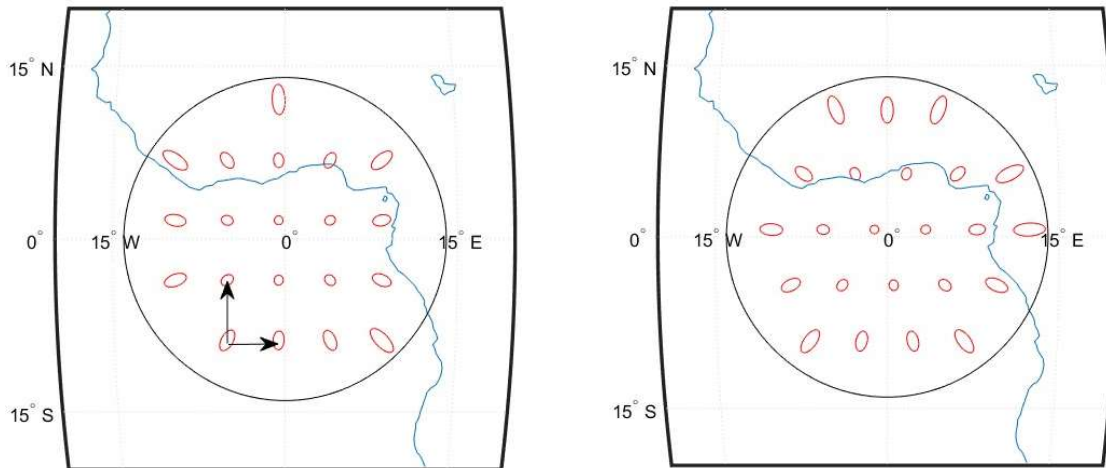


Figure 11: -3 dB beam contours of beams mapped to beam index #0 for $nIndex=64$ (left) and $nIndex=61$ (right).

This multiplexing scheme assumes that the signals transmitted simultaneously in different beams are independent. If this was not the case, the beamformer matrix would become a single beamformer vector generating multiple beams, which can be calculated as the superposition of the beamformers pointing to each of the beam centres associated with the same SSB index. Although each individual beamformer use same amplitude coefficients for all antennas, it is well known that their superposition provides amplitude variations due to constructive and destructive additions. Consequently, the overall transmit power needs to be reduced to fulfil the power constraint in the antenna experiencing maximum constructive additions. Therefore, the independent signal assumption is taken here based on the fact that the free subcarriers within the OFDM symbols of the SSB are filled with beam specific data. Moreover, different cell ID would be used in each beam impacting on the SSB sequences as well.

0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7
8	9	10	11	12	13	14	15	8	9	10	11	12	13	14	15
16	17	18	19	20	21	22	23	16	17	18	19	20	21	22	23
24	25	26	27	28	29	30	31	24	25	26	27	28	29	30	31
32	33	34	35	36	37	38	39	32	33	34	35	36	37	38	39
40	41	42	43	44	45	46	47	40	41	42	43	44	45	46	47
48	49	50	51	52	53	54	55	48	49	50	51	52	53	54	55
56	57	58	59	60	0	1	2	56	57	58	59	60	0	1	2
3	4	5	6	7	8	9	10	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	11	12	13	14	15	16	17	18
19	20	21	22	23	24	25	26	19	20	21	22	23	24	25	26
27	28	29	30	31	32	33	34	27	28	29	30	31	32	33	34
35	36	37	38	39	40	41	42	35	36	37	38	39	40	41	42
43	44	45	46	47	48	49	50	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	51	52	53	54	55	56	57	58
59	60	0	1	2	3	4	5	59	60	0	1	2	3	4	5
6	7	8	9	10	11	12	13	6	7	8	9	10	11	12	13
14	15	16	17	18	19	20	21	14	15	16	17	18	19	20	21
22	23	24	25	26	27	28	29	22	23	24	25	26	27	28	29
30	31	32	33	34	35	36	37	30	31	32	33	34	35	36	37
38	39	40	41	42	43	44	45	38	39	40	41	42	43	44	45
46	47	48	49	50	51	52	53	46	47	48	49	50	51	52	53
54	55	56	57	58	59	60	0	54	55	56	57	58	59	60	0
1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
9	10	11	12	13	14	15	16	9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24	17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32	25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56	49	50	51	52	53	54	55	56
57	58	59	60	0	1	2	3	57	58	59	60	0	1	2	3
4	5	6	7	8	9	10	11	4	5	6	7	8	9	10	11
12	13	14	15	16	17	18	19	12	13	14	15	16	17	18	19
20	21	22	23	24	25	26	27	20	21	22	23	24	25	26	27
28	29	30	31	32	33	34	35	28	29	30	31	32	33	34	35
36	37	38	39	40	41	42	43	36	37	38	39	40	41	42	43
44	45	46	47	48	49	50	51	44	45	46	47	48	49	50	51
52	53	54	55	56	57	58	59	52	53	54	55	56	57	58	59
60	0	1	2	3	4	5	6	60	0	1	2	3	4	5	6
7	8	9	10	11	12	13	14	7	8	9	10	11	12	13	14
15	16	17	18	19	20	21	22	15	16	17	18	19	20	21	22
23	24	25	26	27	28	29	30	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	31	32	33	34	35	36	37	38
39	40	41	42	43	44	45	46	39	40	41	42	43	44	45	46
47	48	49	50	51	52	53	54	47	48	49	50	51	52	53	54
55	56	57	58	59	60	0	1	55	56	57	58	59	60	0	1
2	3	4	5	6	7	8	9	2	3	4	5	6	7	8	9
10	11	12	13	14	15	16	17	10	11	12	13	14	15	16	17
18	19	20	21	22	23	24	25	18	19	20	21	22	23	24	25
26	27	28	29	30	31	32	33	26	27	28	29	30	31	32	33
34	35	36	37	38	39	40	41	34	35	36	37	38	39	40	41
42	43	44	45	46	47	48	49	42	43	44	45	46	47	48	49
50	51	52	53	54	55	56	57	50	51	52	53	54	55	56	57
58	59	60	0	1	2	3	4	58	59	60	0	1	2	3	4
5	6	7	8	9	10	11	12	5	6	7	8	9	10	11	12
13	14	15	16	17	18	19	20	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	21	22	23	24	25	26	27	28
29	30	31	32	33	34	35	36	29	30	31	32	33	34	35	36
37	38	39	40	41	42	43	44	37	38	39	40	41	42	43	44
45	46	47	48	49	50	51	52	45	46	47	48	49	50	51	52
53	54	55	56	57	58	59	60	53	54	55	56	57	58	59	60

Figure 12: Example of index mapping for $nIndex=61$.

3.1.1 Beam widening by phase tapering

Beam widening has been well studied in the literature. There are three main solutions: aperture truncation (i.e. switching off some elements making the resulting antenna aperture smaller), amplitude tapering and phase tapering. Aperture truncation and amplitude tapering cause EIRP reduction since not all amplifiers are used at their maximum power. In contrast to this,

phase tapering keeps EIRP values constant but at the price of increased gain ripples. Here we focus on phase tapering to keep maximum transmitted power.

In phase tapering or phase only beam synthesis, trading off gain ripples and sidelobe levels requires complex non-convex optimization presenting large number of local optima, but making difficult to find the global one, as detailed in [8]. In practice, it limits the possibility of having a systematic approach able to achieve the ripple and side lobe level (SLL) requirements for different beam synthesis problems. Moreover, such optimizations often provide results similar or worse than those obtained with closed-form phase tapering, [8]. Therefore, and considering also the computational complexity of running non-convex optimizations for each beam as the satellite moves over the Earth-fixed grid, we focus on closed-form phase tapering leveraging on the fact that the Fourier transform of a complex linearly frequency modulated pulse can be approximated by a rectangular window. In particular, we follow the approach in [9], so the beamforming vector writes as

$$\mathbf{w} = \sqrt{P_a} e^{i \frac{2\pi}{\lambda} (x \sin(\theta) \cos(\varphi) + y \sin(\theta) \sin(\varphi))} \circ e^{i \frac{2\pi}{\lambda} \left(\frac{x_r^2 \cos(w_x) \sin(w_y) + y_r^2 \sin(w_x)}{D} \right)},$$

where $\circ$ represents the Hadamard product. The first term of $\mathbf{w}$ is the traditional phased array beamformer with amplitude equal to the available power per antenna P_a and the phase being the steering vector towards the direction of the centre of the beam (θ, φ) . The second term corresponds to the phase tapering, with w_x and w_y being the target angular widening in the x and y dimensions. Note however that we not only need to apply beam widening in the principal axes of the antenna but in the angles determined by the position of the beam centres (θ, φ) . For this, as a difference to [9], we apply the phase tapering to a rotated version of the antenna positions, so that the principles axes of the rotated positions align with the required widening directions. The rotated positions are calculated as

$$x_R = x \cos(\varphi) - y \sin(\varphi)$$

$$y_R = x \sin(\varphi) + y \cos(\varphi).$$

Figure 13 compares this beam widening method with aperture truncation and the kind of upper bound used in [1] consisting of aperture truncation but keeping the total transmit power, so just considering the loss in directivity. The beam without widening is shown also as a reference. The beam points to Nadir and the target beamwidth is 11.7° . The phase tapering provides a pattern similar to the upper bound but with the foreseen gain ripple and an EIRP increase of 10 dB compared to the aperture truncation.

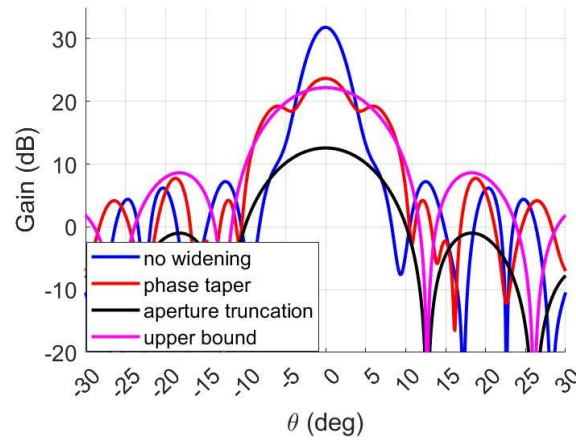


Figure 13: Comparison of beam widening solutions

The step-by-step process for calculating the beamformer for all widened beams in beam designs A and C is as follows:

1. Calculate the contour at -3 dB of the reference beam footprint. This could be at the edge ($\theta = 46^\circ, \varphi = 0^\circ$), at the middle of the coverage ($\theta = 26^\circ, \varphi = 0^\circ$) or at the 75% ($\theta = 32.25^\circ, \varphi = 0^\circ$).
2. Calculate the radius R of the beam footprint corresponding to the $\varphi = 0^\circ$ axis.
3. Calculate the distance between centres of the hexagonal grid of beams on Earth as $d = \sqrt{3}R$.
4. Generate the hexagonal grid and create circles of radius R in each grid point
5. Transform the circles in geodetic coordinates to u-v plane
6. Calculate w_x and w_y as the lengths of the major and minor axes of the resulting ellipses
7. Calculate the beamforming weights

Figure 14 exemplifies this process for the case of designing uniform beams (i.e. beam design A). On the left, Figure 14 shows the beam footprints in geodetic coordinates. The magenta beam is the reference non-widened beam in step 1. Black dots and circles are the hexagonal grid and the ideal beam footprints in step 4. Red curves are the calculated footprints through phase tapering. They are plot at -5dB since -3 dB contours are narrower due to gain ripple. On the right, Figure 14 shows the target contours on the u-v plane in black as derived in step 5, and the synthesized beams in red with the phase tapering.

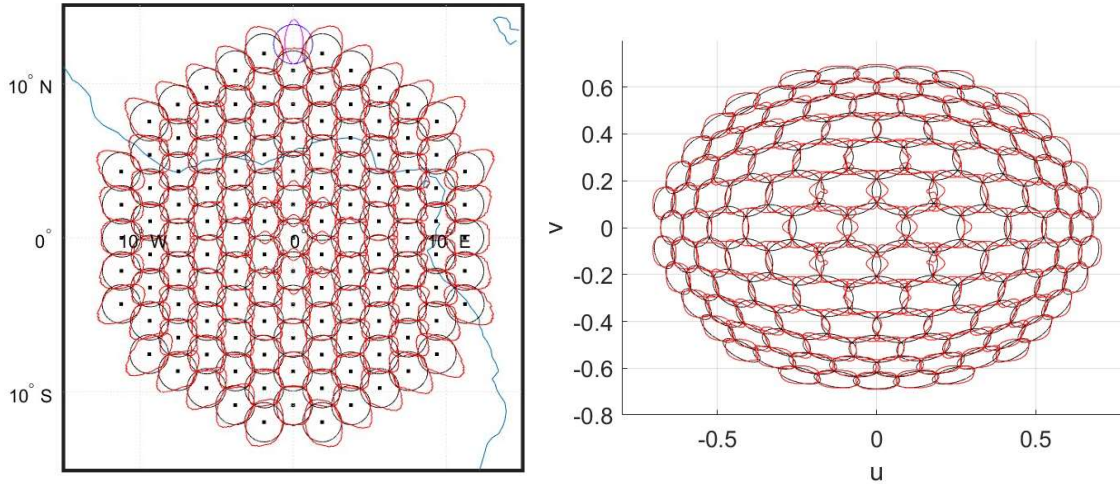


Figure 14: Beam design for alternative A (uniform beams). Left: Beam footprints over Earth; right: Beam contours in u-v plane.

3.1.2 Power equalization

With the beams already designed and the multiplexing scheme decided, it is possible to equalize the power per beam so that all the beams transmitted simultaneously provide similar SINR to the users, at least at beam centres. This equalization allows compensating the effects of larger pathloss and scan loss at coverage edge by allocating more power to beams on those regions. Besides, it may also allocate more power to those affected by large widening, which has a significant impact on EIRP. Two equalization schemes are proposed and evaluated. The first is intended for beam designs B and D, to balance the low SNR at coverage edge. The power per beam is calculated as

$$P_k = P_a N \frac{c_k}{\sum_k c_k}$$

with

$$c_k = \frac{\text{trace}(\mathbf{H}\mathbf{H}')} {\mathbf{h}_k \mathbf{h}_k'},$$

being $\mathbf{H}$ the channel, matrix between all antennas and K beam centres and $\mathbf{h}_k$ the channel vector to the centre of beam k .

For designs A and C, equalizing considering only the channel is not the best choice since SNR differences come also from the beamformers due to beam widening. Therefore, in this case the power per beam is calculated according to P_k , but c_k is calculated as

$$c_k = \frac{\text{trace}(|\mathbf{H}\mathbf{W}|^2)}{|\mathbf{h}_k \mathbf{w}_k|^2},$$

where $\mathbf{W}$ is the beamforming matrix and $\mathbf{w}_k$ is the beamforming vector to beam centre k .

3.1.3 Numerical evaluation

The numerical evaluation is carried out following this procedure:

1. Design the beams and the Earth-fixed uniform grid of footprint centres. For beam designs A and C the method explained in section 3.1.1 is used. For designs B and D, the method is even simplified. First, the tilt angle at beam crossover is projected on Earth to calculate the beam footprint radius R . Then, the uniform hexagonal grid of beams is calculated with a separation between points of $d = \sqrt{3}R$. Finally, the steering vectors to the positions of the grid points are calculated to form the beamforming matrix.
2. Set a dense uniform grid on Earth to sense the received SINR. In particular, a hexagonal grid with distances between points of 15 km, for a total number of more than 32.000 points in the coverage area, is constituted.
3. Assign each point on the sensing grid to the closest beam footprint according to the geodetic distance between the sensing grid point and the beam footprint centres
4. Calculate the SINR in the sense grid points associated to beams assigned to hop index ih , which are the only ones simultaneously illuminated.
5. Point 4 is repeated for all beam indexes, so the SINR to all sensing grid points are calculated.
6. If the percentile 5 of the SINR over the whole coverage area is below 3 dB, then increase N_h and repeat Step 4 and 5 until SINR= 3dB is achieved.

Figure 15 exemplifies the steps 4 to 6 for the design with maximum EIRP beams (design B) on the left, and the design with maximum EIRP beams with beam crossovers at 6° (design D) on the right. It plots the Cumulative Distribution Function (CDF) of the SINR perceived across the whole coverage. As the number of beam index is increased, the SINR increases since less simultaneous beams are transmitted for each beam index. Moreover, Figure 15 also evaluates the effects of beam power equalization. It shows that for maximum EIRP beams with -3B beam crossovers, the power equalization provides more balanced SINR levels that translate to clear improvements in the low percentile region. In contrast, as the beam crossover is increased, this benefit is reduced since lower SINR are observed not only at coverage edges but at the edge of beams close to Nadir. We analysed these effects for the different beam designs and just report on the best performing solutions, with or without equalization.

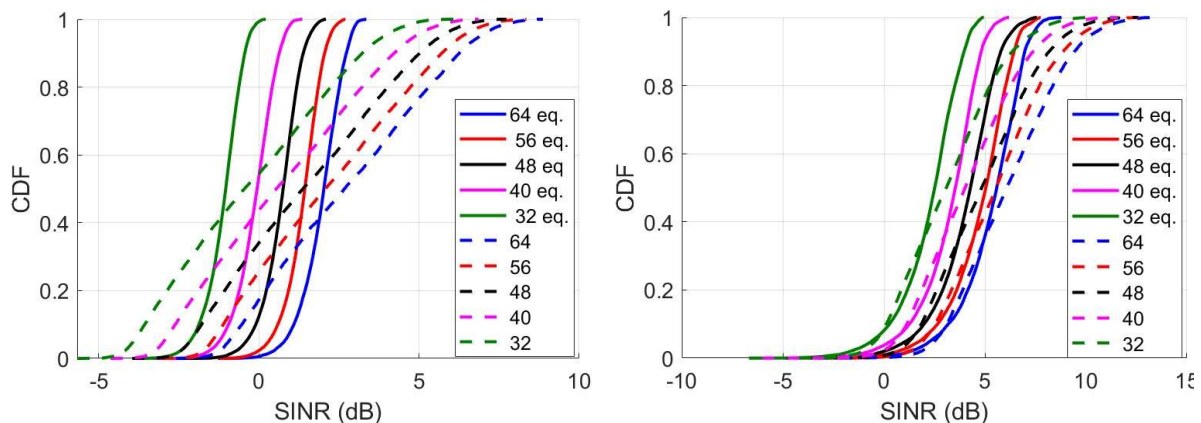


Figure 15: CDF of SINR over the whole coverage. Left: Maximum EIRP beams (design B); right: Maximum EIRP beams with beam crossover at 6° (design D).

The optimization of step 6 is carried out for 9 different designs as summarized in Table 5. They include Uniform beams (design A), Maximum EIRP beams (design B), two cases of Non-

Uniform widened beams (design C), widening according to the beam footprint radius at 50% (i.e. 23°) or 75% (i.e. 32.25°) of the maximum tilt angle, and 5 cases of Maximum EIRP beams with reduced beam crossovers levels (design D) corresponding to tilt angles from 4.8° to 9° . Figure 16 shows the CDF of the SINR obtained when fulfilling the requirement of having 95% of the coverage above 3 dB. All curves cross at 3 dB but the SINR distributions are different. The equalized solutions (B, D1, D2, C1, C2) provide more balanced SINR, whereas within the non-equalized, larger imbalances are observed as the number of total beams is decreased either through widening (A) or extending the beam cross over (D6).

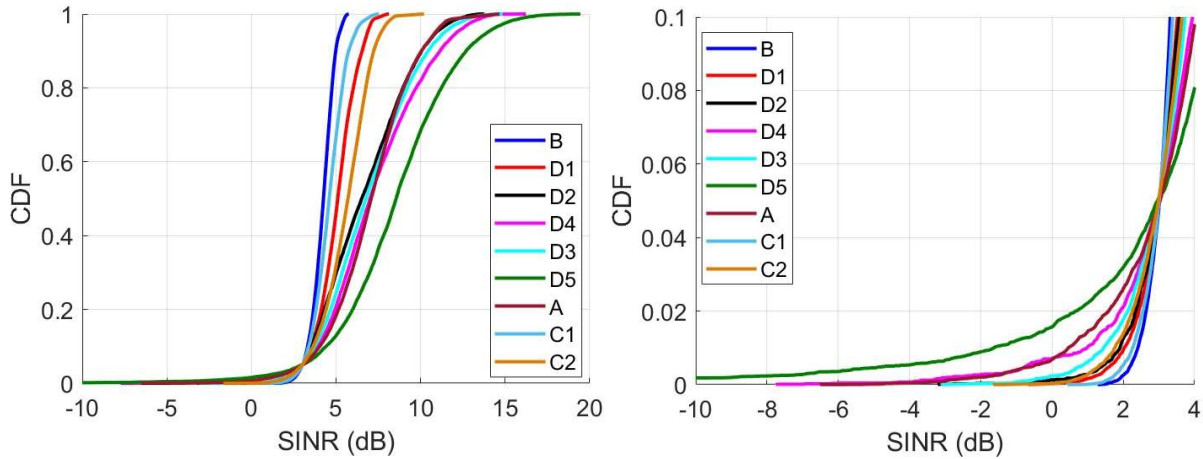


Figure 16: CDF of SINR over the whole coverage for the designs achieving 95% coverage over 3 dB, summarized in Table 5. Left: full plot; right: zoom to low percentile range.

Figure 17 plots the performance metric, i.e. the number of beam indexes required to achieve that 95% of the coverage perceives an SINR above 3 dB. Table 5 complements the results by showing not only the numerical values but also the total number of beams required for each design and the maximum and minimum number of beams transmitted simultaneously. For designs based on Maximum EIRP beams (B and D), an optimum beam cross over point at 7° minimizing the number of beam indexes is observed. Therefore, reducing the number of beams by enlarging the separation between them is beneficial until the SINR at the edges of the beams close to Nadir is too low. In contrast, for beam widening solutions (A and C), the larger the reduction on the number of beams the better. Interestingly, the best designs of both techniques (**D3 and A**) provide similar results, but for a very different total number of beams. Figure 18 depicts the SINR distribution across the coverage area for these two designs. In D3, the region below 3 dB are the beam edges of beams at the coverage centre and areas at the coverage edge. In contrast, in A the beam edges of beams at the coverage centre present good SINR.

Recommendation: To sweep the coverage area with the minimum number of resources, the best candidate is D3, which requires 62 beam indexes.

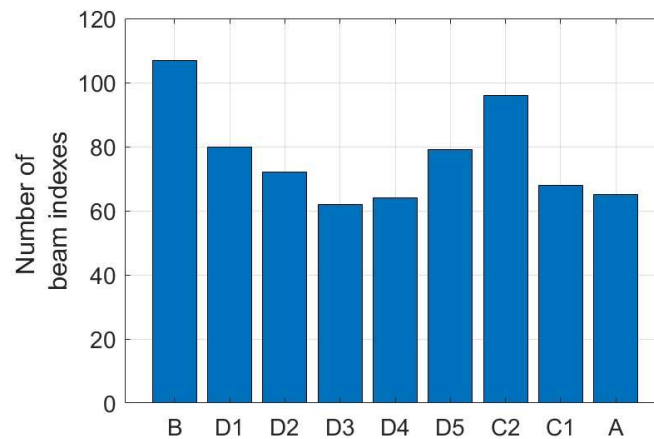


Figure 17: Number of beam indexes (time resources) required to illuminate the whole coverage for the designs selected in Table 5.

Table 5: Summary of simulation results

Design	Total number of beams	Required number of beam indexes (n_{Index}) for 95% coverage above 3 dB	Minimum number of beams per beam index	Maximum number of beams per beam index
A: Uniform beams	151	65	1	4
B: Maximum EIRP beams with beams crossover at 3.64° with power equalization	1723	107	15	18
C1: Non uniform widened beams with reference at 23° and equalization	1237	96	11	15
C2: Non uniform widened beams with reference at 32.25° and equalization	517	68	5	10
D1: Maximum EIRP beams with beams crossover at 4.8° with power equalization	979	80	10	15
D2: Maximum EIRP beams with beams crossover at 6°	637	72	6	10

D3: Maximum EIRP beams with beams crossover at 7°	451	62	6	10
D4: Maximum EIRP beams with beams crossover at 7.5°	397	64	4	8
D5: Maximum EIRP beams with beams crossover at 9°	279	79	2	5

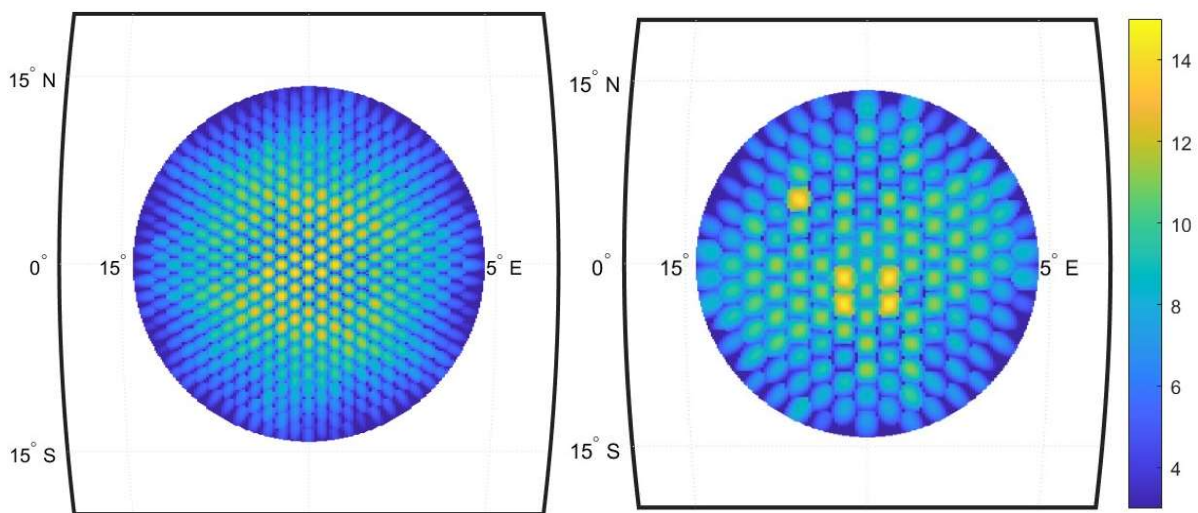


Figure 18: SINR distribution across the coverage area for the two most efficient designs: D3 (left) and A (right).

As a final observation, let us highlight that only designs D3 and D4 results in a required number of beam indexes below or equal to 64. In these cases, a single SSB burst would be enough to sweep over the whole coverage, which can be performed in a half subframe. Note however that this only applies to SSB transmissions, the requirements for other broadcast signalling are further analysed in Section 3.2.

3.2 BEAM ILLUMINATION PLAN

As outlined in Section 3.1, the beam design to transmit and receive common control information to all users is driven by the number of resources that are needed to sweep the coverage area. In this context, Design D3 (defined in Table 5) is considered the most suitable candidate. For reference, Design B will also be evaluated, as it ensures maximum EIRP across the coverage area. To assess the benefits of Design D3, this section begins with an overview of broadcast and RACH signalling. The results of this analysis are then used to propose a frame structure for both UL and DL. Finally, based on this design, a system-level evaluation of beam hopping is carried out.

3.2.1 Broadcast signalling

To design the broadcast beams in the downlink, this section analyses the following signals: SSB, PDCCH, SIB1, other SIBs, paging and msg2/msg4 during the random access procedure.

SSB

The SSB is used by the UE to acquire the downlink synchronisation and to detect the PCI. The SSB is confined within a 5 ms window and the default periodicity is 20 ms. To enhance the downlink coverage, the default periodicity can be extended beyond 20ms. In such a case, legacy UEs perform cell search as if the SSB is transmitted with 20 ms periodicity, which could affect its initial access and connection establishment. Due to the SSB periodicity extension, there could be a backward compatibility issue. The 160 ms value that can be considered for the SSB periodicity extension. It is worth highlighting that extending SSB periodicity will also have a significant impact on the latency during DL synchronization and initial cell access.

The SSB frequency position depends on the operating band, which is specified in [10]. The SSB spans 240 subcarriers (20 resource blocks) and 4 OFDM symbols, as shown in Figure 19. Remarkably, the SSB is formed by two synchronization signals, i.e., primary synchronisation signal (PSS) and secondary synchronisation signal (SSS), as well as the physical broadcast channel (PBCH).

- PSS is used for the detection of the initial symbol boundary, Cyclic Prefix (CP), sub-frame boundary, and initial frequency synchronisation to the cell; it consists of one out of 3 possible known m-sequences BPSK modulated.
- SSS is used for the identification of the radio frame boundary and of the PCI; it is one out of 336 sequences (together with the PSS, there are thus $3 \cdot 336 = 1008$ PCIs), obtained as the combination of two BPSK modulated m-sequences.
- PBCH contains the MIB, which provides also information on how to detect the SIB1, as previously mentioned. To decode the MIB, the Demodulation Reference Signal (DM-RS) is decoded, based on the PCI detected from the combination of PSS and SSS; this allows the UE to estimate the channel coefficients and exploit such information for the decoding of the MIB. Once the MIB is decoded, information on where the SIB1 is transmitted (and how to decode it) is available at the UE.

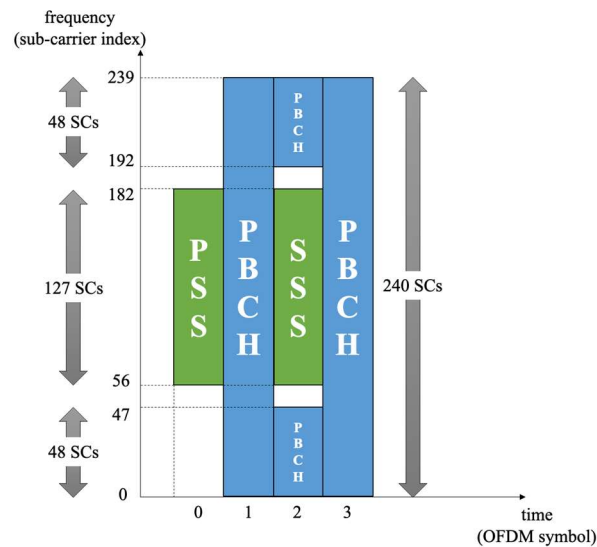


Figure 19: NR SSB structure

It shall be noticed that each SSB corresponds to a specific beam (this is how the UE performs the initial beam sweep), beamformed in a different direction. A group of SSBs forms a SS burst that always spans 5 ms, which corresponds to half-frame. In FR2, up to 64 SSBs can be transmitted in one burst of 5ms. This highlights that the maximum number of beams that can be managed in one cell is 64.

The SSB location in time is determined by the OFDM symbol index in a half-frame. For carrier frequencies within FR2-NTN frequency and 120 KHz SCS, the SSB indexes are given by $\{4,8,16,20\} + 28 \cdot n$, where $n = 0, \dots, 18$. The rest of configurations are provided in 3GPP TS 38.213.

The SSB location in frequency is parametrized by a frequency offset with respect to the frequency of the first RB.

It shall also be mentioned that, as part of RAN1 agreements, 3GPP concluded that SSB channel enhancement other than SSB periodicity extension is not considered. It has been agreed that

- For cell search procedure a UE may assume that half frames with SSBs occur with a periodicity of 20ms, 40ms, 80ms and 160ms;
- DL frequency compensation on the service link is not performed at the gNB, while it is left for implementation at the UE;
- Any frequency compensation on the feeder link or management of transponder frequency errors is operated at the gateway, thus left for system implementation.

PDCCH

The control resource set (CORESET) is defined by a set of time and frequency resources that are used to carry the physical downlink control channel (PDCCH). The PDCCH bits are created from a downlink control information (DCI) message after performing CRC attachment, channel coding and rate matching. These coded bits are scrambled and converted to complex modulated symbols.

The number of resource elements (REs) of a CORESET that are required to carry a PDCCH DCI message is called aggregation level, which is expressed in terms of control channel elements (CCEs). Hence, the number of CCEs that compose a PDCCH candidate is determined by the CCE Aggregation level. The larger the CORESET, the more resources are available to send DCI messages.

The search space defines the PDCCH placement within a CORESET. Hence, it defines a set of resources that can contain a specific PDCCH. A CORESET can have multiple search spaces. To decode the signalling that is common to all UEs, the standard defines different types of common search spaces, as detailed in 3GPP TS 38.213, [11]. The search space that the UE has to monitor depends on the radio network temporary identity (RNTI). The summary is provided in Table 6

Table 6: Set of common search spaces.

Common search space	RNTI type	Configuration
Type0-PDCCH	System Information RNTI (SI-RNTI)	For SIB1 acquisition
Type0A-PDCCH	System Information RNTI (SI-RNTI)	For other SIB acquisition
Type1-PDCCH	Radio Access RNTI (RA-RNTI), Temporary cell RNTI (TC-RNTI), cell RNTI (C-RNTI)	For random access procedure
Type2-PDCCH	Paging RNTI (P-RNTI)	For paging

CORESETs are configured by the network and communicated to the UE via RRC reconfiguration messages. Control resource set zero (CORESET#0) is a special type of CORESET that is configured to the UE by using information conveyed in the MIB. Upon decoding the MIB, the UE can use the tables predefined in [11] to determine:

- The number of RBs and OFDM symbols that are assigned to CORESET#0;
- The CORESET#0 time and frequency location.

Table 7 provides some configurations for different SSB and PDCCH subcarrier spacing. For the sake of clarity, since there are a lot of permutations, only a few examples are provided.

The CORESET#0 frequency location is determined by a frequency offset with respect to the lowest frequency of the SSB.

The CORESET#0 time location depends on the multiplexing pattern with respect to the SSB. The pattern 1 refers to time domain multiplexing. As a consequence, the CORESET#0 and the SSB could be transmitted in different time slots. To reduce the number of broadcast beams it is convenient to allocate the SSB and the CORESET#0 in the same time slot. This is allowed by multiplexing patterns 2 and 3.

Table 7: Set of physical resources assigned to CORESET#0.

SSB SCS (KHz)	PDCCH SCS (KHz)	(Multiplexing pattern, number of RBs, Number of symbols)
240	120	(1,48,1), (1,48,2), (2,24,1), (2,48,1)
240	60	(1,96,1), (1,96,2)
120	120	(1,24,2), (1,48,1), (1,48,2), (3,24,2), (3,48,2)
120	60	(1,48,1), (1,48,2), (1,48,3), (1,96,1), (1,96,2), (2,48,1), (2,96,1)

Recommendation: To simplify the beam hopping pattern, the CORESET#0 must be transmitted in the same time slot as the SSB. To make this possible, the most suitable approach is based on adopting the multiplexing pattern 3. An example is illustrated in Figure 20. In such a case, the dwell time is dictated by the SSB, which spans 4 OFDM symbols. Under this configuration, a broadcast beam including the SSB and the CORESET#0, should be active for 4 OFDM symbols. This pattern allows the system to define the beam illumination pattern associated with the CORESET#0 and the SSB.



Figure 20: CORESET#0 and SSB scheduling for multiplexing pattern 3

SIB1

The PDCCH scheduling SIB1, which is transmitted in the CORESET#0, provides time and frequency resource allocation for PDSCH carrying SIB1.

The SIB1 is the first broadcast message, which transmits cell-specific information. Hence, the SIB1 is essential for proper network operation. The SIB1 periodicity is 160 ms. The transmission repetition periodicity within that 160 ms is determined by the network implementation.

The SIB1 provides details about the search space as well as periodicity and windowing for other SIBs, i.e., SIB2 and beyond.

SIB19

The SIB1 provides information regarding the search space to monitor the PDCCH scheduling for other SIBs. The search space is defined within a predefined CORESET. To attach to the network, SIB19 must be decoded in IDLE state. Hence, SIB19 must be scheduled over Type0A-PDCCH. By setting the search space for other SIBs to zero, the UE monitors the same PDCCH occasions for SIB1 and SIB19.

The SIB19 is carried in system information (SI) messages. The periodicities for SIB19 are chosen from {80, 160, 320, 640, 1280, 2560} ms. SIBs having the same periodicity can be mapped to the same SI message if the maximum SI messages size allows. The maximum size of an SI message is 2976 bits. Each SI message is transmitted within a time window, which is referred to as SI-window. The SI-windows of different SI messages do not overlap. To enhance the downlink coverage, each SI message can be repeated within the SI-window. The details are provided in [12].

PAGING

The paging messages are sent from the gNB to the UE in IDLE state in order to initiate calls. In IDLE state, the gNB does not know the exact position in which the user is located so paging messages are transmitted in a tracking area basis, to ensure they reach the target user. From SIB1 the UE will extract the search space, which determines the physical resources within a CORESET that shall be monitored to detect the PDCCH scheduling the incoming paging messages.

MSG2 AND MSG4

Upon detecting msg1 in the selected time-frequency resources, the gNB generates a response called msg2. At this stage, all UEs requesting access shall monitor the Type1-PDCCH search space to receive the corresponding response message. If UE successfully decodes the PDCCH, it gets the random access response data, which carries critical information such as the time advance command for timing adjustment and an initial uplink grant for the UE. The gNB also assigns a temporary identifier, the RA-RNTI, to the UE.

Using the initial uplink grant provided in msg2, the UE transmits msg3 on the PUSCH to request connection. After processing msg3, the gNB sends msg4 to the UE. msg4 is the contention resolution message that contains the UE's identity, confirming that the gNB has correctly identified the UE. At this step, the network provides the UE with a C-RNTI.

The PDCCH for scheduling msg2 and msg4 can use the same search space as the SIB1.

SCHEDULING OF COMMON SIGNALS

The aim now is to further analyse how to efficiently schedule all common signalling for the downlink in the 5G frame structure. To this end, all the common signalling is packed together. If feasible, this would be advantageous to reduce the number of broadcast beams, so that more resources are available for data beams. Accordingly, in the same frame, the UE might be able to:

- Perform SSB measurements;
- Perform initial cell search by decoding SIB1;

- Acquire or update satellite ephemeris by decoding the SIB19;
- Monitor the PDCCH scheduling msg2 and msg4;
- Monitor the PDCCH scheduling incoming paging messages.

Since the 3GPP standard already specified the timing of a beam sweep to transmit SSB burst across a cell in a half frame of 5 ms, the proposed approach is to co-schedule all common signalling within the slots devoted to SSB transmissions, so a single sweep may allow broadcasting all common signalling across the coverage, reducing the amount of time the beam pattern is designed under common signalling a criteria. We focus on FR2 assuming a subcarrier spacing of 120 kHz. In this case, the half frame contains 40 slots, 32 of them carrying two SSB beams each and the remaining 8 slots not carrying SSB signalling. The target is to have one paging and one RA occasion per beam each SSB period, which by default can be either 20 ms or 160 ms. Note that in the case of 20 ms, SIB1 and SIB19 do not need to be allocated in each frame carrying SSB since their minimum periodicities are larger.

In RAN1#117 meeting, NR-NTN downlink coverage enhancements have been discussed in [13]. The results related to the required SNR for different signals are provided in Table 8. The required SNR is the minimum value to meet a BLER of 1%. This information will be taken as reference to design the beam layout and the beamforming. In this regard, it can be inferred that broadcast beams can convey common signals as well as traffic, if the resulting SINR is equal or higher than -4.1 dB. This requirement comes from the correct detection of the PDSCH.

Recommendation: In alignment with Section 3.1, broadcast beams have to be designed to achieve a minimum SNR of 3 dB to ensure a sufficient margin.

Table 8: Required SNR to meet the target BLER.

Signal	Required SNR (dB) to meet the target SNR
SSB	-6.3
PDCCH	-6
PDSCH msg2	-10.9
PDSCH msg4	-5.2
PDSCH SIB1	-5.8
PDSCH SIB19	-6.9
PDSCH for VoIP	-11
PDSCH for 1Mbps data rate	-4.1

To assess whether the resources in one half frame are sufficient, we first derived the amount of resource elements required to transmit each common signal, as reported in Table 9. It

assumes that the 3 dB SINR target allows using a Modulation and Coding Scheme (MCS) index of 6 [14], which corresponds to QPSK modulation and code rate 449/1024 [15]. The required resources for SSB and CORESET0 are specified by 3GPP. We resort to SSB-CORESET0 multiplexing pattern 3, so SSB uses 20 PRB and 4 symbols whereas CORESET0 is scheduled in the first two symbols of SSB and occupies 48 PRBs. Therefore, a search space Type 0 (SS0) is already co-scheduled with each SSB block. Figure 21 illustrates two scheduling alternatives by showing a reasonable allocation for the first two slots of a half frame transmitting SSBs.

Table 9: Required time/frequency resources

Signal	Size in bits	PRBs	OFDM symbols
SSB	-	20	4
CORESET0/SS0	-	48	2
CORESET1/SS1	-	132	1
SIB1	1280	70	2
		33	4
SIB19	616	34	2
		16	4
Paging	1280 (32 UE)	70	2
		33	4
		22	6
	490 (7 UE)	27	2
MSG2	560 (8 UE)	15	4
	630 (9 UE)	11	6
		57	2
MSG4	1040	27	4
		18	6
		8	13

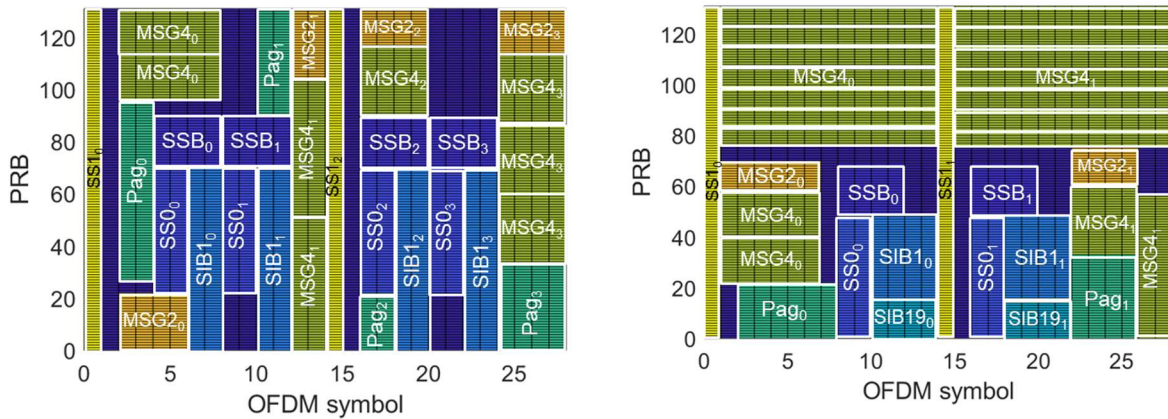


Figure 21: Common signalling allocation assuming two SSB transmissions per slot (left) and a single transmission per slot (right)

The first alternative (Figure 21 left) is to co-schedule all common signalling in the slots devoted to SSB transmissions, so the hop duration becomes half-slot. Therefore, the beams associated with up to 64 beam indexes are sequentially illuminated within a half frame. According to the standard, the starting position of each SSB and SS0 within each slot varies with the SSB index (SSBI), being 4, 8, 2 and 6 for SSBI mod 4=0,1,2,3, respectively. When the SS0 starts at symbols 4 and 2 only two symbols after it remain available for common signalling accommodation. In view of this, we propose to set a CORESET1 occupying one OFDM symbol and the whole band used by a Search Space 1 (SS1) allocated at the first symbol of each slot. This enables PDSCH resources for random access and paging to be scheduled

before the SS0. However, for SSBs starting at symbols 8 and 6 there is no room to include such SS1. The proposed solution is then to map the SSBI to different cell types. Cells type A will use only SSB blocks fulfilling $SSBI \bmod 4 = 0, 2$, and will map common search spaces Type 0 (SIB1) and 0A (SIB19) to SS0 and Type 1 (random access) and 2 (paging) to SS1. In contrast, cells type B will use the rest of SSB blocks mapping all common search spaces to SS0. The drawback is that a cell with this configuration only supports up to 32 beams, and not the 64 provided by the standard.

In Figure 21 left, the worst case is observed for the half slot corresponding to SSB2, which only allows for 1 msg4 and a paging message addressing 15 UE. In consecutive SSB periods, the resources used for SIB1, may be used for SIB 19 or for an extra msg4 exploiting the larger required SIB1 periodicity. Assuming SSB periodicity of 20 ms and SIB1 and SIB 19 periodicities of 160 ms, in the worst case, the minimum msg4 and paging capacity achieved in a beam would be 87.5 UE/s and 750 UE/s, respectively. Note that the proposed allocation is based on the default time resource allocation tables specified by 3GPP [15]. Customized tables would allow a better utilization of the resources within each slot, but still the room for msg4 would be scarce.

To increase paging and random access capacity and to reduce the differences between beams mapped to different SSBI, we propose to co-schedule all common signalling with SSB slots, but to transmit a single SSB per slot, thus enlarging the hop duration to one slot. In this case, the beams associated with up to 32 beam indexes are sequentially illuminated within a half frame, since from the 64 SSB blocks supported by the standard, only 32 are used. As depicted in Figure 21 right, this solution may allow the allocation of up to 9 msg4, with paging messages addressing 32 UE, which is the maximum allowed by the standard. For SSB periodicity of 20 ms, the msg4 and paging capacity are increased to 450UE/s and 1600 UE/s, respectively. Moreover, this scheme allows repetitions of SIB1 and SIB19 every 20 ms. Nevertheless, cells with this configuration only support 32 beams.

The comparison between the two scheduling schemes and their combination with different beam layout and beam multiplexing solutions is made in terms of common signalling (CS) coverage ratio and efficiency. Coverage ratio was defined in 3GPP as the number of beam footprints broadcasting common signaling illuminated within a SSB period over the total number of beam footprints. The CS efficiency is defined here as the number of slots not scheduling CS over the total number of slots within the SSB periodicity.

The beam hopping system parameters used for the numerical assessment are gathered in Table 10. It is worth mentioning that two different designs are compared.

Table 10: Beam hopping specifications

	Maximum EIRP beams	5G-STAR DUST SOLUTION
Number of beams	$N_T = 1723$	$N_T = 451$
Average number of active beams	$N_A = 16$	$N_A = 7$
Percentage of simultaneous active beams (%)	0.92	1.55
Maximum bandwidth (MHz)	200 MHz	200 MHz

Subcarrier spacing (KHz)	120	120
Total resource blocks	132	132
Number of slots	$N_s = 80$	$N_s = 80$
Minimum beam radius (km)	41	64
Minimum beam gain (dBi)	28	28

Figure 22 shows a timing comparison between the two scheduling solutions assuming a SSB periodicity of 20 ms and two different beam hopping designs from Section 3.1.3:

5G-STAR DUST solution (Design D3). This is the most efficient solution with $nIndex=62$. The beams are widened with respect to the maximum EIRP approach by increasing the crossover level from 3.64° to 7° . Consequently, the separation between adjacent grid points is increased, which translated into a less dense grid of beams.

Maximum EIRP beams (Design B). This design is treated as benchmark since it does not involve any layout optimization. This solution aims to maximize the EIRP across the coverage area with $nIndex=107$. To this end, a dense grid of beams is created according to the beam footprint radius at Nadir, which implies deploying a significantly large number of beams and having large beam overlapping as we move towards coverage edge.

For $nIndex=62$, the two solutions can broadcast CS through the whole coverage within a SSB period, resulting in a coverage ratio of 100%. However, the half slot solution only requires a half frame to broadcast all CS in the 62 hops, whereas the full slot requires 2 half frames. This translates to CS efficiencies of 80.6% and 61.3% for half slot and full slot, respectively. Without layout optimization, i.e. for $nIndex=107$, the number of half frames required to complete the beam hopping sweep over the whole coverage is larger. Consequently, CS efficiencies are reduced to 66.3% and 33.1%, respectively

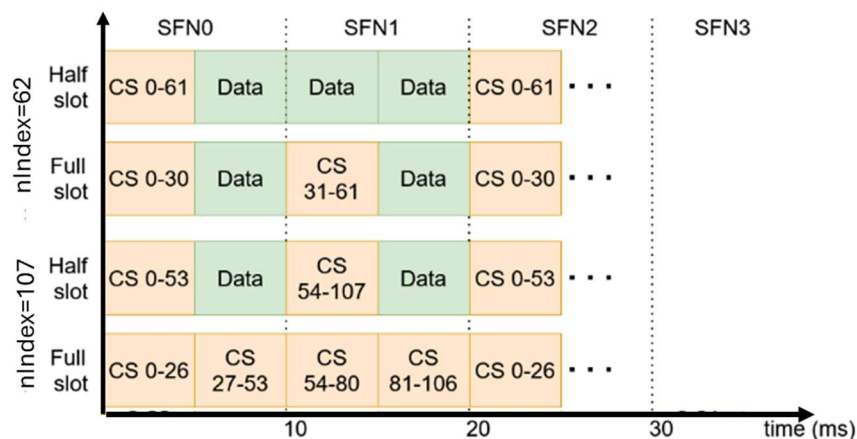


Figure 22: CS scheduling timing comparison for SSB periodicity of 20 ms and designs D3 and B in Section 3.1.3. Green subframes are devoted to data transmission, oranges are shared between CS and data transmission, and reds are exclusively used for CS.

Recommendation: Based on the beam hopping evaluation in FR2, the SSB default periodicity does not need to be extended beyond 20 ms to ensure 100% coverage ratio.

3.2.2 RACH signalling

In the uplink, the need to cover the FoV applies to the RACH signalling. More precisely, for reception of msg1 during the random access procedure. To be consistent with the broadcast signalling, the beam illumination plan in the uplink is designed using the same beam pattern that is conceived for the downlink. Accordingly, for both uplink and downlink, the coverage area is divided into N_T beams and each cell controls N_B beams.

In order to mimic the scheduling of common downlink signals, the idea is to allocate random access resources associated to multiple beams in the same uplink frame.

MSG1

The msg1 is transmitted in the initial access of the random access procedure to access the network. To this end, the UE uses a set of configured RACH resources. This information is conveyed in SIB1. The configuration of PRACH transmission parameters is described in [11, 16], which include the preamble format and the time and frequency resources of RACH occasions (ROs). A specific mapping between the SSB and the RO is specified in the standard so that the network can figure out which SSB beam is selected by the UE, as detailed in [11]. The mapping indicates:

- How many ROs are allocated in frequency domain at the same time location;
- How many SSBs can be mapped into one RO and how many preamble index can be mapped to single SSB.

The standard supports one-to-one, one-to-many and many-to-one association between the SSB and the random access occasion. Some examples are depicted in Figure 23, Figure 24 and Figure 25.

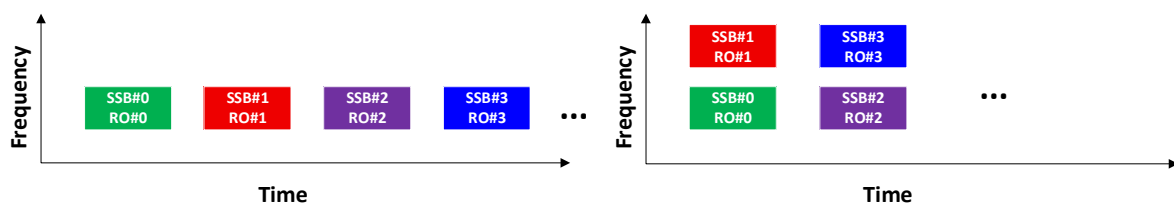


Figure 23: One SSB per RACH occasion: (left) one PRACH transmission in one time instance, (right) two PRACH transmissions in one time instant.

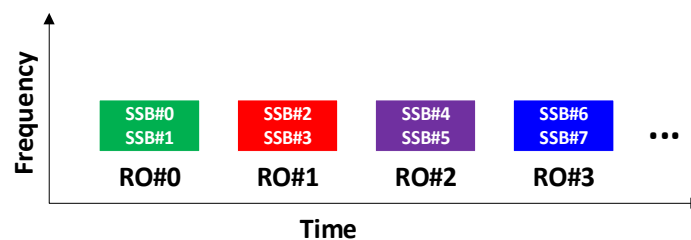


Figure 24: 2 SSBs per RACH occasion.

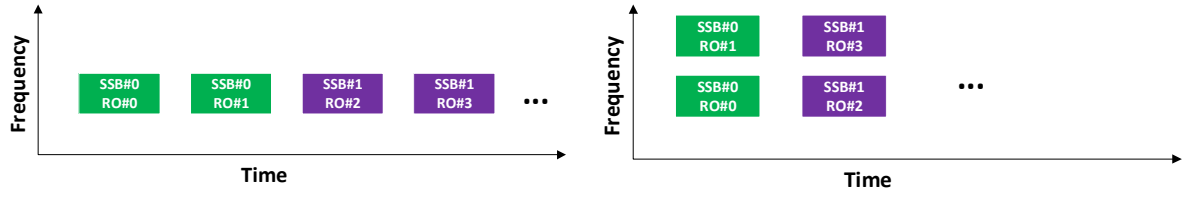


Figure 25: One SSB in 2 RACH occasions: (left) one PRACH transmission in one time instance, (right) two PRACH transmissions in one time instant,

Recommendation: In FR2, up to 64 SSBs can be transmitted in a 5ms burst. To reduce the initial access delay, one approach is to associate multiple SSBs with the same RO in the same UL frame. According to [16], the number of SSB indices per frame can be configured as $M = 1, 2, 3, 6, 7$, depending on the preamble configuration. To configure $M > 1$, the processing must occur at sub-slot granularity. This implies that the dwell time in the uplink would be lower than a full slot duration. To relax the processing, it is proposed to reduce the beam switching rate, so that the UL dwell time is fixed to one frame duration. To this end, each SSB index has to be associated with a dedicated RO.

Even under this constraint, initial access delay can still be reduced by frequency-multiplexing multiple ROs. The concept is illustrated in Figure 23. The standard allows to frequency multiplex up to 8 ROs in the same uplink frame, i.e., $msg1 - FDM = 1, \dots, 8$. To align with the beam to cell mapping scheme designed in the DL, it is considered that each cell handles N_B beams. Consequently, the number of UL frames that is required to allocate all the ROs of a single cell is given by

$$F_{RO} = \lceil N_B / msg1 - FDM \rceil.$$

To reduce as much as possible the access delay, the FDM parameter is selected as $msg1 - FDM = 8$. Consequently, during F_{RO} frames, $N_A N_B$ beams are served for the random access purpose. Here N_A represents the average number of beams that can be simultaneously illuminated with a 200 MHz channel bandwidth. To support more than N_A beams, the channel bandwidth must be reduced accordingly.

The schematic view of the random access signalling frame structure is shown in Figure 26. In this example $F_{RO} = 2$.

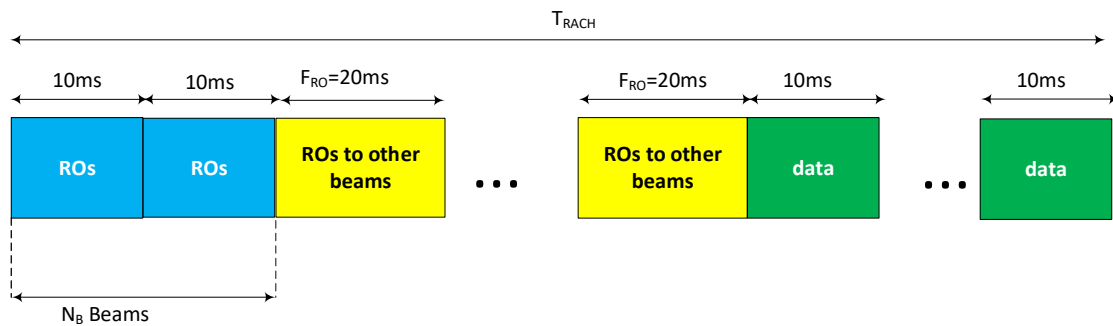


Figure 26: UL frame structure for random access signalling and data transmission

RACH OCCASIONS

The objective is to schedule ROs to achieve 100% coverage ratio within the RACH periodicity, which is referred to as T_{RACH} . The RACH configuration specified in [11] allows the network to set $T_{RACH} \in \{10, 20, 40, 80, 160\} ms$. To reduce the initial access delay, multiple RO are transmitted in a frequency division multiplexing manner in the same UL frame. According to [16], the number of ROs per frame can be configured to $msg1 - FDM = 1, \dots, 8$.

Interestingly, the standard supports flexible mapping. The proposed methodology to establish the mapping between the SSBs and the ROs is as follows:

1. Set the number of beam footprints that can be served for the random access purpose in a single UL frame as $msg1 - FDM = 8$;
2. Set the number of beams per cell N_B , according to the DL CS scheduling illustrated in Figure 22;
3. Compute the number of frames required to allocate all the ROs of a single cell as $F_{RO} = \lceil N_B / msg1 - FDM \rceil$;
4. Compute the number of beams served in F_{RO} as $N_A \times N_B$;
5. Compute the minimum RACH periodicity to serve all beams as $T_{min} = \left\lceil \frac{N_T}{N_A \times N_B} \right\rceil \times F_{RO} \times 10 ms$;
6. Set the RACH periodicity T_{RACH} to any of the values in the list $\{20ms, 40ms, 80ms, 160ms\}$;
7. Compute the coverage ratio within T_{RACH} ;

Adopting this methodology, different configurations and beam designs are evaluated in Table 11.

Table 11: Beam hopping system level evaluation for uplink

	Design B				Design D3			
Total beams	$N_T = 1723$				$N_T = 451$			
ROs per frame	$msg1 - FDM = 8$				$msg1 - FDM = 8$			
Beams per cell	$N_B = 27$				$N_B = 31$			
Frames required to allocate the ROs of one cell	$F_{RO} = 4$				$F_{RO} = 4$			
Beams served in F_{RO}	$N_A N_B = 432$				$N_A N_B = 217$			
Minimum RACH periodicity	$T_{min} = 160 ms$				$T_{min} = 120 ms$			
RACH periodicity (ms)	20	40	80	160	20	40	80	160

Coverage ratio (%)	14.86	29.72	59.43	100	24.83	49.67	99.33	100
--------------------	-------	-------	-------	-----	-------	-------	-------	-----

Recommendation: Based on the beam hopping evaluation in FR2, the RACH periodicity should be set to 160 ms to ensure 100% coverage ratio, for both beam designs.

3.3 EARTH-MOVING BEAMS

Section 3.1 showed that the design of Earth-fixed beam layout becomes inefficient in terms of coverage. The main reason is that a fixed regular grid on Earth does not match well with the beam footprint widening caused by the Earth curvature. Depending on the strategy, it may result in large beam count and large beam overlapping, reduced beam count and overlapping but with coverage holes close to coverage centre, or in reduced beam gains if phase tapering is used to synthesize uniform beam footprints.

Earth-Moving beams present two important benefits with respect to Earth-fixed. First, fixing a regular grid on the angular domain of the satellite antenna translate to a non-regular grid of beam footprints on Earth, with the distance between beams increasing as the beams get widened due to the Earth curvature. This enables a full coverage of a region of interest with a reduced number of beams and with a fixed beam overlapping throughout the whole field of view. Second, beamforming weights can be fixed and do not need to be frequently updated as the satellite pass over the region of interest.

On the drawback side, Earth-Moving beams may trigger frequent hand overs by satellite movement. To avoid that, our proposal is to align satellite cells with millimetre wave terrestrial cells, so manage the satellite mobility through beam management. In this case, a single 5G cell would be mapped to the whole satellite field of view. As the satellite pass over, active users periodically measure and report SSB indexes (SSBI) to inform the satellite in which beam they are located. Such reporting can be made with negligible SNR degradation with a periodicity up to 60 ms [17], though SSB periodicities of 160 ms would require a dedicated assessment. The report of a single SSB index requires $\log_2(N_B)$, being N_B the number of beams.

Currently, the 3GPP standard only allows up to 64 SSB indexes for a FR2 cell, which is well below the beams required to cover a target region of interest limited by a user minimum elevation of 30°. Our recommendation is not to change that, but to allow for the spatial multiplexing of such SSB indexes, as presented in Section 3.1. In this way, the beam suited for a given user would be identified by combining the reported SSBI and the spatial information. In this case, a 5 bits report would be sufficient.

Let us remind here that the satellite capacity is not related to the number of cells but to the number of beams that can be simultaneously illuminated, since all cells share the same beams. Moreover, for user centric beamforming, beam to cell mapping loose its meaning. The only drawback of having a single cell for the whole field of view is the need to broadcast paging messages everywhere.

The design of an Earth-moving beam hopping sequence minimizing the number of hops to illuminate the coverage with a constraint of a minimum SINR of 3 dB for the 95% of the coverage is an NP-hard problem that we divide in three stages: beam layout design, power equalization and beam hopping/multiplexing design. They are further discussed next.

3.3.1 Beam layout

We consider a fixed hexagonal grid of beams in the uv plane of the satellite antenna with an angular spacing between beams of $abs = \sqrt{3}\sin(\alpha/2)$. To cover a FoV of 46° , we include the beam centre's that satisfy $d \leq \sin(46\pi/180) + \beta\sin(\alpha/2)$, where d is the distance to the center of the coverage and β is a scalar that controls the number of beams at the edge. Making α equal to the half power beamwidth (i.e. 3.64°) the beam crossover level is -3 dB [18]. Specifically, we evaluate the following schemes:

- E: $\alpha = 3.64^\circ$ and $\beta = 0.5$
- F: $\alpha = 3.64^\circ$ and $\beta = 1$
- G: $\alpha = 4^\circ$ - 5° and $\beta = 0.5$

Scheme E, depicted in red in Figure 27 is the baseline with -3dB beam cross overs and beams at the edge with most of the -3 dB footprint inside the target FoV. Scheme F is similar to E but tries to avoid some coverage holes at the very edge of the FoV adding the beams depicted in blue in Figure 27. Finally, scheme G tries to reduce the total number of beams reducing the beam cross over levels.

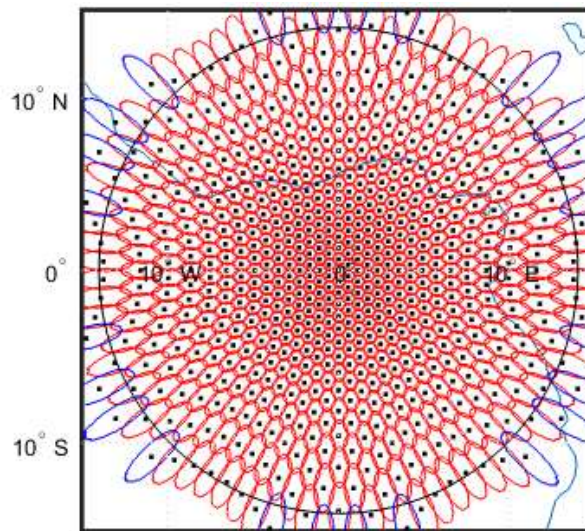


Figure 27: -3 dB beam contours for schemes E (red) and F (red and blue)

3.3.2 Beam hopping/multiplexing design

Another key advantage of Earth-moving beams is that beam footprints move with the satellite but they remain static from a satellite point of view so, as the satellite pass over a given region the assigned beam indexes do not need to be updated. This was not the case for Earth-fixed beams, for which every time a new footprint centre entered the FoV, it needed to be assigned to a new beam index. To avoid a frequent optimization and update of the index assignment, in Section 3.1 we resorted to a deterministic assignment based on the Earth grid and maximum distance criteria. Its main drawback is that the number of beams assigned to a given index becomes not uniform, so the resulting SINR levels are lower when the number of beams is higher, since the power is shared among more beams. This is illustrated in Figure 28, which plots the probability of having $SINR < 3dB$ for the different beam groups created by the beam index assignment process of Section 3.1 with $nIndex=41$. The probability grows as the number of beams per index increases.

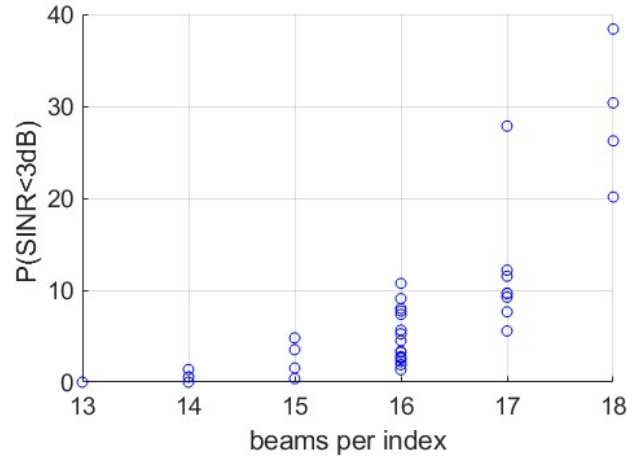


Figure 28: Probability of SINR<3dB per beam index as a function of the number of beams simultaneously illuminated with the same index for nIndex=41

In view of this, we propose a heuristic optimization of the beam index assignment based on a hexagonal grid beam grouping with balanced beam counts. It aims to assign groups (i.e. beam, indexes) to satellite beams arranged in a hexagonal grid, satisfying the following requirements:

1. No neighbouring beams share the same group (interference avoidance).
2. The number of beams per group is approximately equal (load balancing).
3. Beam assignment follows a spiral order from the centre outward, resulting in a uniform distribution within the coverage area (i.e. the number of edge beams, affected by larger pathloss is similar in each group).

Let $\mathcal{B} = \{b_1, b_2, \dots, b_N\}$ denote the set of beams with positions $u_i \in \mathbb{R}^2$, G the number of groups and d_{neigh} the maximum neighbor distance. We construct an adjacency matrix defined as:

$$A_{ij} = \begin{cases} 1 & \text{if } \|u_i - u_j\|_2 \leq d_{neigh} \\ 0 & \text{otherwise} \end{cases} \quad i, j = 1, \dots, N$$

being N the number of beams and u the beam centre coordinates in the uv plane. The assignment problem is then formulated as a graph colouring problem with balanced group sizes:

$$\begin{aligned} & \text{assign groups}[i] \in \{1, \dots, G\}, \forall i = 1, \dots, N \\ & \text{subject to groups}[i] \neq \text{groups}[j], \forall i, j: A_{i,j} = 1 \\ & \text{and minimize } \Delta = \max |B_g| - \min |B_g| \end{aligned}$$

being B_g the number of beams assigned to group g .

We start computing the radial distance of each beam from the coverage centre set to $u_{center} = [0, 0]$ as

$$r_i = \|u_i - u_{center}\|_2 \quad i = 1, \dots, N$$

Beams are then grouped into K concentric rings. Within each ring, beams are sorted by polar angle relative to the center

$$\theta_i = \text{atan} \left(\frac{v_i - v_{center}}{u_i - u_{center}} \right)$$

The spiral order is then defined as

$$order = [b_{center}, ring_1(sorted), \dots, ring_K(sorted)]$$

For each beam b_i we calculate the set of neighbouring beams as

$$\mathcal{N}_i = \{j \mid A_{ij} = 1\}$$

Now we follow a greedy group assignment by iterating over the beams in the defined spiral order. For beam b we identify the allowed groups

$$allowed_b = \{g \in \{1, \dots, G\} \mid g \notin groups[\mathcal{N}_b]\}$$

We select the group with the fewest beams assigned so far

$$g^* = \arg \min_{g \in allowed_b} |\mathcal{B}_g|$$

and we assign the beam and update the group definitions

$$groups[b] = g^*, \quad |\mathcal{B}_{g^*}| = |\mathcal{B}_{g^*}| + 1$$

3.3.3 Numerical evaluation

The evaluation of the proposed beam layout schemes in Section 3.3.1 and of the beam multiplexing solution in Section 3.3.2 is carried out following the same procedure as in Section 3.1.3. We also resort to the system configuration and parameters of Section 3.1.3 and to the power normalization that considers the beamformer used, unless otherwise stated.

Using the scheme E as a baseline, we start evaluating the impact of the proposed optimized beam assignment (opt. BA) with respect to the deterministic solution (det. BA) presented in Section 3.1.3. Besides, we also compare the performance of the power equalization in Section 3.1.3 and the reference with equal power allocation to all simultaneous beams. The results are depicted in Figure 29.

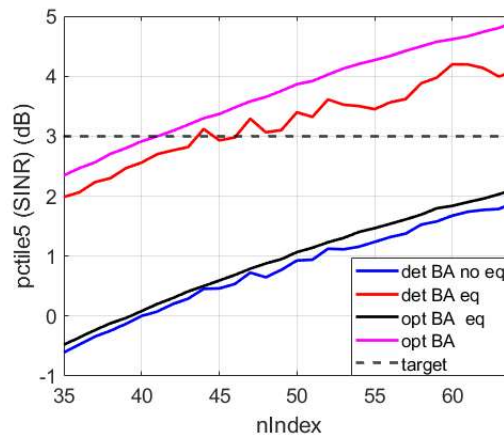


Figure 29: Percentile 5 of SINR across the coverage as a function of the number of beam indexes (groups) nIndex for scheme E.

It can be observed that the power allocation equalizing beam gains provides a significant improvement over the equal power allocation, which does not meet the 3 dB target even for 64 indexes (i.e. beam groups). The main reason is that equalizing SNR levels in all beams avoids having large low-SNR regions at coverage edge. On top of it, the optimized beam assignment provides an extra improvement over the deterministic one reducing the minimum number of beam groups from $nIndex=44$ to $nIndex=42$. Remarkably the curve for the deterministic beam assignment is not smooth due to the differences on the number of beams per group/index.

Let us now focus on the comparison of the different beam layout schemes, using power allocation for beam gain equalization and the optimized beam assignment. Figure 30 depicts the simulation results.

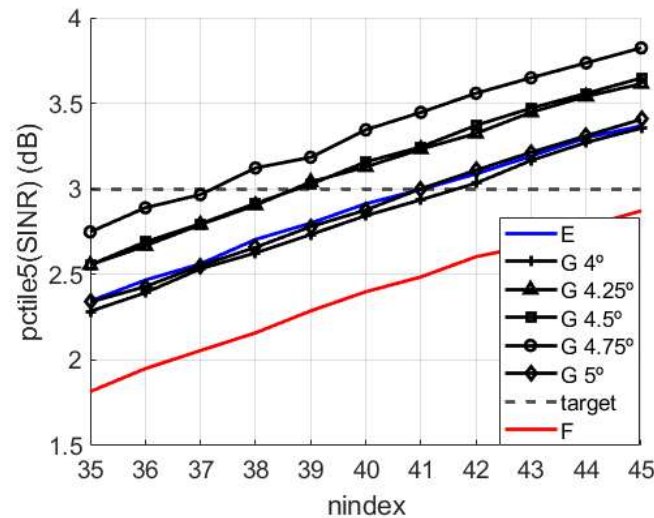


Figure 30: Percentile 5 of SINR across the coverage for different beam layout schemes.

Using scheme E with -3dB beam, cross overs as a baseline, it can be observed that using extra beams at the edge to avoid coverage holes as in scheme F has a detrimental effect, with a clear increase in the number of beam groups required to meet the 3 dB target. In contrast, as in the Earth-fixed case, increasing the distance between beam footprints, thus reducing the total number of beams allows reducing the required number of groups. However, an optimum point at $\alpha = 4.75^\circ$ is observed requiring $nIndex=38$, beyond which further separation increases the number of groups required due to the low beam cross over levels. Table 12 summarizes the numerical results.

Table 12: Summary of simulation results

Design	Total number of beams	Required number of beam indexes ($nIndex$) for 95% coverage above 3 dB	Minimum number of beams per beam index	Maximum number of beams per beam index
E: $\alpha = 3.64^\circ$ (-3dB beam crossovers)	649	42	15	16
F: $\alpha = 3.64^\circ$ extended coverage	673	46	14	15
G 4°: $\alpha = 4^\circ$	547	42	12	14
G 4.25°: $\alpha = 4.25^\circ$	475	39	12	13

G 4.5°: $\alpha = 4.5^\circ$	433	39	11	12
G 4.75°: $\alpha = 4.75^\circ$	379	38	9	11
G 5°: $\alpha = 5^\circ$	349	41	8	9

Figure 31 shows the distribution of the coverage point not fulfilling the target $SINR \geq 3$ dB. They mainly correspond to the edge of the beams, both at coverage center and at coverage edge.

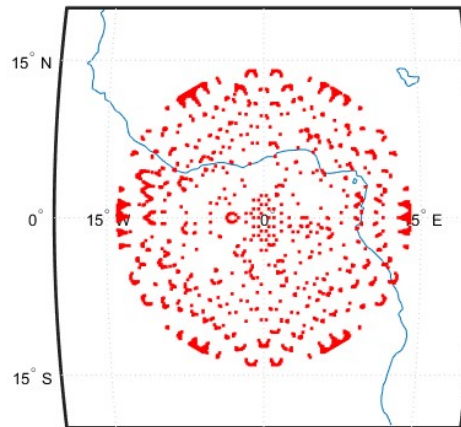


Figure 31: Distribution of coverage points with $SINR < 3$ dB

3.3.4 Earth-fixed vs Earth-moving

To compare Earth-fixed and Earth-moving schemes we resort to the CS efficiency, defined in Section 3.2.1 as the number of slots devoted to common signalling broadcast over the total number of slots within a SSB period. We focus on the best schemes obtained for both fixed and moving beams, hence schemes D3 and G 4.75°, respectively. Table 13 shows the comparison for the CS scheduling solutions in Section 3.2.1, with more aggressive one including all CS in half slot, and a more conservative using a full slot. The improvement of Earth-Moving with respect to Earth-Fixed is below 10% for the half-slot case, but it grows to a significant 24.5 for the full slot.

Table 13: Comparison of CS efficiency

Design	Half slot	Full slot
D3: Maximum EIRP beams with beams crossover at 7°	80.6%	61.3%
G 4.75°: $\alpha = 4.75^\circ$	88.1%	76.3%
Improvement	9%	24.5%

Recommendation: Allow for the spatial multiplexing of SSB indexes to enable the mapping of a single 5G cell to the whole coverage area, so that satellite mobility can be handled at beam management level and the efficiency of Earth-moving beam layout can be exploited.

4 USER-CENTRIC BEAMFORMING

The previous sections focused on the design of broadcast beams for delivering signalling and common control information. In contrast, this section addresses the design of data beams used to transmit user-specific information through user-centric beamforming.

In this Section, we focus on the design and performance of user-centric beamforming solutions by: i) completing the design of the system-level procedures and computing an estimate of the overhead introduced by the transmission of ancillary information needed for the beamforming algorithms; ii) designing a methodology to generate non-uniform traffic databases for residential broadband and air traffic scenarios; iii) designing two schedulers for user-centric beamforming, in which priority metrics can be flexibly defined; iv) assessing the performance with ideal and non-ideal ancillary information in more than 1000 different combinations of traffic and scheduler configurations; v) designing a framework for federated user-centric beamforming and assessing the performance; vi) developing a framework for user-centric beamforming with OFDM and OTFS waveforms, and assessing their performance under different channel and impairment conditions.

4.1 SYSTEM DEFINITION

In this sub-section, we summarise the selected architecture configuration and the main assumptions, also referring to what already discussed in the preliminary report [1].

4.1.1 Architecture and network elements

We focus on **Distributed Beamforming Management (DBFM) architectures**, which are based on the assumption that all beamforming related operations¹ are performed on-board the serving NTN node(s). Figure 32 provides an extension of this architecture, initially designed in [1] for a standalone NTN node, with $N_{sat} \geq 1$ NTN nodes organised into a swarm that creates a **distributed antenna system in space**.

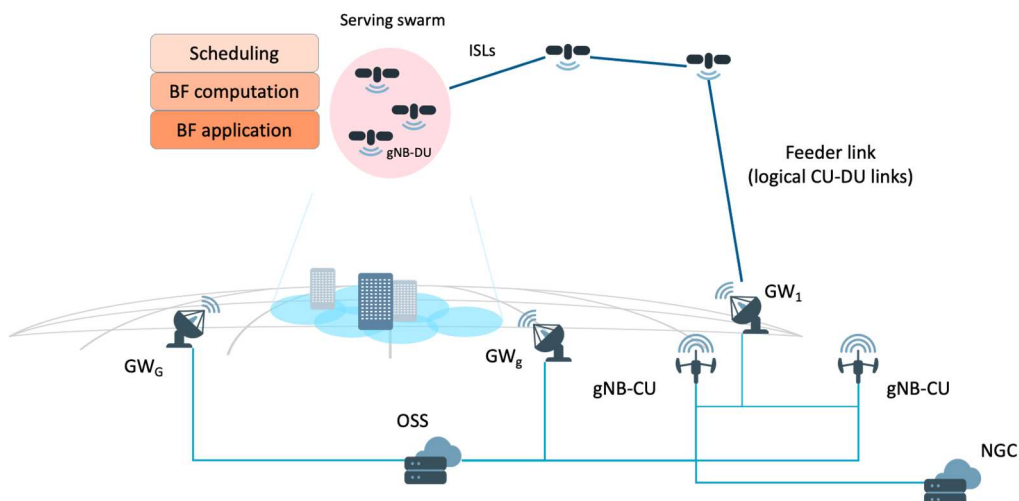


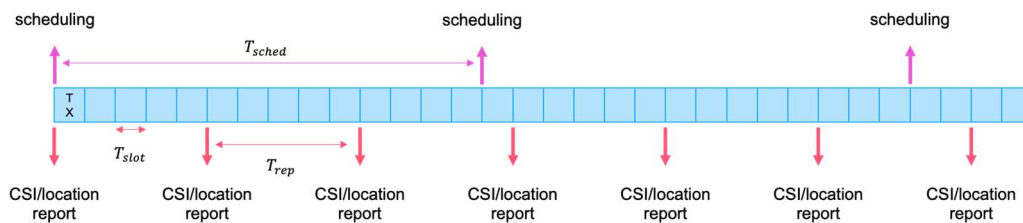
Figure 32: Distributed Beamforming Management (DBFM) architecture with one or more NTN nodes.

¹ As discussed in [1], these include: the **computation of the scheduling matrix**, the **computation of the beamforming coefficients**, and the **application of the beamforming matrix** to the signals to be transmitted.

This architecture assumes that Option 4 (or higher layer splits, up to Option 1 right above PDCP/SDAP) is considered for functional split, *i.e.*, the gNB-Distributed Unit (gNB-DU) on-board at least one of the swarm nodes implement at least up to the High MAC layer (as requested for Option 4). It shall be noticed that not necessarily all of the swarm nodes require the same level of on-board capabilities, as all computations can be performed on-board a master (more computationally advanced) NTN node while the others only implement lower layer functions, *e.g.*, a Radio Unit (RU). Notably, this approach calls for the implementation of intra-swarm Inter-Satellite Links (ISLs), which in any case are needed since, to support federated beamforming in space, the swarm shall be tightly synchronised in the frequency and time domain.

Similarly to the standalone node scenario, the following assumptions and considerations hold throughout this document, unless otherwise stated:

- The ground segment includes G Gateways (GWs) providing NTN access to the terrestrial infrastructure, in addition to including the Operation Support Systems (OSS), in charge of the overall satellite system management.
- The Non-Terrestrial (NT) access segment provides Earth-fixed coverage, but no impact is expected on any consideration in this document in case Earth-moving systems are deployed. In the considered DBFM architecture, the serving gNB-DU(s) might be in direct visibility of the GW/gNB-CU(s), thus not requiring inter-swarm Inter-Satellite Links (ISLs); however, once over the visibility of the serving GW, one or more ISLs will be needed to guarantee the connection to the ground segment and access to the core network or other data networks. When $N_{sat} \geq 1$, *i.e.*, in the federated beamforming case, at least ISLs at swarm level (intra-swarm) are needed in order to coordinate the swarm (which requires a tight time and frequency synchronisation) and to allow the exchange of information such as the users' reports or the beamforming coefficients².
- In the swarm, as previously mentioned, at least one node (master node) implements at least up to the High MAC layer on-board; this node is the one in charge of computing the scheduling matrix and the beamforming coefficients, which are then provided to the other nodes³. The other swarm nodes can, in principle, be simpler and only implement the RU, to support the distributed antenna array in space. Clearly, in case more swarm nodes have advanced on-board capabilities (again up to the High MAC), the computational burden can be more efficiently and fairly split among them. These considerations are in line with what was designed in [1], with the on-board Network Element (NE) in charge of computing the scheduling and beamforming coefficients for DBFM architectures.



² Compared to other sources of latency in the system, the over-the-air latency introduced by intra-swarm ISLs can be neglected due to the reduced distances between the cooperating NTN nodes.

³ Please note that, in principle, the master node can also send to the other swarm nodes the beamformed signals, similarly to what is done in On-Ground Beamforming (OGBF) systems. However, this distinction does not impact the generality of the observations and analyses in this document.

Figure 33: High-level relationship between scheduling and reporting intervals, [1].

In general, the master node(s) shall perform the following operations:

- Implement the Radio Resource Management (RRM) algorithm based on the traffic requests from the User Equipments (UEs). As shown in Figure 33, we denote the periodicity with which the scheduling matrix is computed as T_{sched} , while T_{slot} and T_{rep} denote the slot duration and the periodicity of ancillary information reporting, respectively. Assuming that $N_{sched}^{(slot)}$ time slots are included in a scheduling period T_{sched} , with N_{sched} being the number of scheduling periods, the scheduling matrix $\mathbf{S}^{(m)}$ for the a generic scheduling period is a $N_B \times N_{sched}^{(slot)}$ matrix, where N_B denotes the number of separate user-centric beams per transmission slot. Denoting by $\mathbf{S} = [\mathbf{S}^{(1)}, \dots, \mathbf{S}^{(m)}, \dots, \mathbf{S}^{(N_{sched})}]$ the overall scheduling matrix, obtained as concatenation of those of each scheduling period: i) the b -th row $\mathbf{s}_{b,:}$ provides the indexes of each user to be served through the generic b -th beam in each time slot; and ii) the t -th column $\mathbf{s}_{:,t}$ contains the indexes of the users to be served in the generic t -th time slot, with $t = 1, \dots, N_{tot}^{(slot)}$ and $N_{tot}^{(slot)}$ being the total number of time slots in which the satellite is passing over the service area. Thus, the satellite is visible over the service area for a time $N_{tot}^{(slot)} T_{slot}$.
- Computing the beamforming coefficients for each time slot, based on the scheduling matrix defined by the RRM algorithm. In the generic t -th time slot, the master node then computes a complex beamforming matrix $\mathbf{B}^{(t)}$ with size $N_R \times N_B$, where N_R denotes the total number of radiating elements in the distributed antenna system. When $N_{sat} = 1$, we are in the scenario already described in [1] and the number of total radiating elements is equal to those on a single satellite, $N_R = N_{R,sat}$; in the federated beamforming case, $N_R = N_{R,sat} N_{sat}$ (for instance, $N_{R,sat} = 512$ as designed in 5G-STARDUST leads to $N_R = 1024$ assuming $N_{sat} = 2$, which is the case we will evaluate later in this document).
- The computation of the beamforming coefficients and the RRM algorithm are based on the channel vectors of each user, which are obtained either as estimated Channel State Information (CSI) or inferred from the estimated location by each visible user. In the former case, we thus assume that the UEs directly provide their CSI reports to the master NTN node; in the latter, the master node shall interact with the NGC to obtain the location of each visible UE, at least based on current 3GPP specifications. The location information, as introduced in [1], is exploited to extrapolate all non-stochastic terms in the channel coefficients, leaving all stochastic terms out of the estimated CSI (e.g., atmospheric loss and shadowing). As such, in clear-sky conditions and assuming ideal ancillary information, CSI-based and location-based solutions are equivalent.

The reports provided by the users shall include the following information: i) the requested traffic type and request; ii) the ancillary information (CSI or location estimates); and iii) the terminal type/class, to derive information to estimate the noise power at the receiving equipment (required for the beamforming algorithms). The latter might be classified by the terminal manufacturers; in such cases, estimates shall be exploited.

In the designed architecture, we can now detail the system-level procedures supporting both standalone and federated user-centric beamforming.

4.1.2 System-level procedures

In terms of system-level procedures, no modification is needed with respect to what already introduced in [1]. The only difference is that we may now have multiple satellites for user-centric beamforming, which calls for more overhead as we will discuss soon, but in terms of

the flow of information no impact arises, and the same considerations made for the standalone case still hold. Thus, extending the procedure for standalone NTN nodes to the case of NTN swarms, and referring to Figure 34, the following functional operations are performed:

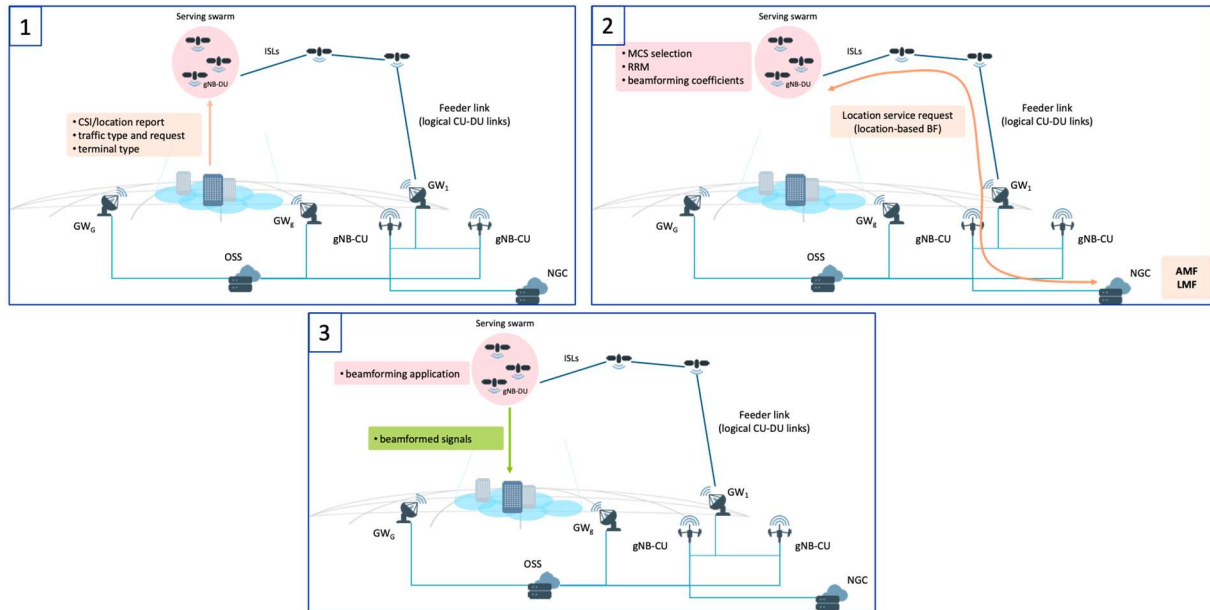


Figure 34: DBFM system-level procedures.

1. Initially, it can be assumed that the NE in charge of computing the beamforming coefficients (the swarm master node) knows the radiation pattern at radiating element level (required only for location-based beamforming) and the total available power on-board (required for both CSI-based and location-based beamforming).
2. The UEs in visibility of the NTN nodes in the swarm requiring connectivity provide to the master NTN node(s) the estimated location or CSI reports, the traffic request and type, and the type of terminal, if available.
3. For CSI-based solutions, the NTN node applies the RRM algorithm to compute the scheduling, the beamforming algorithm to compute the coefficients, and then selects the best Modulation and Coding Scheme (MCS) per user. For location-based solutions, these operations can only be performed after the location of the UEs requesting a service has been retrieved from the Location Management Function (LMF) in the core network.
4. The NTN master node sends the beamforming coefficients to the other swarm nodes and the beamformed signals are then sent to the UEs via a distributed antenna array in space.

Below, we discuss how the master node(s) can obtain the required ancillary information.

4.1.3 Beamforming algorithms and ancillary information

To support user-centric beamforming, the UEs need to provide ancillary information to the master NTN node(s) in the swarm, and the type of information depends on the category of the implemented beamforming algorithm, [19]. In general, these can be categorised as: i) CSI-based, when the complex channel coefficients between each on-board radiating element and the UE receiving antenna shall be estimated and provided; or ii) location-based, when the latitude and longitude coordinates of the UE are required.

CSI-BASED

First focusing on CSI-based solutions, from 3GPP specifications we know that the CSI can be provided by the UE with periodic, aperiodic, or semi-persistent approaches. Focusing on the periodic case, which is required in our system, the periodicity can range from 4 time slots to up to 640 slots. Assuming a sub-carrier spacing of 120 kHz (240 kHz), as typical and allowed in Ka-band, this leads to at most 80 ms (40 ms) periodicities.

As shown in Figure 35, the UE estimates the CSI based on the CSI Reference Signal (CSI-RS) transmitted by the satellite and then determines the best CSI reporting parameters, in addition to encoding them. In general, there are two types of beamforming algorithms depending on the type of feedback that is provided: i) **codebook-based**, in which a predetermined set of beam patterns (beamforming matrices) is used; and ii) **non-codebook-based**, in which the beam patterns (beamforming matrices) are dynamically computed. In both cases, CSI-related feedback shall be provided by the UEs to let the master node(s) select the best beamforming matrix (codebook-based) or explicitly compute the beamforming matrix (non-codebook-based).

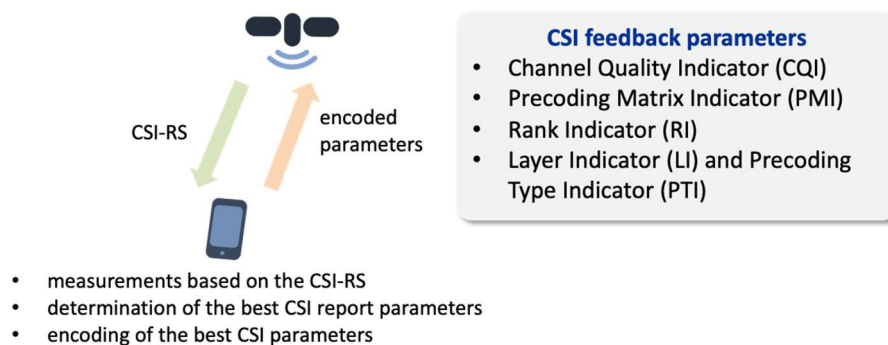


Figure 35: Ancillary information estimation and reporting in CSI-based beamforming.

The CSI feedback that is provided by the user can be one of the following:

- **Channel Quality Indicator (CQI)**: the quality of the downlink channel, derived from the received Signal-to-Noise Ratio (SNR) and the interference experienced on the downlink channel, which can be related to a combination of MCS and transport block size.
- **Precoding Matrix Indicator (PMI)**: an indicator allowing the UE to signal the best beam to the transmitter. This indicator contains two parts: a beam indicator, allowing the selection of the best vertical and horizontal beam, and a phase indicator, allowing the fine-tuning of the phase.
- **Rank Indicator (RI)**: a parameter that indicates to the transmitter the number of spatial streams that can be supported for data transmission, allowing the identification of the appropriate precoding and scheduling strategies.
- **Layer Indicator (LI)**: a parameter that indicates to the transmitter the number of spatial layers that the UE can utilise for data transmission⁴.

All these types of CSI reporting options are essential for codebook-based precoding and they are all supported by 3GPP specifications; however, they are all based on an

⁴ It shall be noticed that LI specifies the number of layers available for data transmission, which is related to the channel conditions, while data streams refer to the actual flow of data being transmitted over those layers.

approximation/quantisation of the actual CSI coefficients, *i.e.*, they do not provide the actual coefficients as the focus is on codebook-based solutions. This clearly has the advantage of reducing the overhead and to allow pre-computing the beamforming/precoding matrices, to some extent, but it also negatively impacts the performance due to the lack of precision for the reported CSI values (the beamforming matrix is in fact based on an approximation of the actual channel matrix); moreover, in particular with rapidly moving Non-Geosynchronous Orbit (NGSO) nodes, the database containing the codebooks might easily become very large since it also shall take into account the impact of the elevation angle. Non-codebook-based solutions, *i.e.*, reporting options providing the exact CSI values, would yield a much better performance and not require databases, but they suffer from two major drawbacks: i) there is currently no way to obtain the CSI values at radiating element level, as the specifications (*i.e.*, the CSI-RSs) allow the channel coefficients to be estimated only after beamforming; and ii) they introduce a significant overhead in the reporting phase.

Figure 36 shows an estimation of the required overhead per user as a function of the number of bits per coefficient and the number of satellites. For instance, in 5G-STARUST we have $N_{R,sat} = 512$; thus, assuming 32 bits per complex coefficient (which means 16 bits per component, not even particularly accurate), we can easily deduce that 32 Mbit or 81.92 Mbit of data would be transmitted from the UEs to the NTN segment assuming 1000 and 5000 users, respectively; these values apply for each single NTN node in the swarm, *i.e.*, with two satellites and 5000 UEs, we would have a 163.84 Mbit overhead every 80 ms.

Clearly, this amount of overhead is not practical, and it calls for different solutions. In particular, location-based beamforming algorithms can be considered, as in this report, but also more advanced solutions based on Artificial Intelligence (AI) can be designed. These aspects will be discussed in Deliverable D4.7, including, specifically, a solution for estimating the expected beamformed Signal-to-Interference-plus-Noise Ratio (SINR) via Neural Networks (NNs).

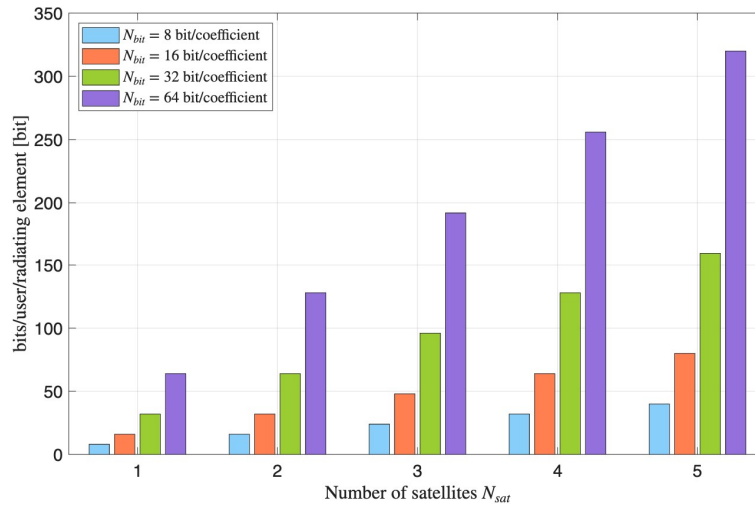


Figure 36: Estimation of the overhead per user, per radiating element, per reporting period, to provide the CSI feedback to the NTN segment as a function of the bits per complex channel coefficient.

LOCATION-BASED

In location-based solutions, the users shall provide their location during the reporting period, as shown in Figure 37; then, the master NTN node exploits this information to infer the channel coefficients based on channel emulators or AI solutions. In this context:

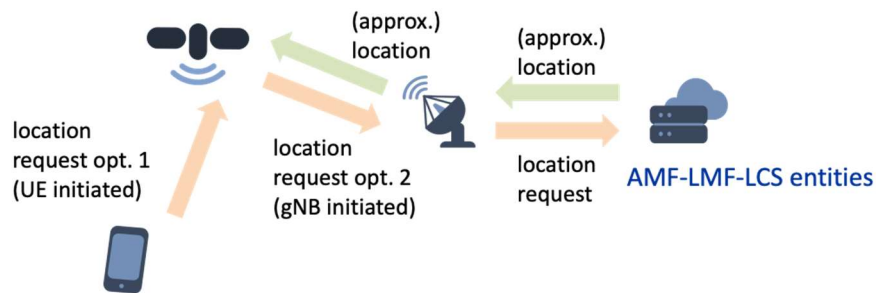


Figure 37: Ancillary information estimation and reporting in location-based beamforming.

- In 5G or 5G-Advanced NTN systems, the UEs are mandatorily equipped with Global Navigation Satellite System (GNSS) capabilities and, thus, are able to estimate and report their location. However, it shall be mentioned (please see [1] for a detailed discussion on this topic) that the location information is not usually available in the RAN but only in the core network; thus, as shown in the above figure, the NTN master node shall request the location to the 5GC. Having the exact users' locations in the RAN, *i.e.*, in the NTN segment, would call for modifications to the current specifications in terms of privacy.
- For future 6G systems, network-based location estimation will be required as GNSS-free solutions are being explored. One of the initial assumptions will be that 1 or 3 satellites will be in visibility for this estimation, which is expected to lead to an accuracy of 100 m.

In this context, focusing on 5G/5G-Advanced systems, we can consider 32 bits per coordinate as in GPS or Galileo. For a single satellite, this leads to 64 kbit or 320 kbit with 1000 or 5000 users, respectively, which is significantly below the amount of overhead in CSI-based solutions. Thus, despite the challenges (due to privacy and specifications, not technical) that need to be faced in this case, the massive reduction in terms of overhead makes location-based algorithms a truly valuable option.

4.2 USER LOCATION AND TRAFFIC DISTRIBUTION

In the following, we design and implement a procedure to model non-uniform distributions for the users' locations and traffic requests. These will be then used for the performance evaluation of the different scheduling algorithms designed for user-centric beamforming.



Figure 38: Considered areas for the elaboration of the JRC GHSL dataset.

It shall be noticed that, while below we also introduce a model for air traffic scenarios, the performance evaluation of the scheduling algorithms will be based on the residential broadband case. The impact of a moving receiver, as on-board an aircraft, is not considered critical for system-level performance assessment for two main reasons: i) information on the movement of the aircraft can be easily made available to predict/adjust the location or CSI information; and ii) previous studies, [19], [20], [21], showed that errors on the location of users, due for instance to movement, have a relatively limited impact on the system-level performance. Thus, we focus on the residential broadband scenario as it leads to systems with ultimately larger capacity requests from many more locations on the service area, *i.e.*, harsher scheduling conditions as we will extensively discuss later in this document.

4.2.1 Residential broadband scenario

4.2.1.1 Population dataset

The traffic characteristics for this scenario are based on the information made available by the **Joint Research Center (JRC) Global Human Settlement Layer (GHSL)**, [22], [23]. It contains a geographical raster of the estimated population count per grid point, with grid points obtained at different resolutions (2 meters, 10 meters, 100 meters, 1 km, 3 arcsec, or 30 arcsec) and with various coordinate systems (Mollweide, WGS84, and UTM). Moreover, both past population data (since 1975) and its predictions (up to 2030) are available. In the following, we refer to the population predictions for 2030 with a resolution of 30 arcsec (which corresponds to grid points of approximately 1 km²). As shown in Figure 38, the data can be accessed from the website selecting a specific area identified by a row and column index.

Figure 39 shows the obtained data, with the population being clipped at 5000 people per square km, since the figure would have been otherwise more difficult to be understood (there are areas with more than 16000 people per square km, as shown in the Cumulative Distribution Function, CDF, on the right). From the CDF, we can also observe that:

- 98.74% of the grid points have at most 2000 inhabitants;
- 97.26% of the grid points have at most 1000 inhabitants;
- 95.18% of the grid points have at most 500 inhabitants.

These values suggest that areas with extremely high population densities are relatively limited.

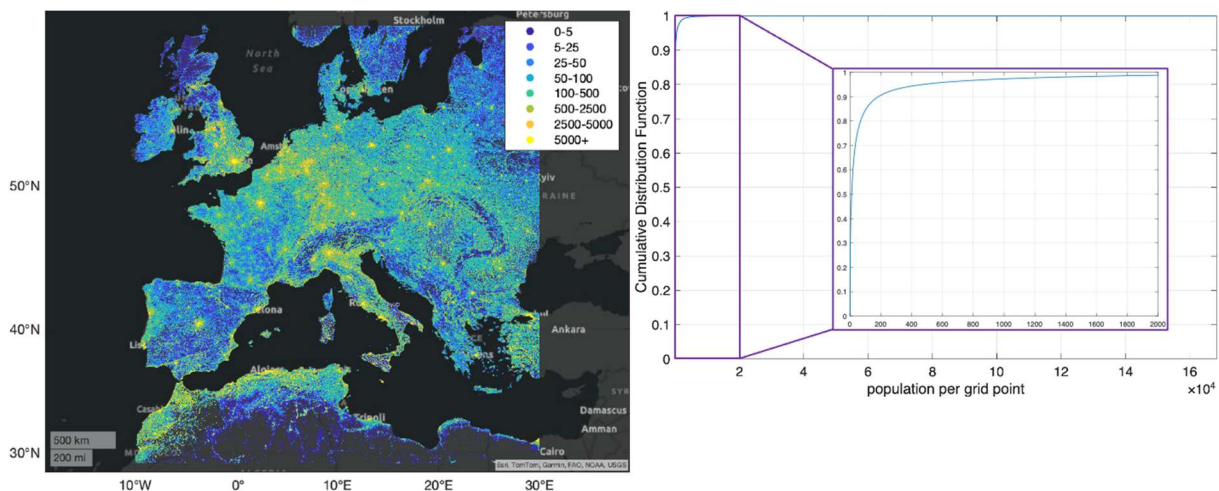


Figure 39: Population data for the selected area of interest.

4.2.1.2 Traffic dataset generation

Figure 40 shows the methodology that was designed to obtain parametric non-uniform traffic distributions for user-centric beamforming analyses, which is detailed below.

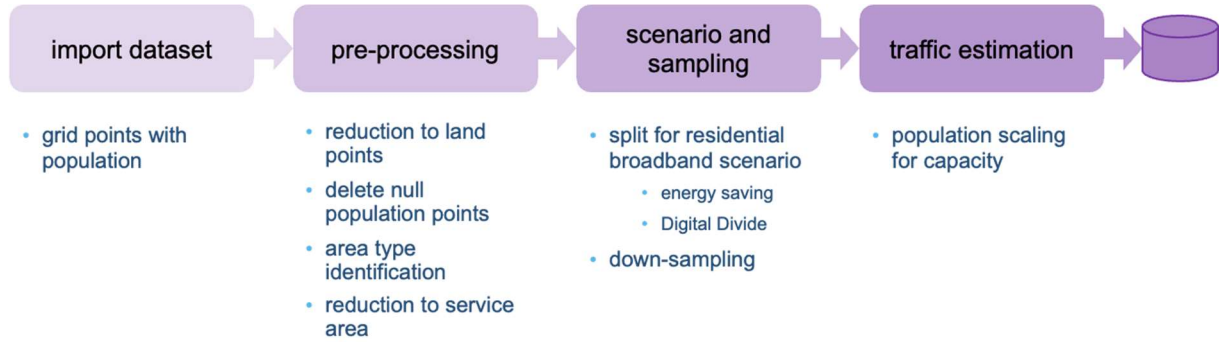


Figure 40: Methodology for the generation of the non-uniform traffic dataset.

PRE-PROCESSING

The first step to be performed is to pre-process the dataset containing the population information. This includes reducing the dataset to land areas only and to delete any entry with a null population, which also slightly relaxes the computational burden of the following steps.

Once the dataset has been reduced to meaningful grid points, we need a system to classify each grid point (approximately 1 square km) based on the population it contains: such mapping will allow us to distinguish, for instance, rural areas and dense urban ones. To this aim, we implement a minor modification to the procedure reported in [24] and [25], adding one area class; in the procedure proposed in the literature, each grid point is classified taking into account the neighbouring areas as follows:

- We denote by P_i the population in the generic i -th grid point, with $i = 1, \dots, N_{grid}$ and N_{grid} being the total number of grid points. Each grid point is preliminary classified as follows:
 - $P_i < 100$: potential rural area point;
 - $100 \leq P_i < 300$: potential sub-urban point;
 - $300 \leq P_i < 1500$: potential urban point;
 - $1500 \leq P_i$: potential dense-urban point.

The procedure implemented in [24]- [25] considers a single area for the first two points, while in the modified procedure we split them so as to better separate rural areas from sub-urban areas. Once each grid point has been classified, clusters are defined as continuous grid points with the same preliminary classification.

- Each cluster is classified as sub-urban, urban, or dense-urban based on the total population in it. In particular, denoting as $\mathcal{C}_c$ the set of indexes of the grid points in the c -th cluster and as $P_{T,c} = \sum_{i \in \mathcal{C}_c} P_i$ the total population:
 - $P_{T,c} \geq 50000$: dense-urban area;

- $5000 \leq P_{T,c} < 50000$: urban area;
- $500 \leq P_{T,c} < 5000$: sub-urban area;

all other areas are automatically classified as rural.

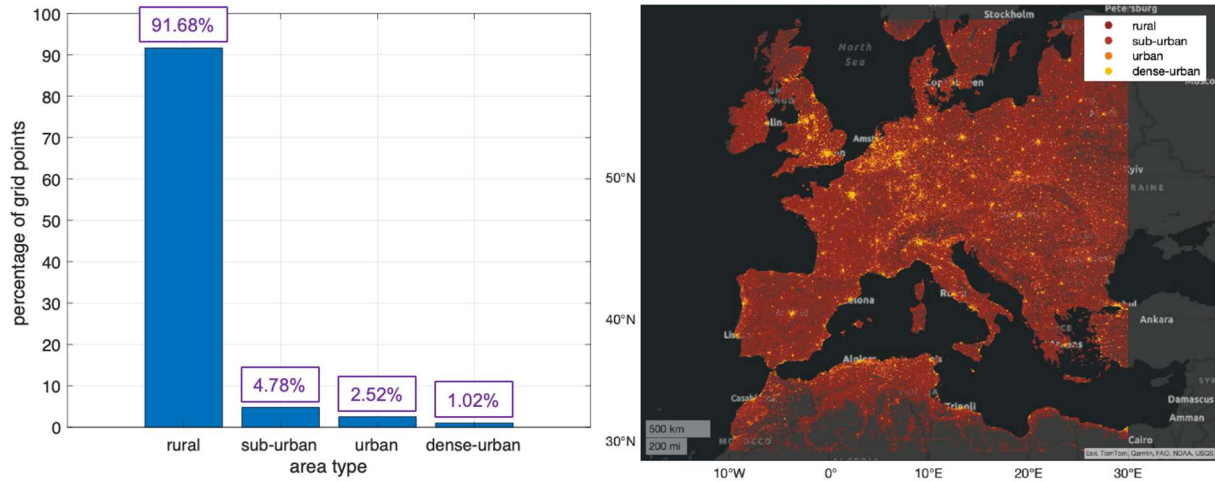


Figure 41: Classification of the grid points based on the population.

Figure 41 shows the considered grid points and their classification with the above procedure (right), in addition to the distribution of their classes (left). It can be noticed that the vast majority of the grid points belongs to rural areas.

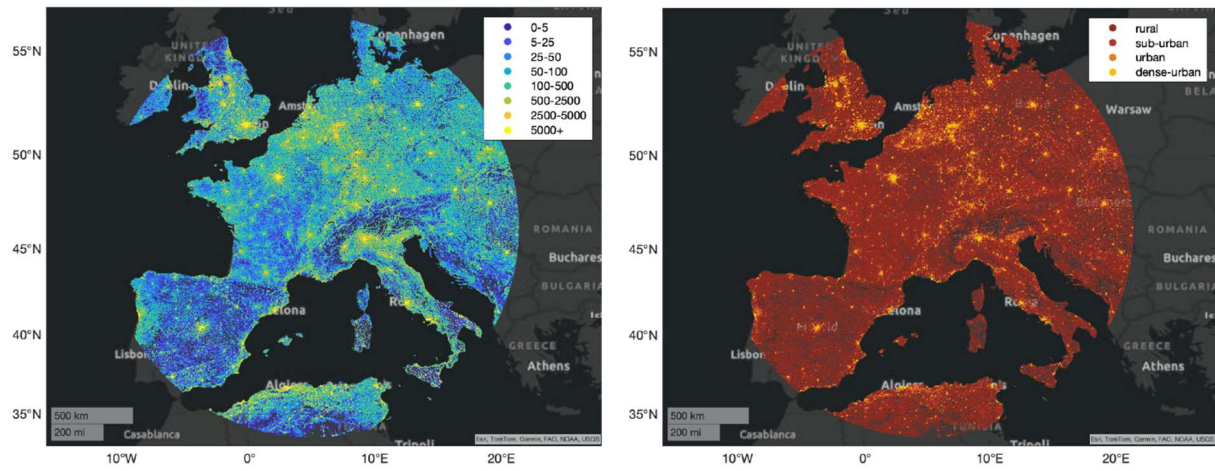


Figure 42: Population per grid point (left) and classification (right) in the service area for the polar orbit constellation (1000 km) and $\varepsilon_s = 30^\circ$.

The last step in the pre-processing phase is to reduce the grid points to those in the service area. As an example, let us consider the polar constellation of the 5G-STARDUST system at 1000 km altitude, [2], and a service area defined as follows, [1]: i) centre in $(\text{lat}_s, \text{lon}_s) = (45^\circ\text{N}, 5^\circ\text{E})$; and ii) radius defined based on the minimum user elevation angle, $\varepsilon_s = 30^\circ$ (as per 3GPP specifications). Figure 42 shows the population data and the resulting point classification in this scenario.

SCENARIO AND SAMPLING

At this stage, we obtained a full dataset of all land grid points in the service area with non-null population and a resolution of 1 km². The next step requires an elaboration of this data in order to consider two aspects: i) in the residential broadband category, we have two specific scenarios (energy saving and large sparsely populated areas); and ii) not all grid points will be active (*i.e.*, requesting traffic) at any moment, so there will be some kind of activity probability to be taken into account. With respect to the former, we can differentiate the cases as follows:

- **energy saving scenario:** all area types are considered (from rural to dense-urban), as energy saving is independent of the area, but only night periods are retained (which means that a reduction on the activity probability shall be introduced to account for this);
- **large sparsely populated areas** (in the following denoted as Digital Divide for the sake of conciseness): only rural areas are considered, for the entire day.

Based on these observations, the probability that a given grid point is active and thus requesting traffic, is computed as follows:

$$p_A = \alpha \cdot \beta_{NTN} \cdot \beta_{SNO} \cdot \beta_{POL} \cdot \beta_{SAT} \cdot \beta_{D2N} \quad (1)$$

where:

- α is the system activity factor, usually in the order of 20%-40%;
- β_{NTN} is the expected percentage of NTN users by 2030, [26];
- $\beta_{SNO} = 1/N_{SNO}$ takes into account that we are focusing on a single Satellite Network Operator (SNO) out of a total of N_{SNO} potentially available in the service area;
- $\beta_{POL} = 1/N_{POL}$ takes into account that we focus on a single polarization out of N_{POL} ;
- $\beta_{SAT} = 1/(N_{sat}^{(incl)} + N_{sat}^{(pol)})$ allows us to take into account that, for the moment being, we focus on a single satellite out of a total of $N_{sat}^{(incl)} + N_{sat}^{(pol)}$ simultaneously visible from a user, with $N_{sat}^{(incl)}$ and $N_{sat}^{(pol)}$ denoting those from the inclined and polar constellation shells, respectively;
- finally, β_{D2N} is only considered for the energy saving scenario and it accounts for the reduction in the activity factor when moving from day traffic to night traffic.

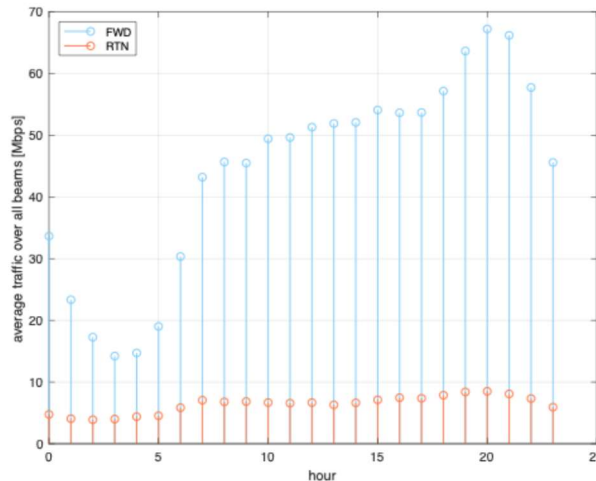


Figure 43: Average traffic over all beams on the forward and return links as a function of the time of the day, obtained from the Hispasat dataset.

With respect to β_{D2N} , Figure 43 shows the average (per hour) traffic over all beams on the forward (FWD) and return (RTN) links as a function of the time of the day, obtained from the Hispasat dataset. Since we are interested in the day-to-night traffic ratio, considering as day all hours between 7:00 and 23:00 we obtain:

- FWD: $\beta_{D2N} = \frac{21.79}{53.39} = 0.4082$;
- RTN: $\beta_{D2N} = \frac{4.49}{7.15} = 0.6280$.

Table 14 summarises the values of the parameters in eq. (1) that will be considered in the following, unless otherwise stated. Figure 44 shows an example of population distribution (which is non-uniform in both the location and the population per location) obtained with the proposed procedure. As described below in the next, and final, database generation step, the population per grid point will be used to obtain a non-uniform distribution of the traffic request.

Table 14: Summary of the parameters to compute the activity probability per grid point.

Parameter	Value
Activity factor α	0.3, 0.4
Percentage of NTN users in 2030 β_{NTN}	7.5%
Percentage of serving SNOs β_{SNO}	25% ($N_{SNO} = 4$)
Percentage of available polarisations β_{POL}	50% ($N_{POL} = 2$)
Percentage of simultaneously visible satellites β_{SAT}	16.67% ($N_{sat}^{(incl)} + N_{sat}^{(pol)} = 5 + 1$)
Day-to-night traffic ratio β_{D2N}	40.82%

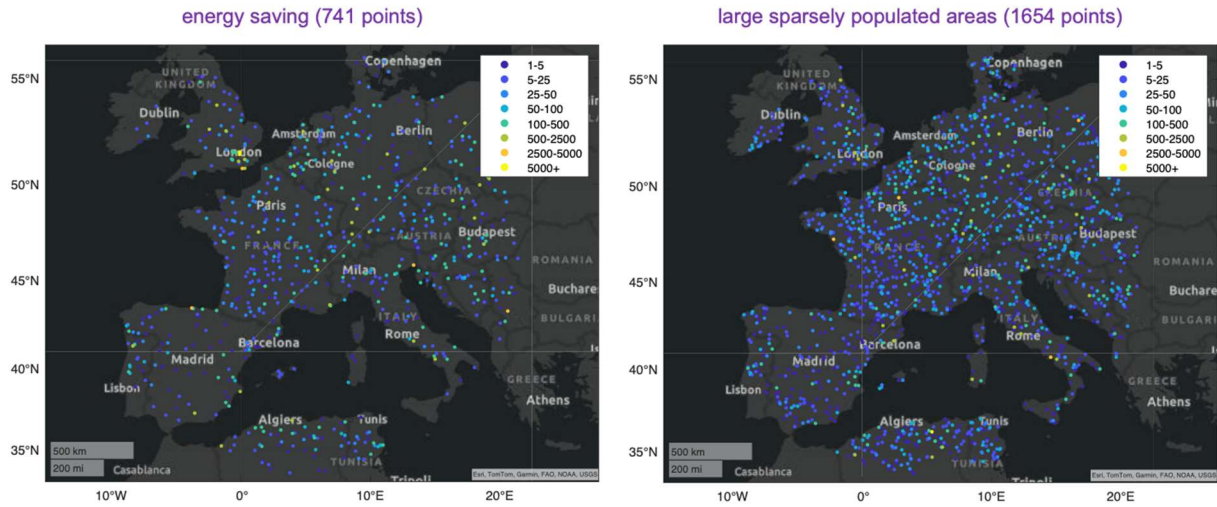


Figure 44: Example of population distribution for energy saving (left) and digital divide (right) scenarios, polar orbit constellation (1000 km) and $\varepsilon_s = 30^\circ$.

TRAFFIC ESTIMATION

In the following, we assume that each grid point is associated with a single Very Small Aperture Terminal (VSAT). As a consequence, the traffic request that we associate to each grid point actually represents the aggregate request in that area (1 km²), rather than the request of a single user. This assumption is based on the following considerations. Notably, the beamforming performance is negatively impacted by a scheduler selecting too similar CSI vectors, which usually (but not always) correspond to geographically close users; in fact, in these situations the channel matrix becomes ill-conditioned. In this context, modelling more than one receiver in an area of 1 km² would lead to the presence of many receivers that cannot be scheduled together, either increasing the scheduler computational complexity or leading to a poor performance in case these situations cannot be properly managed by the scheduler. Since the CSI vectors, as seen from the orbit, are almost identical for users separated by at most a few tens of meters, we can easily model multiple receivers as a single VSAT, delegating to the aggregate capacity the modelling of multiple users requesting traffic in that area.

The aggregate traffic request per grid point is obtained by introducing two additional parameters that will lead to different scenario configurations during the performance assessment: the minimum and the maximum capacity requests, C_{min} and C_{max} , respectively. For a given scenario obtained by randomly down-sampling the grid points as discussed above, we have a population distribution with values between $[P_{min}, P_{max}]$, where $P_{min} = \min_i P_i$ and

$P_{max} = \max_i P_i$. This population data is rescaled in order to provide values in the range $[C_{min}, C_{max}]$, i.e., the aggregate capacity per grid point. This leads to a non-uniform traffic request in which the grid point with the least population is requesting the minimum capacity and the grid point with the most population is requesting the maximum capacity. As already mentioned, the minimum and maximum capacity requests are parameters that will be varied to lead to different traffic scenarios; Figure 45 shows the CDF of a realisation of the traffic request per VSAT receiver (grid point) in the digital divide scenario, based on the proposed procedure and assuming 5 or 20 Mbps for the minimum capacity and 100 or 500 Mbps for the maximum capacity request.

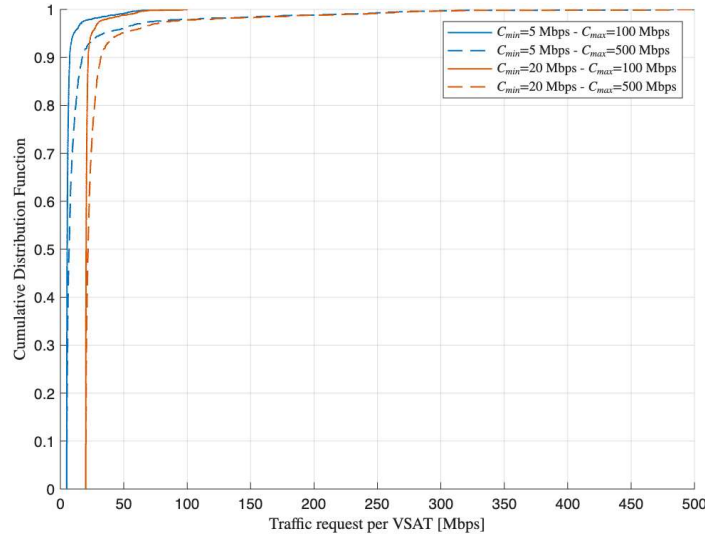


Figure 45: Example of traffic request distribution with $C_{min} = 5$ (blue), 20 (orange) Mbps and $C_{max} = 100$ (solid), 500 (dashed) Mbps, polar orbit constellation (1000 km) and $\varepsilon_s = 30^\circ$.

4.2.2 Air traffic scenario

For the air traffic scenario, we have to consider two specific use cases as well: in-flight entertainment and switch at boarding. For the former, we implement the following procedure:

- consider the number of NGSO air terminals in 2026, which is approximately 3000, provided in [26];
- equally split the estimated capacity in 2026, again provided in [26] leading to approximately 26.33 Mbps per plane. The plane movement can be modelled as detailed in [1].

With respect to the switch at boarding scenario, we have to consider airport traffic. To this aim, we can proceed as follows:

- select the 100 busiest airports in Europe (e.g., from this [website](#));
- add grid points to each location within the identified service area;
- associate to each airport the ratio of flights per day and:
 - scale it in order to obtain a total of 3000 flights (the value we consider for in-flight entertainment);

- multiply by the capacity per flight, again already obtained above;
- scale the capacity per airport by the percentage of time in which the satellite is visible;
- Note: with user-centric beamforming, planes in the same airport must be served in different time slots to avoid ill-conditioned channel matrices. As such, the airport location is assumed to approximate all of its planes.

Figure 46 shows an example of the proposed procedure for the switch at boarding case.

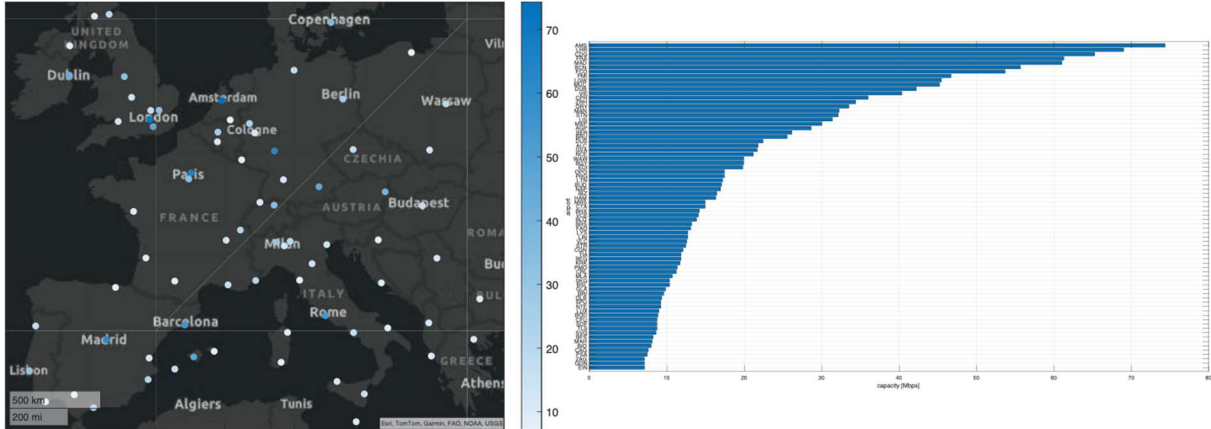


Figure 46: Example of traffic distribution per airport in the switch at boarding use case.

4.3 SYSTEM-LEVEL ANALYSES

4.3.1 Standalone beamforming

In this sub-section, we design two scheduling algorithms and extensively assess their performance assuming a single NTN node. Before introducing them, for the sake of completeness and clarity, we summarise the system model and assumptions from [1]. As we will see, these concepts are instrumental for a complete understanding of the designed schedulers and the corresponding performance.

4.3.1.1 System model and KPIs

The standalone NTN node is equipped with a Direct Radiating Array (DRA) antenna containing $N_{R,sat} = 512$ radiating elements, [1], [2]. Assuming an Earth-fixed coverage, the service area is defined in terms of: i) the coordinates of its centre, denoted as (lat_s, lon_s) ; and ii) its extension, defined through the NTN node Field of View (FoV), which is adjusted to guarantee a minimum elevation angle ε_s from altitude h_{sat} and assuming the Sub-Node Point (SNP) at (lat_s, lon_s) . The standalone node is considered the serving satellite for the duration of a complete pass over the service area, *i.e.*, the period during which its ground track is fully contained within the service area boundary. Figure 47 depicts the considered service area and the NTN node ground track with: i) a satellite of the 5G-STARUST polar constellation ($i = 99^\circ$) at $h_{sat} = 1000$ km altitude; ii) a minimum elevation angle of $\varepsilon_s = 30^\circ$ (as per 3GPP), leading to a FoV of 48.46° ; and iii) the center of the service area located at $(lat_s, lon_s) = (45^\circ N, 5^\circ E)$.

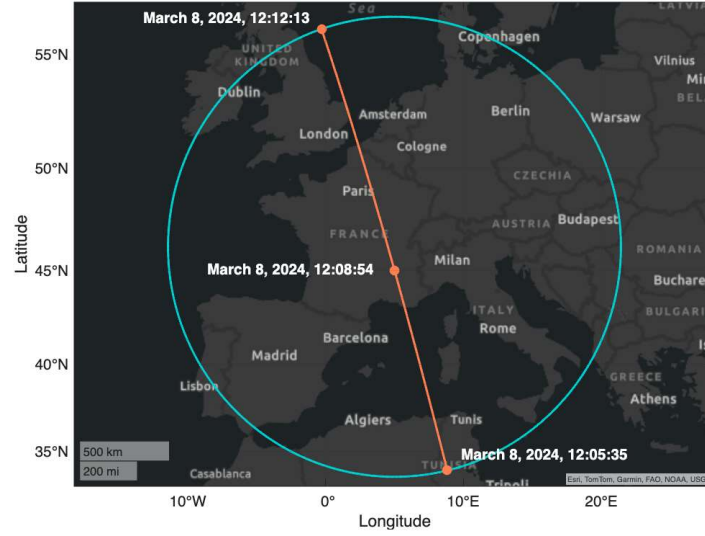


Figure 47: Service area and ground track for standalone user-centric beamforming with $(lat_S = 45^\circ N, lon_S = 5^\circ E)$, $\varepsilon_S = 30^\circ$, $h_{sat} = 1000$ km (5G-STARUST polar constellation, $i = 99^\circ$).

As outlined in Section 4.1.3, in a reporting instant the NTN node receives the ancillary information (CSI or location, depending on the beamforming algorithm category) from the UEs and it uses this information to compute the scheduling matrix. Based on $\mathbf{S}$, and again relying on the received ancillary information, the beamforming matrix for each of the $N_{sched} N_{sched}^{(slot)}$ time slots managed by the scheduler is computed. In line with [1] and 3GPP TR 38.811, [27], and TR 38.821, [18], the channel coefficient in a generic time slot t between the i -th on-ground UE and the n -th on-board radiating element can be computed as:

$$h_{i,n}^{(t)} = \frac{g_{i,n}^{(tx,t)} g_{i,n}^{(rx,t)}}{\frac{4\pi d_i^{(t)}}{\lambda} \sqrt{L_i^{(t)} \kappa B_i T_i}} e^{-j \frac{2\pi}{\lambda} d_i^{(t)}}, i = 1, \dots, N_{UE}, n = 1, \dots, N_R \quad (2)$$

where:

- $L_i^{(t)}$ is a term including all stochastic losses (shadowing, atmospheric losses, scintillation, and clutter loss in Line-of-Sight, LoS, or Non-LoS, NLoS, conditions, as per 3GPP TR 38.811 and TR 38.821);
- $g_{i,n}^{(tx,t)}$ and $g_{i,n}^{(rx,t)}$ represent the transmitting and receiving complex antenna patterns between the n -th element and the i -th UE, respectively (the antenna model is extensively detailed in [1]);
- $\kappa B_i T_i$ represents the thermal noise power, with B_i being the user bandwidth, κ Boltzmann's constant, and T_i the equivalent noise temperature;
- $e^{-j \frac{2\pi}{\lambda} d_i^{(t)}}$ represents the phase shift due to the slant range $d_i^{(t)}$.

In CSI-based techniques (assuming proper modifications have been introduced, as discussed in Section 4.1.3), the UEs directly estimate the channel coefficients in eq. (2) as $\hat{h}_{i,n}^{(t_R)}$ and provide them to the NTN node to compute the scheduling matrix and the beamforming coefficients; t_R denotes the generic latest reporting slot to highlight that reports are not continuously provided to the NTN node, but they are only available with specific periodicities

(e.g., every 80 ms). In location-based solutions, the UEs estimate their location and the NTN node derives the channel coefficients based on this information and the current NTN node position on its orbit, which is known on-board; notably, this approach leads to an estimated coefficient that does not include the stochastic terms, $L_i^{(t)}$, as only geometry-related (i.e., deterministic) terms can be inferred, i.e.:

$$\hat{h}_{i,n}^{(t_R)} = \frac{\hat{g}_{i,n}^{(tx,t_R)} \hat{g}_{i,n}^{(rx,t_R)}}{\frac{4\pi \hat{d}_i^{(t_R)}}{\lambda} \sqrt{\kappa B_i T_i}} e^{-j \frac{2\pi}{\lambda} \hat{d}_i^{(t_R)}} \quad (3)$$

where the hat symbol over the parameters indicates their estimate based on the location, and again we refer to a generic reporting instant t_R .

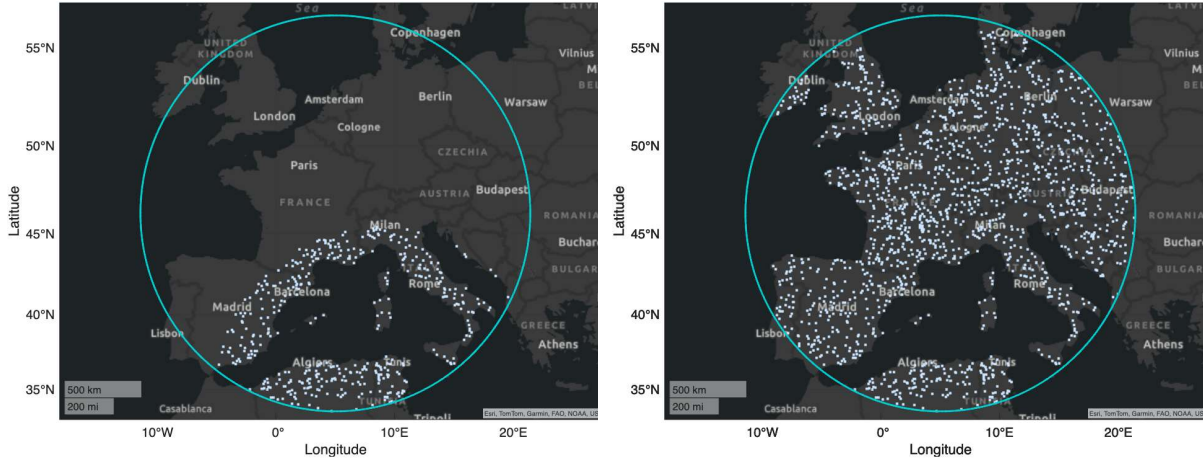
Let us denote by N_{UE} the total number of users in the service area, obtained from one of the databases extensively discussed in Section 4.2. In the generic reporting instant t_R , i.e., when users shall estimate the ancillary information and report it, only a subset of the N_{UE} users is in the visibility of the NTN node, i.e., with an elevation angle above a system target ε_t . If we denote as $\varepsilon_i^{(t_R)}$ the elevation angle of the i -th user in the generic reporting instant t_R , then the number of users that are in visibility and can report ancillary information, i.e., the UEs that can potentially be scheduled by the RRM algorithm as they have a valid report, is given by:

$$N_{UE}^{(t_R)} = \left| \left\{ i: \varepsilon_i^{(t_R)} \geq \varepsilon_t, i = 1, \dots, N_{UE} \right\} \right| \quad (4)$$

An example of visible users depending on the NTN node position is shown in Figure 48. Based on the latest received information and the specific RRM algorithm (designed in the next sections), in the generic scheduling instant $t_S > t_R$ (this condition ensures that scheduling is based on the latest report) the NTN node defines the $N_B \times N_{sched}^{(slot)}$ scheduling matrix $\mathbf{S}^{(m)}$, which indicates the visible users to be served in each of the upcoming $N_{sched}^{(slot)}$ time slots, up to N_B per time slot. The (b, t) -th element $s_{b,t}$ of the scheduling matrix contains the index of the user to be served via the generic b -th user-centric beam in the generic t -th time slot.

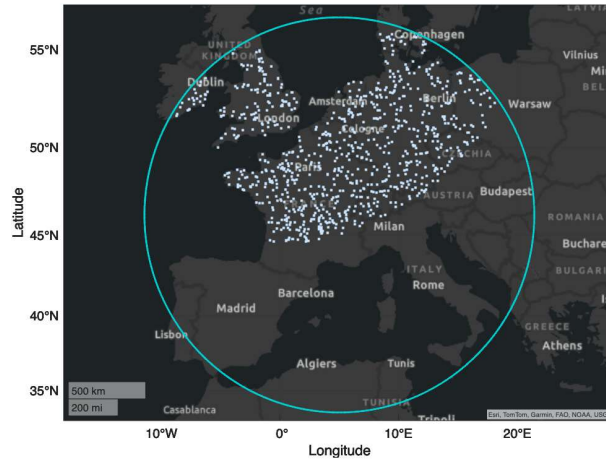
Assuming that the RRM algorithm provided the scheduling matrix (see Section 4.3.1.2 below and Figure 49 for an example of user-centric scheduled users), in the generic t -th transmission time slot the NTN node computes the beamforming coefficients by: i) selecting the N_B -dimensional t -th column of the scheduling matrix $\mathbf{S}$, $\mathbf{s}_{:,t}$; and ii) creating the estimated channel matrix at time slot t , $\hat{\mathbf{H}}^{(t)}$, from the available ancillary information. The estimated channel matrix provides, on each row, the estimated CSI vector of the users to be served in the t -th time slot; thus, denoting by $\hat{\mathbf{h}}_{s_{b,t},:}^{(t)} = [\hat{h}_{s_{b,t},1}^{(t)}, \dots, \hat{h}_{s_{b,t},n}^{(t)}, \dots, \hat{h}_{s_{b,t},N_R}^{(t)}]$ the complex channel coefficients between the user scheduled for the b -th beam in the t -th time slot, provided by $s_{b,t}$, and the $N_R = N_{R,sat}$ radiating elements on-board, we have that:

$$\hat{\mathbf{H}}^{(t)} = \left[\left(\hat{\mathbf{h}}_{s_{1,t},:}^{(t)} \right)^T, \dots, \left(\hat{\mathbf{h}}_{s_{b,t},:}^{(t)} \right)^T, \left(\hat{\mathbf{h}}_{s_{N_B,t},:}^{(t)} \right)^T \right]^T \quad (5)$$



(a) Visible users at 12:05:35

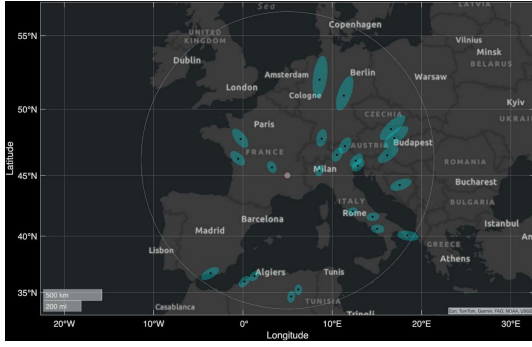
(b) Visible users at 12:08:54



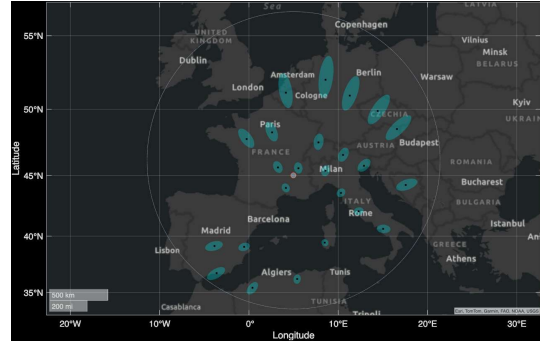
(c) Visible users at 12:12:13

Figure 48: Example of visible users according to the NTN node position on its orbit with $(lat_S = 45^\circ N, lon_S = 5^\circ E)$, $\varepsilon_S = 30^\circ$, $h_{sat} = 1000$ km (5G-STARUST polar constellation, $i = 99^\circ$).

This matrix is the input to the beamforming algorithm, which we generically denote as a function $\mathcal{B}(\cdot)$ providing the $N_R \times N_B$ beamforming matrix $\mathbf{B}^{(t)} = \mathcal{B}(\hat{\mathbf{H}}^{(t)})$. The details on the considered beamforming algorithms are available in [1] and summarised in Annex A. In this deliverable, we consider both the **Minimum Mean Square Error (MMSE)** and **Conventional Beamforming (CBF)** algorithms; as for the former, we consider legacy MMSE (CSI-based) and Location Based MMSE (LB-MMSE) solutions. Finally, in terms of the power distribution, based on the extensive performance assessment in [1] and Annex A to this document, we consider only the **Per Antenna Constraint (PAC)**, in which each radiating element is assumed to transmit the same amount of power $\frac{P_{av}}{N_{R,sat}}$, with P_{av} being the total available power on-board the NTN node. In fact, in the considered system PAC is the power distribution strategy providing the best performance when taking into account the achievable spectral efficiency, the exploitation of the on-board available power, and the limitation of the power at the input of the on-board amplifiers to avoid non-linearities. For the sake of completeness, Table 15 summarises the equations of the considered beamforming algorithms and the power distribution strategy.



(a) $\rho_{GEO} = 1$



(b) $\rho_{GEO} = 3$

Figure 49: Example of scheduled users at 12:07:38 with different values of the geographical scheduling constraint (see Section 4.3.1.2.)

The normalised beamforming matrix $\tilde{\mathbf{B}}^{(t)}$ is the one used to transmit the beamformed signals to the scheduled users. Denoting by $\mathbf{x}^{(t)} = [x_1^{(t)}, \dots, x_{N_B}^{(t)}]^T$ the N_B -dimensional vector of unit-variance symbols to be transmitted to the scheduled users in the generic t -th slot, the signal received by the generic k -th user can be written as:

$$y_k^{(t)} = \underbrace{\mathbf{h}_{k,:}^{(t)} \tilde{\mathbf{b}}_{:,k}^{(t)} x_k^{(t)}}_{\text{intended}} + \underbrace{\sum_{\substack{\ell=1 \\ \ell \neq k}}^{N_B} \mathbf{h}_{k,:}^{(t)} \tilde{\mathbf{b}}_{:, \ell}^{(t)} x_{\ell}^{(t)}}_{\text{interfering}} + z_k^{(t)} \quad (6)$$

Table 15: Summary of the considered beamforming algorithms and power distribution strategy.

Algorithm	Category	Equation	Estimated channel coefficient
MMSE	CSI-based	$\mathbf{B}^{(t)} = (\hat{\mathbf{H}}^{(t)})^H (\hat{\mathbf{H}}^{(t)} (\hat{\mathbf{H}}^{(t)})^H + \alpha \mathbf{I}_{N_B})^{-1}$	Estimation of all terms in eq. (2)
LB-MMSE	Location-based		Estimation of all non-stochastic terms in eq. (3)
CBF	Location-based	$\mathbf{b}_{:,k}^{(t)} = \frac{1}{\sqrt{N_R}} \sum_{n=1}^{N_R} e^{-jk_0 \mathbf{r}_n \mathbf{p}_i^{(t)}} = \frac{1}{\sqrt{N_R}} \Phi^* (u_i^{(t)}, v_i^{(t)})$	Estimation of all non-stochastic terms in eq. (3)

PAC power distribution: $\tilde{\mathbf{B}}^{(t)} = \sqrt{\frac{P_{av}}{N_R}} \text{diag} \left(\frac{1}{\|\mathbf{b}_{1,:}^{(t)}\|}, \dots, \frac{1}{\|\mathbf{b}_{N_R,:}^{(t)}\|} \right) \mathbf{B}^{(t)}$

NOTE 1: $\alpha = \frac{N_R}{P_{av}}$ is the MMSE beamforming regularisation factor

NOTE 2: in CBF, $\mathbf{r}_n$ represents the position of the n -th radiating element on the array and $\mathbf{p}_i^{(t)}$ the (u, v) coordinates (direction cosines) of the i -th user at the t -th time instant

where $z_k^{(t)}$ is a circularly symmetric Gaussian random variable (r.v.) with zero mean and unit variance, which is licit observing that the channel coefficients defined above are normalised to the noise power. It shall be noticed that, in eq. (6), the beamforming matrix is computed referring to the ancillary information obtained in the latest reporting period, while the channel encountered during the actual transmission is clearly, more or less, different. More specifically, the larger the aging interval $t - t_R$, the less aligned the beamforming matrix $\tilde{\mathbf{B}}^{(t)}$ will be to the actual channel conditions during the transmission, *i.e.*, to the $N_B \times N_R$ actual complex channel matrix $\mathbf{H}^{(t)} = \left[\left(\mathbf{h}_{s_1,t,:}^{(t)} \right)^T, \dots, \left(\mathbf{h}_{s_b,t,:}^{(t)} \right)^T, \dots, \left(\mathbf{h}_{s_{N_B,t,:}}^{(t)} \right)^T \right]^T$. Notably, the aging interval is always upper

bounded by the reporting periodicity, *i.e.*, $\Delta t = t - t_R \leq T_{rep}$. Equation (6) can be easily extended to obtain the column vector of all received symbols during the t -th time slot:

$$\mathbf{y}^{(t)} = \mathbf{H}^{(t)} \tilde{\mathbf{B}}^{(t)} \mathbf{x}^{(t)} + \mathbf{z}^{(t)} \quad (7)$$

From eq. (7), we can compute the $N_B \times N_B$ **power transfer matrix** as, [1]:

$$\mathbf{A}^{(t)} = |\mathbf{H}^{(t)} \tilde{\mathbf{B}}^{(t)}|^2 \quad (8)$$

which provides, in the generic (i, j) -th element, the power transmitted from the NTN node to the j -th user and received by the i -th user, normalised to the noise thermal power. Thus, all diagonal elements, $i = j$, represent the normalised **intended power** level, while all off-diagonal elements represent the normalised **interfering power** levels; in particular, all off-diagonal elements on a given row provide the interference received by the i -th user due to the other $N_B - 1$ transmissions scheduled in the same time slot. Bearing this in mind, it is straightforward to obtain the following Key Performance Indicators (KPIs) for the generic k -th user:

- The **Signal-to-Noise Ratio (SNR)** of the k -th user in the t -th time slot, given by the k -th diagonal element of the power transfer matrix:

$$\text{SNR}_{k,t} = a_{k,k}^{(t)} \quad (9)$$

- The **Interference-to-Noise Ratio (INR)** of the k -th user in the t -th time slot, given by the sum of the off-diagonal elements of the k -th row in the power transfer matrix:

$$\text{INR}_{k,t} = \sum_{\substack{\ell=1 \\ \ell \neq k}}^{N_B} a_{k,\ell}^{(t)} \quad (10)$$

- The **SINR**, obtained from the SNR and INR in equations (9) and (11), respectively:

$$\text{SINR}_{k,t} = \frac{\text{SNR}_{k,t}}{1 + \text{INR}_{k,t}} = \frac{\|\mathbf{h}_{k,:}^{(t)} \tilde{\mathbf{b}}_{:,k}^{(t)}\|^2}{1 + \sum_{\ell=1, \ell \neq k}^{N_B} \|\mathbf{h}_{k,:}^{(t)} \tilde{\mathbf{b}}_{:,\ell}^{(t)}\|^2} \quad (11)$$

From the above SINR, the rate achieved by the k -th user can be evaluated either from the Shannon bound formula or from the adopted MCS. In this framework, 3GPP TR 38.803, [28], indicates that the spectral efficiency for system-level simulations can be computed as:

$$\eta_{k,t} = \begin{cases} 0, & \text{SINR}_{k,t} < \text{SINR}_{\min} \\ \rho \cdot \log_2(1 + \text{SINR}_{k,t}), & \text{SINR}_{\min} \leq \text{SINR}_{k,t} \leq \text{SINR}_{\max} \\ \rho \cdot \log_2(1 + \text{SINR}_{\max}), & \text{SINR}_{k,t} > \text{SINR}_{\max} \end{cases} \quad (12)$$

where $\text{SINR}_{\min} = -10$ dB and $\text{SINR}_{\max} = 30$ dB are the minimum and maximum SINR values, respectively, and ρ is a factor representing the implementation losses. Since this factor is a multiplicative term outside the Shannon formula, we can assume $\rho = 1$ as a different value would simply act as scaling factor of all the results, without impacting the observations and the relationships among the various techniques and scenarios.

It shall be noticed that the spectral efficiency computed in eq. (12) represents the **asymptotic spectral efficiency**, *i.e.*, the spectral efficiency achieved assuming that the generic k -th user is granted a resource (user-centric beam) in all the available time slots and not just a subset

of them (depending on the RRM strategy). To take the performance of the specific RRM algorithm into account, we need to consider the **experienced spectral efficiency** per user, which is the sum of achieved spectral efficiencies weighted by the percentage of channel use (*i.e.*, a weighted average of the asymptotic spectral efficiencies). To compute this quantity, let us denote by $\boldsymbol{\eta}_k$ the vector of spectral efficiencies achieved by the k -th user during the satellite pass, *i.e.*, $\boldsymbol{\eta}_k = [\eta_{k,1}, \dots, \eta_{k,t}, \dots, \eta_{k,N_{tot}^{(slot)}}]$ where $N_{tot}^{(slot)}$ is the total number of time slots (as previously stated, the satellite is passing over the service area for a time period equal to $T = N_{tot}^{(slot)} T_{slot}$). Let us also denote by $T_{vis,k} = N_k^{(slot)} T_{slot}$ the total visibility time for the generic k -th user; $T_{vis,k}$ represents the total amount of time in which the user can be served by this satellite (in other periods, it will be scheduled by another NTN node in the constellation). The experienced spectral efficiency of the k -th user is thus obtained as:

$$\bar{\eta}_k = \sum_{t=1}^{N_{tot}^{(slot)}} \frac{\eta_{k,t} T_{slot}}{T_{vis,k}} = \frac{1}{N_k^{(slot)}} \sum_{t=1}^{N_{tot}^{(slot)}} \eta_{k,t} \quad (13)$$

Finally, since we are operating with Full Frequency Reuse (FFR), all the users are allocated the entire bandwidth B in each time slot. Thus, we can compute the total **achieved capacity** of the generic k -th user as:

$$C_k^{(A)} = B \cdot \bar{\eta}_k = \frac{B}{N_k^{(slot)}} \sum_{t=1}^{N_{tot}^{(slot)}} \eta_{k,t} = \sum_{t=1}^{N_{tot}^{(slot)}} C_{k,t}^{(A)} \quad (14)$$

where in the last formulation we brought the bandwidth in the summation to obtain the achieved capacity as the sum of all the achieved capacities of the user in its allocated time slots. This metric provides the total achieved capacity at system-level during a full NTN node pass over the service area, for the generic k -th user. Differently from [1], in which random scheduling algorithms were considered because the traffic request from the users was uniform, here we consider **non-uniform traffic requests**. In particular, denoting as $C_k^{(REQ)}$ the capacity requested by the generic k -th user (see Section 4.2), we introduce the following metrics:

- The **achieved sum capacity**, providing the total achieved capacity for all users (*i.e.*, the total capacity provided at system-level by the NTN node in a service period T):

$$C^{(A)} = \sum_{k=1}^{N_{UE}} C_k^{(A)} = B \sum_{k=1}^{N_{UE}} \bar{\eta}_k = \frac{B}{N_k^{(slot)}} \sum_{t=1}^{N_{tot}^{(slot)}} \sum_{k=1}^{N_{UE}} \eta_{k,t} \quad (15)$$

- The **unmet capacity of the k -th user**, measuring the amount of capacity that the user requested but was not satisfied by the NTN node in the service period T :

$$C_k^{(U)} = \max(C_k^{(R)} - C_k^{(A)}, 0) \quad (16)$$

This metric is null when $C_k^{(R)} - C_k^{(A)} < 0$, which corresponds to a scenario in which the k -th user was allocated at least the capacity that was requested. From this metric, which will be extensively exploited by the scheduling algorithms designed in the next section, we can derive the **unmet sum capacity**, which provides the total amount of capacity that remained unmet by the NTN node after the service period T :

$$C^{(U)} = \sum_{k=1}^{N_{UE}} C_k^{(U)} = \sum_{k=1}^{N_{UE}} \max(C_k^{(R)} - C_k^{(A)}, 0) \quad (17)$$

It is convenient to define an additional metric to be exploited in designing the scheduling algorithms below, namely the **unmet capacity of the k -th user at time t** :

$$C_{k,t}^{(U)} = \max\left(C_k^{(R)} - \sum_{i=1}^{t-1} C_{k,i}^{(A)}, 0\right) \quad (18)$$

- The **excess sum capacity**, which provides a measure of the amount of capacity that has been exceeded compared to the actual requests of the users:

$$C^{(E)} = \max(C^{(A)} - C^{(R)}, 0) = \max\left(\sum_{k=1}^{N_{UE}} (C_k^{(A)} - C_k^{(R)}), 0\right) \quad (19)$$

This metric might be non-null because the RRM algorithms, as extensively discussed in the next sections, will reintroduce users for which the capacity request was already met in the scheduling loop, in case resources are still available.

These metrics, and in specific cases the corresponding CDFs, will be used as the main KPIs to assess the performance of the designed scheduling algorithms, which we now introduce.

4.3.1.2 Scheduling

In this section, we design two scheduling algorithms for user-centric beamforming, namely the **Priority Queue Scheduler (PQS)**, the benchmark, and the **Priority Rank Scheduler (PRS)**.

Notably, as already discussed in [1] and assessed in the corresponding performance evaluation in Annex A, the system performance is deeply impacted by the similarity among the CSI vectors of the users scheduled in the same time slot. This means that, when building the beamforming matrix $\tilde{\mathbf{B}}^{(t)}$ for the generic t -th time slot based on the estimated channel matrix $\hat{\mathbf{H}}^{(t)}$, schedulers leading to rows in $\hat{\mathbf{H}}^{(t)}$ that are more correlated to each other will lead to a worse performance in terms of the SINR as per eq. (11), due to the ill-conditioning of the channel matrix; as such, they will ultimately lead to a poor performance in terms of achieved capacity. In the literature, these aspects have been considered in the past by designing pretty complex schedulers based on graph theory, integer programming, or iterative solutions requiring for each tentative allocation the computation of the beamforming matrix (leading to poorly efficient solutions), among the others. In this framework, the objective of the schedulers designed in the following is to exploit ancillary information on the users (channel or location) to improve the performance of user-centric beamforming, while maintaining a relatively low complexity of the schedulers that shall be implemented on-board.

In this framework, before describing the PQS and PRS solutions, we introduce two constraints to be applied to the scheduling algorithms, aiming at improving the beamformer performance based on the ancillary information. Based on the above observations, and taking into account that users close in terms of geographical position tends to have similar channel vectors, we introduce two different **scheduling constraints aimed at avoiding ill-conditioning in $\hat{\mathbf{H}}^{(t)}$** :

- **geographical scheduling constraint** of the i -th user in the t -th time slot:

$$\Gamma_{GEO,i}^{(t)}(\mathcal{W}): d_{i,j} \geq \rho_{GEO} r_{uv}, \forall j \in \mathcal{W} \quad (20)$$

where: i) $\mathcal{W}$ is a generic test set containing the indexes of users with respect to which the condition shall be verified; ii) $d_{i,j}$ represents the distance between the i -th reference user and the j -th test user in the comparison set $\mathcal{W}$; iii) r_{uv} is the beam radius in the nadir direction computed in (u, v) coordinates; and iv) ρ_{GEO} is the geographical scheduling constant that will be a scheduler parameter. If we consider $\mathcal{W} = \mathbf{s}_{:,t}$, *i.e.*, the test set contains the users currently scheduled in the t -th time slot, the geographic constraint $\Gamma_{GEO,i}^{(t)}$ is equal to 1 if and only if the reference user i is sufficiently distant from all the users already included in the scheduler in this time slot: in this case, the reference user i can be safely added to the time slot. Notably, the distance $d_{i,j}$ can be computed with any function, but in the following we will consider the Euclidean distance. It shall be noticed that, being the distance in (u, v) coordinates that represent the direction cosines from the NTN node, we are actually measuring the distance between users in a space that implicitly considers their directions rather than the actual geographical distance on the spherical Earth. In any case, in the following we will refer to this method as “geographical,” to distinguish it from the next one based on the channel coefficients.

- **CSI scheduling constraint** of the i -th user in the t -th time slot:

$$\Gamma_{CSI,i}^{(t)}(\mathcal{W}): \gamma_{i,j} = \frac{|\hat{\mathbf{h}}_{i,:}^{(t)} \hat{\mathbf{h}}_{j,:}^{(t)}|}{\|\hat{\mathbf{h}}_{i,:}^{(t)}\| \|\hat{\mathbf{h}}_{j,:}^{(t)}\|} \leq \rho_{CSI}, \forall j \in \mathcal{W} \quad (21)$$

where ρ_{CSI} is the CSI scheduling constant that will be a scheduler parameter. In this case, we are measuring the CSI Coefficient of Correlation (CoC) between the i -th and the j -th user, $\gamma_{i,j}$, and comparing it to a threshold: if the coefficient of correlation is too large, the users' channels are too similar and they should not be scheduled in the same time slot.

It shall be noticed that, since close users tends to exhibit similar (*i.e.*, correlated) channel vectors, the two approaches might seem similar. However, in some cases even users that are distant from a geographical perspective (even considering direction cosines) might still experience similar channels, due to the peculiarities of the NTN channel including the periodicity introduced by the slant range. As such, we might have users meeting the geographical scheduling constraint in (20) but not the CSI scheduling constraint in (21), and *vice versa*. This behaviour can be observed in Figure 50 and Figure 51, which show the reference user, the potential users, and the barred users⁵ in the polar constellation with different values of the geographical or CSI scheduling constants, respectively. It can be observed that: i) with the geographical scheduling constraint, we are restricting the potential users to those outside a protection area in the surrounding of the test user (not necessarily a circle, as we are considering the direction cosines, but rather an ellipses elongated according to the current NTN node position on its orbit); and ii) with the CSI scheduling constraint, even users that are far away from the test one might exhibit correlated CSI vectors and shall, thus, be identified as barred and excluded from potential scheduling in the same slot.

⁵ In the rest of the document, the **set of barred users** refers to the set of users that are prohibited from being scheduled in the current time slot (the time slot being defined by the scheduler).

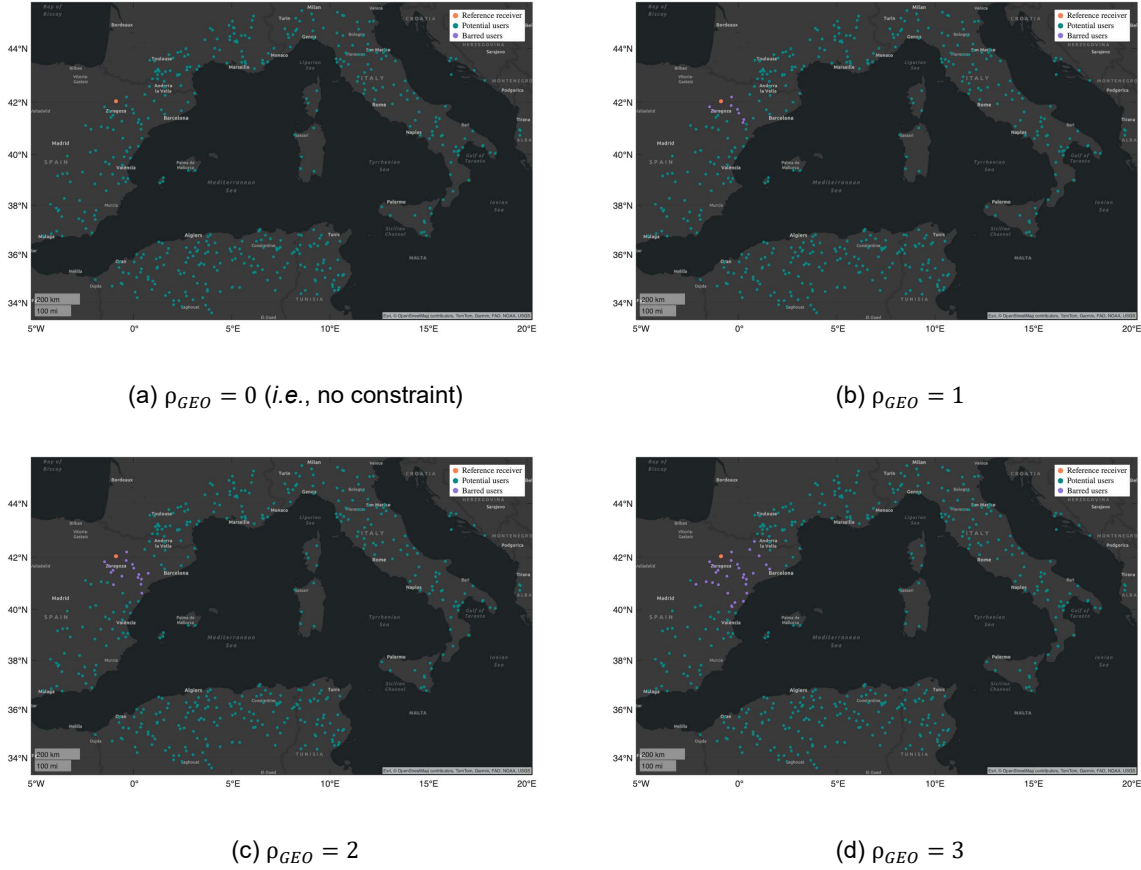


Figure 50: Example of geographical scheduling constraint in eq. (20) with potential users and barred users in the polar constellation (initial NTN node position on its orbit).

The following nomenclature and symbols are used throughout the next sections, if not otherwise stated:

- $\mathbf{v}_{vis}^{(t)}$ denotes a $N_{UE} \times 1$ column vector indicating if the user is visible at time slot t (i.e., there is a valid CSI or location report on-board provided before the t -th time slot and the elevation angle is still above the minimum value for service provisioning) or not. If $v_{vis,i}^{(t)} = 1$, $i = 1, \dots, N_{UE}$, the i -th user can be considered for scheduling in the t -th time slot, otherwise it is not considered visible and $v_{vis,i}^{(t)} = 0$. Notably, this vector is a function of time, as it depends on both the users' and NTN node's locations.
- $\mathbf{a}$ denotes a $N_{UE} \times 1$ column vector containing the number of allocated slots for each user, i.e., a_i , $i = 1, \dots, N_{UE}$, provides the number of slots in which the i -th user was scheduled during the entire satellite pass over the service area.
- $\mathcal{I}(\mathbf{x})$ denotes a function providing the indexes (element position) of the non-null elements in vector $\mathbf{x}$. Thus, $\mathcal{I}(\mathbf{v}_{vis}^{(t)})$ corresponds to a set containing the N_{vis} indexes of the users that are visible because $\mathbf{v}_{vis}^{(t)}$ is either 1 or 0, with $N_{vis}^{(t)} = \sum_{i=1}^{N_{UE}} v_{vis,i}^{(t)}$ being the total number of visible users in time slot t .

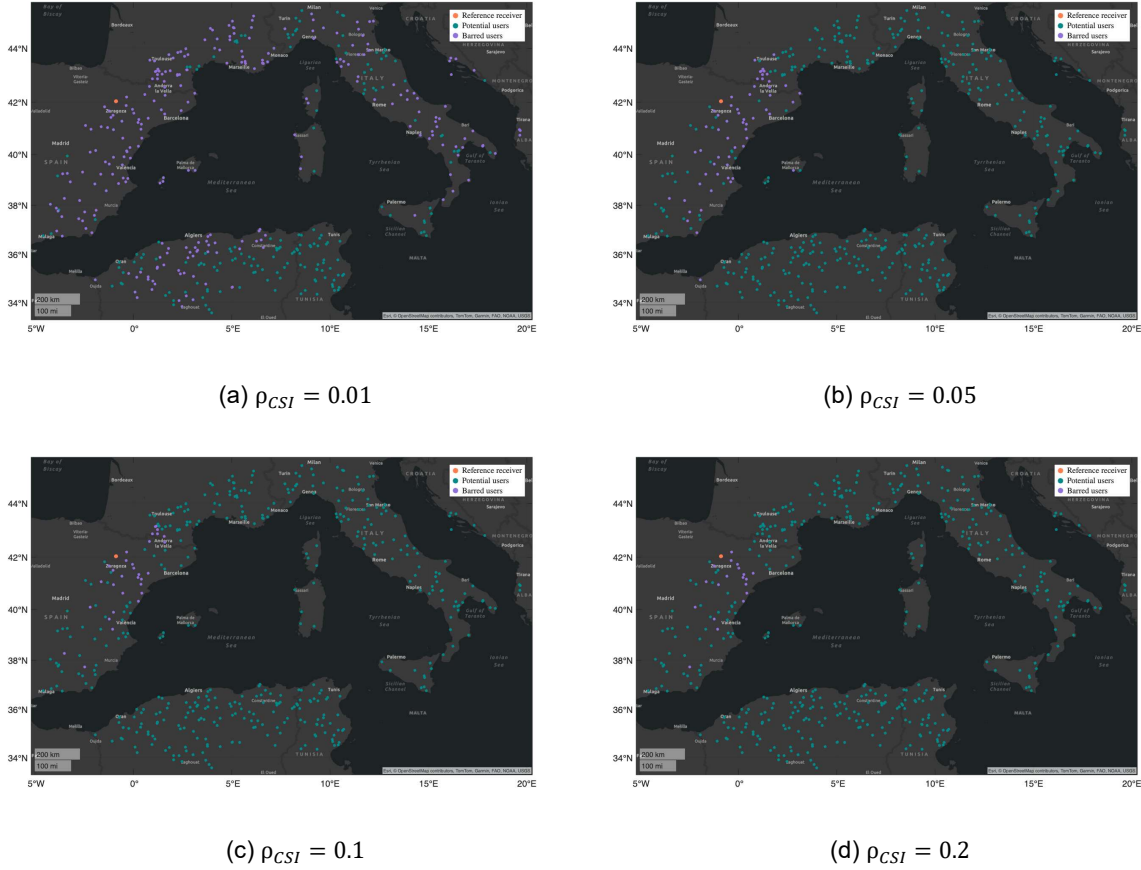


Figure 51: Example of CSI scheduling constraint in eq. (21) with potential users and barred users in the polar constellation (initial NTN node position on its orbit).

- $\mathbf{S}^{(m)}$, $m = 1, \dots, N_{sch}$ denotes the $N_B \times N_{sched}^{(slot)}$ scheduling matrix in the m -th scheduling period. As such, the global scheduling matrix is given by $\mathbf{S} = [\mathbf{S}^{(1)}, \dots, \mathbf{S}^{(m)}, \dots, \mathbf{S}^{(N_{sched})}]$. Assuming $N_{sch}^{(slot)}$ slots per scheduling period, in the m -th scheduling period time slots from $N_m = N_{sched}^{(slot)}(m - 1)$ to $mN_{sched}^{(slot)}$ are defined in $\mathbf{S}^{(m)}$.
- $\text{Card}(\mathbf{A})$ denotes a function providing the cardinality of the generic set or matrix $\mathbf{A}$, which can have any number of dimensions.
- $\omega(\mathbf{A}, n)$ denotes a function randomly (with uniform distribution) selecting n elements from $\mathbf{A}$, which can be a set or a matrix with any number of dimensions.
- $\mathbf{C} = \mathbf{A} \odot \mathbf{B}$ and $\mathbf{C} = \mathbf{A} \oslash \mathbf{B}$ are the Hadamard product and division between matrices $\mathbf{A}$ and $\mathbf{B}$: $c_{i,j} = a_{i,j}b_{i,j}$ and $c_{i,j} = a_{i,j}/b_{i,j}$, $\forall i, j$, respectively. Both matrices can have any number of dimensions and elements per dimension, as long as they are coherent to support element-wise multiplication.

PRIORITY QUEUE SCHEDULER (PQS)

The Priority Queue Scheduler (PQS) is the **benchmark solution** that we design; it is a form of Priority Queueing algorithm, with some slight differences to take into account the

peculiarities of beamforming performance and the geographic or CSI scheduling constraint introduced in the previous section. The high-level diagram representing PQS is provided in Figure 52, in which we can notice that the scheduler: i) **classifies the packets** (or data in general) at its input based on two or more Priority Classes (PCs); ii) **updates the priority queues** based on the class of each packet; iii) **randomly select packets** from the queues, in priority order, until the maximum number of user-centric beams per time slot, N_B , is reached; and iv) the selection of each packet from the scheduling queues is subject to the satisfaction of the geographic or CSI **scheduling constraint**. Please note that, in the following, scheduling a packet or its user are considered as synonymous, since, for the sake of simplicity, each user only has one type of traffic; this does not affect the generality of the solutions or discussions, and the extension to more complex scenarios with multiple packet/traffic types per user can in any case be implemented with minor modifications.

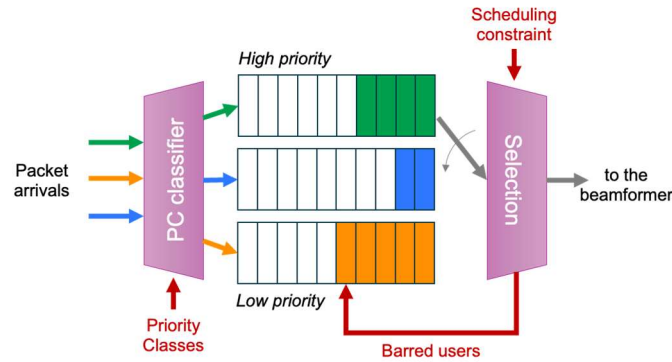


Figure 52: High-level diagram of the Priority Queue Scheduler.

The high-level steps in the PQS algorithm, formalised in Algorithm 1, are discussed hereafter:

- First, the scheduler shall verify that, in the t -th time slot, users requesting data are visible; in fact, it might happen that since the last report sent by a user, due to the satellite movement, visibility has been lost. After the last reporting period, the NTN node is assumed to be able to estimate the visibility period of each of the users that provided a report (this can be obtained from the ephemeris and the ancillary information). As previously stated, we denote by $\mathbf{v}_{vis}^{(t)}$ the visibility vector in the generic time slot t .

In addition to visibility, we have to also consider only users requesting traffic when the scheduling period begins; from eq. (18), this condition corresponds to $C_{k,N_m}^{(U)} > 0$.

Based on these considerations, the condition that the generic k -th user is still visible in time slot t and requesting capacity when the scheduling period starts is $\mathbf{v}_{vis,k}^{(t)} C_{k,N_m}^{(U)} > 0$. If we denote by $\boldsymbol{\xi}^{(U)}$ the $N_{UE} \times 1$ vector containing the unmet capacity values of all users, $\boldsymbol{\xi}^{(U)} = [C_{1,N_m}^{(U)}, \dots, C_{k,N_m}^{(U)}, \dots, C_{N_{UE},N_m}^{(U)}]^T$, we get the set of visible users with unmet capacity during the m -th scheduling period through the $N_{UE} \times 1$ vector $\hat{\mathbf{v}}_{vis}^{(t)} = \boldsymbol{\xi}^{(U)} \odot \mathbf{v}_{vis}^{(t)}$. This vector has non-null entries only for visible users ($\mathbf{v}_{vis,k}^{(t)} \neq 0$) requesting traffic ($C_{k,N_m}^{(U)} > 0$).

Algorithm 1: Priority Queue Scheduler algorithm.

Input

- visibility vector $\mathbf{v}_{vis}^{(t)}$, with $t = 1 + N_{sched}^{(slot)}(m - 1), \dots, mN_{sched}^{(slot)}$
- $N_{UE} \times 1$ allocation vector $\mathbf{a}$
- set of mutually exclusive Priority Class conditions $\{\Omega_q\}_{q=1}^{N_{PC}}$
- current unmet capacity per user $C_{k,N_m}^{(U)}$, $k = 1, \dots, N_{UE}$
- beam radius at nadir r_{uv} and ρ_{GEO} (only for $\Gamma_{GEO,i}^{(t)}$)
- $N_{UE} \times N_R$ estimated channel matrix and ρ_{CSI} (only for $\Gamma_{CSI,i}^{(t)}$)

Assuming that m is the scheduling period to address, $N_m = N_{sched}^{(slot)}(m - 1)$

Output

- $N_B \times N_{sched}^{(slot)}$ scheduling matrix $\mathbf{S}^{(m)}$, $m = 1, \dots, N_{sched}$

for $t = 1 + N_{sched}^{(slot)}(m - 1), \dots, mN_{sched}^{(slot)}$:

1	Build $\xi^{(U)} = [C_{1,N_m}^{(U)}, \dots, C_{k,N_m}^{(U)}, \dots, C_{N_{UE},N_m}^{(U)}]^T$ Compute $\tilde{\mathbf{v}}_{vis}^{(t)} = \xi^{(U)} \odot \mathbf{v}_{vis}^{(t)}$ If $\text{Card}(\tilde{\mathbf{v}}_{vis}^{(t)}) = 0$, then $\tilde{\mathbf{v}}_{vis}^{(t)} = \mathbf{v}_{vis}^{(t)}$	$\xi^{(U)}$ is a vector containing the residual unmet capacity of all users Identify the visible users with non-null residual unmet capacity If there is no visible user with unmet capacity, re-insert all visible users
2	for $q = 1, \dots, N_{PC}$: for $k = 1, \dots, N_{UE}$: $\tilde{v}_{vis,i}^{(t,q)} = 1$ iff $\mathcal{J}(\tilde{\mathbf{v}}_{vis,i}^{(t)}) \in \Omega_q$, 0 otherwise end for end for	Classification into Priority Classes
3	Initialise beam counter: $b = 1$ Initialise Priority Class counter: $q = 1$ Initialise the set of barred users $\delta_{vis}^{(t,q)} = \mathbf{0}$, $\forall q = 1, \dots, N_{PC}$ while ($b \leq N_B$ or $q \leq N_{PC}$): if $\text{Card}(\tilde{\mathbf{v}}_{vis}^{(t,q)}) > 0$: $i = \omega(\mathcal{J}(\tilde{\mathbf{v}}_{vis}^{(t,q)}) \setminus \mathcal{J}(\delta_{vis}^{(t,q)}), 1)$ $\Gamma_{x,i}^{(t)}(\mathbf{s}_{1:b-1,t})$, $x = GEO, CSI$ if $\Gamma_{x,i}^{(t)}(\mathbf{s}_{1:b-1,t}) = 1$: $\mathbf{s}_{b,t} = i$ $b = b + 1$ else: $\delta_{vis,i}^{(t,q)} = 1$ else: $q = q + 1$ end while	Iterative addition of users in the scheduling slot as long as there are user-centric beams or PCs available We exclude from the current PC the users that have already been considered and did not satisfy the selected scheduling constraint If the constraint is met, the user is added to the scheduling slot If the constraint is not met, the user is added to the barred set for this time slot If the current queue is empty, we move to the next PC
4	update the allocation number vector $\mathbf{a}$: $a_k = a_k + 1, \forall k \in \mathbf{s}_{:,t}$	We increment the number of allocations per user based on the scheduled slot
end for		The scheduling matrix for the current scheduling period, $\mathbf{S}^{(m)}$, is complete

The PQS algorithm thus requires the unmet capacity to proceed. This can clearly be obtained from the users, via reporting on the uplink. However, should this not be feasible

(due to overhead, signalling, resources, etc.), alternative solutions can be designed. For instance, it might be possible to compute an estimate of the achieved capacity (and thus the unmet capacity) based on the beamforming matrix (which in any case is computed to transmit the signals to the users) or to exploit AI based predictions (discussed in D4.7).

Finally, in case there is no visible user with residual unmet capacity, $\xi^{(U)} = \mathbf{0}$ and, thus, $\mathbf{v}_{vis}^{(t)} = \mathbf{0}$ by definition: this means that the queues are all empty. To avoid wasting available time slots, shall this scenario arise we re-populate the priority queues requests from all visible users, *i.e.*, $\tilde{\mathbf{v}}_{vis}^{(t)} = \mathbf{v}_{vis}^{(t)}$ and we schedule again users for which the capacity request was already satisfied previously (this leads to an excess capacity as per eq. (19)).

- Classification: the PQS is designed to classify the packets (in the following we directly refer to users⁶) into N_{PC} Priority Classes (PCs), which must be mutually exclusive, *i.e.*, a user cannot belong to more than one PC. Denoting as $\{\Omega_q\}_{q=1}^{N_{PC}}$ the set of all PCs, with Ω_q being the generic q -th class, condition is formalised as:

$$\bigcap_{q=1}^{N_{PC}} \Omega_q = \emptyset \quad (22)$$

This condition implicitly means that a generic user cannot belong to more than one PC. These classes can be clearly defined based on any metric that is deemed relevant for the system performance, as long as mutual exclusivity is satisfied. As discussed in the performance evaluation below, we will consider $N_{PC} = 2$ and the PCs defined based on both the residual unmet capacity and the residual visibility time of each user. However, it shall be noticed that the algorithm and considerations reported below apply to any set of classes meeting the condition in eq. (22). The classification process provides N_{PC} vectors $\tilde{\mathbf{v}}_{vis}^{(t,q)}$, in which the generic element is 1 if and only if the corresponding element in $\tilde{\mathbf{v}}_{vis}^{(t)}$ is 1 (visible user with unmet capacity) and the user belongs to the q -th PC.

- At the output of the classification, we have N_{PC} priority queues providing the users that are visible and requesting traffic. The PQS selects the highest priority queue ($q = 1$) and, if it is not empty, selects a random user from the queue; if the user that has been selected guarantees that the selected scheduling constraint (geographic in eq. (20) or CSI in eq. (21)) is still met considering the users already present in the slot, then it is added to the current time slot, otherwise it is inserted in the barred users set and a new random user is selected. The barred user set is maintained during the definition of the current time slot: when moving to the next one, the barred user set is emptied as the scheduling constraint shall be verified from scratch. Let us formalise these steps:

- Let us denote by $\delta_{vis}^{(t,q)}$ the barred user set for the q -th PC, initialised to the empty set. Thus, $\mathcal{J}(\delta_{vis}^{(t,q)})$ denotes the set of users (their indexes) that are prevented from transmission in this time slot due to a non-satisfied scheduling constraint. The set $\mathcal{J}(\tilde{\mathbf{v}}_{vis}^{(t,q)})$ denotes the visible users (their indexes) that belong to the q -th PC and have unmet residual capacity. In this context, $i = \omega(\mathcal{J}(\tilde{\mathbf{v}}_{vis}^{(t,q)}) \setminus \mathcal{J}(\delta_{vis}^{(t,q)}), 1)$ denotes the selection of a random index from the

⁶ As mentioned above, this is a valid assumption as long as no user changes the type of traffic being requested during a satellite pass. This assumption does not impact the discussions reported in this document.

set of visible users that belong to the q -th PC, have unmet residual capacity, and have not been previously discarded based on the condition in (20) or (21).

- The selected user, i , then shall be tested as well to see if its geographic location or CSI vector meets the geographic or the CSI scheduling constraint, respectively. Thus, if $\Gamma_{x,i}^{(N+t)}(\mathbf{s}_{1:b-1,t}^{(m)}) = 1$, with $x = GEO, CSI$, the user can be added to the time slot, otherwise it is added to the barred user set and a new one is picked and verified.
- Once all beams have been allocated for the current time slot, or if the priority queues have been emptied due to the insertion in the scheduling matrix or the barred set, the number of allocations per user is updated and the scheduler moves to the next time slot. Recalling that matrix $\mathbf{S}^{(m)}$ is the scheduling sub-matrix during the m -th scheduling period, $\mathbf{s}_{:,t}^{(m)}$ provides the indexes of the users that have been scheduled in the t -th time slot of the m -th scheduling period. As such, the update of the allocations per scheduled user is simply given by: $a_k = a_k + 1, \forall k \in \mathbf{s}_{:,t}^{(m)}$.

As a concluding remark, it is worthwhile highlighting that the PQS implements a strict and strong prioritisation of the packets: as long as a higher priority queue has traffic, no packets are considered from lower order queues. To partially smooth this behaviour, a **Weighted PQS (WPQS)** algorithm can be introduced: each priority queue is associated to a weight, representing the maximum fraction of resources (or time slots) it can be allocated over a predefined adjustment interval. This guarantees that, in the worst case, even if all PCs have packets, each of them will receive at least some resources.

PRIORITY RANK SCHEDULER (PRS)

The definition of multiple and mutually exclusive priority classes in PQS, even with a relatively small value of N_{PC} , risks to rapidly become challenging. In fact, to optimise the performance, the definition of the boundaries of the PCs shall depend on the specific conditions of the system, which might clearly fluctuate in terms of users, capacity requests per user, achievable capacity per user, etc., due to not just the users' activities, but also the overhead and channel conditions on the satellite links. Another potential issue is that the unmet capacity vector provides the values of unmet capacity at the beginning of the scheduling period, but it is not updated during the scheduling period; as such, it is possible that some users for which the capacity request was met are still maintained in the priority queues, using resources that could have been dedicated to users still having unmet capacity.

To circumvent these issues, a different approach is proposed: the Priority Rank Scheduler (PRS). The underlying principle, shown in Figure 53, is that of not defining multiple PCs to perform an initial classification, but rather to define a **priority rank**, χ , that allows to directly rank the users from highest to lowest priority. Notably, different metrics can be used while exploiting the PRS algorithm described below; in the following, we compute it as:

$$\chi = \psi_C \tilde{\chi}^{(U)} + \psi_V \tilde{\chi}_{vis} \quad (23)$$

where $\tilde{\chi}^{(U)}$ and $\tilde{\chi}_{vis}$ denote two metrics related to the amount of unmet capacity and residual visibility of the users, respectively, while ψ_C and ψ_V , with $\psi_C + \psi_V = 1$, denote the priority weights. Assuming that the support of both $\tilde{\chi}^{(U)}$ and $\tilde{\chi}_{vis}$ is limited to $[0,1]$, it is straightforward to obtain that $\chi \in [0,1]$. Please note that more complex metrics can be clearly used, even adding other terms to (23), with the only requirement being that the final priority rank χ provides values in $[0,1]$ (this also implicitly means that all weights shall sum to 1). We can now compute

the metrics dedicated to identifying the priority of each user in terms of unmet capacity and residual visibility:

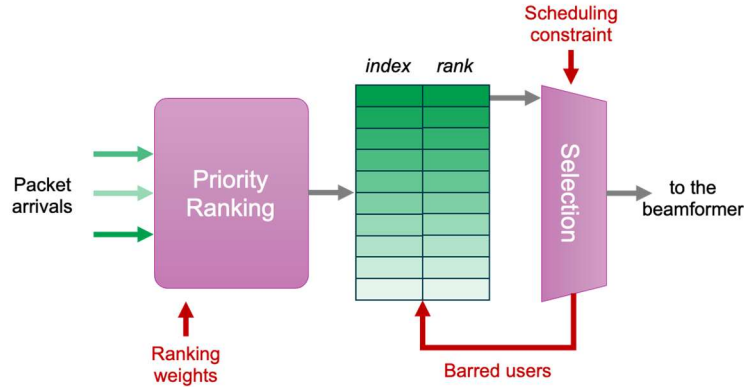


Figure 53: High-level diagram of the Priority Rank Scheduler.

- Let us initially include in $\mathbf{x}^{(U)}$ the unmet capacity of the visible users in the current time slot, i.e., $\chi_k^{(U)} = C_{k,t}^{(U)} \forall k \in \mathcal{J}(\mathbf{v}_{vis}^{(t)})$. The larger this value, the larger is the unmet capacity of the user and, thus, it shall lead to a higher position in the priority rank. To obtain a metric that is bounded to $[0,1]$, we can normalise this vector as:

$$\tilde{\mathbf{x}}^{(U)} = \mathbf{x}^{(U)} \oslash \max_{k \in \mathcal{J}(\mathbf{v}_{vis})} \chi_k^{(U)} \quad (24)$$

where $\oslash$ denotes the Hadamard division, i.e., the element-wise matrix division. When the generic j -th user is the one with the largest unmet capacity, $\tilde{\chi}_j^{(U)} = 1$, while for all other values it will be bounded to $[0,1]$ as the unmet capacity cannot be negative by definition in equation (17).

- With respect to $\mathbf{x}_{vis}$, let us initially include in the generic k -th element the currently residual visibility of that user measured in number of time slots. The larger this value, the lower shall be the impact on the priority rank, since the user will have more opportunities to be served in the future. With this in mind, we normalise this vector as:

$$\tilde{\mathbf{x}}_{vis} = \min_{k \in \mathcal{J}(\mathbf{v}_{vis})} \chi_{vis,k} \oslash \mathbf{x}_{vis} \quad (25)$$

focusing on the j -th user with the least residual visibility time, we can see that the value of $\tilde{\chi}_{vis,j}$ is 1 by definition. For all other users, the residual visibility time will be larger than or equal to that of the j -th user, leading to smaller values of this metric. Thus, $\tilde{\mathbf{x}}_{vis} \in [0,1]$.

Based on the above observations and definitions, the element providing the larger value of the priority metric computed as per eq. (23) is the user that shall be prioritised, taking into account (as in this document) the residual unmet capacity and the residual visibility.

As already highlighted above, other parameters can be included in this computation, with additional weights, as long as all leads to a normalised priority rank. As such, this method can be flexibly adapted to any need that might arise from a scheduling perspective (e.g., adding a factor that takes into account the users' subscriptions, their terminal types, or the number of additional NTN nodes simultaneously visible for the user, just to name a few).

Algorithm 2: Priority Rank Scheduler algorithm.

Input		
- visibility vector $\mathbf{v}_{vis}^{(t)}$, with $t = 1 + N_{sched}^{(slot)}(m-1), \dots, mN_{sched}^{(slot)}$ $N_{UE} \times 1$ allocation number vector $\mathbf{a}$ - capacity and visibility weights ψ_C, ψ_V with $\psi_C + \psi_V = 1$ - unmet capacity per user $C_k^{(U)}$, $k = 1, \dots, N_{UE}$ - beam radius at nadir r_{uv} and ρ_{GEO} (only for $\Gamma_{GEO,i}^{(t)}$) $N_{UE} \times N_R$ estimated channel matrix and ρ_{CSI} (only for $\Gamma_{CSI,i}^{(t)}$)		
Output		
$N_B \times N_{sched}^{(slot)}$ scheduling matrix $\mathbf{S}^{(m)}$, $m = 1, \dots, N_{sched}$		
for $t = 1 + N_{sched}^{(slot)}(m-1), \dots, mN_{sched}^{(slot)}$:		
1	Obtain $\mathbf{x}^{(U)}, \chi_k^{(U)} = C_{k,t}^{(U)} \forall k \in \mathcal{J}(\mathbf{v}_{vis}^{(t)})$	Build a vector containing the unmet capacity of the visible users and compute their residual visibility time
	Obtain $\mathbf{x}_{vis}$, $\forall k \in \mathcal{J}(\mathbf{v}_{vis}^{(t)})$	Normalise $\mathbf{x}^{(U)}$ and $\mathbf{x}_{vis}$ as per eq. (24) and (25).
	Normalise: $\tilde{\mathbf{x}}^{(U)} \leftarrow \mathbf{x}^{(U)} \oslash \max_{k \in \mathcal{J}(\mathbf{v}_{vis}^{(t)})} \chi_k^{(U)}$	
	Normalise: $\tilde{\mathbf{x}}_{vis} \leftarrow \min_{k \in \mathcal{J}(\mathbf{v}_{vis}^{(t)})} \chi_{vis,k} \oslash \mathbf{x}_{vis}$	Compute the priority rank as per eq. (23) and construct the sorted priority table
	Compute $\mathbf{T} = [\mathcal{J}(\mathbf{v}_{vis}^{(t)}), \tilde{\mathbf{x}}]$ and $\mathbf{T} \leftarrow \text{sort}(\mathbf{T}, \chi)$	
2	Initialise beam counter: $b = 1$	Iterative addition of users in the scheduling slot as long as there are beams and users in the priority table
	Initialise $\delta_{vis}^{(t)} = \mathbf{0}$	
	while ($b \leq N_B$ and $\text{Card}(\mathbf{T}) > 0$):	Down-sampling of the priority table based on the selected constraint. Since it was already ordered, the down-sample deletes also the scheduled user and no other sorting is needed.
	$\mathbf{s}_{b,t}^{(m)} = \mathbf{T}_{1,1}$	
	$\delta_{vis}^{(t)} = \{k \mid \Gamma_{x, \mathbf{s}_{b,t}^{(m)}}^{(t)}(\mathbf{T}_{k,1}) = 1, k = 1, \dots, \text{Card}(\mathbf{T})\}$	
	$\mathbf{T} \leftarrow \mathbf{T}_{\delta_{vis}^{(t)}, :}$ $b = b + 1$	
	end while	
3	update the allocation number vector $\mathbf{a}$:	We increment the number of allocations per user based on the scheduled slot
	$a_k = a_k + 1, \forall k \in \mathbf{s}_{:,t}^{(m)}$	
end for		The scheduling matrix for the current scheduling period, $\mathbf{S}^{(m)}$, is complete

The high-level steps in the PRS algorithm are formalised in Algorithm 2 and detailed hereafter:

- Similarly to the PQS, also in this case the first step involves the identification of users that are still in visibility in a set $\mathcal{J}(\mathbf{v}_{vis}^{(t)})$. The vector of residual unmet capacities of these users, $C_{k,t}^{(U)}$, is retrieved to build the unmet capacity vector $\mathbf{x}^{(U)}$. For the identified visible users, the residual visibility time is also computed based on their location and the satellite ephemeris in $\mathbf{x}_{vis}$. These vectors are then normalised according to equations (24) and (25), respectively, and then used to compute the priority rank via eq. (23).

- We can now create the $\text{Card}\left(\mathcal{I}\left(\mathbf{v}_{vis}^{(t)}\right)\right) \times 2$ priority rank table, $\mathbf{T}$, in which the first column contains the set of visible users $\mathcal{I}\left(\mathbf{v}_{vis}^{(t)}\right)$ and the second column the priority rank. This table is then sorted in descending order based on the priority rank; as such, the first row provides the index and priority rank of the user with highest priority (rank).
- The iterative identification of the users to be served can then be initiated. As highlighted above, the first row of the priority rank table $\mathbf{T}$ provides, by definition, the user with highest priority and, thus, we can already schedule it as $\mathbf{s}_{1,t}^{(m)} = \mathbf{T}_{1,1}$.
- To guarantee that we meet the selected scheduling constraint (geographic or CSI), we identify a set of barred users $\delta_{vis}^{(t)}$, i.e., users to be discarded from the list of available ones of the current time slot. This set shall include the users for which the selected constraint requirement computed with respect to the last scheduled user is not met. From a formal point of view, this means that we compute $\Gamma_{x,s_{b,t}}^{(t)}(\mathbf{T}_{k,1})$ for all $k = 1, \dots, \text{Card}(\mathbf{T})$: this represents the scheduling constraint between the last scheduled user, $\mathbf{s}_{b,t}^{(m)}$, and all the potentially available users still in the priority rank table. By identifying these users and removing them from the priority rank table, we are implicitly maintaining in the list of potentially available users for the current time slot only those which would meet the scheduling constraint. This down-sampling of the priority rank table is formalised as $\mathbf{T}_{\delta_{vis}^{(t)}}$, with $\delta_{vis}^{(t)}$ denoting the complement of the set of barred users (those that can still be considered for scheduling). This approach guarantees that no further check is needed in the following iteration, as the priority rank table already includes users that are meeting the required scheduling constraint. The down-sampled table is then sorted.
- After all users have been scheduled in the considered time slot, we need to update the priority ranks. This is done, in the considered implementation of this metric, by:
 - Updating the visibility time by simply reducing all values in χ_{vis} by 1.
 - Updating the residual unmet capacity in $\chi^{(U)}$, if possible, in line with what already discussed for (W)PQR. PRS, differently from (W)PQR, updates the unmet capacity at the beginning of each time slot to be scheduled, not only at the beginning of the scheduling period. This approach has been designed to solve the potential issue related to users with unmet capacity that was satisfied during the definition of the scheduling period. It shall be noticed that this approach can also be implemented in (W)PQR, simply taking the unmet capacity vector at time t at the beginning of each time slot. However, PRS is still a different design as no PCs are introduced to identify the priorities for the scheduler.

The last step, also with PRS, consists in updating the allocation vector $\mathbf{a}$.

4.3.1.3 Computational complexity

In this section, we estimate the computational complexity of the designed PQS and PRS algorithms. Both algorithms, to define a scheduling time slot, include a non-deterministic number of iterations since the condition that one or more users do not meet the geographic or CSI scheduling constraint is not predictable. As such, we estimate the computational complexity from two perspectives: i) the **fixed phase** of the algorithm, which denotes the operations to be performed before and after iterating on the potential users; and ii) the **iteration phase** of the algorithm, in which we select and test users for inclusion in the time slot.

Algorithm 3: PQS with big-o computational complexity.

for $t = 1 + N_{sched}^{(slot)}(m - 1), \dots, mN_{sched}^{(slot)}:$	
$\xi^{(U)} = [c_{1,t}^{(U)}, \dots, c_{k,t}^{(U)}, \dots, c_{N_{UE},t}^{(U)}]^T$	$\mathcal{O}(N_{UE})$
$\tilde{\mathbf{v}}_{vis}^{(t)} = \xi^{(U)} \odot \mathbf{v}_{vis}^{(t)}$	$\mathcal{O}(N_{UE})$
If $\text{Card}(\tilde{\mathbf{v}}_{vis}^{(t)}) = 0$, then $\tilde{\mathbf{v}}_{vis}^{(t)} = \mathbf{v}_{vis}^{(t)}$	$\mathcal{O}(N_{UE})$
for $q = 1, \dots, N_{PC}:$	
for $k = 1, \dots, N_{UE}:$	
$\tilde{\mathbf{v}}_{vis,i}^{(t,q)} = 1$ iff $\mathcal{J}(\tilde{\mathbf{v}}_{vis,i}^{(t)}) \in \Omega_q$, 0 otherwise	$\mathcal{O}(N_{UE}N_{PC})$
end for	
end for	
$b = 1$	$\mathcal{O}(1)$
$q = 1$	$\mathcal{O}(1)$
$\delta_{vis}^{(t,q)} = \mathbf{0}, \forall q = 1, \dots, N_{PC}$	$\mathcal{O}(N_{PC})$
O_1	$\mathcal{O}(N_{UE}N_{PC})$
while $(b \leq N_B \text{ or } q \leq N_{PC}):$	
if $\text{Card}(\tilde{\mathbf{v}}_{vis}^{(t,q)}) > 0:$	$\mathcal{O}(N_{UE})$
$i = \omega(\mathcal{J}(\tilde{\mathbf{v}}_{vis}^{(t,q)}) \setminus \mathcal{J}(\delta_{vis}^{(t,q)}), 1)$	$\mathcal{O}(N_{UE})$
$\Gamma_{x,i}^{(t)}(\mathbf{s}_{1:b-1,t}), x = GEO, CSI$	$\mathcal{O}(N_B)$
if $\Gamma_{x,i}^{(t)}(\mathbf{s}_{1:b-1,t}) = 1:$	$\mathcal{O}(1)$
$\mathbf{s}_{b,t} = i$	$\mathcal{O}(1)$
$b = b + 1$	$\mathcal{O}(1)$
else:	$\mathcal{O}(1)$
$\delta_{vis,i}^{(t,q)} = 1$	
else:	$\mathcal{O}(1)$
$q = q + 1$	
end while	
O_2	$\mathcal{O}(N_{UE})$
$a_k = a_k + 1, \forall k \in \mathbf{s}_{:,t}^{(m)}$	$\mathcal{O}(N_B)$
O_3	$\mathcal{O}(N_B)$
end for	

For the complexity analysis, we consider the **big-o computational complexity**, which provides a measure of the time and space complexity of an algorithm in terms of its upper bound, i.e., a worst-case scenario related to complexity. Algorithm 3 shows the PQS algorithm introduced in Algorithm 1 with the inclusion of the big-o computational complexity of each step. The fixed phase is given by the sum of complexities O_1 and O_2 , while the iteration phase is given by O_3 . More specifically:

- In the fixed phase, we have the following operations:

- copying the unmet capacity vector into $\xi^{(U)}$, with complexity driven by the number of elements in the vector assignment, *i.e.*, $\mathcal{O}(N_{UE})$;
- the computation of $\tilde{\mathbf{v}}_{vis}^{(t)} = \xi^{(U)} \odot \mathbf{v}_{vis}^{(t)}$ requires N_{UE} multiplication with cost $\mathcal{O}(1)$ and an assignment of N_{UE} elements: $N_{UE}\mathcal{O}(1) + \mathcal{O}(N_{UE}) = \mathcal{O}(N_{UE})$;
- the check on the cardinality of a set or vector requires, typically, a complexity again driven by the number of elements, *i.e.*, $\mathcal{O}(N_{UE})$. Depending on the result, it is possible that we need to perform the assignment $\tilde{\mathbf{v}}_{vis}^{(t)} = \mathbf{v}_{vis}^{(t)}$, with cost $\mathcal{O}(N_{UE})$. In conclusion, we have a complexity $\mathcal{O}(N_{UE}) + \mathcal{O}(N_{UE}) = \mathcal{O}(N_{UE})$;
- the assignment of users to priority classes, which has a cost of $\mathcal{O}(1)$ per comparison, leading to a total of $N_{UE}N_{PC}\mathcal{O}(1) = \mathcal{O}(N_{UE}N_{PC})$;
- the initialisation of the indexes, each with complexity $\mathcal{O}(1)$;
- the initialisation of the barred set with cost $\mathcal{O}(N_{PC})$.

This leads to a first term in the fixed phase of the complexity equal to:

$$O_1 = \mathcal{O}(N_{UE}) + \mathcal{O}(N_{UE}) + \mathcal{O}(N_{UE}) + \mathcal{O}(N_{UE}N_{PC}) + 2\mathcal{O}(1) + \mathcal{O}(N_{PC}) \sim \mathcal{O}(N_{UE}N_{PC}) \quad (26)$$

In the above equation, we considered that typically $N_{UE} \gg N_{PC}$ and that, in big-o notation, a function that can be written as a sum of functions has a complexity determined by the fastest growing one. In the fixed terms, we also have to consider the update of the allocation vector, which has a cost of $\mathcal{O}(N_B)$ as each scheduling vector has N_B elements. Thus, the total complexity for the fixed phase of the PQS algorithm is:

$$\mathcal{O}_{fixed}^{(PQS)} = O_1 + O_3 = \mathcal{O}(N_{UE}N_{PC}) + \mathcal{O}(N_B) \sim \mathcal{O}(N_{UE}N_{PC}) \quad (27)$$

since $N_{UE} \gg N_B$.

- In the iteration phase, we have several operations as reported hereafter:
 - the check on the cardinality can again be considered as $\mathcal{O}(N_{UE})$;
 - the set difference between to identify the users that are not barred requires, in the worst-case with all users in the visible set, $\mathcal{O}(N_{UE})$;
 - the computation of the geographical or CSI scheduling constraint can be considered with a complexity of $\mathcal{O}(N_B - 1) = \mathcal{O}(N_B)$ in the worst case (when we are selecting the very last user in the slot);
 - a series of assignments of single elements, all with a cost of $\mathcal{O}(1)$.

In conclusion, a single iteration of the PQS algorithm requires:

$$\mathcal{O}_{iter}^{(PQS)} = O_2 = 5\mathcal{O}(1) + 2\mathcal{O}(N_{UE}) + \mathcal{O}(N_B) \sim \mathcal{O}(N_{UE}) \quad (28)$$

where we considered again that, typically, $N_{UE} \gg N_B$. As mentioned above, it is possible that this iteration step is computed once or multiple times depending on the result of the scheduling constraint check. Then, also this step is repeated N_{sche} times.

Algorithm 4: PRS with big-o computational complexity.

for $t = 1 + N_{sched}^{(slot)}(m - 1), \dots, mN_{sched}^{(slot)}:$	
$\mathbf{x}^{(U)}, \chi_k^{(U)} = \mathbf{c}_{k,t}^{(U)} \forall k \in \mathcal{J}(\mathbf{v}_{vis}^{(t)})$	$\mathcal{O}(N_{UE})$
$\mathbf{x}_{vis}, \forall k \in \mathcal{J}(\mathbf{v}_{vis}^{(t)})$	$\mathcal{O}(N_{UE})$
$\tilde{\mathbf{x}}^{(U)} \leftarrow \mathbf{x}^{(U)} \oslash \max_{k \in \mathcal{J}(\mathbf{v}_{vis}^{(t)})} \chi_k^{(U)}$	$\mathcal{O}(N_{UE}) + \mathcal{O}(1)N_{UE} = \mathcal{O}(N_{UE})$
$\tilde{\mathbf{x}}_{vis} \leftarrow \min_{k \in \mathcal{J}(\mathbf{v}_{vis}^{(t)})} \chi_{vis,k} \oslash \mathbf{x}_{vis}$	$\mathcal{O}(N_{UE}) + \mathcal{O}(1)N_{UE} = \mathcal{O}(N_{UE})$
$\mathbf{T} = [\mathcal{J}(\mathbf{v}_{vis}^{(t)}), \mathbf{x}]$	$\mathcal{O}(N_{UE})$
$\mathbf{T} \leftarrow \text{sort}(\mathbf{T}, \mathbf{x})$	$\mathcal{O}(N_{UE} \log_2 N_{UE})$
$b = 1$	$\mathcal{O}(1)$
$\delta_{vis}^{(t)} = \mathbf{0}$	$\mathcal{O}(1)$
O_1	$\mathcal{O}(N_{UE} \log_2 N_{UE})$
while ($b \leq N_B$ and $\text{Card}(\mathbf{T}) > 0$):	
$\mathbf{s}_{b,t}^{(m)} = \mathbf{T}_{1,1}$	$\mathcal{O}(1)$
$\delta_{vis}^{(t)} = \{k \mid \Gamma_{x, \mathbf{s}_{b,t}^{(m)}}^{(t)}(\mathbf{T}_{k,1}) = 1, k = 1, \dots, \text{Card}(\mathbf{T})\}$	$\mathcal{O}(N_{UE})$
$\mathbf{T} \leftarrow \mathbf{T}_{\delta_{vis}^{(t)}, \cdot}$	$\mathcal{O}(N_{UE})$
$b = b + 1$	$\mathcal{O}(1)$
end while	
O_2	$\mathcal{O}(N_{UE})$
$a_k = a_k + 1, \forall k \in \mathbf{s}_{\cdot, t}^{(m)}$	$\mathcal{O}(N_B)$
O_3	$\mathcal{O}(N_B)$
end for	

For the PRS scheduler, the complexities reported in Algorithm 4 are discussed hereafter:

- In the fixed phase, we have the following operations:
 - the creation of vectors $\mathbf{x}^{(U)}$ and $\mathbf{x}_{vis}$, which, assuming all N_{UE} are visible, can be associated to a big-o complexity of $\mathcal{O}(N_{UE})$ each;
 - the computation of the maximum of $\mathbf{x}^{(U)}$ and the normalisation, with a total complexity equal to $\mathcal{O}(N_{UE}) + \mathcal{O}(1)N_{UE} = \mathcal{O}(N_{UE}) + \mathcal{O}(N_{UE}) \sim \mathcal{O}(N_{UE})$;
 - the computation of the minimum of $\mathbf{x}_{vis}$ and the normalisation, with a total complexity equal to $\mathcal{O}(N_{UE}) + \mathcal{O}(1)N_{UE} = \mathcal{O}(N_{UE}) + \mathcal{O}(N_{UE}) \sim \mathcal{O}(N_{UE})$;
 - the assignment to the priority rank table, costing $2\mathcal{O}(N_{UE}) \sim \mathcal{O}(N_{UE})$;
 - the sorting of the priority rank table, with a complexity of $\mathcal{O}(N_{UE} \log_2 N_{UE})$;
 - the initialisation of the beam counter and the barred users, both with cost $\mathcal{O}(1)$.

Thus, the total complexity of the first step in the fixed phase for PRS is given by:

$$O_1 = 5O(N_{UE}) + 2O(1) + O(N_{UE} \log_2 N_{UE}) \sim O(N_{UE} \log_2 N_{UE}) \quad (29)$$

After the iterations, we only have the update of the allocation vector (the table will be updated again in the next iteration):

$$O_3 = O(N_B) \quad (30)$$

since $N_{UE} \gg N_B$. Thus, the total cost of the fixed phase of PRS is given by:

$$O_{fixed}^{(PRS)} = O_1 + O_3 = O(N_{UE} \log_2 N_{UE}) + O(N_B) \sim O(N_{UE} \log_2 N_{UE}) \quad (31)$$

- In the iteration step, we have to consider:
 - the assignment of the first user in the priority rank table to the scheduling slot and the beam counter update, which both have a cost of $O(1)$;
 - the computation of the geographical or CSI scheduling constraint, which now has to be computed for all rows in the table. In the worst case, this leads to $O(N_{UE})$;
 - down-sampling the table based on the scheduling constraints, with $O(N_{UE})$.

In conclusion:

$$O_{iter}^{(PQS)} = O(1) + 2O(N_{UE}) \sim O(N_{UE}) \quad (32)$$

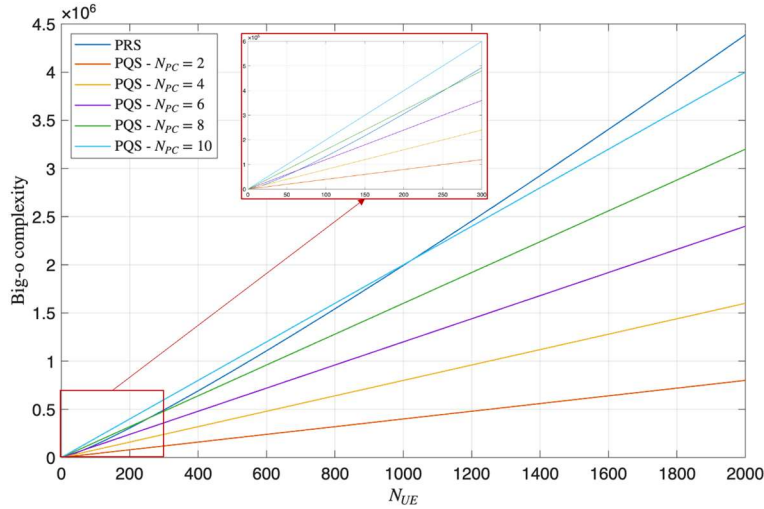


Figure 54: Big-o computational complexity of the fixed phase in PRS and PQS with $N_{PC} = 2, 4, 6, 8, 10$ and $N_{sched} = 200$.

Based on the above analysis, PQS and PRS have the same big-o complexity in each iteration of the scheduler, which scales with the number of users. Again, please note that this is the cost of a single iteration and the actual number of iterations cannot be predicted; in addition, the iteration phase is also repeated N_{sche} times as the fixed phase, clearly. When looking at the preparation phase, their complexities are different: $O_{fixed}^{(PQS)} = O(N_{sche} N_{UE} N_{PC})$ and $O_{fixed}^{(PRS)} = O(N_{sche} N_{UE} \log_2 N_{UE})$. Clearly, the PRS algorithm does not have priority classes, so this factor shall not be considered in the complexity; however, the construction and sorting of the table

poses an additional computational burden as a function of the number of UEs. Figure 54 shows the big-o computational complexity assuming $N_{PC} = 2, 4, 6, 8, 10$ and $N_{sche} = 200$ (please note that N_{sched} , being present in both complexities, is not impacting the generality of the analysis). It can be noticed that as long as the number of priority classes is limited, e.g., 8 or below, the complexity of the PRS is basically always much larger due to the table operations. However, when the PCs are increased, the PRS approach starts to also have an advantage in terms of complexity (up to 1000 users). In general, this increase in the complexity for PRS is the cost for having a better classification of the user (we basically drop the granularity related to discrete classes and directly rank the users in a more general and finer way) and, thus, performance.

It can be observed from (27) and (31) that PRS becomes more efficient of PQS in the fixed phase when $N_{PC} > \log_2 N_{UE}$, or equivalently $2^{N_{PC}} > N_{UE}$. Referring to Figure 54, these thresholds values for $N_{PC} = 2, 4, 6, 8, 10$ are equal to 4, 16, 64, 256, 1024.

Assuming that the iterative phase of both algorithms is repeated, on average, N_{iter} times per scheduling slot, we finally obtain the following big-o complexities:

$$\begin{aligned} \mathcal{O}_{PQS} &= N_{sched} N_{sc}^{(slot)} \left(\mathcal{O}_{fixed}^{(PQS)} + N_{iter} \mathcal{O}_{iter}^{(PQS)} \right) \\ &= \mathcal{O} \left(N_{sched} N_{sche}^{(slot)} N_{UE} N_{PC} \right) + \mathcal{O} \left(N_{sched} N_{sche}^{(slot)} N_{iter} N_{UE} \right) \\ &= \mathcal{O} \left(N_{sched} N_{sche}^{(slot)} N_{iter} N_{UE} \right) \end{aligned} \quad (33)$$

$$\begin{aligned} \mathcal{O}_{PRS} &= N_{sched} N_{sc}^{(slot)} \left(\mathcal{O}_{fixed}^{(PRS)} + N_{iter} \mathcal{O}_{iter}^{(PRS)} \right) \\ &= \mathcal{O} \left(N_{sched} N_{sche}^{(slot)} N_{UE} \log_2 N_{UE} \right) + \mathcal{O} \left(N_{sche} N_{sche}^{(slot)} N_{iter} N_{UE} \right) \\ &= \mathcal{O} \left(N_{sche} N_{sche}^{(slot)} N_{iter} N_{UE} \right) \end{aligned} \quad (34)$$

where we observed that $N_{UE} \gg N_{PC}$ and, typically, $N_{iter} \gg \log_2 N_{UE}$. Thus, both PQS and PRS have the same big-o computational complexity. When observing only the cost of the fixed phase, PRS is more demanding as discussed above; however, when taking into account the entire algorithm, in both schedulers the dominant term is that corresponding to the iterative phase, with the same complexity as demonstrated above.

4.3.1.4 Performance assessment with ideal ancillary information

In this section, we evaluate the performance of PQS and PRS with ideal ancillary information at the transmitter. The simulation parameters common to both schedulers are in Table 16.

PQS PERFORMANCE

Table 17 provides the parameters defining the PQS scheduler configuration. In particular, we consider two priority classes, i.e., $N_{PC} = 2$. With respect to the choice of the class boundaries, we can observe that: i) we are aiming at providing as much requested capacity as possible, i.e., limiting the unmet capacity; and ii) we are considering a LEO NTN node passing over the service area in approximately 6 minutes, meaning that each user will have a limited visibility period in which it can be served from this node. Taking these two aspects into consideration, we identify the two PCs based on the residual unmet capacity and the residual visibility time.

With respect to the residual unmet capacity, $\tilde{C}_k^{(U)}$ provides its current value for the k -th visible user and we can denote by $\tilde{C}^{(U)} = \max_k \tilde{C}_k^{(U)}$ its current maximum value. We state that a user shall be prioritised if its currently unmet capacity is larger than the maximum value scaled up by a predefined parameter σ_{cap} :

$$\tilde{C}_k^{(U)} \geq \sigma_{cap} \tilde{C}^{(U)} \Rightarrow k \in \Omega_1 \quad (35)$$

As for the residual visibility time, we can denote by σ_k the number of residual visibility slots the k -th user and by σ_{vis} a maximum predefined residual visibility time, in number of slots. A user shall then be prioritised also if its residual visibility time is below the predefined maximum value:

$$\sigma_k \leq \sigma_{vis} \Rightarrow k \in \Omega_1 \quad (36)$$

Thus, the priority classes in PQS are defined by merging the conditions in (35) and (36):

$$k \in \begin{cases} \Omega_1, & \text{if } \sigma_k \leq \sigma_{vis} \text{ or } \tilde{C}_k^{(U)} \geq \sigma_{cap} \tilde{C}^{(U)} \\ \Omega_1, & \text{otherwise} \end{cases} \quad (37)$$

Table 16: Simulation parameters for the standalone case with ideal ancillary information.

Parameter	Value	Comments
Sub-Carrier Spacing (SCS)	120 kHz	In this case, 132 PRBs are available
Operating Frequency Range	FR2 (Ka-band)	See 3GPP TR 38.821
Carrier frequency	20 GHz	See 3GPP TR 38.821
User bandwidth B	190.08 MHz	$132 \cdot 120 \cdot 12$ kHz
Scenario	Digital Divide	Parameters in Table 14 of Section 4.2.1.2
Minimum capacity request, C_{min}	5,10,20 Mbps	See Section 4.2.1.2
Maximum capacity request, C_{max}	100,300,500 Mbps	See Section 4.2.1.2
Number of radiating elements $N_{R,sat}$	512	See Section 2.3.1 of [1]
Power per radiating element $P_{av,el}$	65 mW	See Section 2.3.1 of [1]
Total power on-board P_{av}	15.2218 dBW	$10 \log_{10}(0.0065 \cdot 512)$
Geographic constraint coefficient ρ_{GEO}	1, ..., 6	See Section 4.3.1.2
CSI constraint coefficient ρ_{CSI}	0.01,0.02, ..., 0.1	See Section 4.3.1.2
Minimum user elevation angle ε_t	30°	See Section 4.2.3 of [1]
Service area elevation angle ε_s	30°	See Section 4.2.3 of [1]
Channel type	clear-sky	See Section 2.4.2.1 of [1] and 3GPP TR 38.811
Reporting periodicity T_{rep}	80 ms (8 slots)	See Section 4.3.1.1
Transmission time slot T_{slot}	10 ms	See Section 4.3.1.1
NTN node altitude	1000 km	See Section 2.1 of [1]
Service area centre	(45°N, 5°E)	See Section 4.2.3 of [1]
UE terminal	VSAT	See Section 2.3.2 of [1] and 3GPP TR 38.811

Table 17: Configuration parameters for the PQS algorithm.

Parameter	Value	Comments
Scheduling periodicity T_{sched}	2 s (200 slots)	See Section 4.3.1.1
Maximum residual visibility σ_{vis}	10,50,150	With $T_{slot} = 10$ ms, they correspond to 0.1, 0.5, 1.5 s
Unmet capacity factor σ_{cap}	1.2, 1.5, 2	
Number of Priority Classes N_{PC}	2	See Section 4.3.1.2

Before analysing the performance of PQS, it shall be mentioned that being the focus on clear-sky channels, the performance of MMSE (CSI-based) and LB-MMSE (location-based) is the same. In fact, as discussed in [1] and in Section 4.3.1.1, location-based algorithms estimate the non-stochastic terms in the channel vectors based on the users' location. Being the channel clear-sky, there is no stochastic term and, thus, the performance of the two solutions is identical as long as ideal ancillary information is assumed, as in this initial evaluation.

Let us first consider a specific scenario, to more easily understand the metrics and the system trends. Figure 55 shows the achieved and unmet capacities computed as per equation (15) and (17), respectively, considering the PQS configured with $(\sigma_{vis}, \sigma_{cap}) = (10, 2)$ and the traffic scenario with $(C_{min}, C_{max}) = (5 \text{ Mbps}, 100 \text{ Mbps})$. It can be noticed that, focusing exclusively on the achieved capacity, the optimal value of the geographic scheduling would be $\rho_{GEO} = 2$; however, the objective of the scheduler shall be, first, to meet as much as possible the capacity request of all users. Observing the behaviour of the unmet capacity, we can see that values of ρ_{GEO} larger than 2 provide a null unmet capacity. As such, accepting a slight reduction in the total achieved capacity, we can consider the optimum to be $\rho_{GEO} = 3$, providing $C^{(A)} = 20.8 \text{ Gbps}$ and $C^{(U)} = 0 \text{ Gbps}$, *i.e.*, all the users' requests have been satisfied. It is worthwhile highlighting that 20.8 Gbps is the sum achieved capacity of a single NTN node in the entire polar constellation, provided in roughly 6 minutes over a service area covering almost entirely the European soil. It shall also be mentioned that, in this scenario, if a very limited value of unmet capacity is acceptable, a higher achieved capacity can be obtained with $\rho_{GEO} = 2$.

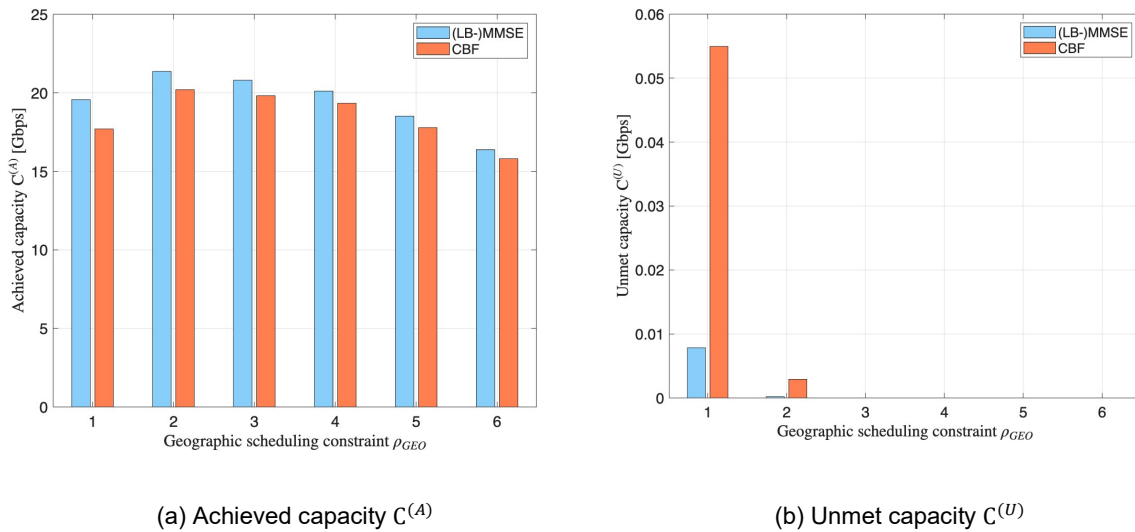


Figure 55: Achieved and unmet capacity with PQS and geographical scheduling constraint, standalone NTN node. Setup: $C_{min} = 5 \text{ Mbps}$, $C_{max} = 100 \text{ Mbps}$, $\sigma_{vis} = 10$, $\sigma_{cap} = 2$.

It is interesting to better observe the impact of the geographic scheduling constraint ρ_{GEO} . Figure 56 shows the CDF of the asymptotic capacity (*i.e.*, the capacity achieved by each user when assuming a channel used for the entire service time) and the number of allocations per user as a function of the geographic scheduling constraint ρ_{GEO} . We focus on these two metrics since these are the parameters leading to the achieved capacity per user as per eq. (15). We can observe that, for both (LB-)MMSE and CBF algorithms, increasing ρ_{GEO} leads to a larger asymptotic achieved capacity per user (left figure) and a reduction in the number of allocations per user (right figure). These trends arise for the following reasons:

- In general, as also observed in [1] and in Annex A to this document, when we increase the distance among the users in the same time slot (increasing ρ_{GEO}), we are usually reducing the correlation among their CSI vectors, thus obtaining a better SINR per user.
- The increase in the geographical scheduling constraint also leads to a reduction in the number of allocations per user. This is an aspect that needs further clarifications. In fact, based on the PQS algorithm (but a similar concept applies to PRS, as we will see), when the scheduler cannot find any more users meeting the geographical (or CSI) scheduling constraint in any priority class, it finalises the time slot, *i.e.*, it does not force the inclusion of other users, leading to less than N_B simultaneous beams. The amount of unused potential beams per time slot tends to increase with ρ_{GEO} , as we are making the scheduling constraint more and more challenging to be met. This effect does not just reduce the number of allocations, but it also increases the SINR (and thus the capacity) since more power per user can be used, on average.

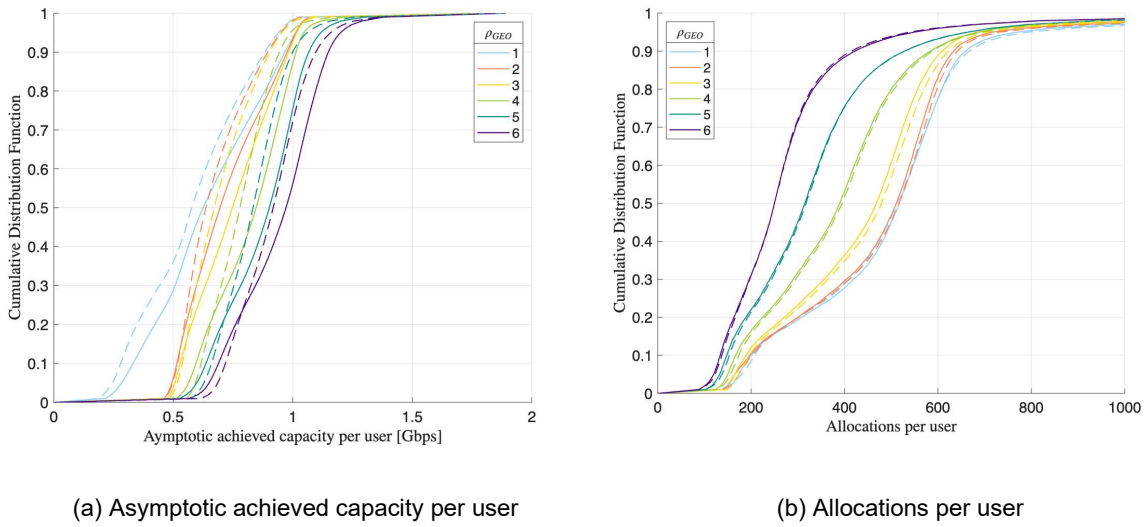


Figure 56: Asymptotic achieved capacity and allocations per user with PQS and geographical scheduling constraint, standalone NTN node. Setup: $C_{min} = 5$ Mbps, $C_{max} = 100$ Mbps, $\sigma_{vis} = 10$, $\sigma_{cap} = 2$. (LB-)MMSE (solid line) and CBF (dashed line).

Figure 57 shows the average percentage of unused beams (assuming a maximum of $N_B = 24$ in each time slot) as a function of the geographical scheduling constraint. It can be noticed that, increasing ρ_{GEO} , we are indeed protecting more the users in each time slot, with a two-fold effect: the channels are less correlated, leading to a larger asymptotic capacity, and we transmit more power per user since we have less beams per slot. Clearly, this is also coming at the expense of less allocations per user, as shown on the right of Figure 56. To have an indication of the effect of these percentages, since the NTN node is serving the users with at most 24 beams, every 10% shown in Figure 57 corresponds to approximately 2 beams that are unused. Thus, having a too strict constraint leads to a significant portion of potential user-centric beams that are not used (almost half of them with $\rho_{GEO} = 6$) with a clear detrimental impact on the performance despite the implicit increase in the SINR per user. Ultimately, the trade-off among all these aspects leads to the optimal performance obtained with $\rho_{GEO} = 3$. Please note that (LB-)MMSE solutions tend to have a higher percentage of unused beams because, by providing a better performance in the achieved capacity, they empty the priority queues slightly in advance compared to CBF.

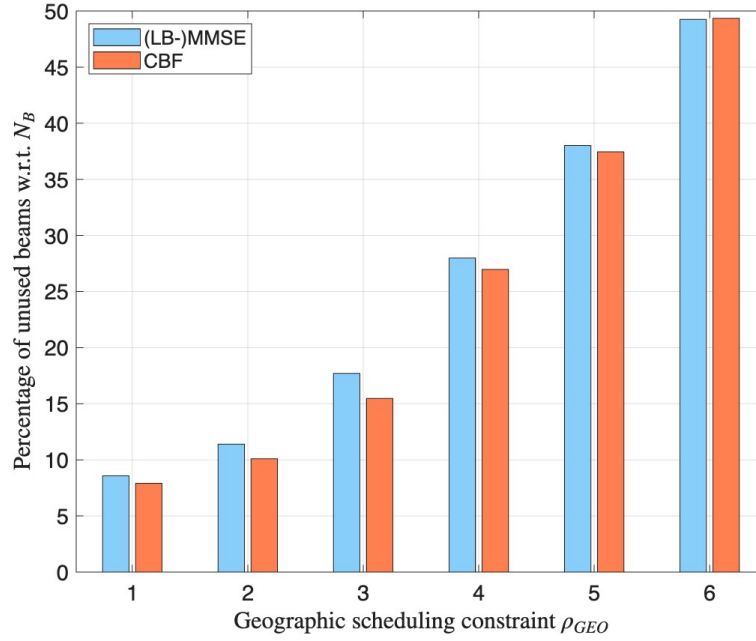


Figure 57: Percentage of unused beams with respect to the maximum N_B with PQS and geographical scheduling constraint, standalone NTN node. Setup: $C_{min} = 5$ Mbps, $C_{max} = 100$ Mbps, $\sigma_{vis} = 10$, $\sigma_{cap} = 2$. (LB-)MMSE (solid line) and CBF (dashed line).

A similar assessment can be made considering the CSI scheduling constraint. Figure 58 shows the achieved and unmet capacities computed as per equation (15) and (17), respectively, considering the PQS configured with $(\sigma_{vis}, \sigma_{cap}) = (10, 2)$ and the traffic scenario with $(C_{min}, C_{max}) = (5 \text{ Mbps}, 100 \text{ Mbps})$; in this case we are considering the CSI constraint in the PQS algorithm, defined by ρ_{CSI} . Before discussing the performance, we shall highlight the difference between the considered scheduling constraints:

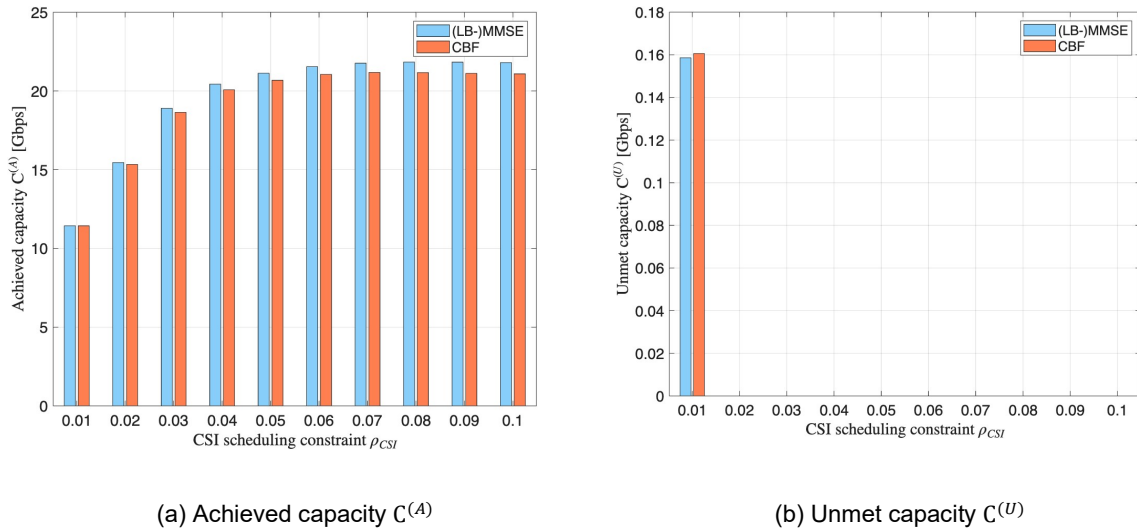


Figure 58: Achieved and unmet capacity with PQS and CSI scheduling constraint, standalone NTN node. Setup: $C_{min} = 5$ Mbps, $C_{max} = 100$ Mbps, $\sigma_{vis} = 10$, $\sigma_{cap} = 2$.

- Increasing ρ_{GEO} leads to a larger distance among the users scheduled in the same time slot, i.e., to increasing the difference in the channels of the users scheduled together.

- ρ_{CSI} is a threshold value in the Coefficient of Correlation: if the users exhibit a channel correlation below this threshold, they can be scheduled together because they are less correlated. Thus, increasing this value means that we become less stringent in the acceptance of users in the same time slot and users with increasingly similar channels can be scheduled together (this is the opposite behaviour compared to ρ_{GEO}).

The optimal value in this scenario is obtained with $\rho_{CSI} = 0.08$. In fact, focusing on the achieved capacity, this configuration provides a slightly better performance compared to other constraint values. Ultimately, we obtain $C^{(A)} = 21.84$ Gbps and $C^{(U)} = 0$ Gbps, approximately 1.04 Gbps more than with the geographical constraint. This is in line with the extensive discussions in Section 4.3.1.2: users located farther away might still show correlation in their channel vectors, so exploiting this metric leads to a more accurate identification of similar users.

While the performance trend seems similar to that obtained with the geographical constraint, due to the different meaning of ρ_{CSI} we have a different underlying system behaviour. Figure 59 and Figure 60 show the asymptotic capacity and the number of allocations, and the percentage of unused beams, respectively, in the CSI scheduling constraint case. Since increasing ρ_{CSI} is equivalent to less stringent scheduling conditions, we have the opposite trends compared to the geographical case: larger ρ_{CSI} leads to more allocations per user, and a reduced asymptotic capacity due to interference and lower SINR values, while the number of unused beams is decreased. Ultimately, the optimal performance with $\rho_{CSI} = 0.08$ leads to a similar situation in terms of unused beams compared to the geographical case.

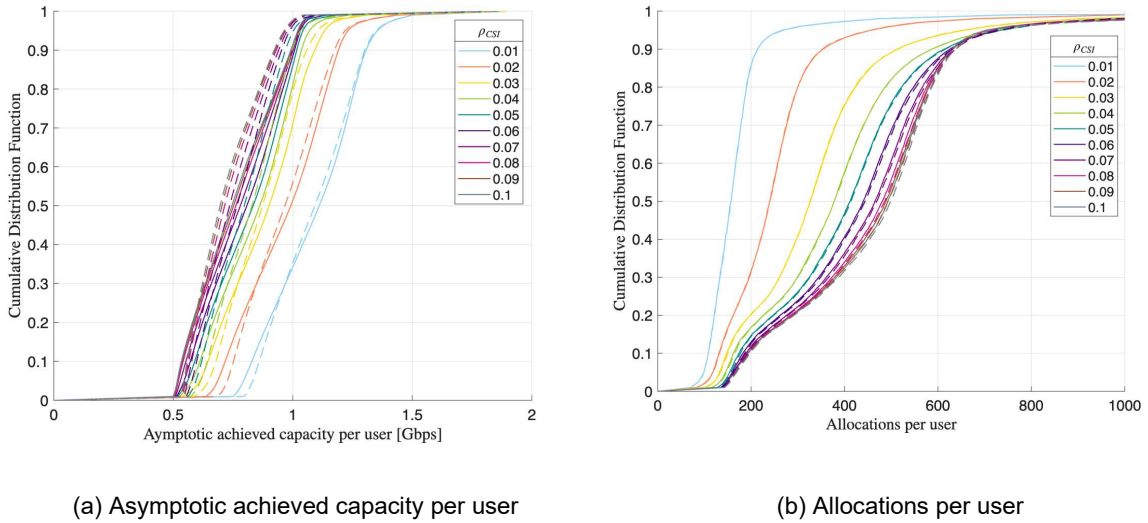


Figure 59: Asymptotic achieved capacity and allocations per user with PQS and CSI scheduling constraint, standalone NTN node. Setup: $C_{min} = 5$ Mbps, $C_{max} = 100$ Mbps, $\sigma_{vis} = 10$, $\sigma_{cap} = 2$. (LB-)MMSE (solid line) and CBF (dashed line).

According to the simulation parameters summarised in Table 16, 9 traffic configurations (combinations of minimum and maximum capacity) with 10 (CSI) or 6 (geographical) constraints have been evaluated, leading to 90 scenarios for CSI scheduling constraints and 54 scenarios in the geographical constraint case. For each generic scenario configuration, from Table 17 we shall consider 9 scheduler setups in terms of σ_{vis} and σ_{cap} ; this ultimately leads to 810 and 486 parameter combinations for CSI and location constraints, respectively. In this context, since we already discussed the major trends as a function of the scheduling constraints (which hold in all configurations), in the following we focus the discussion on the

performance obtained with the optimal geographic or CSI scheduling constraint under different traffic and scheduling configurations, to understand the system trends also related to them.

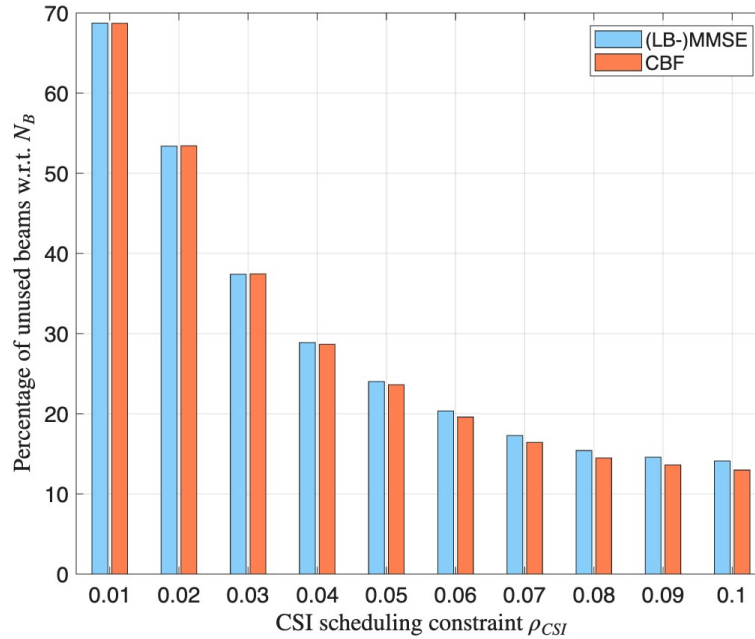


Figure 60: Percentage of unused beams with respect to the maximum N_B with PQS and CSI scheduling constraint, standalone NTN node. Setup: $C_{min} = 5$ Mbps, $C_{max} = 100$ Mbps, $\sigma_{vis} = 10$, $\sigma_{cap} = 2$. (LB-)MMSE (solid line) and CBF (dashed line).

Table 18 shows the optimal value of the geographical scheduling constraint ρ_{GEO} obtained with (LB-)MMSE beamforming in the considered traffic scenarios with PQS configured as (10,2). Figure 61 provides the sum achieved and unmet capacity values and the following observations can be made:

- The optimal value of the scheduling constraint is basically fixed at 4, with the only exception being the scenario with the lowest traffic request that we discussed above. In low traffic scenarios (such as the 5-100 Mbps case), users are in visibility of the NTN node for a sufficient period to allow the scheduler to empty the priority queues. As such, the scheduling constraint can be a little less strict since, reducing the number of users per time slot, the SINR is increased (more power per user and reduced interference in each time slot) while guaranteeing enough allocations per user. This is in line with the previous discussion on the unused beams in Figure 57.

Table 18: Optimal value of the geographical scheduling constraint with (LB-)MMSE beamforming in the standalone scenario. Setup: $\sigma_{vis} = 10$, $\sigma_{cap} = 2$.

Minimum capacity request C_{min}	Maximum capacity request C_{max}		
	100 Mbps	300 Mbps	500 Mbps
5 Mbps	3	4	4
10 Mbps	4	4	4
20 Mbps	4	4	4

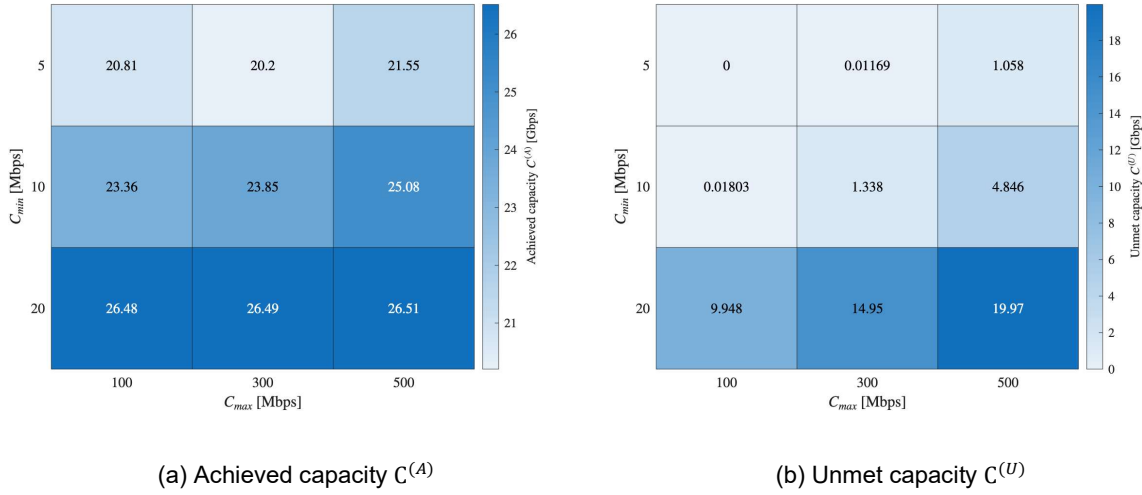


Figure 61: Achieved and unmet capacity with geographical scheduling constraint with (LB-)MMSE beamforming in the standalone scenario. Setup: $\sigma_{vis} = 10$, $\sigma_{cap} = 2$.

- When the traffic requests increase, up to the (very challenging) 20-500 Mbps scenario, the optimal PQS scheduler is the one with a configuration that increases both the average SINR (so the achieved capacity) per time slot and the number of allocations per user.

The former behaviour corresponds to the increase in ρ_{GEO} : the scheduler needs to separate more the users to increase the achieved capacity, but the separation cannot be too large otherwise the allocations per user drop and the unmet capacity increases, thus degrading the performance.

This separation becomes even more relevant when focusing on the second observation: the increase in the allocations per user. Before we highlighted that PQS tends to leave a certain percentage of beams unused in low traffic scenarios: in this case, the capacity request can relatively easily be satisfied (the priority queues are emptied soon) and the best solution is to define less occupied time slots. However, when the capacity request per user is large, the scheduler needs to allocate more users per time slot, in particular as long as there are users satisfying the scheduling constraint in the priority queues. Due to the large requests, the users are not in the visibility of the NTN node for enough time to guarantee a sufficient number of allocations per user and simultaneously satisfy the scheduling constraint. This effect is demonstrated in Figure 62 where we show the percentage of unused beams when $C_{min} = 20$ Mbps and $C_{max} = 100, 500$ Mbps. Compared to the low traffic scenario in Figure 57, the average number of beams per time slot is largely increased: in the optimal configuration, there is an almost negligible number of unused beams, confirming that the scheduler tends to use all resources to meet the challenging requests of the users. Further increasing the scheduling constraint leads to more strict conditions on co-slot users, which ultimately becomes a disadvantage: too many users are discarded from potential service, beams remain unused while there is still a request to be met, and the number of allocations per user drop.

- In the considered scheduler configuration, a minimum capacity of 5 Mbps or at most 10 Mbps with a maximum of 100 or 300 Mbps are still acceptable traffic environments, or in any case scenarios for which the performance in terms of unmet capacity can be improved even by, e.g., a limited power increase. However, as discussed above, scenarios with more traffic requests from the users tend to be more critical and more advanced solutions are needed, such as modifications or re-designs of the scheduler or

more general system-level modifications (e.g., changing the minimum elevation angle, which leads to a modification to the constellation and a reduced number of users).

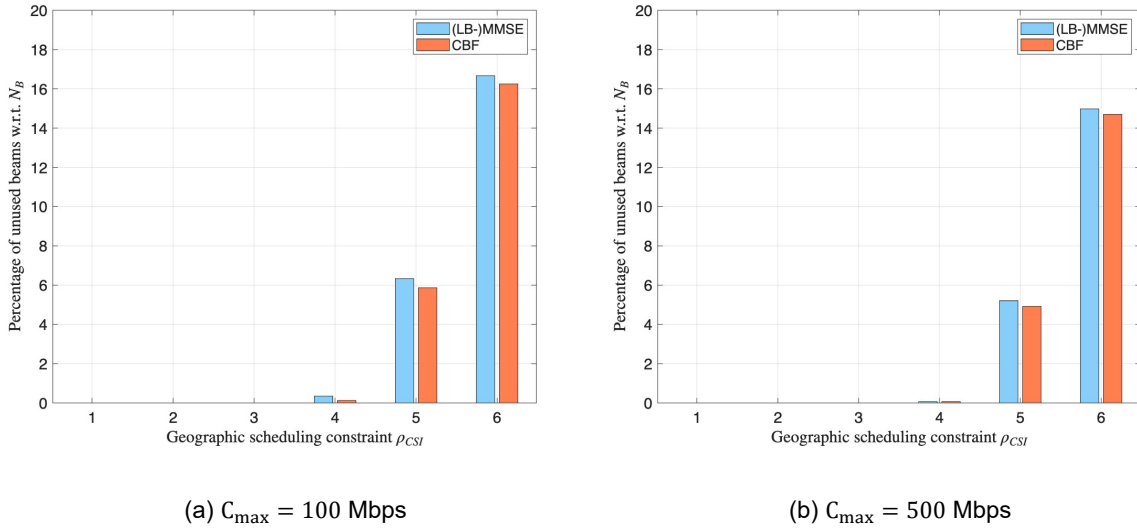


Figure 62: Percentage of unused beams with respect to the maximum N_B with PQS and geographical scheduling constraint, standalone NTN node. Setup: $C_{min} = 5$ Mbps, $C_{max} = 100, 500$ Mbps, $\sigma_{vis} = 10$, $\sigma_{cap} = 2$.

- Finally, the increase in the achieved capacity for increasing traffic conditions is again motivated by the behaviour of the scheduler in terms of unused beams and unmet capacity. In high traffic conditions, the PQS algorithm tends to allocate all N_B beams per time slot, leading to a larger achieved capacity, while in low traffic a sometimes large number of beams is unused, reducing the achieved capacity.

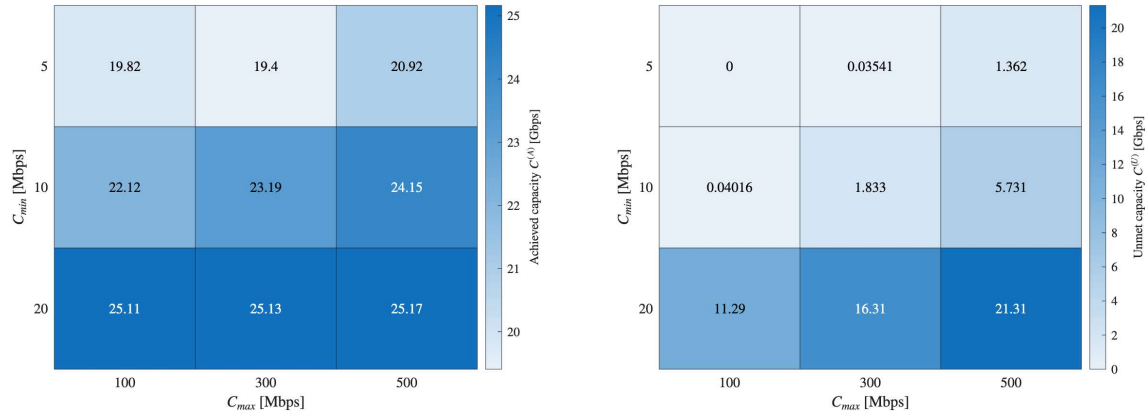
Table 19: Optimal value of the geographical scheduling constraint with CBF beamforming in the standalone scenario. Setup: $\sigma_{vis} = 10$, $\sigma_{cap} = 2$.

Minimum capacity request C_{min}	Maximum capacity request C_{max}		
	100 Mbps	300 Mbps	500 Mbps
5 Mbps	3	4	4
10 Mbps	4	4	4
20 Mbps	4	4	4

The same considerations hold for CBF beamforming, as summarised in Table 19 and Figure 63; we can also notice that the performance of MMSE solutions is better, as expected, but the difference with CBF is not particularly large. This seems to indicate that the system performance is not only limited by the co-slot interference (minimised by MMSE solutions), but also (and more significantly) by the number of allocations per users, which is ultimately driven by the scheduling constraints.

Table 20 and Table 21, with the corresponding Figure 64 and Figure 65, summarise the performance when using the CSI scheduling constraint. The performance, in general, is a bit improved with CSI compared to the geographical case; this is in line with previous discussions: for the CSI vectors of two users to be similar, a low geographical separation between two users is a sufficient condition for similar channels, but not a necessary one. However, the difference

between the two approaches is not massive: up to 1 Gbps in low traffic and 0.5 Gbps in high traffic. Apart from these observations, all the trends extensively discussed above still hold.



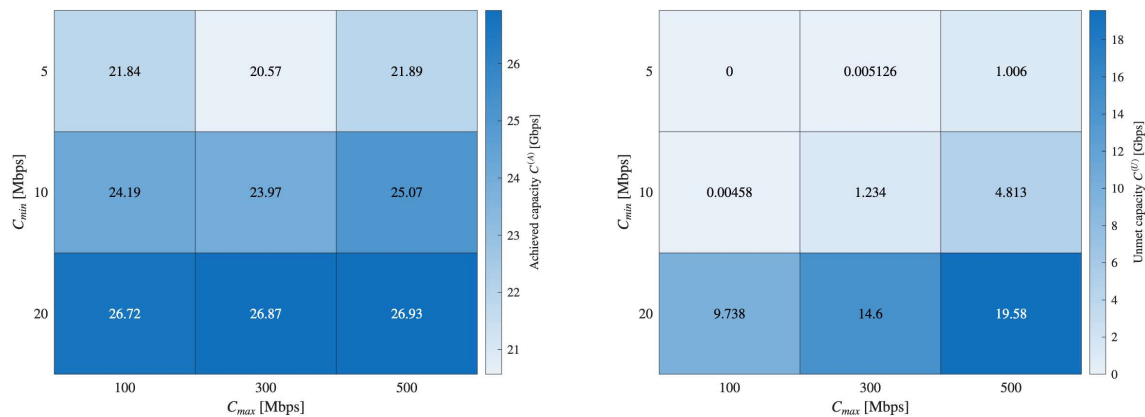
(a) Achieved capacity $C^{(A)}$

(b) Unmet capacity $C^{(U)}$

Figure 63: Achieved and unmet capacity with geographical scheduling constraint with CBF beamforming in the standalone scenario. Setup: $\sigma_{vis} = 10$, $\sigma_{cap} = 2$.

Table 20: Optimal value of the CSI scheduling constraint with (LB-)MMSE beamforming in the standalone scenario. Setup: $\sigma_{vis} = 10$, $\sigma_{cap} = 2$.

Minimum capacity request C_{min}	Maximum capacity request C_{max}		
	100 Mbps	300 Mbps	500 Mbps
5 Mbps	0.08	0.05	0.06
10 Mbps	0.05	0.05	0.05
20 Mbps	0.04	0.04	0.04



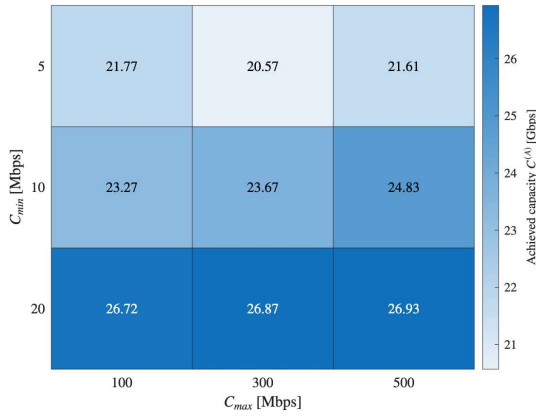
(a) Achieved capacity $C^{(A)}$

(b) Unmet capacity $C^{(U)}$

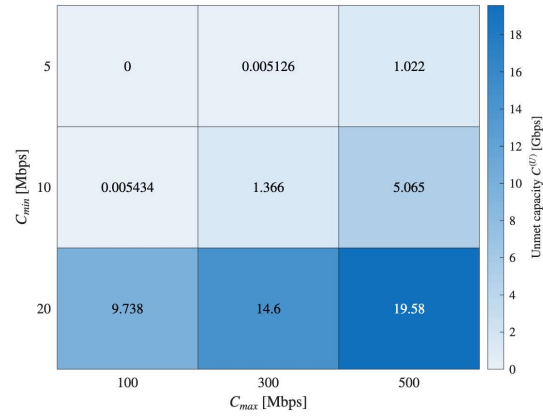
Figure 64: Achieved and unmet capacity with CSI scheduling constraint with MMSE beamforming in the standalone scenario. Setup: $\sigma_{vis} = 10$, $\sigma_{cap} = 2$.

Table 21: Optimal value of the CSI scheduling constraint with CBF beamforming in the standalone scenario. Setup: $\sigma_{vis} = 10$, $\sigma_{cap} = 2$.

Minimum capacity request C_{min}	Maximum capacity request C_{max}		
	100 Mbps	300 Mbps	500 Mbps
5 Mbps	0.07	0.05	0.05
10 Mbps	0.04	0.04	0.04
20 Mbps	0.04	0.04	0.04



(a) Achieved capacity $C^{(A)}$



(b) Unmet capacity $C^{(U)}$

Figure 65: Achieved and unmet capacity with CSI scheduling constraint with CBF beamforming in the standalone scenario. Setup: $\sigma_{vis} = 10$, $\sigma_{cap} = 2$.

Table 22: Optimal value of the geographical scheduling constraint with (LB-)MMSE and CBF beamforming in the standalone scenario. Setup: $\sigma_{vis} = 10, 50, 150$, $\sigma_{cap} = 2$.

Scheduler configuration	C_{min}	(LB-)MMSE			CBF		
		C_{max}					
		100 Mbps	300 Mbps	500 Mbps	100 Mbps	300 Mbps	500 Mbps
$\sigma_{vis} = 10$ ms $\sigma_{cap} = 2$	5 Mbps	3	4	4	3	4	4
	10 Mbps	4	4	4	4	4	4
	20 Mbps	4	4	4	4	4	4
$\sigma_{vis} = 50$ ms $\sigma_{cap} = 2$	5 Mbps	3	4	4	3	4	4
	10 Mbps	4	4	4	4	4	4
	20 Mbps	4	4	4	4	4	4
$\sigma_{vis} = 150$ ms $\sigma_{cap} = 2$	5 Mbps	3	4	4	3	4	4
	10 Mbps	4	4	4	4	4	4
	20 Mbps	4	4	4	4	4	4

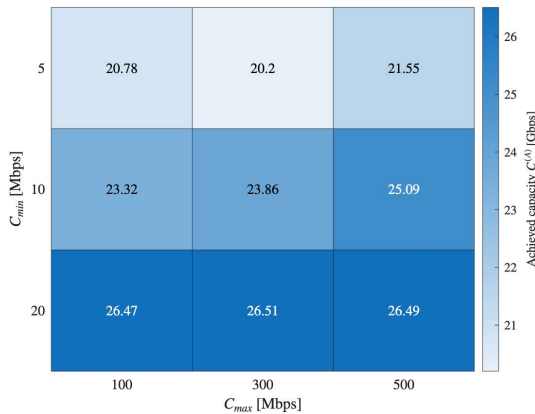
Table 22 and Table 23 report the optimal values of the geographical and CSI scheduling constraints when considering $\sigma_{cap} = 2$ and a variable value for the residual visibility bound σ_{vis}

to identify the two considered priority classes. It can be noticed that significantly reducing ($\sigma_{vis} = 10$ means that the user will be out of visibility within the next 0.1 seconds) or increasing ($\sigma_{vis} = 150$ means that the user will be out of visibility within the next 1.5 seconds) the visibility bound is basically not impacting at all the optimal constraint. The case with $\sigma_{vis} = 10$ is kept into these tables for ease of comparison, but this case was already discussed above, and we highlighted in bold the cases in which a different optimal configuration arises.

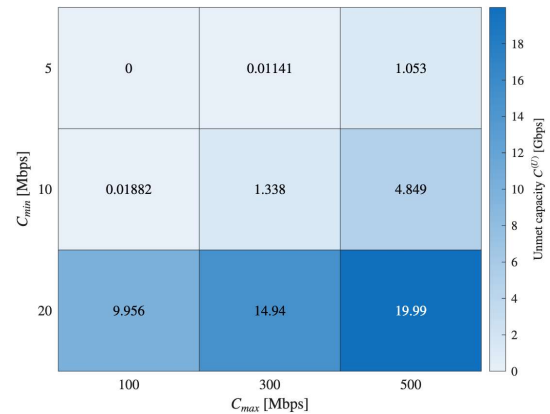
Figure 66 and Figure 67 show the sum achieved and unmet capacity with the optimal geographical scheduling constraint, while Figure 68 and Figure 69 provide the numerical results with the optimal CSI scheduling constraint. It is confirmed that the differences in the capacity values are not significant (they are in the order of a few tens of Mbps, if any). Thus, we can state that, at least for PQS, the impact of the residual visibility time on the optimal scheduler configuration is quite limited.

Table 23: Optimal value of the CSI scheduling constraint with (LB-)MMSE and CBF beamforming in the standalone scenario. Setup: $\sigma_{vis} = 10, 50, 150$, $\sigma_{cap} = 2$.

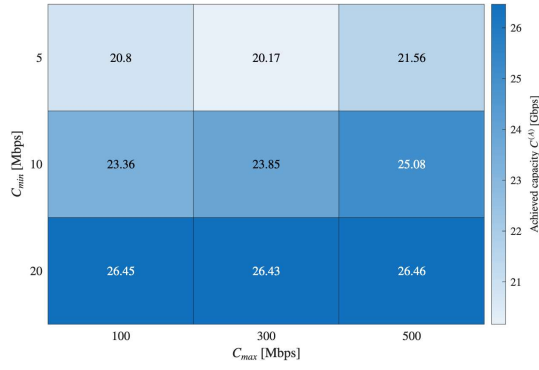
Scheduler configuration	C_{min}	(LB-)MMSE			CBF		
		C_{max}					
		100 Mbps	300 Mbps	500 Mbps	100 Mbps	300 Mbps	500 Mbps
$\sigma_{vis} = 10$ ms $\sigma_{cap} = 2$	5 Mbps	0.08	0.05	0.06	0.07	0.05	0.05
	10 Mbps	0.05	0.05	0.05	0.04	0.04	0.04
	20 Mbps	0.04	0.04	0.04	0.04	0.04	0.04
$\sigma_{vis} = 50$ ms $\sigma_{cap} = 2$	5 Mbps	0.09	0.05	0.06	0.07	0.05	0.05
	10 Mbps	0.05	0.05	0.05	0.04	0.04	0.04
	20 Mbps	0.04	0.04	0.04	0.04	0.04	0.04
$\sigma_{vis} = 150$ ms $\sigma_{cap} = 2$	5 Mbps	0.09	0.05	0.06	0.07	0.05	0.05
	10 Mbps	0.05	0.05	0.05	0.04	0.04	0.04
	20 Mbps	0.04	0.04	0.04	0.04	0.04	0.04



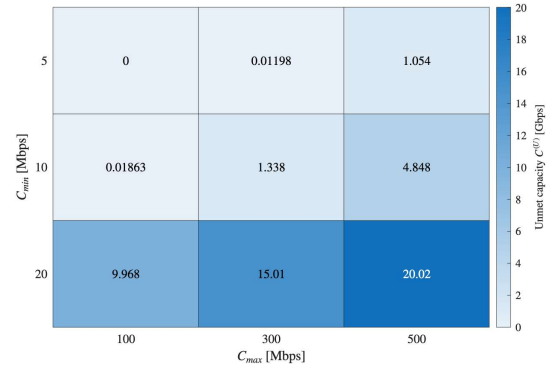
(a) Achieved capacity $C^{(A)}$, $\sigma_{vis} = 50$



(b) Unmet capacity $C^{(U)}$, $\sigma_{vis} = 50$

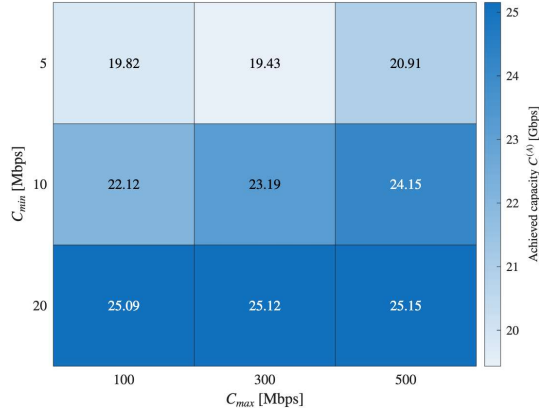


(a) Achieved capacity $C^{(A)}$, $\sigma_{vis} = 150$

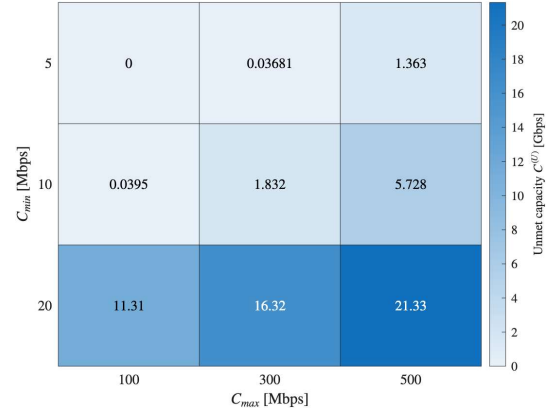


(b) Unmet capacity $C^{(U)}$, $\sigma_{vis} = 150$

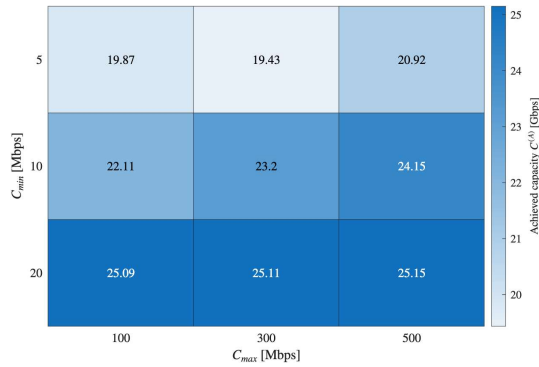
Figure 66: Achieved and unmet capacity with geographical scheduling constraint with (LB-)MMSE beamforming in the standalone scenario. Setup: $\sigma_{vis} = 50, 150$, $\sigma_{cap} = 2$.



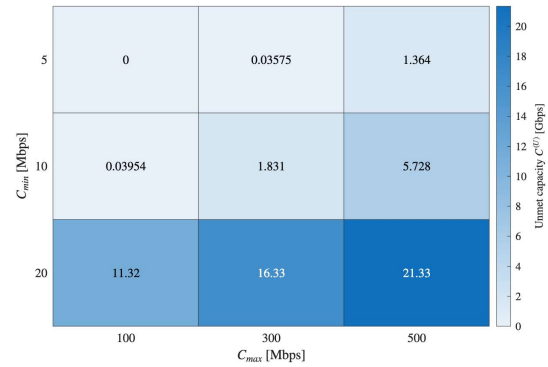
(a) Achieved capacity $C^{(A)}$, $\sigma_{vis} = 50$



(b) Unmet capacity $C^{(U)}$, $\sigma_{vis} = 50$

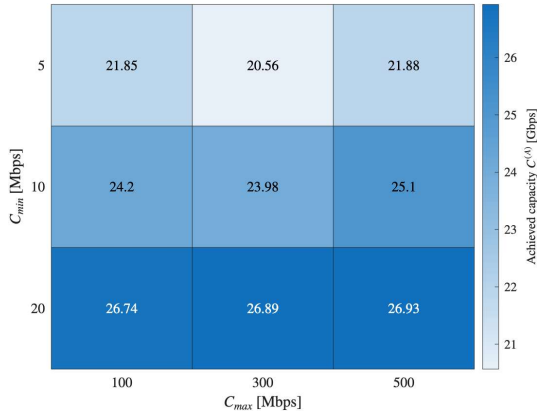


(a) Achieved capacity $C^{(A)}$, $\sigma_{vis} = 150$

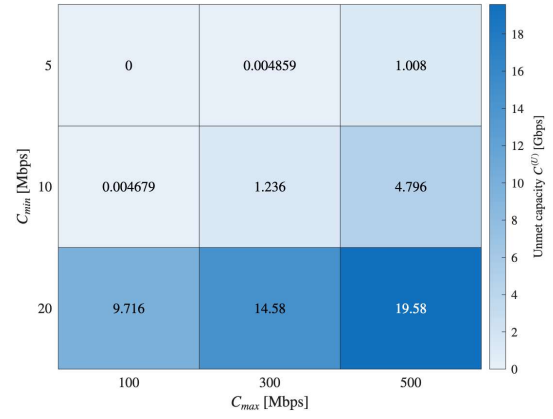


(b) Unmet capacity $C^{(U)}$, $\sigma_{vis} = 150$

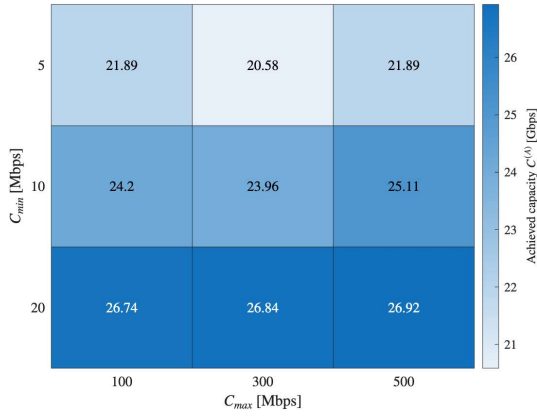
Figure 67: Achieved and unmet capacity with geographical scheduling constraint with CBF beamforming in the standalone scenario. Setup: $\sigma_{vis} = 50, 150$, $\sigma_{cap} = 2$.



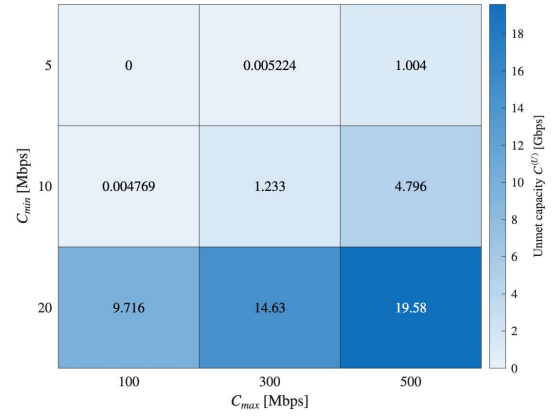
(a) Achieved capacity $C^{(A)}$, $\sigma_{vis} = 50$



(b) Unmet capacity $C^{(U)}$, $\sigma_{vis} = 50$

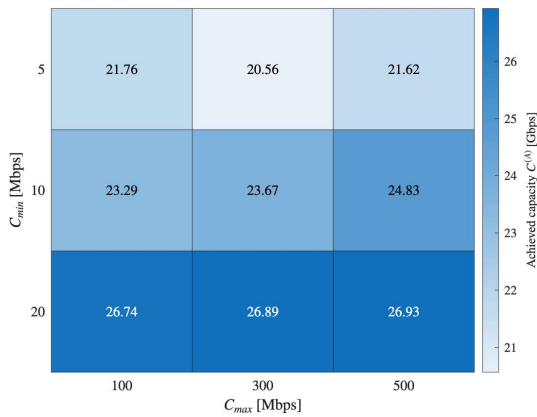


(a) Achieved capacity $C^{(A)}$, $\sigma_{vis} = 150$

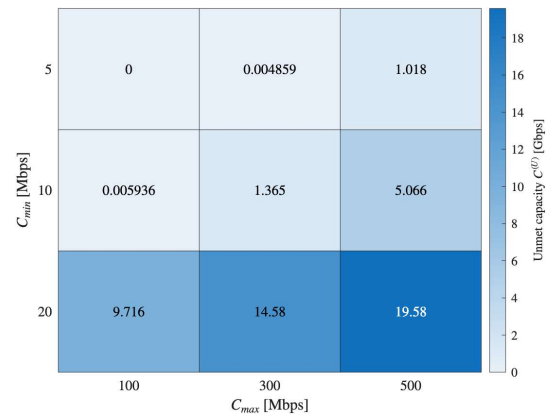


(b) Unmet capacity $C^{(U)}$, $\sigma_{vis} = 150$

Figure 68: Achieved and unmet capacity with CSI scheduling constraint with (LB-)MMSE beamforming in the standalone scenario. Setup: $\sigma_{vis} = 50, 150$, $\sigma_{cap} = 2$.



(a) Achieved capacity $C^{(A)}$, $\sigma_{vis} = 50$



(b) Unmet capacity $C^{(U)}$, $\sigma_{vis} = 50$

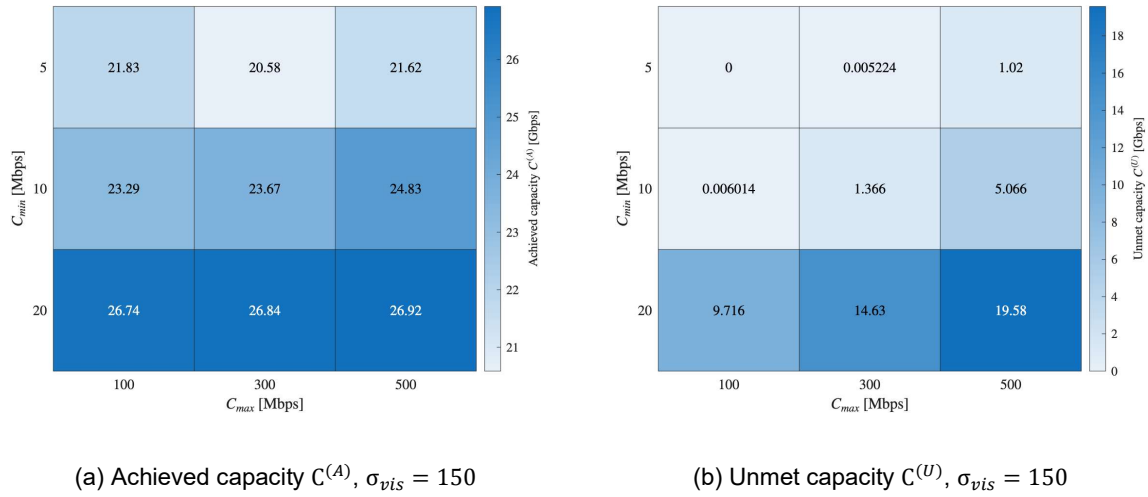


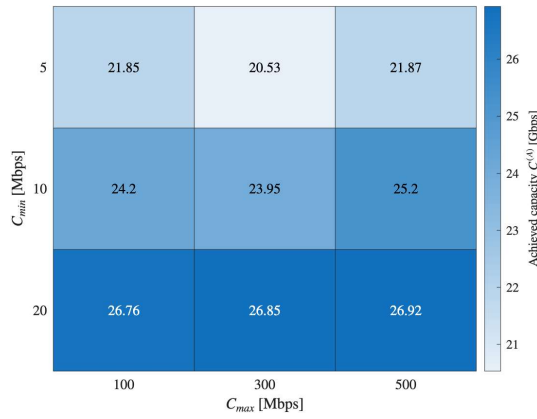
Figure 69: Achieved and unmet capacity with CSI scheduling constraint with CBF beamforming in the standalone scenario. Setup: $\sigma_{vis} = 50, 150$, $\sigma_{cap} = 2$.

In all the above configurations of traffic and scheduler: i) the CSI constraint was outperforming the geographical one; and ii) the impact of the residual visibility in the definition of priority traffic was negligible, if present at all. Based on these considerations, to complete the assessment of PQR with ideal ancillary information, we now evaluate the impact of a variable bound of the unmet capacity, σ_{cap} , while fixing $\sigma_{vis} = 50$ and only considering the CSI scheduling constraint.

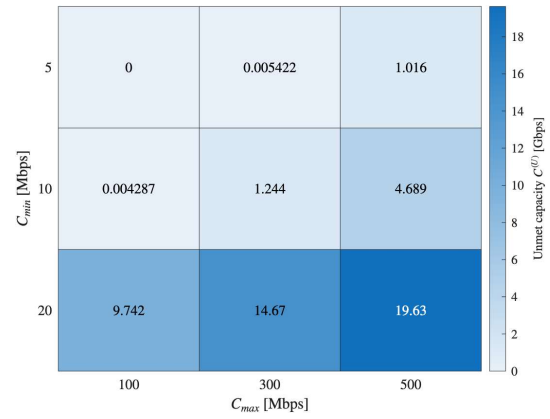
Table 24 reports the optimal value of ρ_{CSI} in the considered scenarios, with the case $\sigma_{cap} = 2$ that is duplicated here to ease the comparison with the analyses discussed above. We can observe that, also varying the bound on the unmet capacity when defining the priority classes, in very limited cases a different optimal configuration is required. Moreover, the differences in the performance are very limited. For instance, with $\sigma_{cap} = 1.2$, $\rho_{CSI} = 0.07$ and $\rho_{CSI} = 0.05$ (the optimal configuration in the reference case with $\sigma_{cap} = 2$) lead to: i) achieved capacity equal to 25.09 Gbps and 24.2 Gbps; and ii) unmet capacity equal to 4.81 Gbps and 4.7 Gbps.

Table 24: Optimal value of the CSI scheduling constraint with (LB-)MMSE and CBF beamforming in the standalone scenario. Setup: $\sigma_{vis} = 50$, $\sigma_{cap} = 1.2, 1.5, 2$.

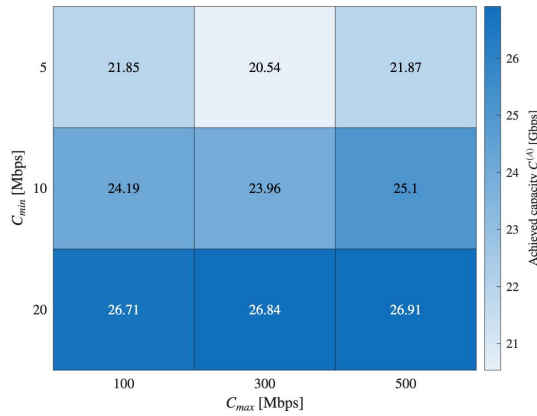
Scheduler configuration	C_{min}	(LB-)MMSE			CBF		
		C_{max}					
		100 Mbps	300 Mbps	500 Mbps	100 Mbps	300 Mbps	500 Mbps
$\sigma_{vis} = 50$ ms $\sigma_{cap} = 1.2$	5 Mbps	0.09	0.05	0.06	0.07	0.05	0.05
	10 Mbps	0.05	0.05	0.07	0.04	0.04	0.04
	20 Mbps	0.04	0.04	0.04	0.04	0.04	0.04
$\sigma_{vis} = 50$ ms $\sigma_{cap} = 1.5$	5 Mbps	0.08	0.05	0.06	0.07	0.05	0.05
	10 Mbps	0.05	0.05	0.05	0.04	0.04	0.04
	20 Mbps	0.04	0.04	0.04	0.04	0.04	0.04
$\sigma_{vis} = 50$ ms $\sigma_{cap} = 2$	5 Mbps	0.09	0.05	0.06	0.07	0.05	0.05
	10 Mbps	0.05	0.05	0.05	0.04	0.04	0.04
	20 Mbps	0.04	0.04	0.04	0.04	0.04	0.04



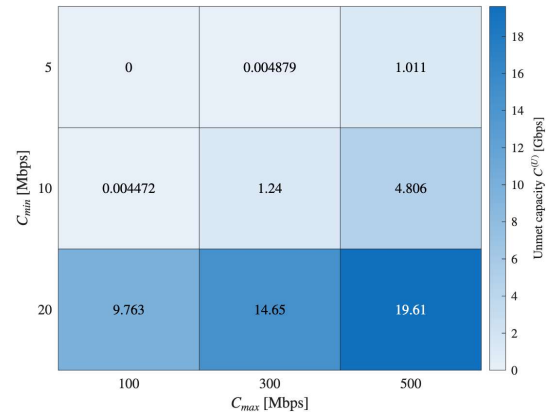
(a) Achieved capacity $C^{(A)}$, $\sigma_{vis} = 1.2$



(b) Unmet capacity $C^{(U)}$, $\sigma_{vis} = 1.2$

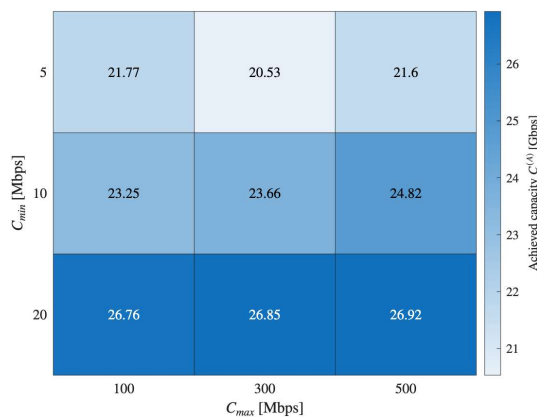


(a) Achieved capacity $C^{(A)}$, $\sigma_{cap} = 1.5$

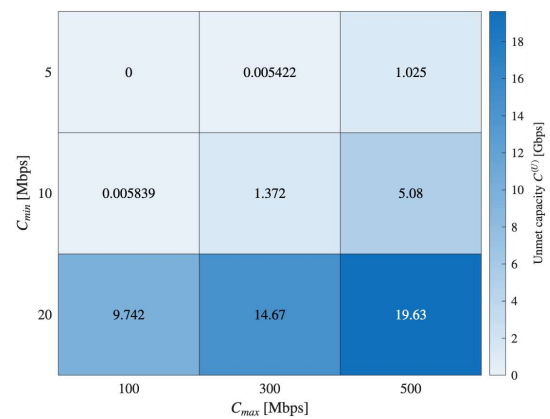


(b) Unmet capacity $C^{(U)}$, $\sigma_{vis} = 1.5$

Figure 70: Achieved and unmet capacity with CSI scheduling constraint with (LB-)MMSE beamforming in the standalone scenario. Setup: $\sigma_{vis} = 50$, $\sigma_{cap} = 1.2, 1.5$.



(a) Achieved capacity $C^{(A)}$, $\sigma_{vis} = 1.2$



(b) Unmet capacity $C^{(U)}$, $\sigma_{vis} = 1.2$

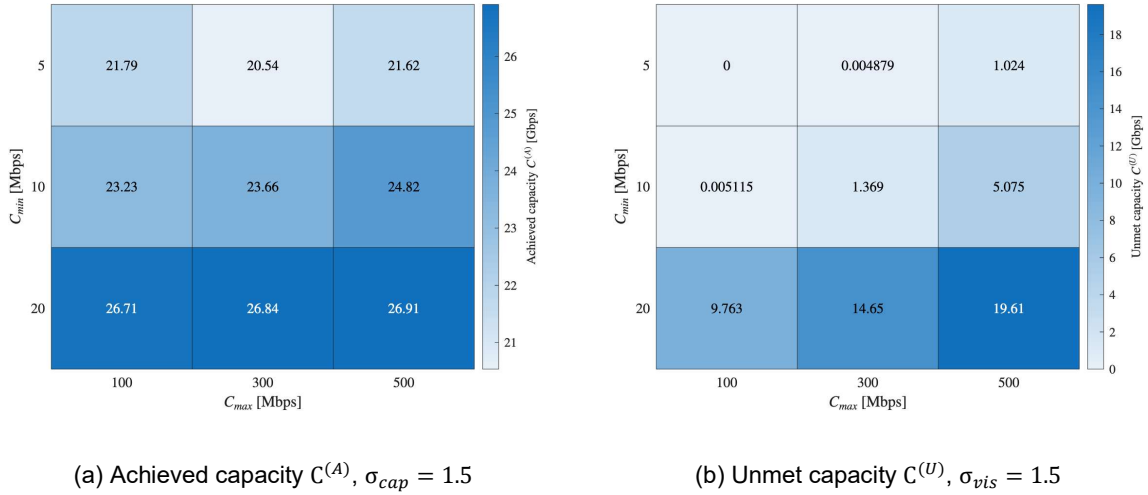


Figure 71: Achieved and unmet capacity with CSI scheduling constraint with CBF beamforming in the standalone scenario. Setup: $\sigma_{vis} = 50$, $\sigma_{cap} = 1.2, 1.5$.

Figure 70 and Figure 71 provide the achieved and unmet capacity in these configurations for (LB-)MMSE and CBF beamforming, respectively. These figures confirm that also the selected values of the unmet capacity bound seem to have a relative impact on the overall performance. Based on the numerical assessment, optimal values of the geographic and CSI scheduling constraint can be obtained; these typically depend on the traffic request of the users, but reasonable trade-offs with a single common value can be derived in almost all medium-to-high traffic scenarios. A limited impact was observed when modifying the capacity or visibility bounds in the definition of the priority classes. This is mainly due to the fact that the traffic distribution is concentrated in a limited range of values (see Figure 45 for an example) and tailoring the PQS configuration to this would call for a larger number of PCs (basically quantizing the range of unmet traffic at each iteration of the algorithm). This might lead to an excessive complexity increase and, in addition, it leads to a reduced flexibility, compared to what can be offered by PRS discussed below.

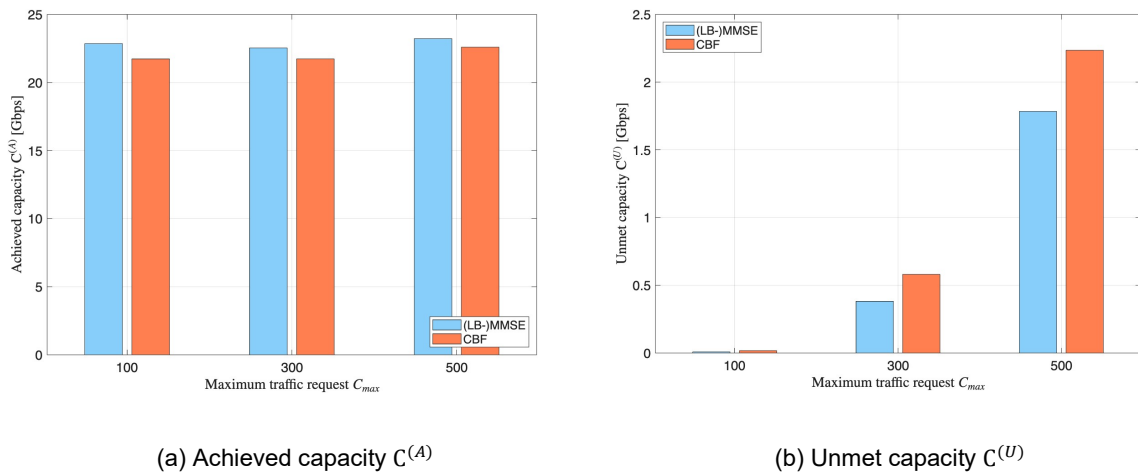


Figure 72: Achieved and unmet capacity with PQS and geographical scheduling constraint, standalone NTN node. Setup: $C_{min} = 20$ Mbps, $\sigma_{vis} = 10$, $\rho_{GEO} = 4$, $\alpha = 0.2$.

The large amount of unmet capacity in high traffic scenarios, which arises in all the considered scheduler and priority class configurations, indicates that the system might not be able to satisfy those requirements. To verify this, we also considered a reduced activity factor, $\alpha = 0.2$ instead of 0.4 as in all other cases, with: i) both geographic and CSI scheduling constraints, in their optimal configuration based on the above analysis; ii) $\sigma_{vis} = 10$, $\sigma_{cap} = 2$; and iii) $C_{min} = 20$ Mbps and $C_{max} = 100, 300, 500$ Mbps.

The results are shown in Figure 72 for the geographical constraint and Figure 73 for the CSI case. It can be observed that the amount of unmet capacity is massively reduced compared to what was happening with a larger activity factor. The achieved capacity is slightly reduced, but this should not worry: we are indeed considering a scenario with a reduced traffic request. It is thus confirmed that, despite the implementation of user-centric beamforming with full frequency reuse, too large traffic requests might be excessive for the considered NTN system. In such conditions, different strategies can be implemented, such as reducing the service area, clipping the maximum traffic per receiver, or reducing the number of served users. These considerations shall be performed by the system designed and are in any case out of the scope of this deliverable, which aims at identifying the most relevant parameters and performance trends for user-centric beamforming with non-uniform traffic requests.

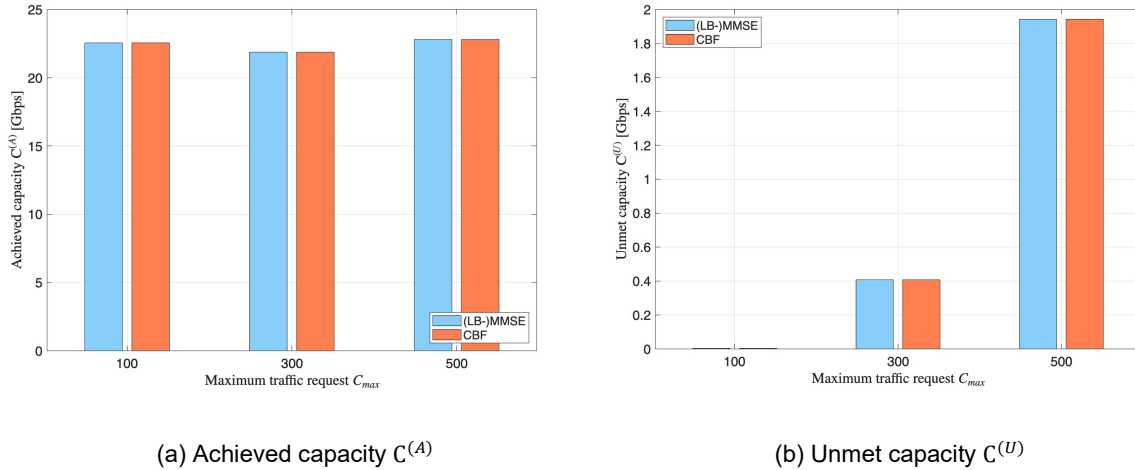


Figure 73: Achieved and unmet capacity with PQS and CSI scheduling constraint, standalone NTN node. Setup: $C_{min} = 20$ Mbps, $\sigma_{vis} = 10$, $\rho_{CSI} = 0.04$, $\alpha = 0.2$.

PRS PERFORMANCE

Table 25 reports the configuration of the PRS scheduler, while the general parameters are still those provided in Table 16 above. As discussed when designing the algorithm, and also justified by the above assessment, tailoring the scheduler operations to the traffic might be challenging with PQS. The PRS approach, on the other hand, does not require the definition of priority classes and specific bounds *a priori*, thus being implicitly more flexible. The complexity spared in the fixed phase (see the complexity assessment above) by not having classes is exploited to implement a more frequent update of the unmet capacity and residual visibility per user to maintain an updated priority rank table.

Since in the above assessment the CSI scheduling constraint was always outperforming the geographic case, in this section we focus on this approach only without affecting the generality of the conclusions or results.

Table 25: Configuration parameters for the PRS algorithm.

Parameter	Value	Comments
Scheduling periodicity T_{sched}	0.1 s (1 slot)	See Section 4.3.1.1
Capacity priority weight ψ_C	0.90, 0.91, ..., 1.00	
Visibility priority weight ψ_V	$1 - \psi_C$	To guarantee that $\psi_C + \psi_V = 1$

In line with the approach followed for PQS, we can first focus on a specific scheduler configuration and then search for the optimal CSI scheduling constraint in each traffic scenario. Then, we will derive the optimal configuration of the scheduler for a general performance assessment. This is motivated by the fact that, based on Table 25, there are 11 values of ψ_C with 5 different configurations of the CSI scheduling constraint and 9 traffic scenarios, leading to a total of 495 cases to be addressed. Let us then first consider the scheduler configured as $(\psi_C, \psi_V) = (0.94, 0.06)$ in the traffic scenario $(C_{min}, C_{max}) = (5 \text{ Mbps}, 100 \text{ Mbps})$.

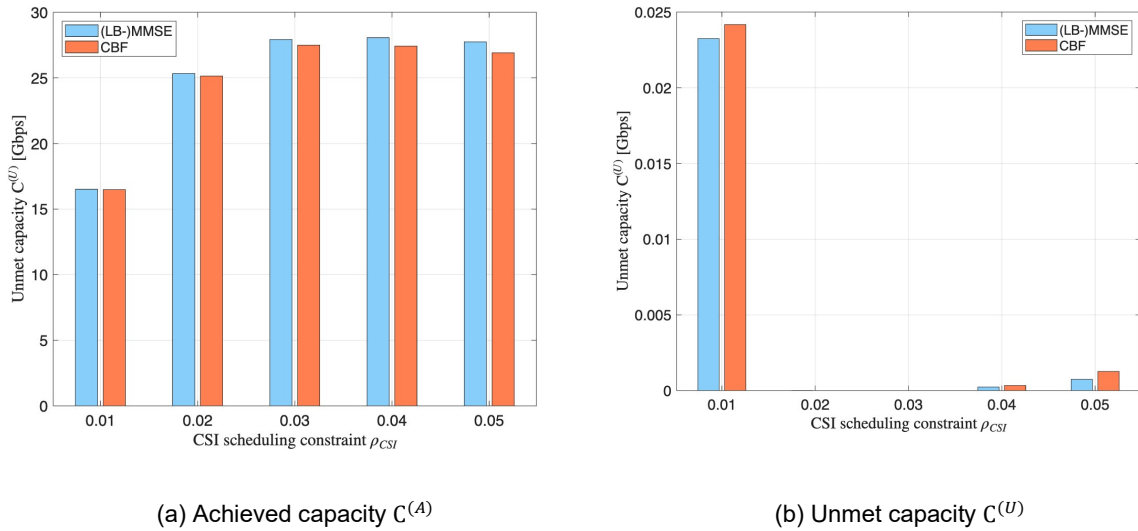

Figure 74: Achieved and unmet capacity with PRS and CSI scheduling constraint, standalone NTN node. Setup: $C_{min} = 5 \text{ Mbps}$, $C_{max} = 100 \text{ Mbps}$, $\psi_C = 0.94$.

Figure 74 shows the achieved and unmet capacity values as a function of the CSI scheduling constraint, that varies between 0.01 and 0.05. The optimal value of the constraint is $\rho_{CSI} = 0.03$, for which we obtain $C^{(A)} = 27.93 \text{ Gbps}$ and $C^{(U)} = 0 \text{ Gbps}$ (6 Gbps more in the achieved capacity compared to PQS).

In line with the trends observed with PQS, the achieved (unmet) capacity increases (decreases) for increasing values of the constraint ρ_{CSI} . When the constraint is too large, we are considering as valid in the same time slot users with too similar channel coefficients and, thus, the performance degrades as expected. These trends are reflected in the performance of the asymptotic achieved capacity per user and the number of allocations per user, provided in Figure 75. Figure 76 shows the percentage of unused beams with a varying CSI constraint and, also in this case, the results are in line with the expectations based on the PQS behaviour. However, compared to PQS, we can observe that the PRS algorithm tends to exploit more the available beams (see Figure 62), and this also leads to a larger achieved capacity in low traffic conditions. As already highlighted previously, PRS has an improved flexibility compared to

PQS, thanks to the definition of a priority table instead of quantizing users into multiple classes, and to the fact that the update of the priority rank table can be performed more frequently.

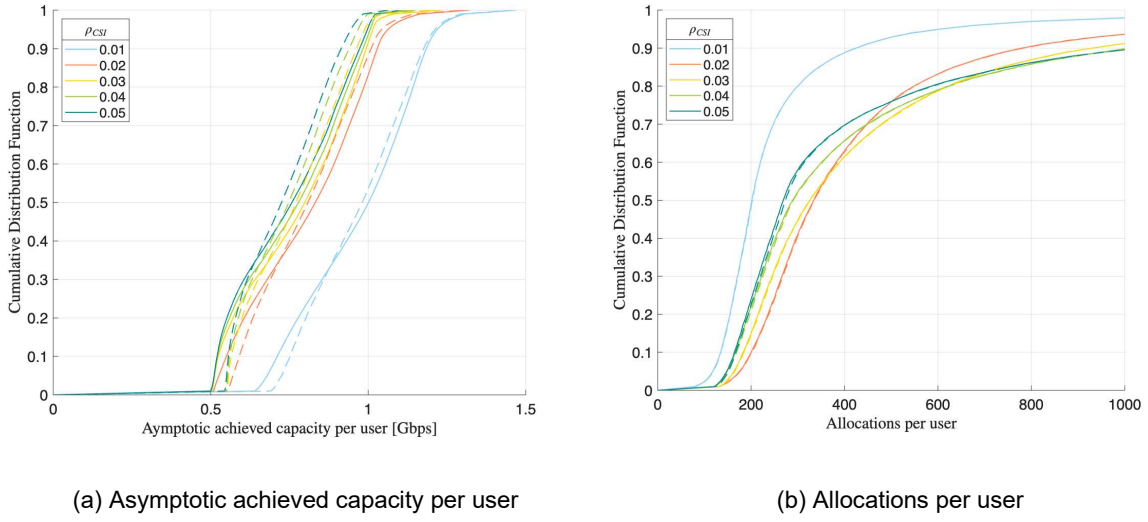


Figure 75: Asymptotic achieved capacity and allocations per user with PRS and CSI scheduling constraint, standalone NTN node. Setup: $C_{min} = 5$ Mbps, $C_{max} = 100$ Mbps, $\psi_C = 0.94$. (LB-)MMSE (solid line) and CBF (dashed line).

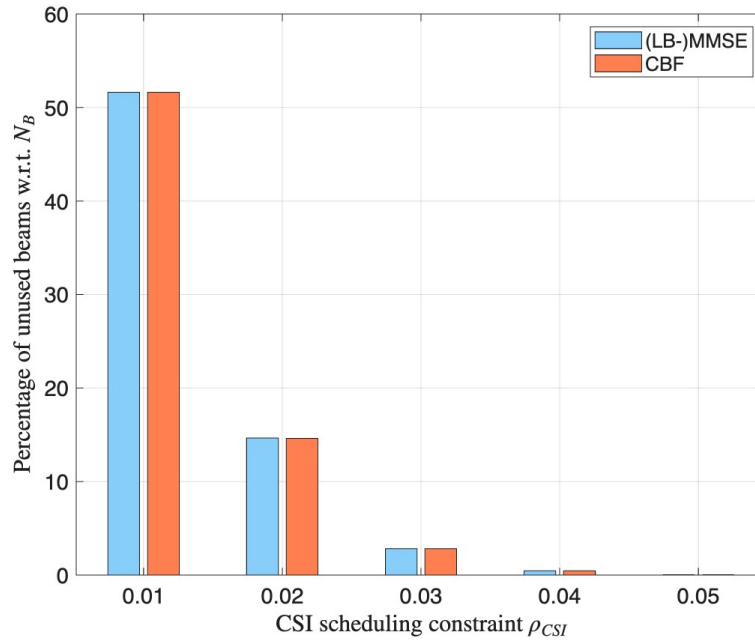


Figure 76: Percentage of unused beams with respect to the maximum N_B with PRS and CSI scheduling constraint, standalone NTN node. Setup: $C_{min} = 5$ Mbps, $C_{max} = 100$ Mbps, $\psi_C = 0.94$.

To identify the best configuration depending on the traffic conditions, in this case we evaluate the optimal combination of scheduling constraint ρ_{CSI} and scheduler weight ψ_C (ψ_T is obtained recalling that $\psi_C + \psi_T$ must be always equal to 1), in terms first of the residual sum unmet capacity and then considering the sum achieved capacity. Table 24 provides these optimal configurations in the form (ψ_C, ρ_{CSI}) . The following observations hold:

- The weight of the unmet capacity, ψ_C , is larger in low to medium traffic scenarios. As a matter of facts, when the minimum capacity request is either 5 or 10 Mbps, the scheduler is optimally configured to either ignore the visibility time or to barely consider it with a 2% impact on the priority rank. This also applied to a minimum capacity of 20 Mbps, as long as the maximum capacity is only 100 Mbps. In such low-medium traffic conditions, the scheduling constraint ρ_{CSI} can be larger so as to increase the allocations per user (higher correlation threshold to discard potential users) even at the expense of a lower average achieved capacity per slot (more users per slot lead to lower SINR per slot); basically, in low-medium traffic scenarios there is no need to provide massive capacities per time slot because the residual visibility of the users is sufficient to serve them through more allocations. This is also reflected in the almost negligible importance of the residual visibility time in the scheduler configuration.

When the traffic requests increase too much, it is more challenging for the scheduler to satisfy the capacity requirement before the users are out of the visibility of the NTN node. Thus, the weight of the residual visibility, ψ_T , is (slightly) increased in order to also take into account the residual visibility of the users.

- The trends with CBF are confirmed when we observe the increase in the minimum capacity, which is a much more impactful factor in defining high traffic scenarios compared to the value of the maximum capacity. For increasing values of the maximum capacity it seems that the trend is opposite compared to what discussed above. However, when evaluating the numerical results, it was observed that both the unmet and the achieved capacities in these scenarios were very close to each other; while the optimal value is indeed the one minimising the unmet capacity, the other configurations provide almost the same values, as shown in Figure 77. As such, this apparently incoherent trend is not a source of concern.

Table 26: Optimal value of the CSI scheduling constraint and scheduler setup ψ_C with (LB-)MMSE and CBF beamforming in the standalone scenario with PRS.

C_{min}	(LB-)MMSE			CBF		
	C_{max}					
	100 Mbps	300 Mbps	500 Mbps	100 Mbps	300 Mbps	500 Mbps
5 Mbps	(1.00, 0.05)	(1.00, 0.04)	(0.98, 0.05)	(1.00, 0.05)	(0.98, 0.05)	(0.98, 0.05)
10 Mbps	(1.00, 0.03)	(1.00, 0.03)	(0.98, 0.04)	(0.90, 0.05)	(0.94, 0.05)	(0.96, 0.05)
20 Mbps	(1.00, 0.03)	(0.94, 0.03)	(0.94, 0.04)	(0.90, 0.05)	(0.90, 0.05)	(0.92, 0.05)

Figure 78 and Figure 79 show the achieved and unmet capacity with (LB-)MMSE and CBF, respectively, obtained with the optimal PRS and scheduling constraint configurations reported in Table 26. In general, similar trends as those already discussed with PQS can be observed. However, compared to PQS, we can notice (as already observed above) that the achieved capacity is maximised under all traffic conditions. As for the unmet capacity, the values are in line with those obtained with PQS and we can further highlight the observation reported previously: the system is already maximising the capacity it can provide and, thus, under high traffic requests more advanced/impacting solutions (reduction of the service area or modification to the constellation) shall be needed. As already mentioned, addressing these solutions is out of the scope of this deliverable.

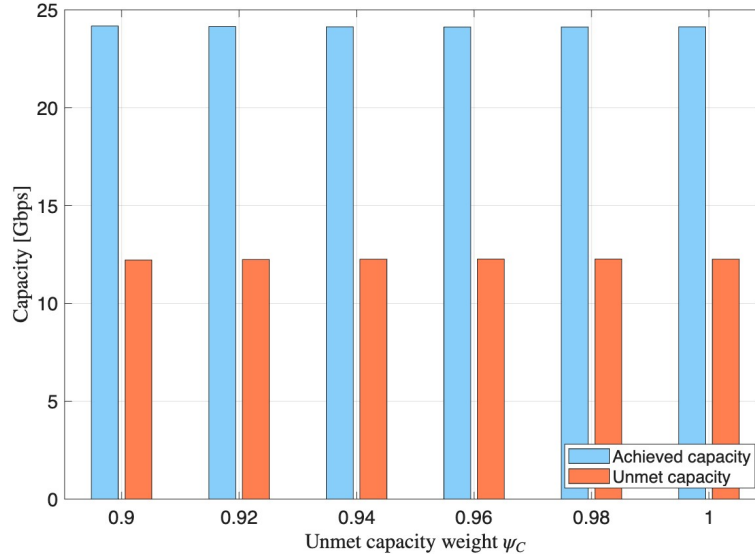
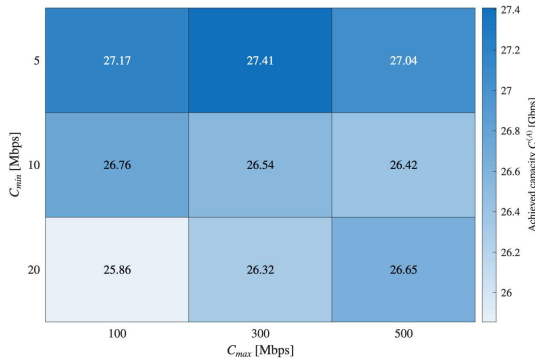
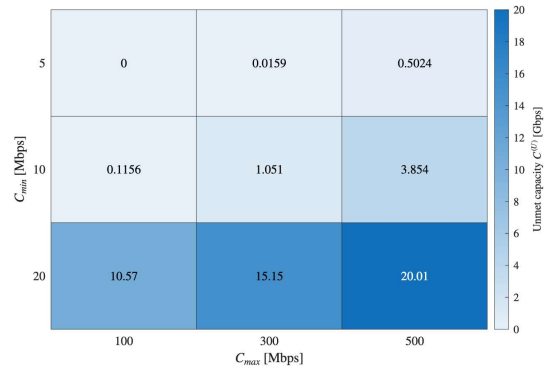


Figure 77: Achieved and unmet capacity with CBF and PRS with the CSI scheduling constraint, standalone NTN node. Setup: $C_{min} = 20$ Mbps, $C_{max} = 100$ Mbps, $\rho_{CSI} = 0.05$.

To conclude the analysis on the PRS solution, we address one last comparison with PQS. In particular, the better performance of the PRS in terms of achieved capacity is motivated by the better flexibility that the approach based on the priority rank provides. However, the performance of the PQS was evaluated with $T_{sche} = 2$ s while PRS with $T_{sched} = 10$ ms; as such, it shall be verified whether the benefit in the achieved capacity (the unmet capacity is basically not changing) is due to the increased scheduling periodicity rather than a better algorithmic solution.

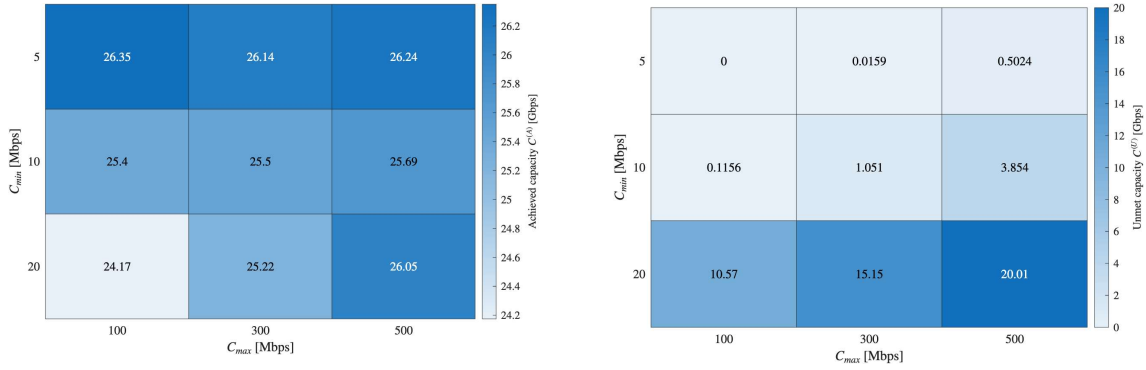


(a) Achieved capacity $C^{(A)}$



(b) Unmet capacity $C^{(U)}$

Figure 78: Achieved and unmet capacity with CSI scheduling constraint and (LB-)MMSE beamforming in the standalone scenario. Optimal PRS configuration.

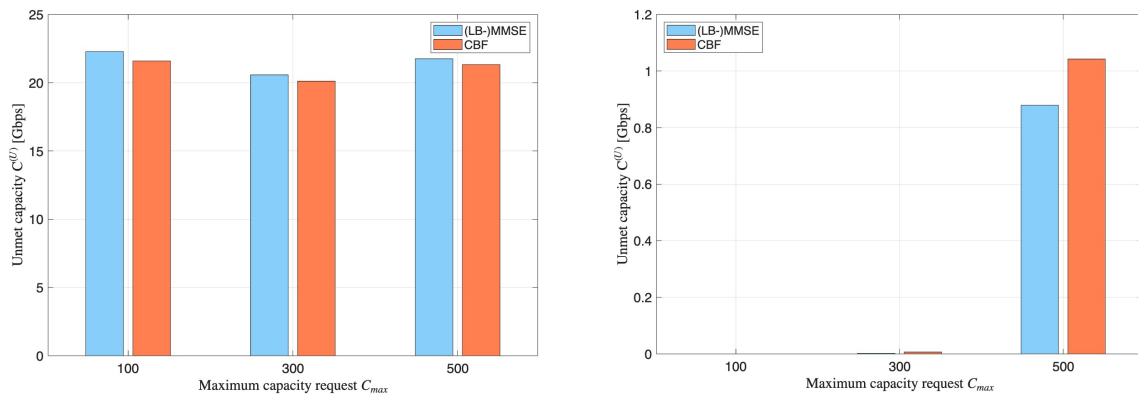


(a) Achieved capacity $C^{(A)}$

(b) Unmet capacity $C^{(U)}$

Figure 79: Achieved and unmet capacity with CSI scheduling constraint and CBF beamforming in the standalone scenario. Optimal PRS configuration.

To address this last aspect, Figure 80 shows the performance of PQS in its optimal configuration (see above) when considering traffic scenarios at $C_{min} = 5$ Mbps; in fact, these are the scenarios in which the most difference between PRS and PQS arises. It can be noticed that the performance of PQS in its most adaptive implementation (scheduling performed each at time slot) is basically identical (a few Mbps difference) to that previously obtained with $T_{sc} = 2$ s. This confirms that the improved performance of the PRS algorithm is motivated by the better flexibility and capability of the priority rank to tailor the scheduling slots to the actual user needs. The identification of priority classes is more difficult, requiring an optimisation on both their number and the boundaries; the latter may not be straightforward with large N_{PC} , since they shall be mutually exclusive, thus calling for advanced optimisation methods. PRS, on the other hand, requires the optimisation of the relative weight of the unmet capacity, which is any case is within to a relatively limited range as discussed above, at least focusing on traffic scenarios for which the final unmet capacity is acceptable (for all other cases, different approaches are needed).



(a) Achieved capacity $C^{(A)}$

(b) Unmet capacity $C^{(U)}$

Figure 80: Achieved and unmet capacity with CSI scheduling constraint and CBF beamforming in the standalone scenario. Optimal PQS configuration with $T_{sched} = 10$ ms and $C_{min} = 5$ Mbps.

4.3.1.5 Performance assessment with non-ideal ancillary information

The extensive performance assessment reported in the previous section is based on the assumption of ideal ancillary information at the transmitter, which means ideal CSI for CSI-based techniques and ideal location estimation for location-based ones. However, both the CSI and the location estimates might be subject to estimation errors. Below, we first introduce the mathematical framework to model the estimation errors and then discuss the impact of non-ideal ancillary information on the performance, [20], [29]. It shall be noticed that, under the assumption of non-ideal ancillary information, even in clear-sky condition the MMSE and LB-MMSE beamforming algorithms provide different performance, since they are differently affected by the errors in the estimation provided by the users.

In 5G/5G-Advanced systems, the estimation of the CSI is based on the CSI-RS signals sent by the NTN node(s), which carry known reference signals to support the estimation of the CSI. Typically, the CSI-RS can occupy 1, 2, or 4 OFDM symbols; assuming that we are transmitting the CSI-RS with a normal Cyclic Prefix (CP) length, we thus have a data-aided estimation algorithm with $\ell_0 = 14, 28, 56$ symbols. In this framework, we can model the CSI errors by exploiting the Cramer-Rao Bound (CRB), which provides the lower bound on the variance of an unbiased estimator. In particular, the CRB of the amplitude and the phase estimates of a generic parameter via unbiased data-aided estimation can be computed as, [30], [31]:

$$\sigma_A^{(CRB)} = \frac{1}{\ell_0} \frac{1}{\gamma} \quad (38)$$

$$\sigma_P^{(CRB)} = \frac{1}{2\ell_0} \frac{1}{\gamma} \quad (39)$$

where γ denotes the SNR at which the estimation has been performed. Let us now denote by $\mathbf{H}_{vis}^{(t_R)}$ the $N_{UE}^{(t_R)} \times N_R$ matrix of channel coefficients between the $N_{UE}^{(t_R)}$ visible in the generic reporting instant t_R and the N_R on-board radiating elements. The RRM algorithm and, based on its decisions, the beamformer rely on this matrix under the assumption of ideal CSI estimation. When the CSI estimation is not ideal, the RRM and the beamforming algorithms exploit the following non-ideal channel matrix $\tilde{\mathbf{H}}_{vis}^{(t_R)}$:

$$\tilde{\mathbf{H}}_{vis}^{(t_R)} = \mathbf{H}_{vis}^{(t_R)} + \mathbf{E} \quad (40)$$

where $\mathbf{E}$ is a $N_{UE}^{(t_R)} \times N_R$ CSI error matrix modelling the CSI errors. Based on equations (38) and (39), for the generic (i, j) -th element $e_{i,j} = |e_{i,j}|e^{-j\angle e_{i,j}}$ we assume that:

- The amplitude $|e_{i,j}|$ is modelled as a zero-mean Gaussian r.v. with variance $\sigma_A^{(CRB)} |h_{i,j}|^2$, i.e., $|e_{i,j}| \sim \mathcal{N}(0, \sigma_A^{(CRB)} |h_{i,j}|^2)$ with $h_{i,j}$ being the generic actual channel coefficient in $\mathbf{H}_{vis}^{(t_R)}$. The scaling of the variance by the amplitude of the actual channel coefficient is motivated by observing that the CRB provides a normalised variance of the estimation error. We thus have to remove such normalisation to obtain the actual error.
- The phase $\angle e_{i,j}$ is modelled as a zero-mean Gaussian r.v. with variance $\sigma_P^{(CRB)}$, i.e., $\angle e_{i,j} \sim \mathcal{N}(0, \sigma_P^{(CRB)})$.

The error matrix $\mathbf{E}$ is thus defined based on the estimation SNR, γ , and the number of symbols for data-aided estimation, ℓ_0 , which will be parameters in the performance evaluation.

With respect to location estimation errors, we can model the positioning error by introducing an uncertainty region, with radius r_{err} , around the correct location of the generic k -th UE, $(\text{lat}_k, \text{lon}_k)$. In this framework, we denote by: i) $\psi_{err,k} \sim \mathcal{U}[0, 2\pi]$ the uniformly distributed random direction of the estimation error, providing the angle between the North direction and the vector connecting the correct and wrong locations; and ii) $\zeta_{err,k} \sim \mathcal{U}[0, r_{err}]$ the distance between the correct and the wrong user location. As shown in [1] and [21], the latter can be translated into an angular distance between the correct and wrong user's location in terms of the Earth central angle as:

$$\delta_{err,k} = \frac{180}{\pi} \frac{\zeta_{err,k}}{3.6R_E} [\text{deg}] \quad (41)$$

with R_E being the Earth's radius. The wrong user location can then be computed as:

$$\begin{aligned} \text{lat}_{err,k} &= \sin^{-1} \tilde{\mathbf{p}}_{err,k}^{(ECEF,z)} \\ \text{lon}_{err,k} &= \text{atan2} \left(\tilde{\mathbf{p}}_{err,k}^{(ECEF,y)}, \tilde{\mathbf{p}}_{err,k}^{(ECEF,x)} \right) \end{aligned} \quad (42)$$

with $\tilde{\mathbf{p}}_{err,k}^{(ECEF)} = [\tilde{\mathbf{p}}_{err,k}^{(ECEF,x)}, \tilde{\mathbf{p}}_{err,k}^{(ECEF,y)}, \tilde{\mathbf{p}}_{err,k}^{(ECEF,z)}]^T$ being the column vector of Earth-Centered Earth-Fixed coordinates of the user affected by the estimation error:

$$\tilde{\mathbf{p}}_{err,k}^{(ECEF)} = \begin{bmatrix} -\sin \delta_{err,i} (\sin \text{lat}_k \cos \text{lon}_k \cos \psi_{err,k} + \sin \text{lon}_k \sin \psi_{err,k}) + \cos \delta_{err,k} \cos \text{lat}_k \cos \text{lon}_k \\ -\sin \delta_{err,i} (\sin \text{lat}_k \sin \text{lon}_k \cos \psi_{err,k} - \cos \text{lon}_k \sin \psi_{err,k}) + \cos \delta_{err,k} \cos \text{lat}_k \sin \text{lon}_k \\ \sin \delta_{err,k} \cos \text{lat}_k \cos \psi_{err,k} + \cos \delta_{err,k} \sin \text{lat}_k \end{bmatrix} \quad (43)$$

The locations affected by the estimation error in (43) are then used by the NTN node(s) to estimate the CSI of each user and apply the RRM and beamforming algorithms. In this case, the estimation error is driven by the dimension of the uncertainty region, r_{err} , which will be a parameter in the performance evaluation.

Table 27: Configuration parameters for the analysis with non-ideal ancillary information.

Parameter	Value	Comments
CSI estimation SNR γ	$-10, -8, \dots, 0$ dB	
Symbols for CSI estimation ℓ_0	14 (1 OFDM symbol) 56 (4 OFDM symbols)	OFDM symbols with normal CP
Uncertainty region radius for location estimation r_{err}	0.1, 1, 10 km	
Scheduler	PRS	See Section 4.3.1.2
Minimum capacity request C_{min}	5, 20 Mbps	
Maximum capacity request C_{max}	100, 500 Mbps	

The performance evaluation with non-ideal CSI or location estimations is performed assuming the PRS algorithm with the CSI scheduling constraint (the best solution, as discussed above). The best combination of scheduling constraint and PRS weights were extensively discussed in the previous section and, thus, we focus only on the impact of different errors on the CSI or location estimates, as summarised in Table 27. It shall be noticed that the performance with errors on the CSI is provided with MMSE only, as this is the algorithm we retained from [1] in the CSI-based category, while the errors on the location estimation are discussed for LB-MMSE and CBF, both location-based solutions. Moreover, it is worthwhile highlighting that an

error on the CSI or location estimation is not just impacting the beamforming coefficients (due to a misalignment with respect to the actual channel condition), but also the identification of users that shall not be scheduled together based on the selected scheduling constraint. The numerical assessment is performed with the configuration providing the best performance, discussed in the previous section and summarised in Table 26.

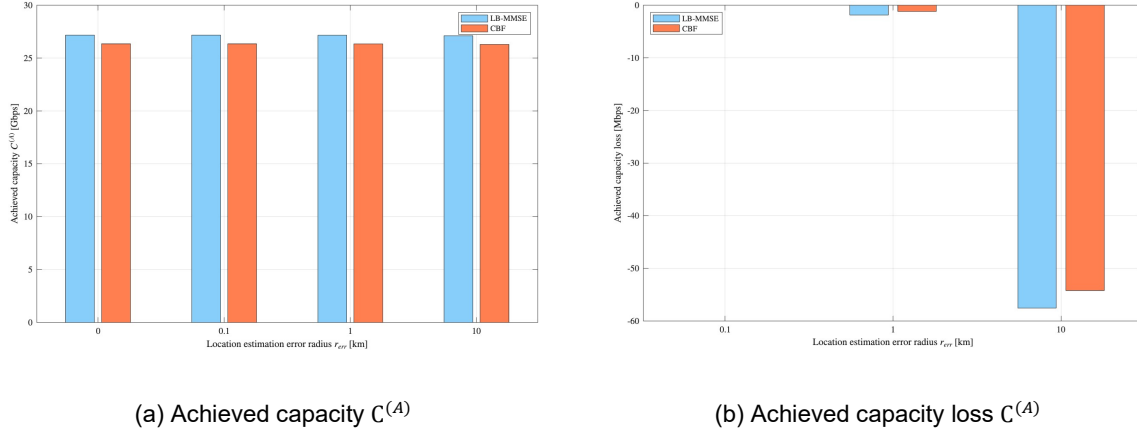


Figure 81: Performance with non-ideal location estimation. Setup: $C_{min} = 5$ Mbps, $C_{max} = 100$ Mbps, PRS with $\psi_{cap} = 1$ and $\rho_{CSI} = 0.05$.

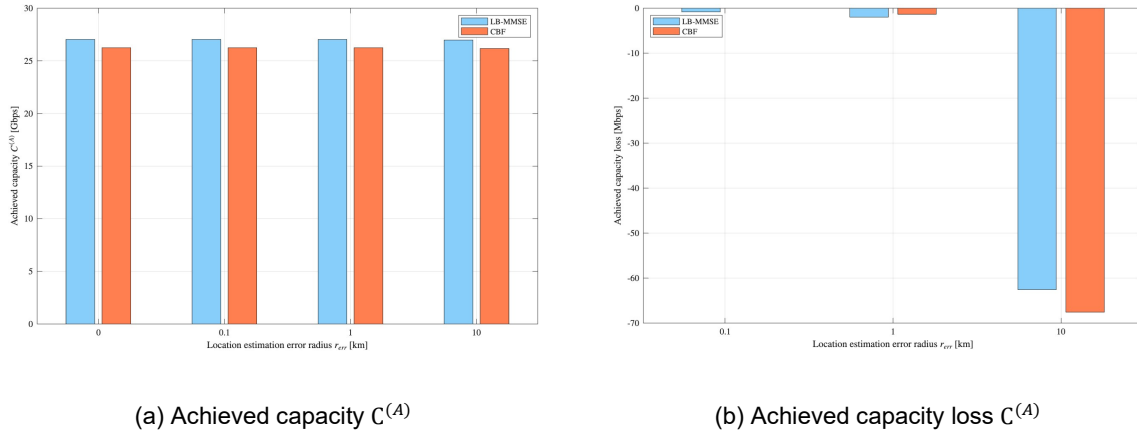
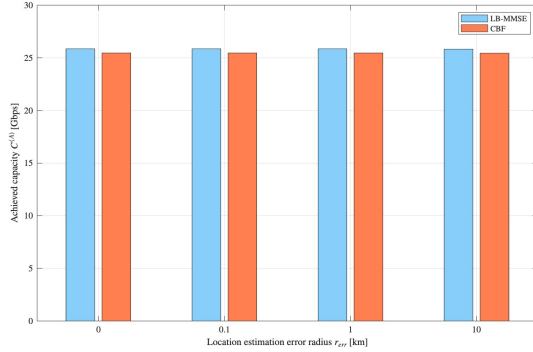


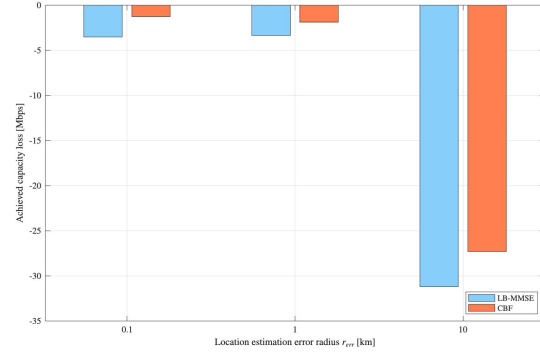
Figure 82: Performance with non-ideal location estimation. Setup: $C_{min} = 5$ Mbps, $C_{max} = 500$ Mbps, PRS with $\psi_{cap} = 0.98$ and $\rho_{CSI} = 0.05$.

Figure 81 and Figure 82 show the achieved capacity in low traffic scenarios, with $C_{min} = 5$ Mbps. In line with the assessments reported in [19], the impact of a non-ideal location estimation is limited on location-based beamforming algorithms. A first motivation for this behaviour is that we are considering clear-sky conditions and, thus, there is no effect related to atmospheric losses or shadowing impacting the performance; in LOS or NLOS conditions, we would have had an impact in case we were also estimating the additional channel impairments via the estimated location. A second motivation is due to geometrical reasons; when estimating the channel from the estimated location, we are deducing the value of the slant range and the radiation pattern at the transmitter (we are always assuming ideal satellite tracking in the standalone scenario, so there is no impact of the radiation pattern at the receiver). With a satellite orbiting at 1000 km, an error of 1-10 km on the ground is definitely

negligible when computing the slant range or the angular difference at which a signal is transmitted. As a consequence, observing the loss in the achieved capacity in the figures on the right we can see that with an error of 10 km we still have a negligible loss (50-60 Mbps).

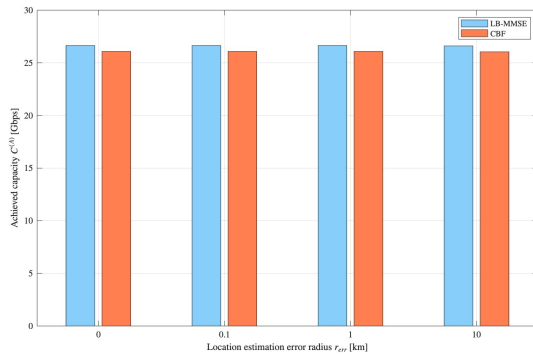


(a) Achieved capacity $C^{(A)}$

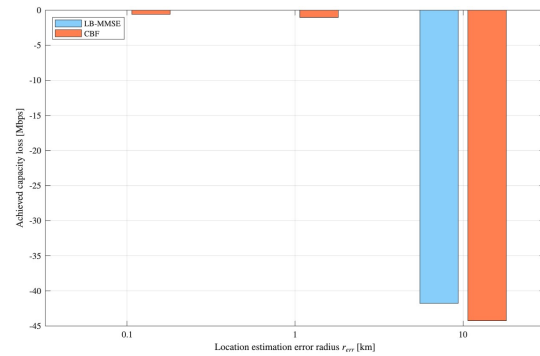


(b) Achieved capacity loss $C^{(A)}$

Figure 83: Performance with non-ideal location estimation. Setup: $C_{min} = 20$ Mbps, $C_{max} = 100$ Mbps, PRS with $\psi_{cap} = 1$ and $\rho_{CSI} = 0.03$.



(a) Achieved capacity $C^{(A)}$

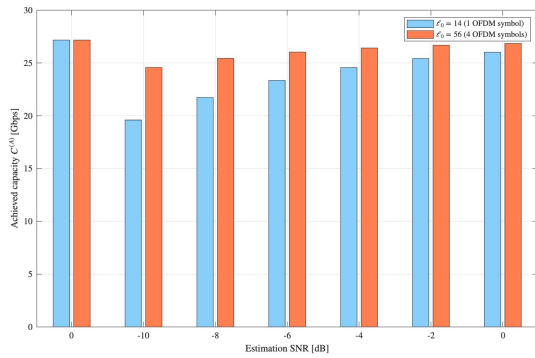


(b) Achieved capacity loss $C^{(A)}$

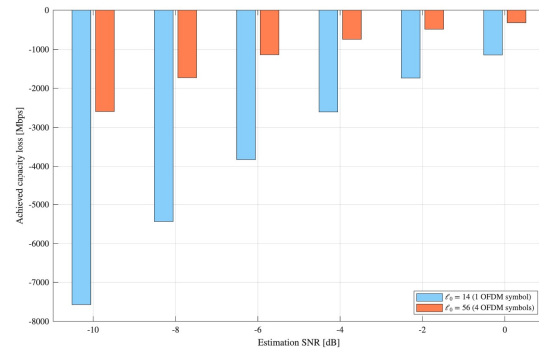
Figure 84: Performance with non-ideal location estimation. Setup: $C_{min} = 20$ Mbps, $C_{max} = 500$ Mbps, PRS with $\psi_{cap} = 0.94$ and $\rho_{CSI} = 0.04$.

Figure 83 and Figure 84 show the performance in high traffic conditions, and it can be noticed that the same considerations still hold. A potential source of larger losses for location-based algorithms could be: i) the non-ideal knowledge of the satellite ephemeris (which in any case is also unlikely), as it would lead to larger effects in the estimated slant range or radiation pattern values; or ii) the non-ideal knowledge of the radiation pattern model, which is required in order to obtain the radiation value in the considered direction.

Thus, location-based beamforming techniques are advantageous in terms of both the resilience to estimation errors at the user and required overhead in the reporting phase; however, they also require more adjustments, in particular in terms of the procedures due to privacy/RAN operation reasons (as discussed above).

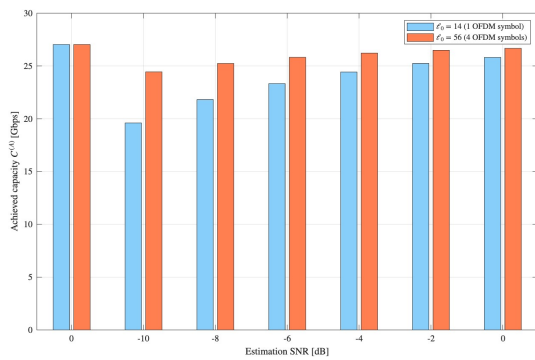


(a) Achieved capacity $C^{(A)}$

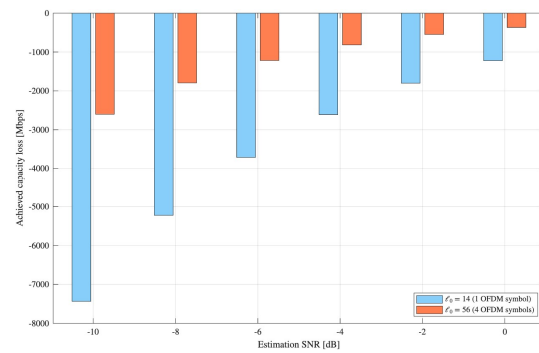


(b) Achieved capacity loss $C^{(A)}$

Figure 85: Performance with non-ideal CSI estimation. Setup: $C_{min} = 5$ Mbps, $C_{max} = 100$ Mbps, PRS with $\psi_{cap} = 1$ and $\rho_{CSI} = 0.05$.

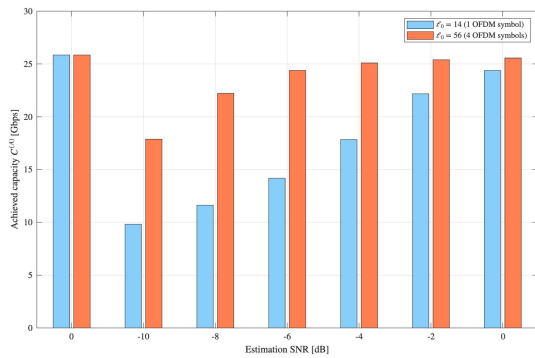


(a) Achieved capacity $C^{(A)}$

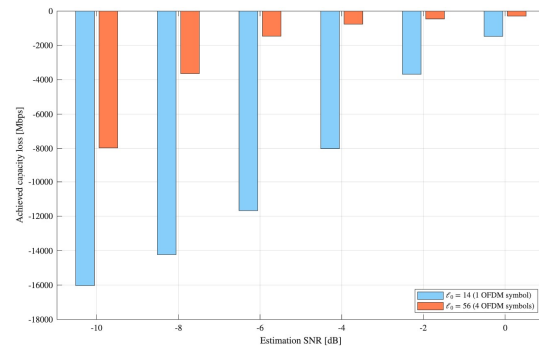


(b) Achieved capacity loss $C^{(A)}$

Figure 86: Performance with non-ideal CSI estimation. Setup: $C_{min} = 5$ Mbps, $C_{max} = 500$ Mbps, PRS with $\psi_{cap} = 0.98$ and $\rho_{CSI} = 0.05$.



(a) Achieved capacity $C^{(A)}$



(b) Achieved capacity loss $C^{(A)}$

Figure 87: Performance with non-ideal CSI estimation. Setup: $C_{min} = 20$ Mbps, $C_{max} = 100$ Mbps, PRS with $\psi_{cap} = 1$ and $\rho_{CSI} = 0.03$.

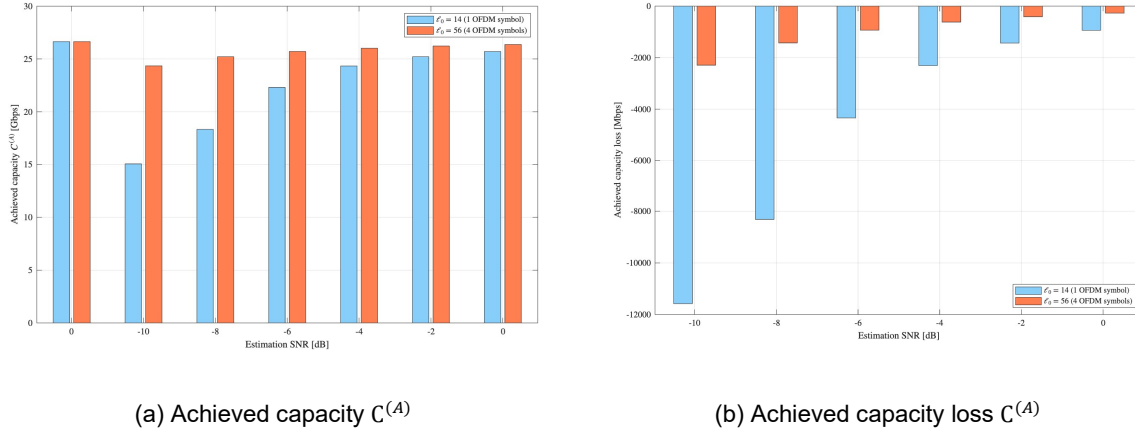


Figure 88: Performance with non-ideal CSI estimation. Setup: $C_{min} = 20$ Mbps, $C_{max} = 500$ Mbps, PRS with $\psi_{cap} = 0.94$ and $\rho_{CSI} = 0.04$.

From Figure 85 to Figure 88, we report the performance of the CSI-based MMSE algorithm with non-ideal CSI estimations at the UE. The performance is evaluated with 1 and 4 OFDM symbols for the estimation via CSI-RS, which leads to 14 and 56 symbols, respectively, assuming a normal CP in the OFDM waveform. It can be noticed that, differently from location errors, the performance loss is significant; the major losses, in the order of 7 Gbps and 18 Gbps with 5 Mbps and 20 Mbps of minimum capacity, respectively, are obtained in low SNR regimes for the estimation and, as expected, with a reduced number of symbols for the data-aided estimation. Increasing the estimation SNR leads to a reduced difference between the use of 1 or 4 OFDM symbols for the estimation, which is in line with equations (38) and (39). When the estimation is performed with 0 dB or more of SNR, the performance loss is almost negligible. Thus, CSI-based techniques are more affected by non-ideal ancillary information compared to location-based solutions; however, this impact can be limited by design, e.g., deploying more satellites with a reduced field of view to boost the SNR, increasing the number of OFDM symbols for the estimation, or developing more advanced estimation techniques.

4.3.1.6 Comparison with newly designed beam lattices

In this Section, we finalise the assessment of the standalone cell-free solution by comparing the performance with the beam layouts designed in Section 3.1. These layouts were designed aiming at grouping the available beams to be illuminated at the same time, in order to minimise the resources (time slots) that are needed to provide service to the entire coverage area.

The objective now is that of understanding if the pre-defined pool of beams defined with this procedure can also be exploited for the implementation of cell-free beamforming solutions. Notably, the techniques assessed in Section 4 assume that either the location or the CSI vector of each user is available and a beam can be created pointing exactly in that direction (i.e., user-centric). When considering the beam layouts defined in Task 4.1, the scenario is slightly different: we have a pre-defined beam lattice and the beams are generated by the NTN node in the directions of the beam centres. Thus, it is not a cell-free solution anymore but rather a beam-centric one: each user is associated to the closest beam centre and the scheduler defines the beams to be illuminated in a given time slot based on the users to be served, i.e., the scheduler provides a mapping from users with unmet capacity to beams to be illuminated.

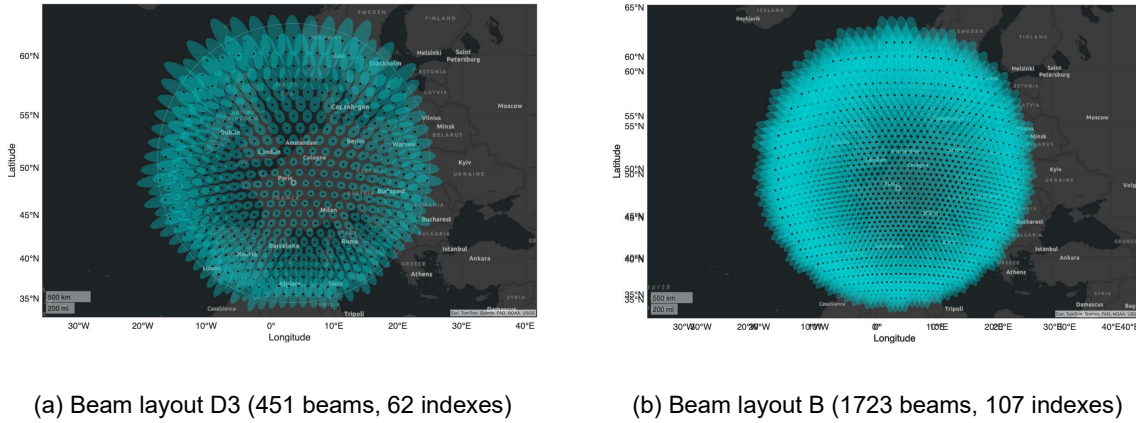


Figure 89: Beam lattices considered in the beam-centric analysis.

Figure 89 shows the two beam lattices that have been considered for this assessment: i) lattice D3, which includes 451 beams and requires 62 indexes for signalling; and ii) lattice B, with 1723 beams and 107 indexes. These layouts were optimised for the inclined orbit at 1300 km altitude and, thus, this is the scenario that is considered in terms of orbital propagation.

With respect to the scheduler, we consider a modified version of the PRS solution extensively discussed and assessed in the previous sections. With respect to the baseline PRS shown in Figure 53, the modifications for the beam-centric PRS (BC-PRS) scheduler shown in Figure 90 consist in: i) not implementing any CSI or geographical scheduling constraint, due to the fact that we have a beam-centric approach now and the beamformer exploits the channel separation implicit in the different directions of the beam centres; ii) while the scheduler still operates considering the user indexes, to monitor the behaviour in terms of achieved and unmet capacity, once the user has been selected the scheduler maps the user to the corresponding beam centre to obtain the pre-computed beamforming coefficients; and iii) rather than barring users based on the CSI or geographical constraint, once a user from beam i has been selected for a time slot, all the remaining users from that i -th beam are excluded from the pool (otherwise we would have two identical columns in the beamformer, leading to an ill-conditioned matrix and a massive performance degradation).

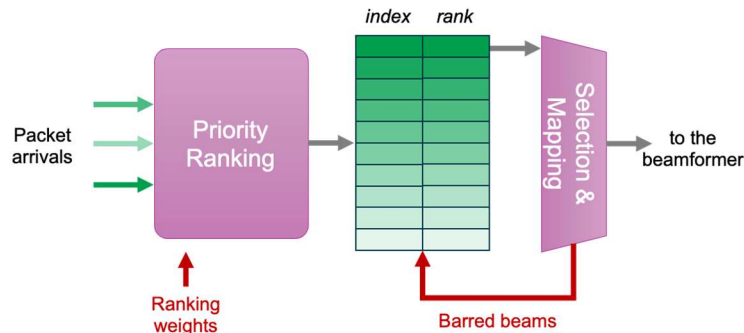


Figure 90: High-level diagram of the beam-centric PRS (BC-PRS).

Based on this architecture, the scheduler first ranks each user based on the weighted metric defined in equation (23), $\chi = \psi_C \tilde{\chi}^{(U)} + \psi_V \tilde{\chi}_{vis}$, where $\tilde{\chi}^{(U)}$ and $\tilde{\chi}_{vis}$ denote the parameters related to the residual unmet capacity and residual visibility time, respectively, and ψ_C , ψ_V denotes the weights to be optimised (as extensively discussed above). Ranking the users from

the largest to the smallest value, at each iteration the top row is selected and then: i) the scheduler identifies the corresponding beam to be illuminated; and ii) deletes all rows referring to co-beam users with respect to the one that was just selected.

Since the BC-PRS scheduler does not consider any geographical or CSI constraint, we implement a fair comparison with the baseline PRS in which no constraint is implemented as well. This means that: i) the geographical scheduling constraint for the i -th user, $\Gamma_{GEO,i}^{(t)}(\mathcal{W})$, in equation (20) becomes $d_{i,j} \geq 0, \forall j \in \mathcal{W}$, which means all users at any distance from the considered one can be still selected; and ii) the CSI scheduling constraint for the i -th user, $\Gamma_{CSI,i}^{(t)}(\mathcal{W})$, in equation (21) becomes $\gamma_{i,j} = \frac{|\hat{\mathbf{h}}_{i,:}^{(t)} \hat{\mathbf{h}}_{j,:}^{(t)}|}{\|\hat{\mathbf{h}}_{i,:}^{(t)}\| \|\hat{\mathbf{h}}_{j,:}^{(t)}\|} \leq 1$, which means that users with any correlation between their CSI vectors can be scheduled together. Notably, this is a worse-condition case for user-centric approaches, which consider the exact user CSI or location to then derive the channel to be used for the computation of the beamforming coefficients; the BC-PRS case considers the beam centres as approximations of the users' location and, thus, implicitly provides a (maybe limited) level of protection from ill-conditioned matrices.

The performance comparison is performed focusing on CBF only, *i.e.*, the beamforming solution based on a pure phase steering. With respect to the PRS scheduling coefficients, we consider a single scenario in which we completely prioritise the capacity, *i.e.*, $\psi_C = 1, \psi_V = 0$. Notably, these coefficients would need an extensive optimisation procedure as shown in the previous sections, which is out of the scope of the comparison reported here.

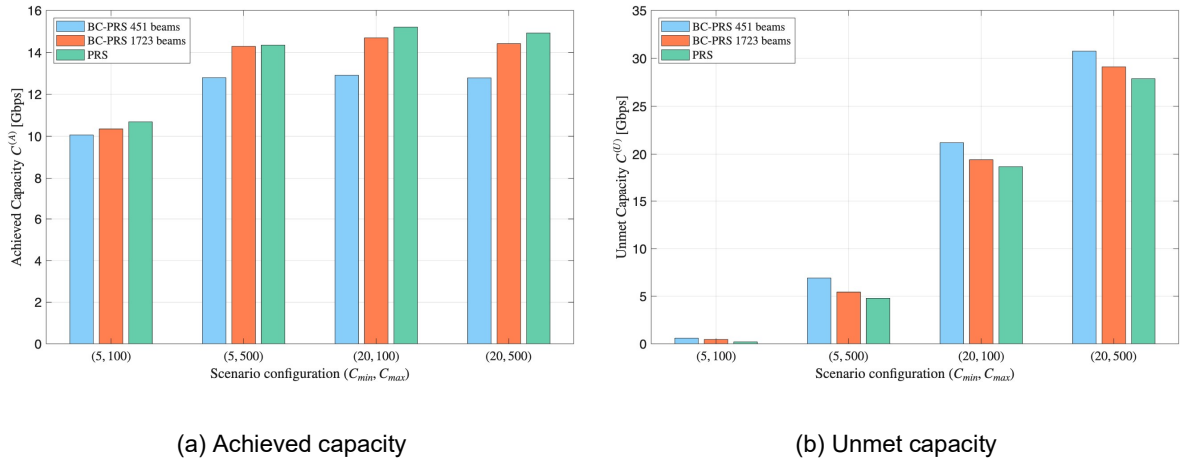


Figure 91: Performance comparison between PRS (user-centric) and BC-PRS (beam-centric).

Figure 91 shows the comparison in terms of achieved and unmet capacity between the beam-centric (BC-PRS) and user-centric (PRS) solutions. It can be noticed that, despite considering a penalising scenario for user-centric solutions (no MMSE beamforming and no scheduling constraints), the performance is still (slightly) better in this case. This is in line with the expectations, since any even minor approximation in the channel coefficients used to compute the beamforming matrix, compared to the actual channel encountered in the transmission phase, is detrimental for the performance. Also as expected, increasing the number of beams leads to a reduced approximation of the channel or position of the users (the beams become smaller) and, thus, the performance with 1723 beams is better compared to 451 beams. The trends are confirmed observing both the achieved capacity (left figure) and the unmet capacity (right figure).

It shall be noticed that, considering the optimised solutions in the previous sections which also took into account the scheduling constraints, the performance is significantly worse (basically half the achieved capacity in some cases). This is indicating that some enhancement in the BC-PRS to also include variants of the scheduling constraints might be beneficial. This can be considered for future developments.

Another development to be considered in the future is the joint optimisation of the beam lattice granularity and the BC-PRS performance. The numerical results suggest (and this is in line also with the scientific literature) that the larger the number of beams with which we are geographically sampling the service area, the better is the approximation of the users' CSI; ultimately, this leads to a better beamforming performance. Thus, the number of beams in the lattice might be not just optimised for signalling purposes but also considering the achieved and unmet capacity that the system is able to produce.

4.3.2 Federated beamforming

In this Section, we focus on the design of a distributed antenna system in space. As introduced in Section 4.1, we consider a swarm of N_{sat} NTN nodes that, by means of a tight cooperation (supported by tight time and frequency synchronisation), allow to create a federated (distributed) user-centric beamforming system in space.

To this aim, before designing the federated beamforming framework, we need to discuss some characteristics of the 5G-STAR DUST constellations to define suitable system configurations and scenarios guaranteeing multiple satellites in simultaneous visibility of the users.

4.3.2.1 Constellation analysis

In 5G-STAR DUST, there are two sub-constellations built according to a Walker Delta design with the parameters listed in Table 28, [2]. To define suitable scenarios in federated beamforming, in the following, we analyse the visibility performance of the two constellations.

To this aim, we focus on a service area with 2025 receives organised on a regular grid centered in (45°N, 5°E), as shown in Figure 92. For both constellations, shown in Figure 93, we consider a full day of orbital propagation starting on June 1, 2024, at 12:00, and a sample time of 1 min.

Table 28: Orbital parameters of the Walker Delta configuration of the 5G-STAR DUST sub-constellations.

Sub-constellation	Parameter					
	Altitude	Inclination	Planes	Satellites per plane	Phasing parameter	Total satellites
Polar	1000 km	99°	5	13	3 (13.85°)	65
Inclined	1300 km	50°	20	11	2 (3.27°)	220

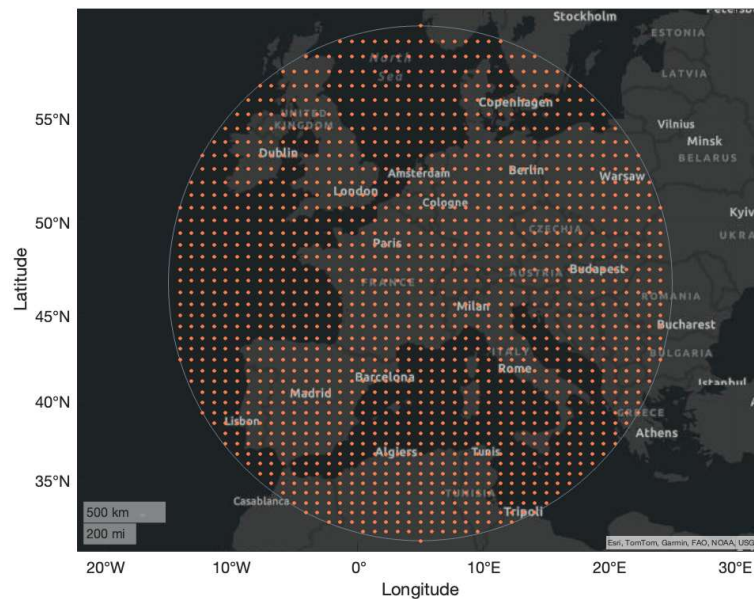
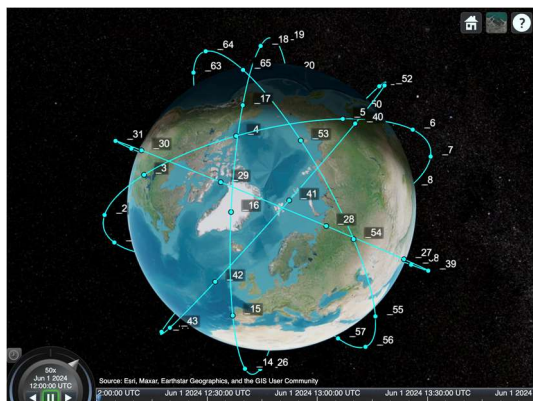
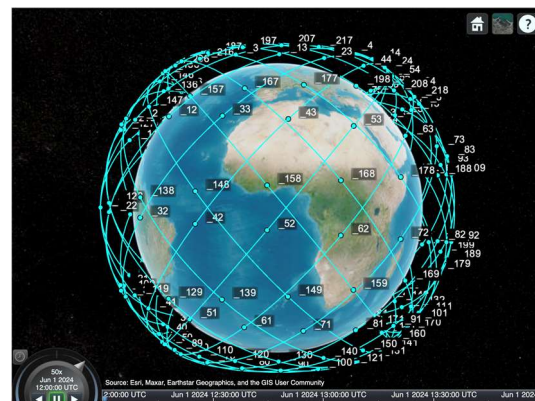


Figure 92: Scenario for constellation visibility assessment.

Figure 94 shows the visibility assessment for the inclined sub-constellation in terms of the average, minimum, and maximum number of visible satellites. This constellation shell is capable of providing continuous coverage in the service area (and, actually, at global level); in fact, the minimum number of visible satellites is never equal to 0: each of the considered grid points has at least one satellite in visibility and up to 8. On average, each grid point is in the visibility of approximately 5 satellites, thus potentially supporting federated beamforming.

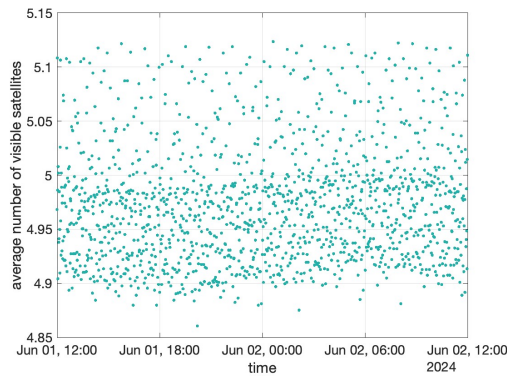


(a) polar sub-constellation

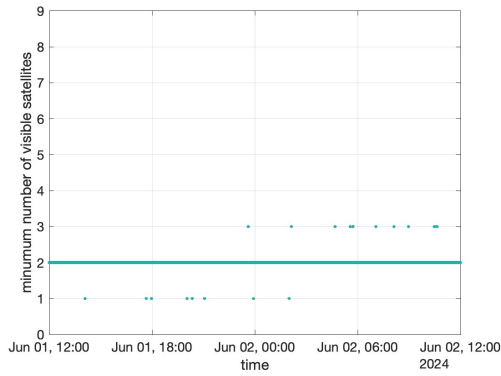


(b) inclined sub-constellation

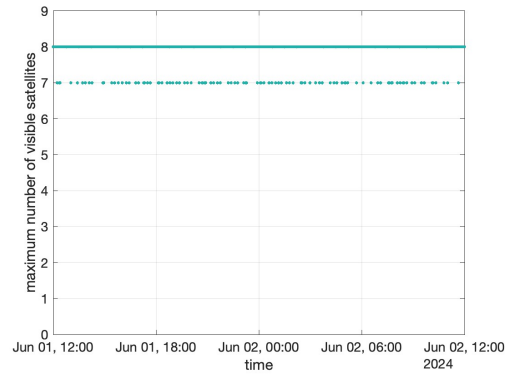
Figure 93: Ground tracks of the two 5G-STARUST sub-constellations.



(a) average number of visible satellites

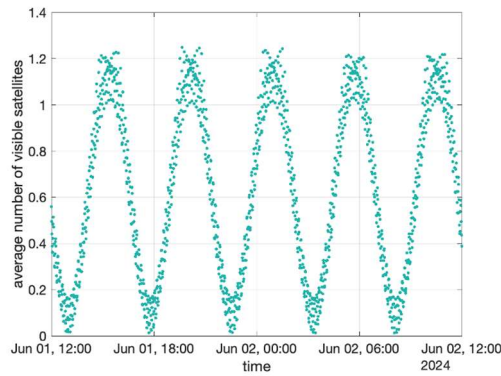


(b) minimum number of visible satellites

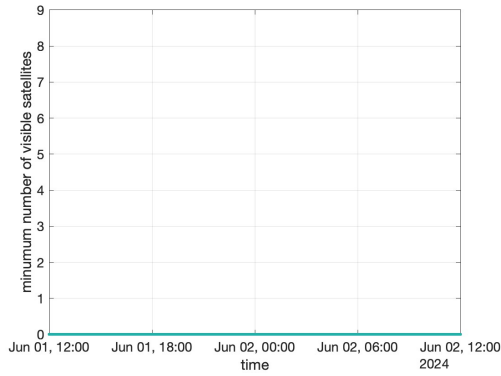


(b) maximum number of visible satellites

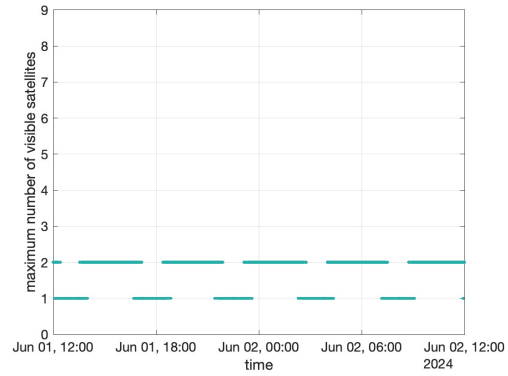
Figure 94: Visibility analysis of the 5G-STARDUST inclined sub-constellation.



(a) average number of visible satellites



(b) minimum number of visible satellites



(b) maximum number of visible satellites

Figure 95: Visibility analysis of the 5G-STARUST polar sub-constellation.

The polar constellation shell, on the other hand, does not provide continuous (and, thus, global) coverage. In fact, the average number of visible satellites is oscillating between 0 and 1.2. Actually, in the considered orbital propagation, 48% of the test points do not see any NTN node. Indeed, observing the ground track in Figure 93, we can see that the limited number of orbital planes does not allow to get continuous global coverage.

Based on these considerations, the focus for the federated beamforming scenario will be on the inclined constellation, as we have to guarantee more than one satellite in visibility.

4.3.2.2 System model and KPIs

We consider $N_{sat} > 1$ NTN nodes organised in a swarm formation; the swarm nodes are all capable of supporting intra-swarm ISLs and equipped with a DRA antenna containing $N_{R,sat} = 512$ radiating elements; the total number of radiating elements in the distributed DRA is thus $N_R = N_{sat} N_{R,sat}$. Also in the federated scenario, we consider Earth-fixed coverage and a service area defined through: i) its center located in (lat_S, lon_S) ; and ii) its extension, defined in terms of the FoV of a theoretical NTN node in the center of mass of the swarm formation and located above the service area center, in order to guarantee an elevation angle of ε_S at the edge of the service area. In the federated scenario, a complete pass of the NTN swarm corresponds to a full pass of all N_{sat} nodes over the service area.

Figure 96 shows an example of the considered service area and swarm ground tracks with: i) $N_{sat} = 2$ NTN nodes in the swarm, from the inclined sub-constellation; ii) a minimum elevation angle of $\varepsilon_S = 30^\circ$ (as per 3GPP); and iii) the center of the service area located as $(lat_S, lon_S) = (48.42^\circ N, 3.68^\circ E)$. The swarm pass is considered to start when the first of the two satellites has a SNP entering into the service area (the satellite with the green ground track in the figure) and it is completed when the last satellite exits the service area (the one with the blue ground track). The total time that is available for service is approximately 9 minutes and 56 seconds.

From a system perspective, the same assumptions and considerations discussed in the standalone case also hold with multiple satellites, with relatively minor adjustments. As outlined in Section 4.1.2, the master NTN node receives the ancillary information from the visible users (eventually supported by the companion satellites in the swarm via ISLs) and exploits this information to compute the scheduling matrix and, from it, the beamforming matrix in each time slot. The RRM algorithms are the same, since they do not explicitly depend on the number of

satellites; thus, the only modifications that we need to introduce are related to the channel coefficients and the computation of the beamforming matrix, as we now have more NTN nodes.

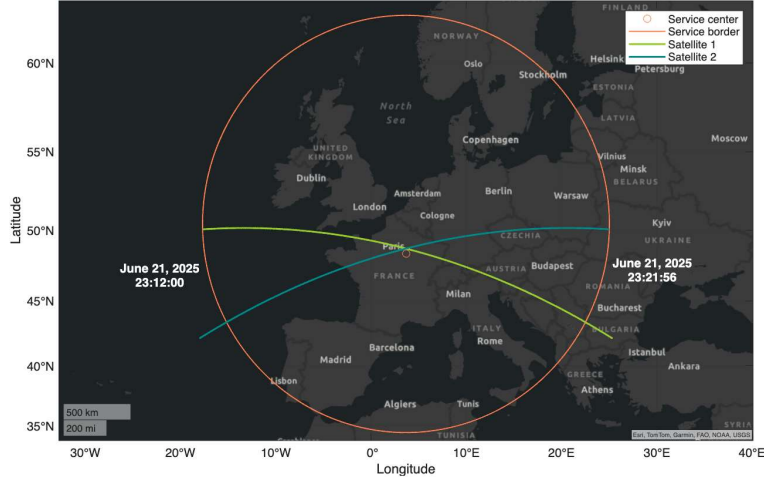


Figure 96: Service area and ground track for federated beamforming with $(lat_s = 48.42^\circ N, lon_s = 3.68^\circ E)$, $\varepsilon_s = 30^\circ$, $h_{sat} = 1300$ km (5G-STARUST inclined constellation, $i = 55^\circ$), $N_{sat} = 2$.

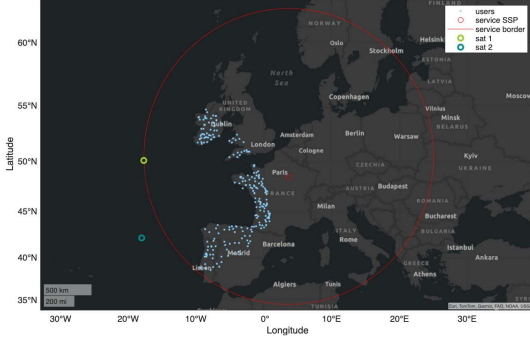
The channel coefficient in a generic time slot t between the i -th on-ground UE and the n -th on-board radiating element of the s -th node is given by:

$$h_{i,n,s}^{(t)} = \frac{g_{i,n,s}^{(tx,t)} g_{i,n,s}^{(rx,t)}}{\frac{4\pi d_{i,s}^{(t)}}{\lambda} \sqrt{L_{i,s}^{(t)} \kappa B_i T_i}} e^{-j \frac{2\pi}{\lambda} d_{i,s}^{(t)}}, i = 1, \dots, N_{UE}, n = 1, \dots, N_{R,sat}, s = 1, \dots, N_{sat} \quad (44)$$

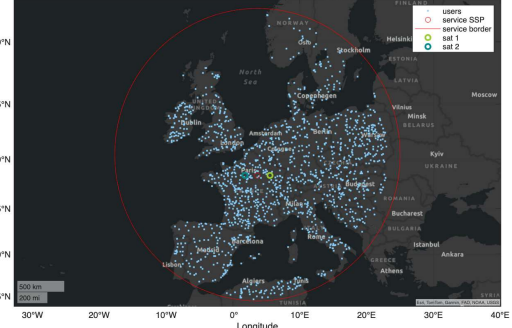
where, compared to the standalone scenario, we can notice that some parameters now depend on the specific node in the swarm, such as the slant range or the radiation pattern, while others still only depend on the user, such as the thermal noise power. With respect to the receiving antenna gain, $g_{i,n,s}^{(rx,t)}$, in the standalone scenario we were assuming an ideal tracking of the single serving satellite; in the federated case, we assume that each user is ideally tracking the best serving satellite (assumed to be the closest one, but other solutions do not affect the generality of the formulation in this document) and, consequently, the signals from the other satellites in the swarm are received with lower radiation pattern values compared to the serving one (with unitary radiation pattern by definition).

When we implement CSI-based techniques, the UEs directly estimate the channel coefficients in eq. (44) as $\hat{h}_{i,n,s}^{(t_R)}$ and provide it to the master NTN node (directly or indirectly) to compute the scheduling matrix and the beamforming coefficients; it is worthwhile mentioning that estimating the CSI coefficients from satellites seen at lower antenna gains is more challenging compared to the reference satellite (or to the standalone scenario). Moreover, as previously discussed, the number of overhead risks to rapidly become excessive. Location-based solutions are not affected by these aspects: the UEs estimate their location, independently from the number of satellites, and the master NTN node derives the channel coefficients based on this information and the current NTN nodes position on their orbit, which must be known on-board. In this case, the estimated channel coefficient is:

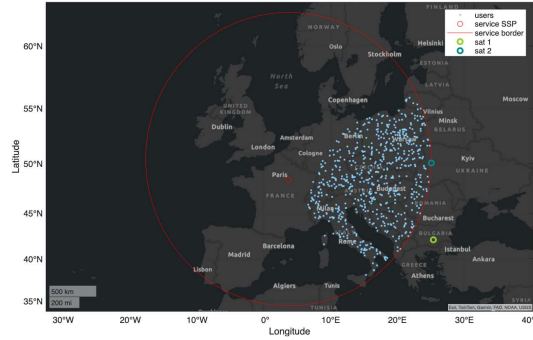
$$\hat{h}_{i,n,s}^{(t_R)} = \frac{\hat{g}_{i,n,s}^{(tx,t_R)} \hat{g}_{i,n,s}^{(rx,t_R)}}{4\pi \hat{d}_{i,s}^{(t_R)} \sqrt{\kappa B_i T_i}} e^{-j \frac{2\pi}{\lambda} \hat{d}_{i,s}^{(t_R)}} \quad (45)$$



(a) Visible users at 23:12:00



(b) Visible users at 23:16:58



(c) Visible users at 23:21:56

Figure 97: Example of visible users according to the NTN nodes position on their orbit with $(lat_s = 48.42^\circ N, lon_s = 3.68^\circ E)$, $\varepsilon_s = 30^\circ$, $h_{sat} = 1300$ km (inclined constellation, $i = 55^\circ$), $N_{sat} = 2$.

In this framework, since we are focusing on federated solutions, it shall be noticed that we consider a user in visibility of the swarm when it is in the **visibility of all satellites of the swarm**. In a real system, beamforming might be implemented combining distributed and standalone approaches according to the actual user visibility in the service area. However, since we are focusing on the performance with multiple satellites, we do not consider this scenario since it would introduce further unnecessary complexities and potential undesired biases in the numerical results. The number of visible users at the generic report instant is:

$$N_{UE}^{(t_R)} = \left| \left\{ i: \varepsilon_{i,s}^{(t_R)} \geq \varepsilon_t, i = 1, \dots, N_{UE}, \forall s = 1, \dots, N_{sat} \right\} \right| \quad (46)$$

where $\varepsilon_{i,s}^{(t_R)}$ denotes the elevation angle of the s -th satellite as seen by the i -th user. An example of visible users depending on the NTN nodes' position is shown in Figure 145 at the beginning, middle, and end of the swarm pass on the service area.

Based on the RRM algorithms (e.g., those designed in Section 4.3.1.2), in the generic t -th transmission time slot the master NTN node computes the beamforming coefficients by: i) selecting the N_B -dimensional t -th column of the scheduling matrix $\mathbf{S}$, $\mathbf{s}_{:,t}$; and ii) creating the

estimated channel matrix at time t , $\hat{\mathbf{H}}^{(t)}$, from the available ancillary information. This estimated channel matrix is now the concatenation of the channel matrices obtained between each visible user and each radiating element on-board the satellites in the swarm. Let us denote by $\hat{\mathbf{h}}_{s_b,t,:s}^{(t)} = [\hat{h}_{s_b,t,1,s}^{(t)}, \dots, \hat{h}_{s_b,t,n_s}^{(t)}, \dots, \hat{h}_{s_b,t,N_{R,s}}^{(t)}]$ the complex channel coefficients between the user scheduled for the b -th beam in the t -th time slot, provided by $s_{b,t}$, and the $N_{R,sat}$ radiating elements on-board the s -th swarm node. From this, we can introduce the channel matrix between the on-ground scheduled users and the s -th swarm node as:

$$\hat{\mathbf{H}}_s^{(t)} = \left[\left(\hat{\mathbf{h}}_{s_1,t,:s}^{(t)} \right)^T, \dots, \left(\hat{\mathbf{h}}_{s_{N_{sat}},t,:s}^{(t)} \right)^T, \left(\hat{\mathbf{h}}_{s_{N_{sat}},t,:s}^{(t)} \right)^T \right]^T \quad (47)$$

The global channel matrix built from all swarm nodes is then obtained as a concatenation:

$$\hat{\mathbf{H}}^{(t)} = \left[\hat{\mathbf{H}}_1^{(t)}, \dots, \hat{\mathbf{H}}_s^{(t)}, \dots, \hat{\mathbf{H}}_{N_{sat}}^{(t)} \right] \quad (48)$$

This is the channel matrix that is used to compute the beamforming matrix based on either the MMSE or the CBF algorithm, as in the standalone case: $\mathbf{B}^{(t)} = \mathcal{B}(\hat{\mathbf{H}}^{(t)})$. The beamforming algorithms do not require modifications; the only difference is that, starting from a channel matrix with dimension $N_B \times (N_{sat}N_{R,sat})$, we obtain a $(N_{sat}N_{R,sat}) \times N_B$ beamforming matrix.

The power distribution strategy, on the other hand, needs some clarifications. In fact, the beamforming matrix is now leading to a transmission power that includes the power emitted by the radiating elements on-board multiple nodes. Consequently, the power distribution strategy shall be adjusted to satisfy the power constraints per node, not at global swarm level. As also discussed in [19], [20], [21], swarm-based power distribution strategies can be defined by observing that, as a consequence of how the channel matrix in (37) is obtained, the beamforming matrix $\mathbf{B}^{(t)}$ is given by:

$$\mathbf{B}^{(t)} = \begin{bmatrix} \mathbf{B}_1^{(t)} \\ \vdots \\ \mathbf{B}_s^{(t)} \\ \vdots \\ \mathbf{B}_{N_{sat}}^{(t)} \end{bmatrix} \quad (49)$$

Based on this concatenation, swarm SPC (sSPC) and swarm MPC (sMPC) distributions can be defined by operating at beamforming sub-matrix level. The PAC strategy, by definition, does not require any modification; in fact, the power distribution approach is based on the normalisation of the power emitted by each radiating element to the same amount, $\frac{P_{av}}{N_R}$, with P_{av} being the total available power on-board the NTN node. The generic row in the beamforming matrix $\mathbf{B}^{(t)}$, or in any of the sub-matrices constituting it, is associated to the corresponding radiating element. Thus, operating on each single row of $\mathbf{B}^{(t)}$ implicitly makes the PAC distribution strategy compatible with swarm operations. In the following, we denote as $\tilde{\mathbf{B}}^{(t)}$ the normalised beamforming matrix according to any of these strategies.

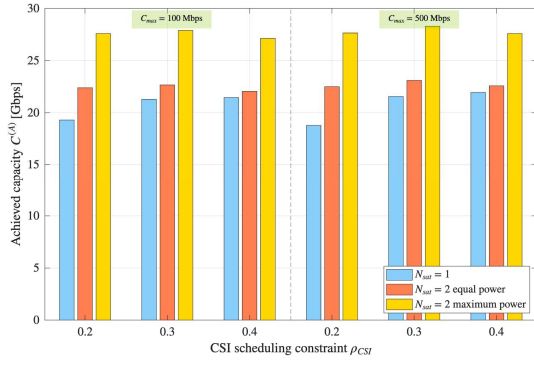
Once the normalised beamforming matrix has been obtained, equations from (6) to (19) providing the received symbols, the asymptotic spectral efficiency, and all other KPIs, still hold without any modification.

4.3.2.3 Performance assessment with ideal ancillary information

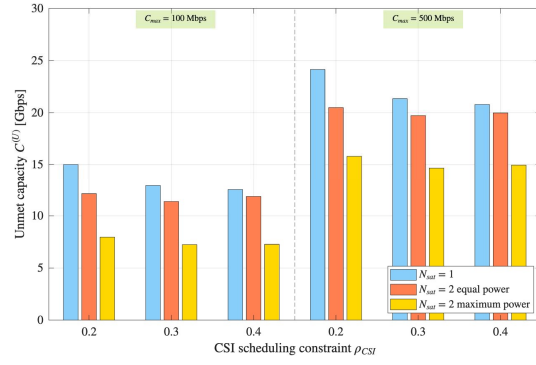
In the context of federated beamforming, it is worthwhile discussing the total amount of power that is used to serve the users. In fact, if we assume that each NTN node in the swarm transmits its total available power, P_{av} , the total amount of power that is exploited in federated scenarios is $N_{sat}P_{av}$; to compare the performance with the standalone beamforming case, we consider two cases: i) an **equal power scenario**, in which the power on-board each NTN node in the swarm is scaled by N_{sat} , so that the total power from the swarm is P_{av} as in the standalone scenario; and ii) a **maximum power scenario**, in which each node in the swarm transmits at maximum power, leading to an additional power gain compared to the standalone scenario.

Table 29: Simulation parameters for the standalone case with ideal ancillary information.

Parameter	Value	Comments
Sub-Carrier Spacing (SCS)	120 kHz	In this case, 132 PRBs are available
Operating Frequency Range	FR2 (Ka-band)	See 3GPP TR 38.821
Carrier frequency	20 GHz	See 3GPP TR 38.821
User bandwidth B	190.08 MHz	$132 \cdot 120 \cdot 12$ kHz
Scenario	Digital Divide	Parameters in Table 14 of Section 4.2.1.2
Minimum capacity request, C_{min}	5,20 Mbps	See Section 4.2.1.2
Maximum capacity request, C_{max}	100,500 Mbps	See Section 4.2.1.2
Number of radiating elements $N_{R,sat}$	512	See Section 2.3.1 of [1]
Number of satellites N_{sat}	2	
Power per radiating element $P_{av,el}$	65 mW	See Section 2.3.1 of [1]
Total power on-board P_{av}	15.2218 dBW	$10\log_{10}(0.0065 \cdot 512)$
Total power per swarm	$P_{av}, 2P_{av}$	
CSI constraint coefficient ρ_{CSI}	0.01, ..., 0.04	See Section 4.3.1.2
Minimum user elevation angle ε_t	30°	See Section 4.2.3 of [1]
Service area elevation angle ε_s	30°	See Section 4.2.3 of [1]
Channel type	clear-sky	See Section 2.4.2.1 of [1] and 3GPP TR 38.811
Reporting periodicity T_{rep}	80 ms (8 slots)	See Section 4.3.1.1
Transmission time slot T_{slot}	10 ms	See Section 4.3.1.1
NTN node altitude	1300 km	See Section 2.1 of [1]
Service area centre	(48.42°N, 3.68°E)	See Section 4.2.3 of [1]
UE terminal	VSAT	See Section 2.3.2 of [1] and 3GPP TR 38.811
Scheduler	PRS	See Section 4.3.1.2
Unmet capacity weight ψ_{cap}	0.98, 1	See Section 4.3.1.2

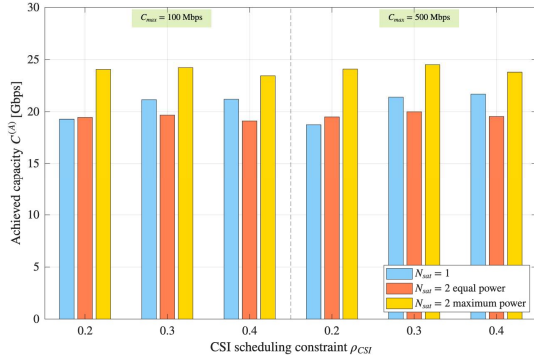


(a) Achieved capacity $C^{(A)}$

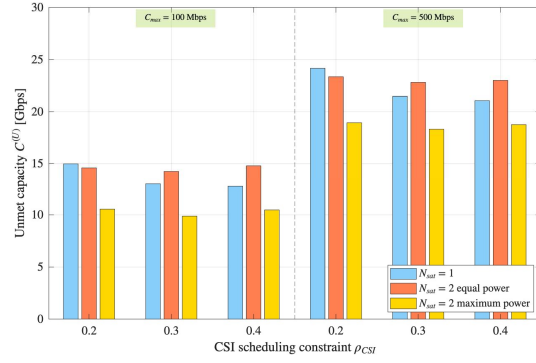


(b) Unmet capacity $C^{(U)}$

Figure 98: Achieved and unmet capacity with federated (LB-)MMSE beamforming, PRS with $\psi_{cap} = 1$, $C_{min} = 20$ Mbps.



(a) Achieved capacity $C^{(A)}$



(b) Unmet capacity $C^{(U)}$

Figure 99: Achieved and unmet capacity with federated CBF beamforming, PRS with $\psi_{cap} = 1$, $C_{min} = 20$ Mbps.

The configuration for the federated beamforming assessment is reported in Table 29. It shall be noticed that we focus on the PRS scheduler since it is the best performing solution. Since in this study the main objective is that of analysing the performance trends, rather than searching the optimal configuration, we then restrict the case studies to an unmet capacity weight ψ_{CSI} of 0.98 and 1. With respect to the traffic distribution, we mainly focus on heavy traffic scenarios, i.e., $C_{min} = 20$ Mbps and $C_{max} = 100,500$ Mbps, and then also observe the performance in a limited traffic case, $C_{min} = 5$ Mbps and $C_{max} = 100$ Mbps.

Figure 98 and Figure 99 show the achieved and unmet capacity with federated beamforming considering both the equal power and maximum power conditions as described above, compared to the case with a single satellite (which was not previously considered for the inclined constellation at 1300 km). The trends in terms of CSI scheduling constraint and traffic intensity are basically the same as those observed for the standalone scenario. Thus, focusing on the federated solution, we can observe that:

- The performance with the maximum power, i.e., exploiting all the available power on-board all the NTN nodes in the swarm, provides the best performance: the achieved

capacity is much larger, and the unmet capacity is significantly reduced in all combinations of traffic or scheduler configuration. While this might have been expected, it is important to remark this point since it is in line with the discussion of high traffic scenarios in standalone configurations; in fact, it was already observed previously that a potential solution to the performance loss with high traffic demands could have been the increase of the available power on-board. Here, we can see that indeed increasing the available power results in several Gbps gains; in particular, the unmet capacity is still not negligible but massively reduced compared to the standalone case.

- With (LB-)MMSE, the performance with a single satellite is always the poorest, in both the achieved and unmet capacity. As for CBF, we can see that, almost always, the performance of the standalone configuration is better than the federated scenario with equal power. The CBF is a pure digital beam steering technique, which is not optimising the beamforming coefficients (the amplitudes in particular) to maximise the intended power while limiting the interfering one as MMSE-based algorithms do. As a consequence, it is more subject to poor SIR conditions; when the power from the best serving satellite is limited, as in equal power configurations, the SIR further drops due to the additional uncoordinated interference that is introduced by the additional satellites. The optimisation implicit in MMSE-based solutions is thus fundamental for the performance of federated beamforming.

Figure 100 and Figure 101 show the performance results with $\psi_{\text{cap}} = 0.98$, while Figure 102 and Figure 103 provide an example in low traffic conditions. The same considerations reported above still apply to this scenario, also considering that the performance is almost the same (the differences are in the order of a few tens of Mbps).

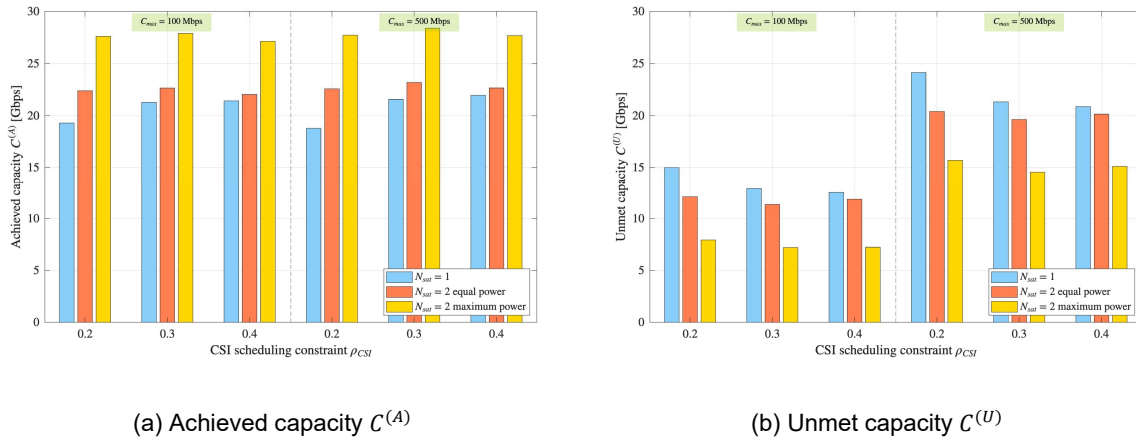
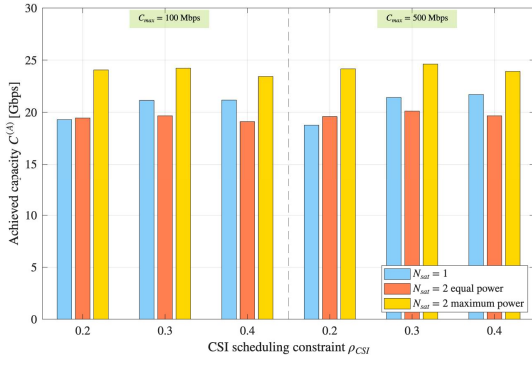
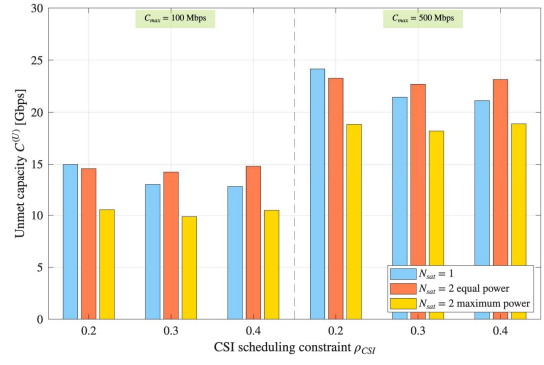


Figure 100: Achieved and unmet capacity with federated (LB-)MMSE beamforming, PRS with $\psi_{\text{cap}} = 0.98$.

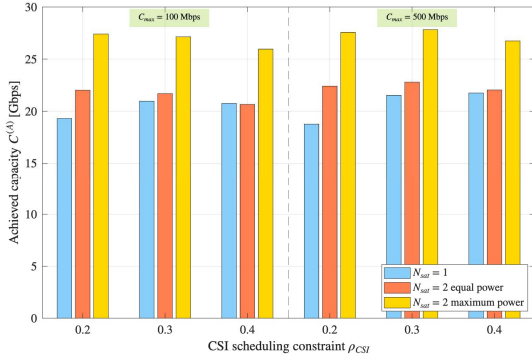


(a) Achieved capacity $C^{(A)}$

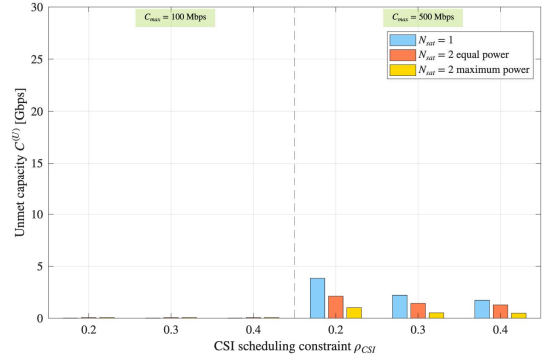


(b) Unmet capacity $C^{(U)}$

Figure 101: Achieved and unmet capacity with federated CBF beamforming, PRS with $\psi_{cap} = 0.98$, $C_{min} = 20$ Mbps.

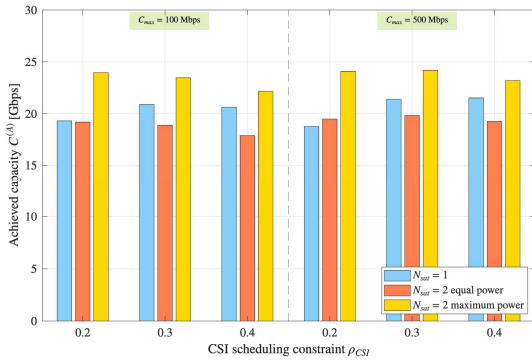


(a) Achieved capacity $C^{(A)}$

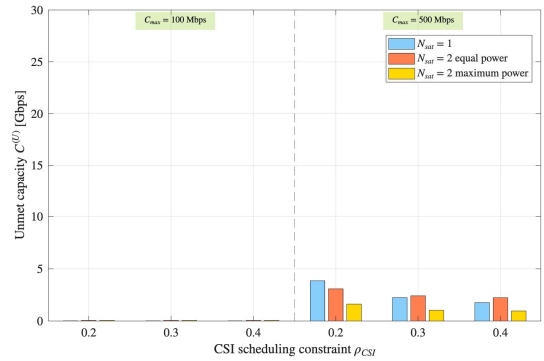


(b) Unmet capacity $C^{(U)}$

Figure 102: Achieved and unmet capacity with federated (LB-)MMSE beamforming, PRS with $\psi_{cap} = 1$, $C_{min} = 5$ Mbps.



(a) Achieved capacity $C^{(A)}$



(b) Unmet capacity $C^{(U)}$

Figure 103: Achieved and unmet capacity with federated CBF beamforming, PRS with $\psi_{cap} = 1$, $C_{min} = 5$ Mbps.

4.4 LINK-LEVEL ANALYSES

In order to compare the performance of the standard 5G NR OFDM waveform with the OTFS waveform described in [32], proper system adaptations were introduced in the NR radio frame so as to fit the delay-Doppler OTFS modulation into the slot architecture of the standard. Different choices are possible, as briefly described in the following, and one was chosen due to complexity and flexibility issues. Then, simulations were run after the identification of a feasible set of system parameters, to provide a meaningful comparison and to outline the possibility to exploit the advantages offered by the cell-free paradigm, when more than one transmitting satellite is employed. In this light, results were drawn by developing a scenario with two transmitting satellites.

4.4.1 OTFS system description

In the following a brief recap of the NR time-frequency radio grid will be provided. The 5G NR downlink physical shared channel (PDSCH) is based on a 10 ms frame which is split into 10 subframes; each subframe is then split into a variable number of slots, depending on the numerology μ , and each slot contains 14 OFDM symbols, as explained in Table 30.

Table 30: Symbols, slots and subframes per frame.

μ	$N_{\text{symbolslot}}$	$N_{\text{slotframe},\mu}$	$N_{\text{slotsubframe},\mu}$
0	14	10	1
1	14	20	2
2	14	40	4
3	14	80	8
4	14	160	16
5	14	320	32
6	14	640	64

The numerology chosen for the simulations was $\mu = 3$, that corresponds to a subcarrier spacing of 120 kHz, since in Frequency Range (FR) 2 of the standard (carriers beyond 6 GHz) only numerologies above $\mu = 3$ are allowed. Thus, the baseline system envisages 80 slots per frame. The OFDM symbol and cyclic prefix durations in this case are, respectively, $8.33\mu\text{s}$ and $0.57\mu\text{s}$. In terms of frequency, depending on the number of transmitted resource blocks (RB), each including 12 resource elements (RE), a bandwidth range from about 50 MHz (32 RB) up to 400 MHz (264 RB) is possible.

The OTFS waveform is instead based on a $M \times N$ delay-Doppler grid, where M is the number of subcarriers and N the number of symbols (time slots). Such a grid can be fitted in the NR frame in many different ways, depending on the complexity that can be afforded, and on the resolution that is required by the system in terms of delay and Doppler, as the delay resolution is given by $\frac{1}{M\Delta f}$ and the Doppler resolution by $\frac{1}{NT}$, where $\Delta f = 1/T$ is the subcarrier spacing and T is the sampling time.

The possible grid size choices that were identified as a potential solution were:

- Full band (FB): same number of subcarriers as the equivalent NR OFDM waveform, and a few symbols as included in a single slot, i.e., 14 symbols.
- Sub-band over a slot (SB): the bandwidth occupied by the NR OFDM waveform is split among different OTFS symbols, i.e., the number of subcarriers of the OTFS symbols is a fraction of the subcarriers of the OFDM symbol, but the number of time symbols is again equal to the symbols in a slot, that is 14.
- Sub-band over multiple slots (xSyS): here the number of subcarriers of the OFDM waveform is again split among different OTFS symbols (defined by the "x"), but the number of time symbols is larger as multiple slots (defined by the "y") can be included in a single OTFS symbol.

The first solution has the advantage of offering a direct mapping of the OFDM waveform into the OTFS waveform; nevertheless, the complexity becomes rapidly unmanageable, as M can take values as large as 4096, whereas N is only equal to 14, which means a poor Doppler resolution. The second solution is more flexible than the first one, as the bandwidth can be split among several OTFS symbols, and the mapping is still quite simple (one NR slot can be mapped directly to the OTFS symbol), but it also still suffers from the poor Doppler resolution. Therefore, the latest solution, namely SMS, can be easily identified as the more flexible and practical, even though more complex to be implemented; the overall bandwidth is split among several OTFS symbols, and the N grid parameter can be chosen as a fraction of the total number of slots per frame, according to the required Doppler resolution. Figure 104, Figure 105 show the comparison of the time-frequency grid of an OFDM waveform over a frame, with 512 subcarriers (288 of which loaded with data symbols) and a corresponding 8S4S OTFS waveform where the bandwidth is therefore split among 8 OTFS symbols and the time dimension includes symbols from 4 slots. Since in this example the numerology μ is equal to 3, the OTFS grid entails $N = 56$. Thus, the NR frame includes 80 OFDM symbols with FFT size equal to 512, and due to the presence of several virtual subcarriers, the occupied bandwidth is approximately 42 MHz (24 RB where set in this example). On the other hand, the NR frame includes 160 OTFS symbols, that occupy the full bandwidth equal to 62 MHz but a shorter time (7 ms per frame vs 10 ms of the NR OFDM frame).

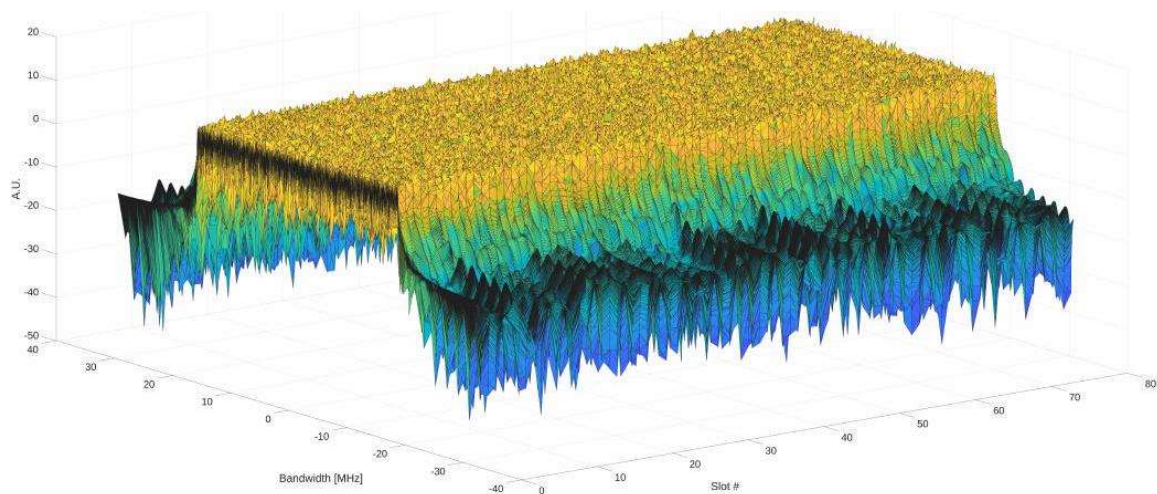


Figure 104: OFDM time frequency grid of one NR frame.

The choice of splitting the bandwidth among several OTFS symbols entails the design of a proper digital FIR filter to transmit the symbols reducing the interference. This filter was therefore designed by exploiting the Parse-McClellan algorithm, where the filter order was set to 202 to obtain sharp edges. The cutoff frequency clearly depends on the number of established sub-bands, and in case of 8 as in the example of Figure 105 it was set around one tenth of the overall signal bandwidth. Figure 106 shows magnitude and phase of such filter.

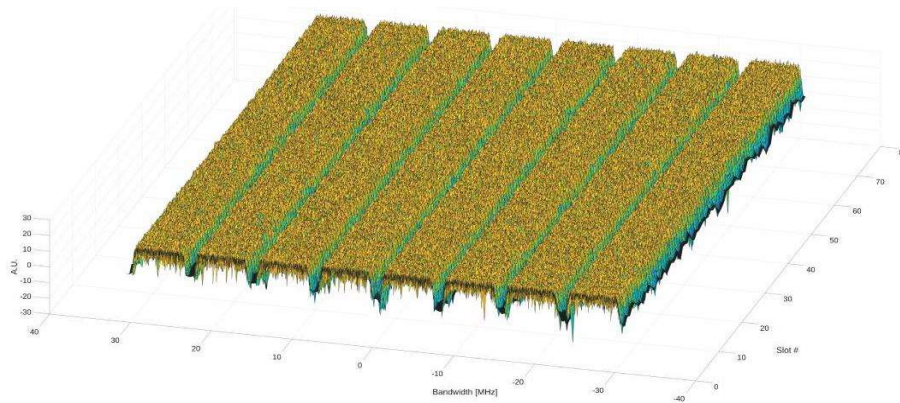


Figure 105: OTFS time frequency grid of one NR frame.

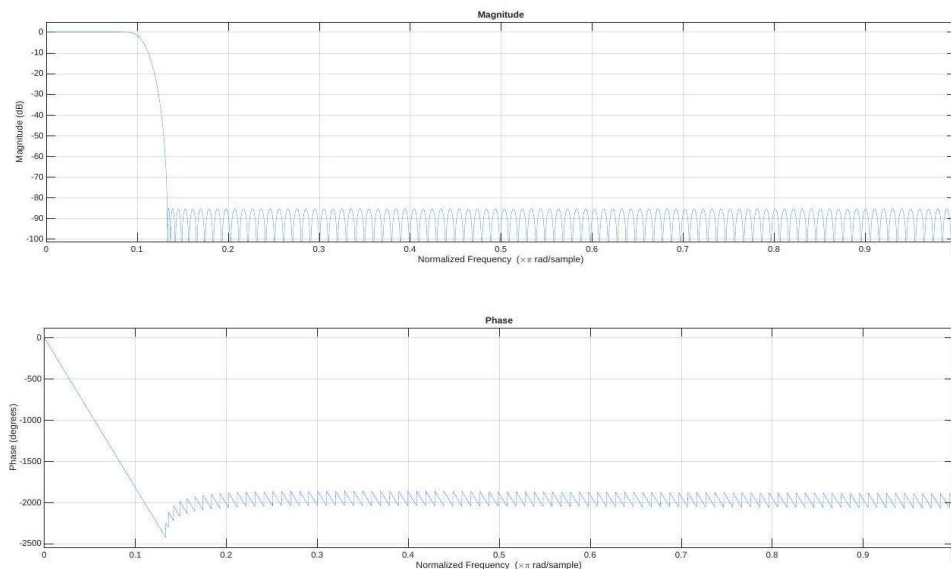


Figure 106: OTFS waveform FIR filter magnitude and phase.

As concerns channel estimation, pilot symbols included in the OFDM waveform (the demodulation reference signal - DMRS) are also clearly configured differently in the OTFS waveform. In the OFDM case these symbols are scattered inside the time-frequency (TF) grid following a few different configurations which are selected by the system according to the scenario at hand. The pilot configuration of the OTFS case is completely different, as outlined in [32]; the delay-Doppler signal representation includes a rectangular region of empty symbols where a single central symbol is transmitted, collecting all the average power of the symbols not transmitted in the empty region.

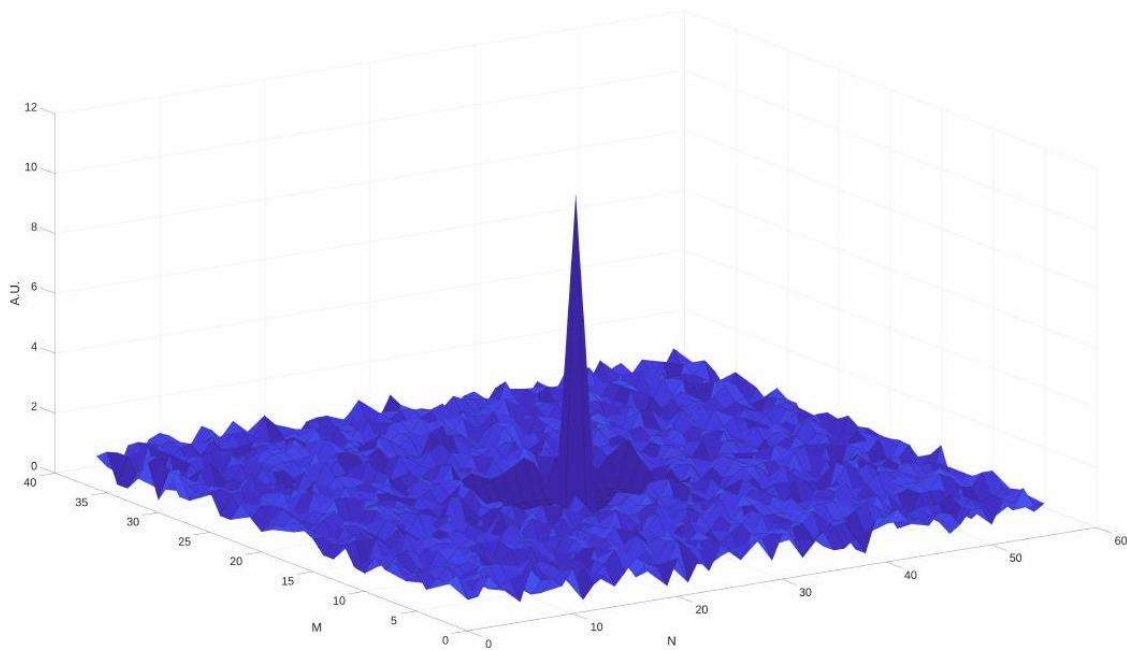


Figure 107: OTFS transmitted symbol, with pilot for channel estimation.

An example is reported in Figure 107. The receiver can estimate the delay and Doppler associated to each scattered signal replica by measuring the distance of the received peak(s) from the expected position (see [32] for a detailed explanation). The size of the empty region, and hence the pilot overhead, can be clearly fixed according to the maximum delay and Doppler expected in the scenario at hand.

4.4.2 OTFS modulation/demodulation via Discrete Zak transform

Two possible implementations of OTFS are possible: one using the Discrete Zak Transform (DZT) and the other using the inverse symplectic finite Fourier transform (ISFFT) [33]. The first one was briefly introduced in [32], whereas the approach based on the DZT, chosen for the implemented system, will be presented in the following, by highlighting the main differences with respect to the ISFFT implementation.

Both SFFT- and DZT-based OTFS implementations are suitable for high-mobility communications. The SFFT-based OTFS architecture is compatible with conventional OFDM systems by concatenating an ISFFT/SFFT operation with an OFDM modulator/demodulator. On the other hand, DZT-based OTFS offers lower computational complexity, requiring only an N -point inverse finite Fourier transform (IFFT) and finite Fourier transform (FFT), along with symbol-wise interleaving and de-interleaving at the OTFS transmitter and receiver. Furthermore, the SFFT-based OTFS design supports efficient TF domain signal processing, enabling improvements such as window function optimization [34] and the development of multiple access schemes [35]. The Input-Output relation for the SFFT implementation, useful for system design, will be also discussed, by presenting the matrix form representation.

4.4.2.1 SFFT implementation

The SFFT-based OTFS transceiver architecture is illustrated in Figure 108. The available bandwidth $B_{\text{OTFS}} = M/T$ and time duration $T_{\text{OTFS}} = NT$, defined in the TF domain, are used to accommodate one OTFS frame. The bandwidth B_{OTFS} is divided into M subcarriers, each

with subcarrier spacing $\Delta f = \frac{B_{\text{OTFS}}}{M} = \frac{1}{T}$. Similarly, the time duration T_{OTFS} is divided into N time slots, each with slot duration $\frac{T_{\text{OTFS}}}{N} = T$.

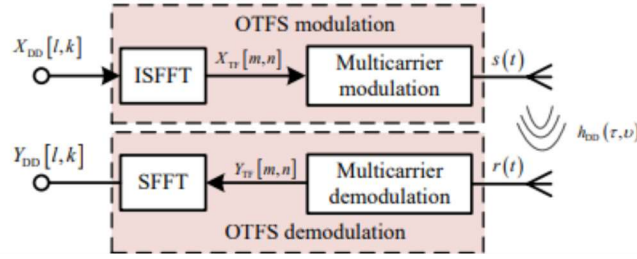


Figure 108: Schematic of the SFFT implementation of the OTFS system.

At the transmitter side, the information symbol for the (l, k) -th delay-Doppler Resource Element is represented as $X_{\text{DD}}[l, k]$, where $l \in \{0, \dots, M-1\}$ is the delay index and $k \in \{0, \dots, N-1\}$ is the Doppler index. The DD domain symbols $X_{\text{DD}}[l, k]$ are then transformed into the TF domain signal $X_{\text{TF}}[m, n]$ using the ISFFT, which is reported below.

$$X_{\text{TF}}[m, n] = \frac{1}{\sqrt{NM}} \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} X_{\text{DD}}[l, k] e^{j2\pi(\frac{nk}{N} - \frac{ml}{M})} \quad (50)$$

here $m \in \{0, \dots, M-1\}$ is the subcarrier index and $n \in \{0, \dots, N-1\}$ is the slot index. The OTFS transmit signal $s(t)$ is obtained by performing multicarrier modulation on the TF domain signal $X_{\text{TF}}[m, n]$, also known as the Heisenberg transform [36], i.e.,

$$s(t) = \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} X_{\text{TF}}[m, n] g_{\text{tx}}(t - nT) e^{j2\pi m \Delta f (t - nT)} \quad (51)$$

where $g_{\text{tx}}(t)$ is the transmit pulse shaping filter. In practice, the TF domain signal $X_{\text{TF}}[m, n]$ is first transformed to a time domain sequence via the inverse finite Fourier transform (IFFT), is then pulse-shaped in the discrete time domain, and is finally transformed to the analog TX signal $s(t)$ by digital-to-analog conversion, as shown in Figure 109.

As already mentioned in [32], the channel is represented using a discretization, where the two-dimensional impulse response is characterized by the spreading function, which is provided below:

$$\gamma(\nu, \tau) = \sum_{p=1}^P h_p \delta(\tau - \tau_p) \delta(\nu - \nu_p) \quad (52)$$

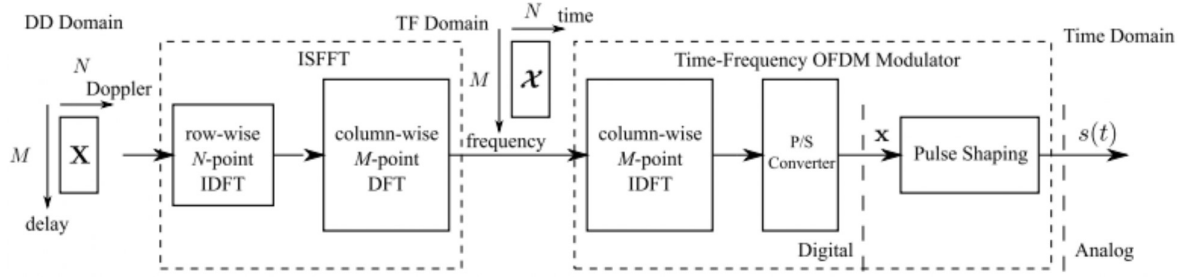


Figure 109: Schematic of the OTFS transmitter.

where h_p represents the channel gain for the p -th path, τ_p is the delay, and ν_p is the Doppler shift associated with the p -th path. Recall that $P \in \mathbb{Z}$ represents the number of paths, and the delays τ_p and Doppler shifts ν_p are bounded. Specifically, the delays satisfy

$$\tau_p \in [0, \tau_{\max}],$$

and the Doppler shifts lie within the range

$$\nu_p \in [-\nu_{\max}, \nu_{\max}].$$

After passing through the linear time-varying (LTV) channel, the received signal at the OTFS receiver is given by:

$$r(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \gamma(\nu, \tau) e^{j2\pi\nu(t-\tau)} s(t-\tau) d\tau d\nu + w(t) \quad (53)$$

which can be expressed as:

$$r(t) = \sum_{i=1}^P h_i s(t - \tau_i) e^{j2\pi\nu_i(t-\tau_i)} + w(t), \quad (54)$$

where $w(t)$ denotes the additive white Gaussian noise (AWGN) in the time domain. At the OTFS receiver, after receive filtering, multicarrier demodulation of $r(t)$ is performed to obtain the TF domain received signal. This operation, also known as the Wigner transform [36], is given by:

$$Y_{\text{TF}}[m, n] = \int_{-\infty}^{\infty} r(t) g_{\text{rx}}^*(t - nT) e^{-j2\pi m \Delta f (t - nT)} dt \quad (55)$$

where $g_{\text{rx}}(t)$ is the receive filter. Subsequently, the TF domain received signal is transformed into the DD domain using the SFFT. The resulting DD domain received signal is:

$$Y_{\text{DD}}[l, k] = \frac{1}{\sqrt{NM}} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} Y_{\text{TF}}[m, n] e^{-j2\pi \left(\frac{kn}{N} - \frac{lm}{M} \right)}. \quad (56)$$

Now, we can combine the previously derived formulas to obtain the Input-Output relation. The resulting expression is as follows:

$$Y_{DD}[l, k] = \sum_{k'=0}^{N-1} \sum_{l'=0}^{M-1} X_{DD}[l', k'] H_{DD}[l, k, l', k'] + W_{DD}[l, k], \quad (57)$$

where $W_{DD}[l, k]$ denotes the effective DD domain noise. The discrete DD domain channel $H_{DD}[l, k, l', k']$ is given by:

$$H_{DD}[l, k, l', k'] = \sum_{i=1}^P h_i g_h(l, k, l', k', l_i, k_i) e^{-j2\pi v_i \tau_i}, \quad (58)$$

where $k_i = v_i NT \in \mathbb{R}$ and $l_i = \tau_i M \Delta f \in \mathbb{R}$. In (58), $g_h(l, k, l', k', l_i, k_i)$ represents the sampling function for the analog DD domain channel response. This function encapsulates the joint effects of delay and Doppler shifts, the transmit pulse shaping filter, and the receive filter. It is given by:

$$g_h(l, k, l', k', l_i, k_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_{rx}^*(t - nT) g_{tx}(t - n'T) e^{-j2\pi(m-m')\Delta f(t-nT)} dt \quad (59)$$

where $g_{tx}(t)$ and $g_{rx}(t)$ are the transmit and receive pulse shaping filters, respectively.

$$\begin{aligned} g_h(l, k, l', k', l_i, k_i) &= \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left[\sum_{m' \neq m}^{M-1} A_{g_{tx}g_{rx}} \left(-\frac{l_i}{M\Delta f}, (m-m')\Delta f - \frac{k_i}{NT} \right) \right. \\ &\quad \left. + e^{-j2\pi \frac{k'}{N}} \sum_{m'=0}^{M-1} A_{g_{tx}g_{rx}} \left(T - \frac{l_i}{M\Delta f}, (m-m')\Delta f - \frac{k_i}{NT} \right) \right] \\ &\quad \times e^{-j2\pi \frac{n(k-k'-k_i)}{N}} e^{j2\pi \frac{(ml-m'l'-m'l_i)}{M}} \end{aligned} \quad (60)$$

where the ambiguity function $A_{g_{tx}g_{rx}}(\tau, \nu)$ is defined as:

$$A_{g_{tx}g_{rx}}(\tau, \nu) = \int_{-\infty}^{\infty} g_{tx}(t) g_{rx}^*(t - \tau) e^{-j2\pi \nu(t-\tau)} dt \quad (61)$$

For the widely adopted rectangular transmit and receive filters, and assuming integer k_i and l_i [36], (57) simplifies to the well-known two-dimensional circular convolution Input-Output relation as follows:

$$Y_{DD}[l, k] = \sum_{i=1}^P h_i X_{DD}[(l - l_i)_M, (k - k_i)_N] \alpha[l, k, l_i, k_i] + W_{DD}[l, k], \quad (62)$$

where $\alpha[l, k, l_i, k_i]$ is a phase rotation term.

$$\alpha[l, k, l_i, k_i] = \begin{cases} e^{j2\pi \frac{(l-l_i)Mk_i}{MN}} & \text{if } l \in \{l_i, \dots, M-1\}, \\ e^{j2\pi \frac{(l-l_i)Mk_i - k_i M - (k-k_i)N}{MN}} & \text{if } l \in \{0, \dots, l_i-1\}. \end{cases}$$

4.4.2.2 Zak implementation

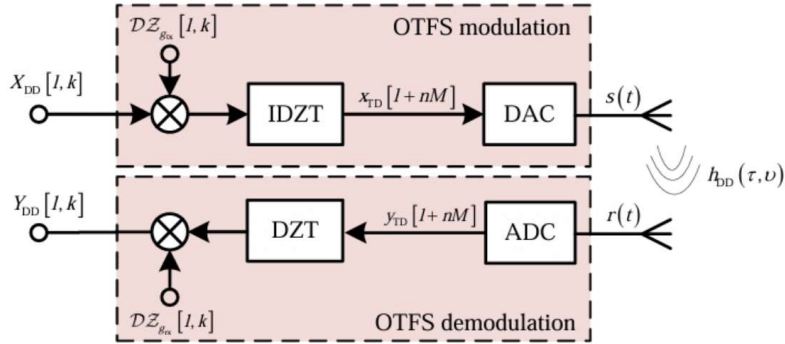


Figure 110: Schematic of the Zak implementation of the OTFS system.

SFFT-based OTFS modulation maps the information symbols from the DD domain first to the TF domain and then to the time domain. In contrast, DZT-based OTFS modulation directly transforms the information symbols from the DD domain to the time domain, as illustrated in Figure 110. The Discrete Zak Transform (DZT) and its inverse (IDZT) establish a mapping between a one-dimensional sequence and a two-dimensional sequence. Their definitions are provided below.

Definition 1 (Discrete Zak Transform): Let x be a periodic sequence with period MN . The DZT of x is defined as:

$$\mathcal{DZ}_x[l, k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[l + nM] e^{-j\frac{2\pi n}{N}k} \quad (63)$$

for $l = 0, 1, \dots, M-1$ and $k = 0, 1, \dots, N-1$. **Definition 2 (Inverse Discrete Zak Transform):** Given the MN -periodic sequence x , whose DZT is given by $\mathcal{DZ}_x[l, k]$, the IDZT is defined as:

$$x[l + nM] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \mathcal{DZ}_x[l, k] e^{j\frac{2\pi n}{N}k} \quad (64)$$

for $l = 0, 1, \dots, M-1$ and $n = 0, 1, \dots, N-1$. As shown in [37], a time-delay (TD) domain sequence can be obtained by applying the IDZT to the DD domain symbols at the OTFS transmitter. Specifically:

$$x_{TD}[l + nM] = \sqrt{M} \sum_{k=0}^{K-1} X_{DD}[l, k] \cdot \mathcal{DZ}_{g_{tx}}[l, k] e^{j\frac{2\pi n}{N}k} \quad (65)$$

Here, $\mathcal{DZ}_{g_{tx}}[l, k]$ represents the DZT of the periodically extended, sampled transmitted shaping pulse $g_{tx}\left(\frac{[k]_{MN}}{M}T\right)$, $k \in \mathbb{Z}$. The OTFS transmit signal $s(t)$ can be obtained by performing a digital-to-analog conversion operation on $x_{TD}[l + nM]$. By interpolating with a rectangular kernel, the OTFS transmit signal is given by:

$$s(t) = x_{TD}[l + nM], t \in \left[\frac{l - 0.5}{MT} + nT, \frac{l + 0.5}{MT} + nT\right] \quad (66)$$

At the OTFS receiver, the channel output $r(t)$ is analog-to-digital converted. The resulting TD domain received signal is given by:

$$y_{TD}[l + nM] = r(t)|_{t=l\frac{T}{M}+nT} \quad (67)$$

The DD domain received signal is obtained by applying the DZT to the TD domain received signal, as follows:

$$Y_{DD}[l, k] = \sqrt{MN} \mathcal{DZ}_{y_{TD}}[l, k] \mathcal{DZ}_{g_{rx}}^*[l, k] \quad (68)$$

Here, $\mathcal{DZ}_{y_{TD}}[l, k]$ represents the DZT of $y_{TD}[l + nM]$, and $\mathcal{DZ}_{g_{rx}}[l, k]$ denotes the DZT of the periodically extended sampled received filter $g_{rx}\left(\frac{[k]_{MN}}{M}T\right)$, $k \in \mathbb{Z}$. A common choice for the transmit and receive filters is rectangular filters. In this case, we have $\mathcal{DZ}_{g_{tx}}[l, k] = \mathcal{DZ}_{g_{rx}}[l, k] = \frac{1}{\sqrt{MN}}$, for all l, k .

4.4.2.3 Input/Output relation in matrix form

To derive a useful Input-Output relation [33], we can rewrite (57) as follows:

$$y[k, l] = \frac{1}{\sqrt{NM}} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} y[n, m] e^{j2\pi\left(\frac{ml}{M} - \frac{nk}{N}\right)} = \sum_{k', l'} x_{k', l'} g_{k, k'}[l, l'] \quad (69)$$

Where the Inter-Symbol Interference (ISI) coefficient of the DD pair $[k', l']$ observed by the sample $[k, l]$ is given by:

$$g_{k, k'}[l, l'] = \sum_p h_p \Psi_{k, k'}^p[l, l'] \quad (70)$$

with $\Psi_{k, k'}^p[l, l']$ defined as:

$$\begin{aligned} \Psi_{k, k'}^p[l, l'] &= \frac{1}{\sqrt{MN}} \sum_{n, n', m, m'} A_{g_{rx}g_{tx}}\left((n - n')T - \tau_p, (m - m')\Delta f - \nu_p\right) \\ &\times e^{j2\pi nT\nu_p} e^{-j2\pi m\Delta f\tau_p} \times e^{j2\pi\left(\frac{nk}{N} - \frac{ml}{M}\right)} e^{-j2\pi\left(\frac{nk'}{N} - \frac{ml'}{M}\right)} \end{aligned} \quad (71)$$

Using rectangular shaping pulses (g_{tx}, g_{rx}) and an approximation [38] of the ambiguity function, $\Psi_{k, k'}^p[l, l']$ can be rewritten as follows:

$$\begin{aligned} \Psi_{k, k'}^p[l, l'] &= \frac{1}{\sqrt{NM}} \frac{1 - e^{-j2\pi(k - k' + \nu_p NT)}}{1 - e^{-j2\pi(k - k' + \nu_p NT)/N}} \frac{1 - e^{-j2\pi(l - l' + \tau_p M\Delta f)}}{1 - e^{-j2\pi(l - l' + \tau_p M\Delta f)/M}} e^{-j2\pi\nu_p l/(M\Delta f)} \\ &\times \begin{cases} 1 & \text{if } l \in [0, M - 1 - \tau_p/(T/M)] \\ e^{-j2\pi(k/N + \nu_p T)} & \text{if } l \in [M - \tau_p/(T/M), M - 1] \end{cases} \end{aligned} \quad (72)$$

Given all possible values of $\Psi_{k, k'}^p[l, l']$, we can construct P matrices, each one indicated as Ψ^p and corresponding to each path, which their dimension is $MN \times MN$. This implies that Ψ^p depends on the DD pair (τ_p, ν_p). The complete Input-Output relation accounts for all paths; hence, we need to determine all such matrices and substitute them into the following equation:

$$\mathbf{y} = \sum_{p=1}^{P-1} h'_p \Psi^p \mathbf{x} + \mathbf{z} = \Psi \mathbf{x} + \mathbf{w} \quad (73)$$

Here, $\mathbf{x}$ is an MN -dimensional vector containing the transmitted symbols, and $\mathbf{z}$, which is also an MN -dimensional vector, represents the AWGN with zero mean and a covariance matrix given by $\sigma^2 \mathbf{I}_{MN}$. $h'_p = h_p e^{j2\pi v_p \tau_p}$ represents the channel gain of the p -th path, and

$$\Psi = \sum_{p=1}^P h'_p \Psi_p. \quad (74)$$

4.4.3 Novel interference cancellation (IC) detector

Before going into the details of the novel detection techniques, it is worth highlighting the signal model that was exploited to derive the algorithms, starting from the channel effect that each symbol in the delay-Doppler grid entails. As an example, consider the transmission over the time-frequency selective channel in (52) of an OTFS block composed by all zero symbols except one (placed anywhere in the grid). Based on this approach the mentioned OTFS pilot symbols configuration, as shown in Figure 107, can be explained more exhaustively. The p -th received component shows the transmitted symbol (circularly) shifted in the grid according to the pair (v_p, τ_p) . Moreover, if the involved channel parameters are integer multiples of the grid resolution, the whole received energy is confined in a single grid position. Otherwise, in the general and realistic scenario of Doppler and delay with a fractional component, an energy leakage, which is dictated by the Dirichlet kernel functions included in the channel matrix Ψ , appears in the positions surrounding the peak value. Due to the system's linearity, the overall effect of the multiple paths results in the superposition of the shifts associated with each component. In Figure 111, an example of the symbol shift in a multipath scenario is shown.

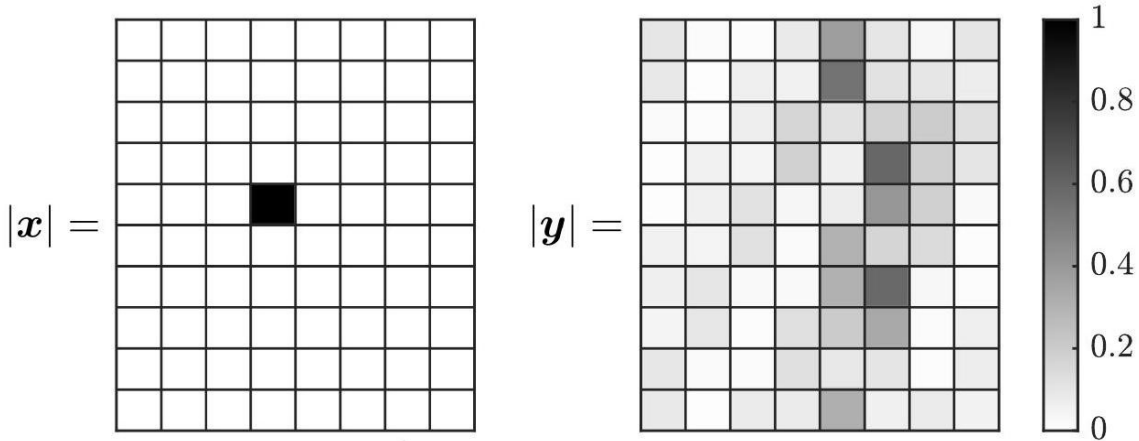


Figure 111: Example of the channel effect on a generic transmitted symbol. In the left subfigure, the transmitted symbol (magnitude) in the delay-Doppler grid. In the right subfigure, the received energy distribution after the propagation through the multipath channel with $P = 3$.

Equation (73) defines the input-output relation according to the Forney observation model. However, an alternative sufficient statistic on which the detectors that will be analysed in Sec. 4.4.3.4 operate is

$$\mathbf{z} = \Psi^H \mathbf{y} = \mathbf{G} \mathbf{x} + \tilde{\mathbf{w}} \quad (75)$$

where $\tilde{\mathbf{w}} = \Psi^H \mathbf{w}$ is the colored additive Gaussian noise, $(\cdot)^H$ stands for the Hermitian, and

$$\mathbf{G} = \Psi^H \Psi. \quad (76)$$

The operation in (75) corresponds to a matched filtering, and the resulting observation model is the Ungerboeck one [39], [40]. The processed channel matrix $\mathbf{G}$ presents a specific structure which can be exploited for the detection process. In fact, it is

- Sparse, as it represents a processed version of the channel matrix Ψ in the delay-Doppler domain;
- arranged in *diagonals*: its entries (i.e., the coefficients responsible for determining the amount of interference) are organized in a relatively ordered way. Specifically, the most relevant ones, i.e., the terms with the largest magnitudes are located along diagonals. Each diagonal consists of elements with comparable magnitude, with only a few exceptions. This structure is also partially present in Ψ , though to a lesser extent.

Specifically, from (76) and (74), we can write $\mathbf{G}$ as

$$\mathbf{G} = \left(\sum_{p=1}^P h'_p \Psi_p \right)^H \left(\sum_{t=1}^P h'_t \Psi_t \right) \quad (77)$$

$$= \underbrace{\sum_{p=1}^P |h'_p|^2 \Psi_p^H \Psi_p}_1 + \underbrace{\sum_{p=1}^P \sum_{t \neq p}^P (h'_p)^* h'_t \Psi_p^H \Psi_t}_2. \quad (78)$$

The first term in (78) identifies the main diagonal. In fact, it can be demonstrated that

$$\Psi_p^H \Psi_p = \mathbf{I}_{MN}, \forall p = 1, \dots, P, \quad (79)$$

where $\mathbf{I}_{MN}$ stands for the identity matrix of size MN . Moreover, equation (79) is verified for any pair (ν_p, τ_p) , no matter if they are or not integer multiples of the grid spacings. The second term in (78) is the sum of $P(P-1)$ matrices which give rise to the subdiagonals.

In the special case where τ_p and ν_p are integer multiples of the symbol grid, the matrix Ψ_p associated with the path p has one nonzero element in each row and column, all placed along (typically incomplete) diagonals, hence it is a permutation matrix. Since the product $\Psi_p^H \Psi_t$ in (78) shifts the diagonal positions without altering the number of elements per row/column, $\mathbf{G}$ results to be composed by $P(P-1) + 1$ nonzero elements in each row and column. From (74), it is clear that, under these unrealistic assumptions, Ψ is sparser than $\mathbf{G}$, as it contains only P nonzero entries per row and column. However, in the general case of real channels where (ν_p, τ_p) are not integer multiples of the grid granularity, the above considerations about the precise matrix structure are not valid anymore. In fact, the energy⁷ is not strictly confined along diagonals but rather spreads to positions near the peak value. Therefore, working on $\mathbf{G}$ can be beneficial because the product of Ψ with its Hermitian allows to obtain a more defined and compact matrix structure. An example of this behaviour can be

⁷ We shall refer to the magnitude of the entries of $\mathbf{G}$ as the energy of the channel response Ψ .

seen in Figure 113, where on the left we reported the structure of the channel matrix Ψ by showing a map of the energy distribution and on the right that of the related matrix $\mathbf{G}$.

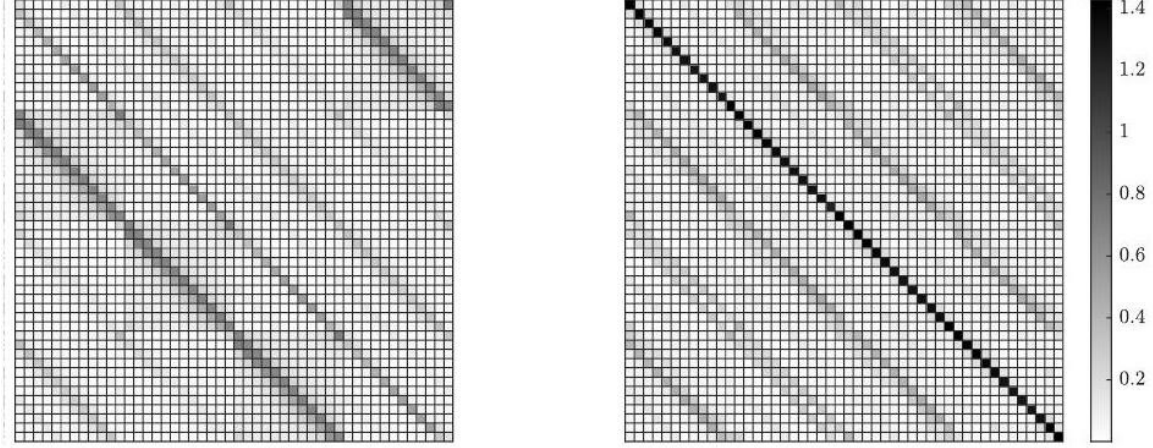


Figure 112: Comparison of the energy distribution for the matrix Ψ (on the left) and $\mathbf{G}$ (on the right) with $P = 5$.

The state-of-the-art detection algorithms and some novel solutions for OTFS-based transmissions are presented in this subsection. First, the baseline MMSE detector for OTFS is presented. Although this approach offers very good performance, its complexity is often impractical. In the following sections, we focus on low-complexity algorithms. Specifically, we consider message-passing (MP) algorithms working on factor-graphs (FG), which are classified into two categories :

- the first one focuses on continuous MP algorithms, where either all the exchanged messages in the FG are supposed to belong to a predefined family [41] or just the interference term is approximated with a continuous distribution while the symbol probability mass function (PMF) is kept discrete, as in [36];
- the second one is based on discrete MP, where the incoming and outgoing messages from a variable node (VN) have size M' , i.e., the constellation size.

Moreover, these two categories also differ in the observation model they work on: the first one includes algorithms derived from the Forney observation model [36], [38], whereas the algorithms belonging to the second category are based on the Ungerboeck model [41], [42]. In general, depending on how the posterior pmf of the modulation symbols $\mathbf{x}$, i.e.,

$$P(\mathbf{x} | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{x})P(\mathbf{x}) \quad (80)$$

is factorized, different algorithms can be derived. Then, a very low-complexity approach based on IC applied on the Ungerboeck observable is proposed. A known issue with suboptimal algorithms is their tendency to produce overconfident estimates. To mitigate this effect, a common strategy is to employ values of AWGN variance that are greater than the actual ones. Consequently, besides the MMSE detector which does not require any parameter tuning, the other analysed algorithms undergo offline optimization of this quantity. Moreover, a widely used technique in literature for iterative algorithms is damping [43], [44]. By computing a convex combination of the new estimate and the previous one, damping can not only accelerate convergence but also enable it in cases where it would otherwise fail. The use of

damping is incorporated into the algorithms discussed in Secs. 4.4.3.2-4.4.3.5, where it has proven to be effective. In general, since suboptimal algorithms benefit from inner detector iterations, in the following we will use the superscript (n) , with $n \in [1, N_{it}]$, on messages and/or parameters to indicate the n -th iteration.

4.4.3.1 Linear MMSE Equalizer

Denoting with x_k the k -th entry of the column vector $\mathbf{x}$ in (73), the detector provides at its output the noisy estimate $\hat{x}_k, k = 1, \dots, MN$. This detector is derived under the assumption that $\mathbf{x}$ is a Gaussian vector whose entries have unit mean square value. The optimal estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ according to the MMSE criterion results to be linear with the observation $\mathbf{y}$ and has expression

$$\hat{\mathbf{x}} = \mathbf{\Psi}^H [\mathbf{\Psi} \mathbf{\Psi}^H + 2\sigma^2 \mathbf{I}_{MN}]^{-1} \mathbf{y} \quad (81)$$

The complexity of this detector is $\mathcal{O}((MN)^3)$, since its implementation requires the inversion of the $MN \times MN$ matrix in (81). In a time-varying scenario, when the channel matrix $\mathbf{\Psi}$ changes at every block, and for typical values of M and N , this complexity cannot be afforded.

4.4.3.2 Continuous MP Algorithms

The algorithms described in this section have been developed to address the complexity issues that arise when applying the SPA to the FG obtained following the Forney observation model, i.e., when working on $\mathbf{\Psi}$. In fact, factorizing the posterior pdf (80) as

$$P(\mathbf{x} | \mathbf{y}) \propto \prod_{k=1}^{MN} p(y_k | \mathbf{x}) P_k(x_k) \quad (82)$$

and defining the FNs

$$Q_k(\mathbf{x}) \triangleq p(y_k | \mathbf{x}) = \exp \left(-\frac{\|y_k - \mathbf{\Psi}(k, :)\mathbf{x}\|^2}{2\sigma^2} \right) \quad (83)$$

where $\mathbf{\Psi}(k, :)$ denotes the k -th row of the channel matrix $\mathbf{\Psi}$, the FG in Figure 113 is obtained. As directly follows from the FNs definition (83), the j -th VN is connected to a number of FNs equal to the number of nonzero elements in the j -th column of $\mathbf{\Psi}$ and the k -th FN is connected to all the symbols corresponding to the positions of the nonzero elements of the k -th row of $\mathbf{\Psi}$. Since in the general case of delay and Doppler shifts with a fractional component the channel matrix is (almost) full, the FG presents an extremely large number of edges. Denoting with ℓ_k the number of nonzero elements in the k -th row of $\mathbf{\Psi}$ which is, therefore, the degree of the FN $Q_k(\mathbf{x})$, the exact application of the SPA would require

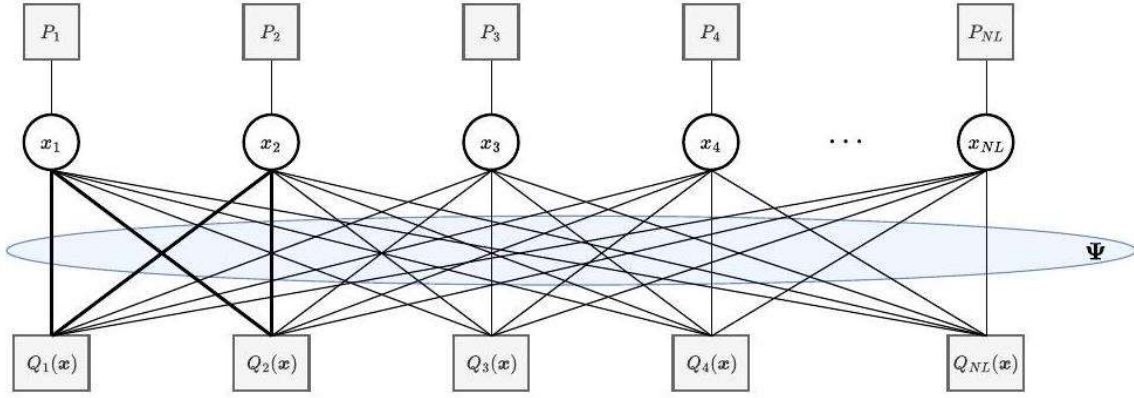


Figure 113: Structure of the FG for the Ψ -based algorithms.

to sum over $\ell_k - 1$ discrete variables taking M' possible values in $\mathcal{A}$. Thus, the computational load is M'^{ℓ_k-1} which, especially for large constellations and/or real channel conditions, is highly impractical. In fact, the message sent from FN $Q_k(\mathbf{x})$ to VN x_j is⁸

$$\mu_{Q_k \rightarrow x_j}(x_j) = \sum_{\sim x_j} Q_k(\mathbf{x}) \prod_{\ell \in n(k) \setminus \{j\}} \mu_{x_\ell \rightarrow Q_k}(x_\ell), \quad (84)$$

where $n(k)$ is the set of VNs connected to FN $Q_k(\mathbf{x})$. Instead, the message transmitted from VN x_j to FN Q_k , applying the SPA rules, is

$$\mu_{x_j \rightarrow Q_k}(x_j) = \prod_{\ell \in m(j) \setminus \{k\}} \mu_{Q_\ell \rightarrow x_j}(x_j), \quad (85)$$

where $m(j)$ is the set of FNs connected to the VN x_j . Another issue with this FG is the presence of cycles of length 4, which hinder the convergence of the SPA, unless the channel matrix is sufficiently sparse.

Due to the impractical complexity, several solutions have been proposed to simplify the exchanged messages. In particular, the algorithm presented in [36], which will be discussed in Sec. 4.4.3.2.1, relies on the approximation of the interference term through a Gaussian RV, whereas in Secs. 4.4.3.3-4.4.3.3.1: other possible approximate MP approaches are described [41], [45]. Specifically, unlike the algorithm in [32], they are based on the assumption that the transmitted symbols x_k are continuous RVs, characterized by a complex Gaussian distribution. For these reasons, they will be referred to as continuous MP algorithms.

4.4.3.2.1 MP on Ψ with Gaussian Approximation of Interference (GIMP $_{\Psi}$)

Among the approximating algorithms based on the Forney observation model, one of the most successful ones is that proposed in [36]. In fact, it relies on the following approximation. When computing the message from FN $Q_k(\mathbf{x})$ to VN x_j , the k -th received sample that can be obtained from (73) as

$$y_k = \Psi_{k,j} x_j + \sum_{\ell \in n(k) \setminus \{j\}} \Psi_{k,\ell} x_\ell + w_k, \quad (86)$$

⁸ In a practical receiver, the SPA rules are implemented in the logarithmic domain to improve accuracy and reduce computational complexity.

where $\Psi_{k,j}$ denotes the element in the k -th row and j -th column of Ψ and $n(k)$ is the set of VNs connected to $\text{FN}Q_k(x)$, is approximated as

$$\tilde{y}_k = \Psi_{k,j}x_j + \zeta_{k \rightarrow j} \quad (87)$$

where $\zeta_{k \rightarrow j}$ is the approximation of the interference term plus thermal noise and is modeled as a Gaussian random variable. In this way, a significant complexity reduction can be obtained, achieving an overall computational load of $\mathcal{O}(N_{it}MNS|\mathcal{A}|)$, where N_{it} denotes the number of allowed inner detector iterations, and S is the number of elements of each row and column of Ψ which are considered for the detection process. It is worth specifying that, in order to increase the sparsity of Ψ and, therefore, decrease the number of edges in the corresponding FG, the detector proposed in [36] considers the delay shifts as integer multiples of the grid resolution. For this reason, the number S of nonzero elements per row/column is limited and much smaller than the channel matrix dimension. However, the above assumption is generally not satisfied by real channels and the matrix Ψ is typically full. Moreover, making the detector work on an approximated channel matrix constructed by rounding the delay shifts to integer multiples of the grid leads to a significant performance degradation. Therefore, in the simulation

results we let the GIMP_{Ψ} detector work on the exact matrix Ψ and we set the value of S according to the complexity constraints and the choice of the interfering terms per symbol is done by selecting the most significant entries in terms of energy.

4.4.3.3 MP on Ψ with Gaussian Symbol Approximation (GSMP $_{\Psi}$)

Differently from the approach of [36], outlined above, the GSMP $_{\Psi}$ algorithm solves the complexity problem of computing (84) and (85) by introducing an approximation on the VNs: the discrete PMF of symbols is substituted with a Gaussian distribution so that only its mean and variance have to be computed and propagated through the graph. According to this approach, each discrete M' -valued message, sent from a VN to a FN is approximated by the Gaussian pdf minimizing the Kullback-Leibler (KL) divergence. The parameters characterizing the Gaussian distribution are, therefore, obtained through moment matching [46].

Substituting the discrete message $\mu_{x_{\ell} \rightarrow Q_k}(x_{\ell})$ in (84) with a Gaussian distribution with mean $\eta_{x_{\ell} \rightarrow Q_k}^{(n)}$ and variance $v_{x_{\ell} \rightarrow Q_k}^{(n)}$, and considering the VNs $x_{\ell}, \ell \in n(k) \setminus \{j\}$ as continuous RVs, we obtain

$$\begin{aligned} & \mu_{Q_k \rightarrow x_j}^{(n)}(x_j) \\ &= \int_{\sim x_j} Q_k(\mathbf{x}) \prod_{\ell \in n(k) \setminus \{j\}} g_{\mathbb{C}}(x_{\ell}, \eta_{x_{\ell} \rightarrow Q_k}^{(n)}, v_{x_{\ell} \rightarrow Q_k}^{(n)}) \\ &\propto g_{\mathbb{C}}(\Psi_{k,j}x_j, \eta_{Q_k \rightarrow x_j}^{(n)}, v_{Q_k \rightarrow x_j}^{(n)}) \end{aligned} \quad (88)$$

where

$$\eta_{Q_k \rightarrow x_j}^{(n)} = y_k - \sum_{\ell \in n(k) \setminus \{j\}} \Psi_{k,\ell} \eta_{x_{\ell} \rightarrow Q_k}^{(n)} \quad (89)$$

$$v_{Q_k \rightarrow x_j}^{(n)} = 2\sigma^2 + \sum_{\ell \in n(k) \setminus \{j\}} |\Psi_{k,\ell}|^2 v_{x_{\ell} \rightarrow Q_k}^{(n)}. \quad (90)$$

Finally, (85) becomes

$$\mu_{x_j \rightarrow Q_k}^{(n)}(x_j = a_s) = \frac{g_{\mathbb{C}}(x_j = a_s, \zeta_{x_j \rightarrow Q_k}^{(n)}, \gamma_{x_j \rightarrow Q_k}^{(n)})}{\sum_{m=1}^M g_{\mathbb{C}}(a_m, \zeta_{x_j \rightarrow Q_k}^{(n)}, \gamma_{x_j \rightarrow Q_k}^{(n)})}, \quad (91)$$

where $a_s \in \mathcal{A}$ and

$$\gamma_{x_j \rightarrow Q_k}^{(n)} = \left(\sum_{\ell \in m(j) \setminus \{k\}} \frac{|\Psi_{\ell,j}|^2}{v_{Q_\ell \rightarrow x_j}^{(n-1)}} \right)^{-1}, \quad (92)$$

$$\zeta_{x_j \rightarrow Q_k}^{(n)} = \gamma_{x_j \rightarrow Q_k}^{(n)} \sum_{\ell \in m(j) \setminus \{k\}} \frac{\Psi_{\ell,j}^* \eta_{Q_\ell \rightarrow x_j}^{(n-1)}}{v_{Q_\ell \rightarrow x_j}^{(n-1)}}. \quad (44) \quad (93)$$

The discrete distribution computed in (91) is then used to find the mean and variance of the approximated Gaussian message transmitted from VN x_j to FN Q_k . The symbols marginal distribution is then computed through the SPA rules by multiplying all the incoming messages in each VN.

4.4.3.3.1 MP on Ψ using Expectation Propagation (EP $_{\Psi}$)

The EP framework [47] offers another viable alternative for reducing the complexity. In this case, the simplification concerns the message in (91). In fact, the symbol belief [41], [45]

$$\beta^{(n)}(x_j = a_s) \triangleq \frac{\prod_{k \in m(j)} \mu_{Q_k \rightarrow x_j}^{(n-1)}(x_j = a_s)}{\sum_{m=1}^M \prod_{k \in m(j)} \mu_{Q_k \rightarrow x_j}^{(n-1)}(x_j = a_m)} \quad (94)$$

where the M discrete values of $\mu_{Q_k \rightarrow x_j}^{(n-1)}(x_j)$ are computed from (88) as

$$\mu_{Q_k \rightarrow x_j}^{(n-1)}(x_j = a_s) \propto \exp \left\{ -\frac{|\Psi_{k,j} a_s - \eta_{Q_k \rightarrow x_j}^{(n-1)}|^2}{v_{Q_k \rightarrow x_j}^{(n-1)}} \right\} \quad (95)$$

is replaced by a Gaussian pdf $\hat{\beta}^{(n)}(x_j) = g_{\mathbb{C}}(x_j, \hat{\eta}_{x_j}^{(n)}, \hat{\tau}_{x_j}^{(n)})$, instead of approximating the message $\mu_{x_j \rightarrow Q_k}(x_j)$. From the approximate symbol belief, it is, then, possible to compute the approximate message $\hat{\mu}_{x_j \rightarrow Q_k}(x_j)$. The parameters $\hat{\eta}_{x_j}^{(n)}$ and $\hat{\tau}_{x_j}^{(n)}$ of $\hat{\beta}^{(n)}(x_j)$ can be computed by minimizing the KL divergence among the exact and the approximated Gaussian distribution, therefore, exploiting moment matching technique, we obtain

$$\hat{\eta}_{x_j}^{(n)} = \mathbb{E}_{\beta^{(n)}(x_j)}[x_j] \quad (96)$$

$$\hat{\tau}_{x_j}^{(n)} = \mathbb{E}_{\beta^{(n)}(x_j)}[|x_j|^2] - |\hat{\eta}_{x_j}^{(n)}|^2 \quad (97)$$

Finally, the parameters of the message from VN to FN directly derive from the ratio among two Gaussian pdfs, i.e.,

$$\hat{\tau}_{x_j \rightarrow Q_k}^{(n)} = \left(\frac{1}{\hat{\tau}_{x_j}^{(n)}} - \frac{|\Psi_{k,j}|^2}{v_{Q_k \rightarrow x_j}^{(n-1)}} \right)^{-1} \quad (98)$$

$$\hat{\eta}_{x_j \rightarrow Q_k}^{(n)} = \hat{\tau}_{x_j \rightarrow Q_k}^{(n)} \left(\frac{\hat{\eta}_{x_j}^{(n)}}{\hat{\tau}_{x_j}^{(n)}} - \frac{\Psi_{k,j}^* \eta_{Q_k \rightarrow x_j}^{(n-1)}}{v_{Q_k \rightarrow x_j}^{(n-1)}} \right) \quad (99)$$

As a general comment, regarding the last two algorithms that were illustrated above, we can state that both GSMP_{Ψ} and EP_{Ψ} pursue similar objectives, albeit referring to a different philosophy. In EP_{Ψ} , a Gaussian approximating family is assumed for the marginal distributions of symbols x_j (although eventually discretized, as in (94)-(95), on the symbol alphabet) while in GSMP_{Ψ} it is the individual messages sent to x_j that are approximated using the Gaussian family, as in (88)-(90). In both cases, the projection operation onto the approximating family is performed by minimizing the KL divergence, using the moment matching technique. These two algorithmic options are typical of linear dynamical systems where discrete (symbols) and continuous (noise) latent variables coexist and generate mixture distributions that are difficult to handle. While EP_{Ψ} is generally considered more accurate, the GSMP_{Ψ} approach avoids the critical division operation between distribution, here implemented in (98)-(99), by transparently removing the VN to FN message before the projection operation of EP_{Ψ} is performed; for such reason this approach is referred to as a *transparent propagation* algorithm in [48] [49] [50].

4.4.3.4 Discrete MP Algorithms

Unlike the strategies discussed in Sec. 4.4.3.2, the discrete MP algorithms analysed here perform detection by operating on the processed channel matrix $\mathbf{G}$ by leveraging the framework based on FGs and the SPA. As observed in Sec. 4.4.3.2, this approach is unfeasible when working on Ψ , since the corresponding factorization involves FNs with high degree. On the contrary, the algorithms described in this section work on FGs with degree-2 FNs.

4.4.3.4.1 Matrix $\mathbf{G}$ -based algorithm ($\text{MP}_{\mathbf{G}}$)

Among the MP algorithms working on FGs, an effective one, here referred to as $\text{MP}_{\mathbf{G}}$, is proposed in [40] and recently applied to OTFS in [38]. The main algorithmic steps are summarized as follows.

The factorization of the conditional pdf in (80) can be rewritten as

$$p(\mathbf{y} | \mathbf{x}) \propto \exp \left(\frac{2\Re\{\mathbf{x}^H \mathbf{z}\} - \mathbf{x}^H \mathbf{G} \mathbf{x}}{2\sigma^2} \right) \quad (100)$$

From (100), it is possible to notice that $\mathbf{z}$ is a sufficient statistic for symbols detection. Expressing in (100) the matrices products explicitly in terms of their elements, it is possible to factorize the conditional pmf in (80) as⁹

$$P(\mathbf{x} | \mathbf{y}) \propto \prod_{k=1}^{MN} \left[P_k(x_k) F_k(x_k) \prod_{\ell=1}^{k-1} I_{k,\ell}(x_k, x_{\ell}) \right] \quad (101)$$

which is based on the assumption that the modulation symbols x_k are treated by the detector as independent and identically distributed (i.i.d.) with given (typically uniform) a-priori pmfs $P_k(x_k)$, $k = 1, \dots, MN$, and we defined

⁹ The inner product on ℓ is formally equal to 1, when $k = 1$.

$$F_k(x_k) \triangleq \exp \left[\frac{1}{\sigma^2} \Re \left\{ z_k x_k^* - \frac{1}{2} G_{k,k} |x_k|^2 \right\} \right] \quad (102)$$

$$I_{k,\ell}(x_k, x_\ell) \triangleq \begin{cases} \exp \left[-\frac{1}{\sigma^2} \Re \{ G_{k,\ell} x_\ell x_k^* \} \right] & \text{if } 1 \leq \ell < k \\ 1 & \text{else} \end{cases} \quad (103)$$

where $G_{k,\ell}$ denotes the element in the k -th row and ℓ -th column of $\mathbf{G}$. The resulting FG has girth 6 and is composed by FNs $F_k(x_k)$ connected to the corresponding VNs $x_k, k = 1, \dots, NL$, and FNs $I_{k,\ell}(x_k, x_\ell)$ representing the interference between the variables x_k and x_ℓ as shown in Figure 114. The resulting MP_G algorithm consists in a direct application of the SPA to the obtained graph. More details can be found in [40].

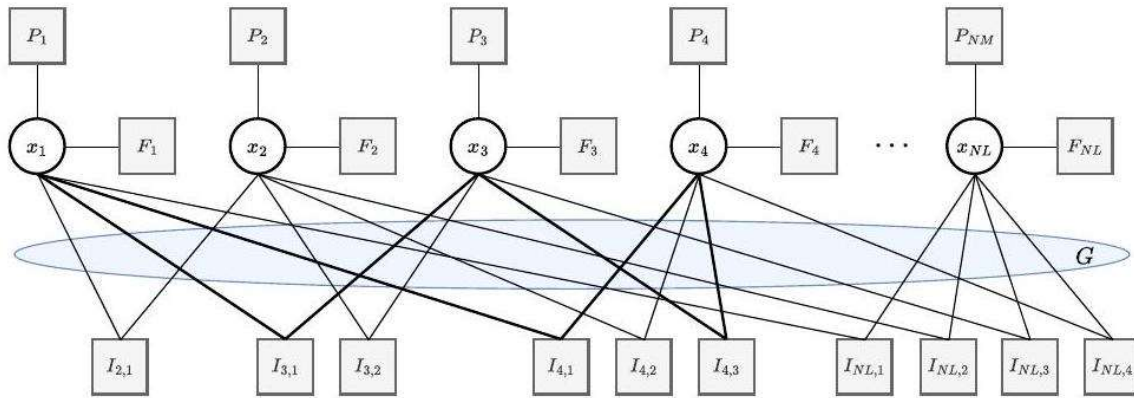


Figure 114: Structure of the FG for the MP_G algorithm.

As further discussed in Sec. 0.1.3.5, the complexity of the MP_G algorithm turns out to be linear in the number of interferers per symbol Z [40] and quadratic in the constellation size M' . In the simulation results we selected the $S < Z$ terms bringing the largest amount of interference, where the value of S is set such that the complexity constraints are satisfied. The overall computational load is thus $\mathcal{O}(N_{it} MNSM'^2)$.

4.4.3.4.2 Hierarchical MP on $\mathbf{G}$ (HBCJR_G)

In this section, we propose a new low-complexity MP algorithm, referred to as HBCJR_G, which strongly outperforms the MP_G by taking advantage of the specific structure of the matrix $\mathbf{G}$.

The first algorithmic step is to compute the energy of each sub-diagonal as

$$e_j = \sum_{k=1}^{MN-j} |G_{j+k,k}|, j = 1, \dots, MN - 1. \quad (104)$$

Notice that, since $\mathbf{G}$ is Hermitian, considering the upper triangular submatrix would yield the same result. Then, the first D diagonals with the largest energy are selected and used in the detection process, where D is chosen in accordance with given complexity constraints and/or optimized in order to achieve the best trade-off between complexity and performance. The largest D energy values are sorted in decreasing order and collected in a vector $\mathbf{e}$. The detection algorithm operates on the reduced Hermitian matrix $\mathbf{G}^r$, which is composed by D

nonzero diagonals in the lower triangular submatrix, plus their adjoint, in the upper triangular, and has the following structure:

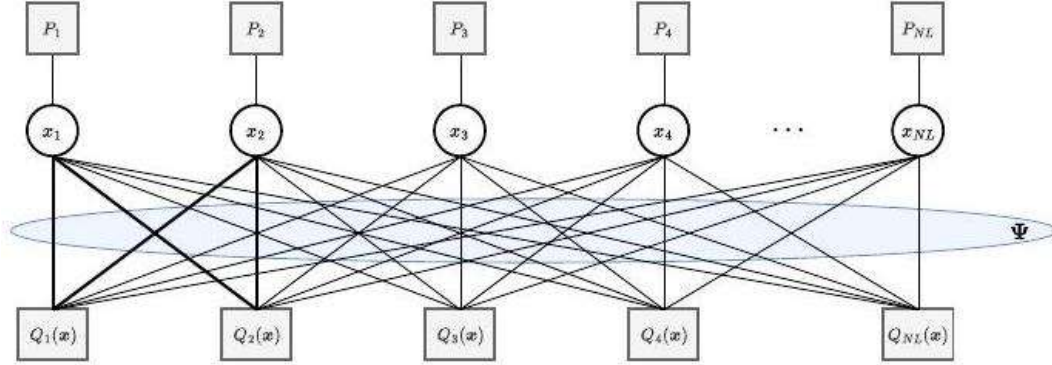


Figure 115: Structure of the hierarchical MP on $HBCJR_G$

where $d_i, i = 1, \dots, D$ is the distance in terms of number of elements between the main diagonal and the i -th subdiagonal. Moreover, the distances $\{d_i\}_{i=1}^D$ are sorted according to the energies in e , so that d_1 denotes the distance of the strongest interferers from the main diagonal, and so on. Analytically,

$$G_{m,n}^r = \begin{cases} G_{m,n}, & \text{if } m = n \vee |m - n| \in \{d_i\}_{i=1}^D \\ 0, & \text{elsewhere} \end{cases} \quad (105)$$

where $G_{m,n}^r$ denotes the element in the m -th row and n -th column of $\mathbf{G}^r$. It is worth underlining that a longer distance from the main diagonal does not imply a reduction in the magnitude of its entries. With reference to the conditional pdf in (100), by substituting $\mathbf{G}^r$ in place of $\mathbf{G}$ and explicitly expressing the matrix operations in terms of their components, we obtain

$$\mathbf{x}^H \mathbf{z} = \sum_{k=1}^{MN} z_k x_k^* \quad (106)$$

$$\begin{aligned} \mathbf{x}^H \mathbf{G}^r \mathbf{x} &= \sum_{k=1}^{MN} G_{k,k}^r |x_k|^2 + \sum_{k=1}^{MN} x_k^* \sum_{\ell \neq k} G_{k,\ell}^r x_\ell \\ &= \sum_{k=1}^{MN} G_{k,k} |x_k|^2 + \sum_{k=1}^{MN} \sum_{j=1}^D 2\Re \{ G_{k,k-d_j} x_{k-d_j} x_k^* \} \end{aligned} \quad (107)$$

Therefore, we can rewrite the approximate a -posteriori pmf as

$$\tilde{P}(\mathbf{x} | \mathbf{y}) \propto \prod_{k=1}^{MN} \left[P_k(x_k) F_k(x_k) \prod_{j=1}^D I_{k,k-d_j}(x_k, x_{k-d_j}) \right] \quad (108)$$

where both factors $F_k(x_k)$ and $I_{k,k-d_j}(x_k, x_{k-d_j})$ are defined in (102)-(103), which account for possibly negative column indices. Considering the generic VN x_k , a portion of the FG corresponding to (108) is shown in Figure 116, where we denote by α_{d_j} and $\bar{\alpha}_{d_j}$ the messages

propagating in the forward direction, and by β_{d_j} and $\bar{\beta}_{d_j}$ those propagating in the backward direction¹⁰. In order to perform detection, we split the interferers into subsets identified by the diagonals of $\mathbf{G}^T$.

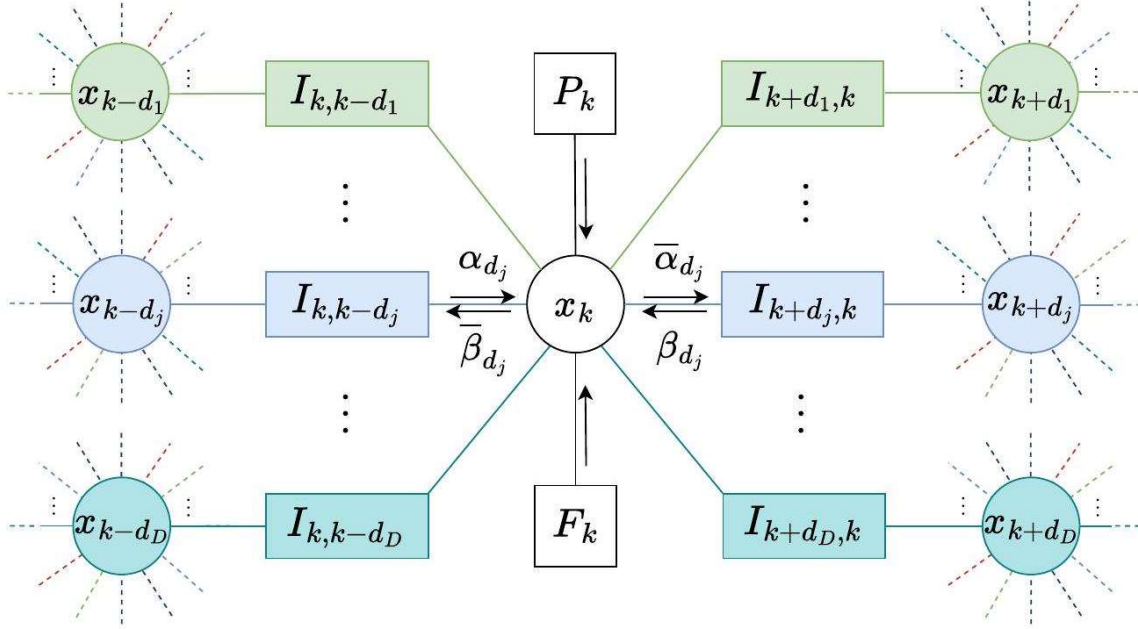


Figure 116: Partial structure of the FG for the algorithm proposed in [42].

Specifically, we first assume that the channel matrix consists only of the interferers located at a distance d_1 from the main diagonal. Then, we implement the optimal detector for this simplified scenario exploiting the well-known BCJR algorithm¹¹ [51]. This approach is then extended to the other diagonals, processed one at a time and ordered according to their energy, as detailed in the following. The overall detection strategy is obtained by allowing these sub-detectors to exchange information. As a result, each symbol x_k will be connected to, at most, $2D$ edges, as shown in Figure 117. The communication among the various detection units is based on the following rule. When considering the BCJR algorithm corresponding to the diagonal at distance d_j , the symbol x_k takes as input information a partial marginal distribution $T_{d_j}^{(n)}(x_k)$ defined as

$$T_{d_j}^{(n)}(x_k) = P_k(x_k)F_k(x_k) \prod_{i=1}^{j-1} \alpha_{d_i}^{(n)}(x_k)\beta_{d_i}^{(n)}(x_k) \cdot \prod_{i=j+1}^D \alpha_{d_i}^{(n-1)}(x_k)\beta_{d_i}^{(n-1)}(x_k) \quad (109)$$

where $\alpha_{d_i}(x_k)$ and $\beta_{d_i}(x_k)$ are the incoming messages to the VN x_k in Figure 116 (from preceding and following VNs, respectively), and their product represents the output information about the k -th symbol generated by the BCJR algorithm corresponding to the diagonal at distance d_i . Notice that in (109) the messages $\alpha_{d_i}^{(n-1)}(x_k)$ and $\beta_{d_i}^{(n-1)}(x_k)$ are the ones computed in the previous iteration because of the scheduling choice, as will be discussed later.

¹⁰ The subscript d_j indicates the position of the involved interferers.

¹¹ With reference to a generic diagonal at distance d from the main one, the optimal detector consists of a set of separate d_j BCJR algorithms with unitary memory, each involving MN/d_j symbols.

Moreover, in order to account for the extrinsic information on the VN x_k , we exclude from the product those messages that refer to the considered j -th BCJR algorithm.

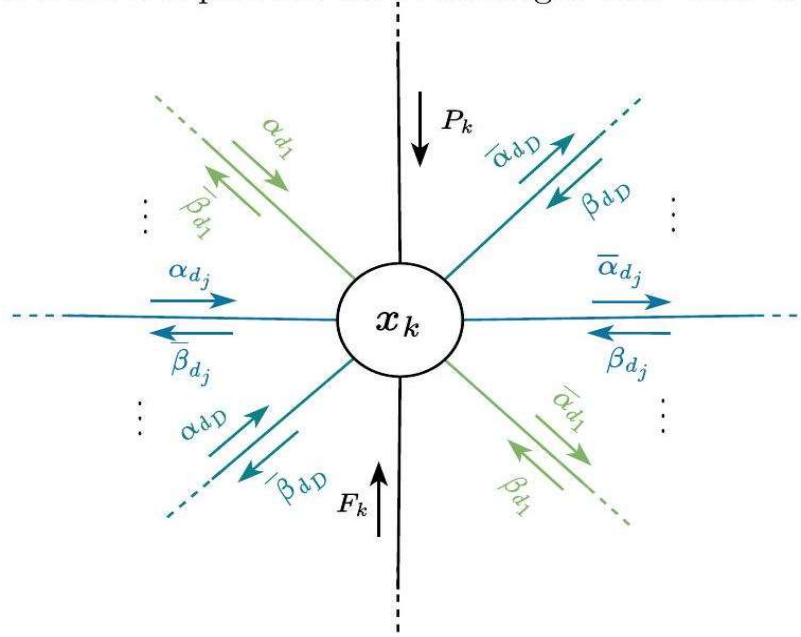


Figure 117: Overview of the outgoing and incoming messages to the VN x_k .

With reference to the messages in Figure 116, by applying the SPA updating rules, $\alpha_{d_j}^{(n)}, \bar{\alpha}_{d_j}^{(n)}, \beta_{d_j}^{(n)}$ and $\bar{\beta}_{d_j}^{(n)}$ can be recursively computed by means of the following forward and backward recursions:

$$\alpha_{d_j}^{(n)}(x_k) = \sum_{x_{k-d_j}} \bar{\alpha}_{d_j}^{(n)}(x_{k-d_j}) I_{k,k-d_j}(x_k, x_{k-d_j}) \quad (110)$$

$$\bar{\alpha}_{d_j}^{(n)}(x_k) = \alpha_{d_j}^{(n)}(x_k) T_{d_j}^{(n)}(x_k) \quad (111)$$

$$\beta_{d_j}^{(n)}(x_k) = \sum_{x_{k+d_j}} \bar{\beta}_{d_j}^{(n)}(x_{k+d_j}) I_{k,k+d_j}(x_{k+d_j}, x_k) \quad (112)$$

$$\bar{\beta}_{d_j}^{(n)}(x_k) = \beta_{d_j}^{(n)}(x_k) T_{d_j}^{(n)}(x_k). \quad (113)$$

Concerning the scheduling, the BCJR algorithm that is activated first, as previously mentioned, is the one involving the largest interference, i.e., $\max(e)$, with initial input information

$$T_{d_1}^{(1)}(x_k) = P_k(x_k) F_k(x_k).$$

Then, we proceed by following the decreasing energy order, i.e., the one dictated by the entries of e . This way, we are prioritizing the strongest interferers, related to the corresponding diagonals of G^T .

This algorithm benefits from inner iterations. However, it achieves convergence very quickly: typically, just 2 iterations are sufficient. as will be demonstrated in Sec 4.4.4. After the allowed iterations have been performed, the final estimate for the modulation symbols is computed as

$$P(x_k | \mathbf{y}) \propto P_k(x_k) F_k(x_k) \prod_{i=1}^D \alpha_{d_i}^{(N_{it})}(x_k) \beta_{d_i}^{(N_{it})}(x_k), \quad (114)$$

where N_{it} denotes the number of allowed iterations. Similarly to the graph of the MP_G [40], [38], the FG on which this algorithm operates has girth 6. However, although the loops in the FG are quite long, a known issue of the SPA when applied to FGs with cycles is the overestimation of messages reliability, due to the suboptimality of the resulting algorithm. As previously mentioned at the beginning of this section, a possible solution is the adoption in (102)-(103) of a value of σ^2 greater than the exact one. The effectiveness of this algorithm comes from its ability to leverage the matrix structure and operate along the diagonals, allowing for different optimizations of σ^2 based on the specific diagonal being considered. In particular, it is sufficient to optimize only two variances: one for the first BCJR and another for the remaining $D - 1$. The complexity of this approach is $\mathcal{O}(N_{it} M N S M'^2)$, where S is the number of interferers per symbol and depends linearly on the number D of selected diagonals.

Despite it is derived from the same factorization and employs the same channel matrix $\mathbf{G}$ as the MP_G algorithm, the $HBCJR_G$ has superior performance and faster convergence, as will be shown in Sec. 4.4.4, thanks to the adopted hierarchical scheduling.

4.4.3.5 The Interference Cancellation Algorithm (IC_G)

In this section, we propose an extremely low-complexity approach based on the Ungerboeck observation model. The compact structure of the processed channel matrix $\mathbf{G}$ enables the effectiveness of this strategy. Specifically, from (75), the k -th received sample can be expressed as

$$z_k = G_{k,k} x_k + \zeta_k + \tilde{w}_k \quad (115)$$

where

$$\tilde{w}_k = \sum_{\ell} \Psi_{k,\ell}^* w_{\ell} \quad (116)$$

$$\zeta_k = \sum_{\ell \neq k} G_{k,\ell} x_{\ell} \quad (117)$$

The proposed algorithm is developed on the principle of successive IC. Through multiple iterations, the interference term $\zeta_k, k = 1, \dots, MN$ is estimated as

$$\tilde{\zeta}_k^{(n)} = \sum_{\ell \neq k} G_{k,\ell} \mu_{\ell}^{(n)}, \quad (118)$$

where $\mu_{\ell}^{(n)}$ is the soft estimate of the ℓ -th symbol and, denoting with $\hat{P}^{(n)}$ the estimated symbols' pmf at the n -th iteration, it can be computed as

$$\mu_{\ell}^{(n)} = \mathbb{E}_{\hat{P}^{(n)}}[x_{\ell}] = \sum_{\alpha_s \in \mathcal{A}} \hat{P}^{(n)}(x_{\ell} = \alpha_s) \alpha_s \quad (119)$$

With reference to the n -th iteration, the symbols' pmf $\hat{P}^{(n)}$ used in (119) is

$$\hat{P}^{(n)}(x_k) \propto g_c(z_k^{(n)}, G_{k,k} x_k, 2\sigma^2 G_{k,k}) \quad (120)$$

$$\begin{aligned}
 & \propto \exp \left\{ -\frac{|z_k^{(n)} - G_{k,k}x_k|^2}{2\sigma^2 G_{k,k}} \right\} \\
 & \propto \exp \left\{ -\frac{|z_k^{(n)}|^2 + G_{k,k}^2 |x_k|^2 - 2\Re\{z_k^{(n)} G_{k,k} x_k^*\}}{2\sigma^2 G_{k,k}} \right\} \\
 & \propto \exp \left\{ \frac{2\Re\{z_k^{(n)} x_k^*\} - G_{k,k} |x_k|^2}{2\sigma^2} \right\}
 \end{aligned} \tag{121}$$

where

- the quantity $2\sigma^2 G_{k,k}$ in (120) is the variance of $\tilde{w}_k$, and therefore it is the entry in position (k, k) of the covariance matrix of the additive colored Gaussian noise, i.e., $\mathbb{E}[\Psi^H \mathbf{w}(\Psi^H \mathbf{w})^H] = 2\sigma^2 \mathbf{G}$;
- $z_k^{(n)}, k = 1, \dots, MN$ are the entries of $\mathbf{z}^{(n)}$ which is the result of IC. Setting $\mathbf{z}^{(1)} = \mathbf{z}$, it follows

$$\mathbf{z}^{(n)} = \mathbf{z}^{(1)} - \tilde{\zeta}^{(n-1)} \tag{122}$$

Furthermore, the convergence process is guaranteed using damping so that, by adopting a damping parameter ξ , the employed expression for the observable update, which is employed in place of (122), becomes

$$\mathbf{z}^{(n+1)} = \xi [\mathbf{z}^{(1)} - \tilde{\zeta}^{(n)}] + (1 - \xi) \mathbf{z}^{(n)} \tag{123}$$

The resulting algorithm which will be referred to as $\text{IC}_{\mathbf{G}}$, is summarized in Algorithm (1).

Algorithm 1 Pseudocode for $\text{IC}_{\mathbf{G}}$

1. **Inputs:** Received samples $\mathbf{z}$, channel matrix $\mathbf{G}$, maximum number of iterations N_{it}
 2. **Initialization:** Initialize $\mathbf{z}^{(1)}$ to $\mathbf{z}$
 3. **for** $n = 1$ to N_{it} **do**
 4. $\forall k = 1, \dots, MN$, compute:
 - the estimate of symbols' pmf $\hat{P}^{(n)}(x_k)$ via (121)
 - the soft symbol estimate $\mu_k^{(n)}$ via (119)
 - the estimate of the interference term $\tilde{\zeta}_k^{(n)}$ by applying (118)
 - $\mathbf{z}^{(n+1)}$ by purifying the observable $\mathbf{z}^{(1)}$ from interference and apply damping as in (122)
 5. **end for**
 6. **Output:** $\hat{P}^{N_{it}}(x_k), \forall k = 1, \dots, MN$
-

The same detection strategy based on IC can of course be applied to the Forney observation model (73), which is based on the channel matrix Ψ . However thanks to the sparser structure of $\mathbf{G}$, which results from the energy compaction property of matched filtering exploited in the Ungerboeck observation model (see Figure 111 for a pictorial representation), the $\text{IC}_{\mathbf{G}}$ approach yields a much superior performance, compared to that based on matrix Ψ and denoted here as IC_{Ψ} .

4.4.3.6 Complexity analysis

The complexity analysis of the tested detectors is reported in Table 31, where we differentiated between two (real and scalar) terms multiplications, additions and accesses to the lookup table (LUT). As shown in Table 31, the algorithms that require the highest computational load are the discrete MP approaches, whose computational complexity grows quadratically with the constellation size M' and linearly with the number of interferers per symbol S . In contrast, the continuous MP algorithms, which rely on message/marginal approximations, allow for a linear growth with the product $M'S$. Finally, a drastic reduction in complexity is achieved by the algorithms based on IC, whose computational complexity is linear with M' and S . This significant gap can be clearly observed in Figure 118, which shows the complexity in terms of number of operations versus the constellation cardinality M' for a fixed number of interferers per symbol S , set to 100 in this case.

Table 31: Complexity comparison per symbol per iteration for the tested detectors.

	Two-terms additions	Two-terms multiplications	LUT accesses
$MP_G, HBCJR_G$	$3M'^2S + 2M'S + 2M'$	0	$2M'^2S$
$GIMP_\Psi$	$9M'S + 7S - M' - 2$	$10M'S + 5S$	$M'S$
$GSMP_\Psi$	$7M'S + 13S + 2M' - 6$	$10M'S + 17S + 4M'$	$M'S$
EP_Ψ	$4M'S + 18S + 5M' - 5$	$6M'S + 23S + 8M' + 2$	M'
IC_G, IC_Ψ	$5M' + 2S - 3$	$7M' + 2S + 2$	M'

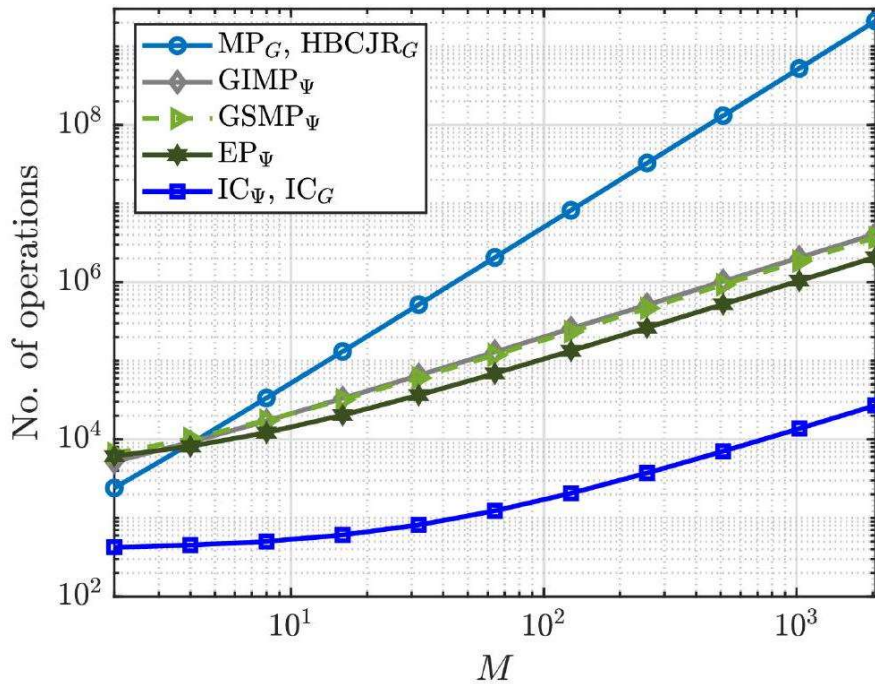


Figure 118: Comparison of the complexity of the analysed algorithms for a fixed number of interferers per symbol ($S = 100$).

4.4.4 Numerical results

OFDM and OTFS were compared in terms of pragmatic capacity (PC, see [32] for details) and bit error rate (BER). The system configuration was described in 4.4.1 and Table 32 shows the main simulation parameters adopted in the case of single transmitting satellite. The scenario was derived by exploiting a

Table 32: Simulation parameters.

Description	Value
Carrier Frequency	20 GHz
Subcarrier spacing, symbol time	120 kHz, 8.33 μ s
Number of NR RB	24
Number of OFDM FFT	512
Number of OTFS subcarriers, M	38
Number of OTFS time slots, N	56
Number of users/beams	24
Number of satellites	1 / 2
Channel model	Clear-sky
Phase noise	See [52]
MCSTable	Table 1
Frequency precomp.	95%
Beamforming	MMSE-PAC (see [1])

system simulator which can provide the beamforming matrices and channel responses given the STARDUST system and antenna configuration and a predefined user placement on the satellite field of view (a random placement was implemented, therefore no specific radio resource management was adopted for the following results). Moreover, the system simulator provides the Doppler shift and the free-space loss for each satellite-user link.

In any case, a 95% precompensation of the true Doppler shift was considered, in order to keep a significant residual frequency offset and then investigate the impact of such impairment on the receivers' performance. Two different scenarios were mainly simulated, with one or two transmitting satellites. In the first case all the 24 beams were exploited to serve users in the field of view, whereas in the second case only the beams of both satellites in visibility of the same users were activated, hence obtaining an average number of active beams around 10. The beamforming technique was set to MMSE with per-antenna constraint (PAC), as this is the solution providing the best results in most cases (actually the sum power constraint was shown to be almost as effective as PAC, concerning link-level simulations). Then, since the NR CSI reporting exploited to set the modulation and coding of the transmitted signal is based on specific antenna configurations and designed for the terrestrial scenario, a few values of SNR margin (namely 0, 4, and 8 dB) were added to the measured value in order to select lower coding rates; the effect of interference and large residual frequency offsets, in particular, turn out to be quite difficult to handle with non-specific feedbacks in the reporting procedure. The overhead due to cyclic prefix, pilot symbols and virtual subcarriers, was accounted for in the computation of the net pragmatic capacity; in particular, the OFDM waveform requires a cyclic prefix per OFDM symbol (for instance, each slot with 14 symbols entails 14 cyclic prefix), whereas OTFS only one per OTFS grid. Given the specific mapping of the OFDM time-frequency grid to the OTFS delay-Doppler grid described in 4.4.1, OTFS offers a lower overhead. On the other hand, the number of pilots were kept equal for both waveforms (DMRS for OFDM and pilot empty sub-grid for OTFS), but the optimization of frequency spacing

between OTFS subchannels allowed to obtain a lower overhead with respect to the presence of many virtual subcarriers in the OFDM waveform. It is worth noting, anyway, that the NR waveform was accurately and extensively designed taking into account a huge number of parameters, constraints and different cases, whereas the proposed OTFS solution only represents a first possible approach to the problem of fitting OTFS into the NR standard. Results are reported for two detectors, namely the standard linear MMSE equalizer and the interference cancellation (IC_G) novel detector described in 4.4.3, for both OFDM and OTFS.

4.4.4.1 One-satellite scenario

Figure 119 shows the performance of OFDM when using the MMSE and IC_G detectors, and with a CSI margin of 0 and 4 dB. It is worth noting that in case of pragmatic capacity, the CSI margin only affects the modulation order choice, since the capacity is computed on the detector soft-outputs without resorting to the decoder. Anyway, a slight increase in the detector performance can be noticed. Also notice that in case of the IC_G detector, the noise variance that is fed to the algorithm for the computation of the soft-outputs can be tuned to optimize the performance and obtain the largest possible value of the capacity. In this light two "options" were envisaged, namely 1 and 2, to be tested; as can be seen, the performance is influenced by this parameter as well, and the best result is obtained by the MMSE detector. This is not surprising, as the IC_G detector was specifically tailored for the OTFS system, and only shortly adapted for the OFDM system. A better refinement of the algorithm and on the parameters configuration could possibly lead to better results. Figure 120 shows the performance of the OTFS system under the same conditions, and in this case the best result is obtained with the IC_G detector with CSI margin 0 dB and option 1. Also notice that these latter figures include the Shannon limit, i.e., the capacity computed by using the SINR values provided directly by the system simulator. As can be seen, there is a recognizable loss, that can be expected and motivated by several explanations (e.g., a faulty frequency offset compensation, the impact of nongaussian inter-beam interference, phase noise, etc.). It is also worth noting that these receivers have many different configuration parameters that should be optimized with the specific channel at hand, but this is a nearly impossible task when simulating such a diverse scenario. For the same reason some links turned out to be in outage.

Figure 121 represents a direct comparison of OTFS and OFDM, where a slight advantage of the former can be noticed. Finally, Table 33 also reports the block-error rates of the same scenarios, where a normalized min-sum decoder with 6 maximum iterations was included. In this case the performance with 0 – dB CSI margin entails a bad performance, but when a margin is added, and therefore lower modulation order and code rates are selected, a near-zero performance can be obtained. Then, results were also drawn in absence of interbeam interference, and, interestingly, the OFDM waveform shows a BLER equal to zero only with a margin greater than zero, whereas the OTFS waveform provides a zero BLER in any case, showing a better robustness against the residual frequency offset.

Figure 122 and Figure 123 show the different distributions of Modulation formats and code rates that are selected by the CSI reporting procedure when the margin is introduced. Clearly, to a higher margin corresponds more lower-order modulations and lower code rates.

4.4.4.2 Two-satellite scenario

Next, a two-satellite scenario was simulated, in order to show the potential benefit of serving the same user with more than one satellite, following the cell-free paradigm; clearly this solution is effective when a fading channel is present, since the main advantage is represented by the possibility to maintain a minimum throughput even in presence of strong fading on one of the links, or, in other words, to avoid outages. In this sense the following results mainly show again a comparison between OTFS and OFDM, whereas a comparison of the one-satellite vs two-satellite scenarios should include a different channel model from the clear-sky that was used

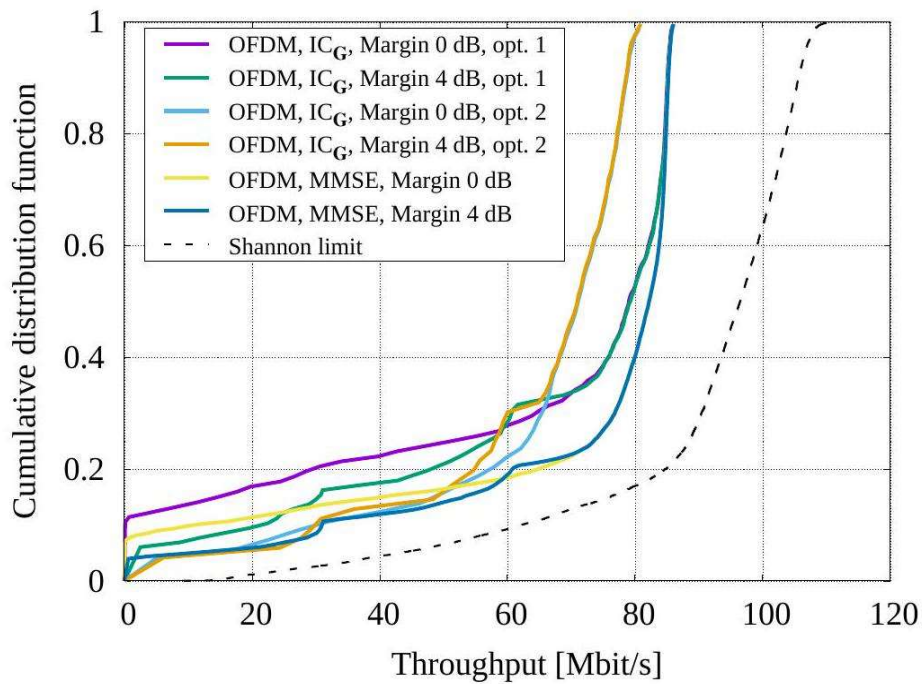


Figure119: Cumulative distribution function of the pragmatic-capacity-based throughput for OTFS.

for these simulations. As before, Figure 124, Figure 125, Figure 126 provide the performance of OFDM, OTFS, and a direct comparison of the two systems with the same parameters mentioned in the one-satellite scenario. Here, the IC_G detector

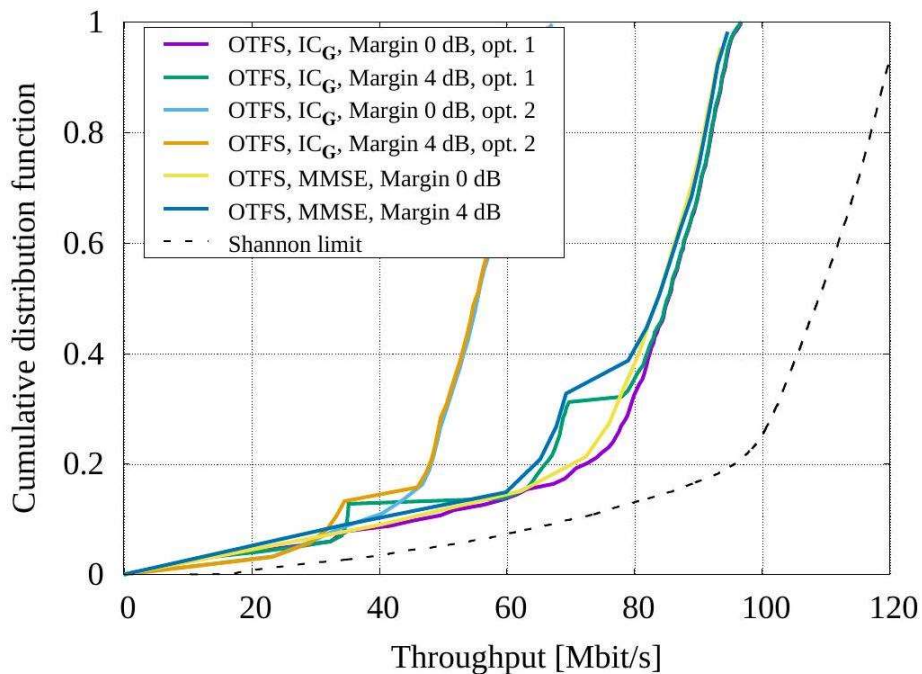


Figure120: Cumulative distribution function of the pragmatic-capacity-based throughput for OTFS.

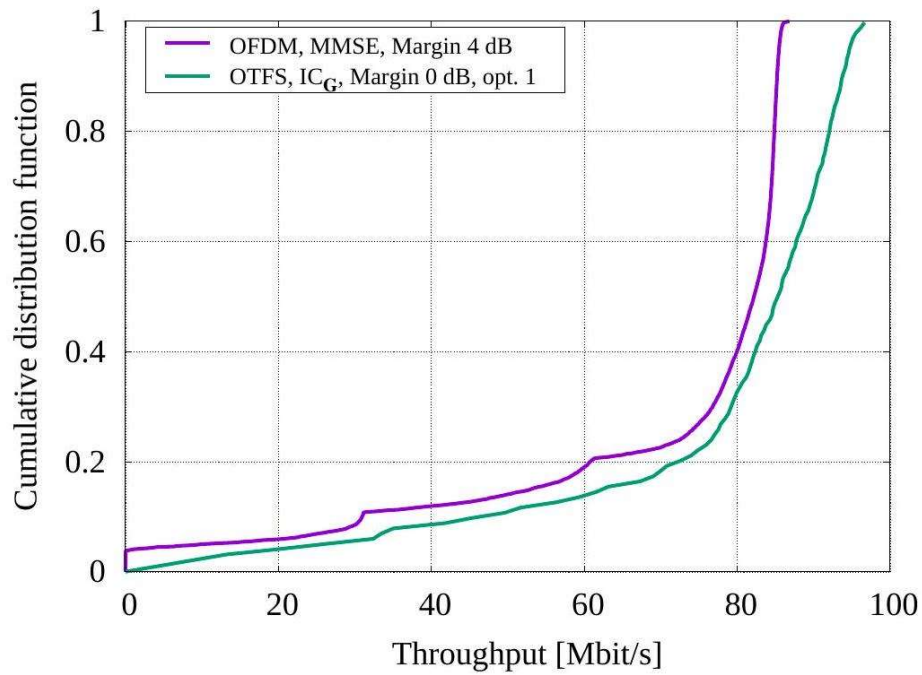


Figure 121: Comparison of the (best) results for OTFS and OFDM.

Table 33: Block error rate performance as a function of CSI margin.

Interbeam interference	Modulation	Detector	CSI Margin [dB]		
			0	4	8
w/	OFDM	IC _G	0.4074	0.0168	0.0156
w/	OFDM	MMSE	0.4006	0.0084	0.0086
w/	OTFS	IC _G	0.0254	0.0101	0.0068
w/	OTFS	MMSE	0.0274	0.0174	0.0070
w/o	OFDM	IC _G	0.2227	0	0
w/o	OFDM	MMSE	0.2448	0	0
w/o	OTFS	IC _G	0	0	0
w/o	OTFS	MMSE	0	0	0

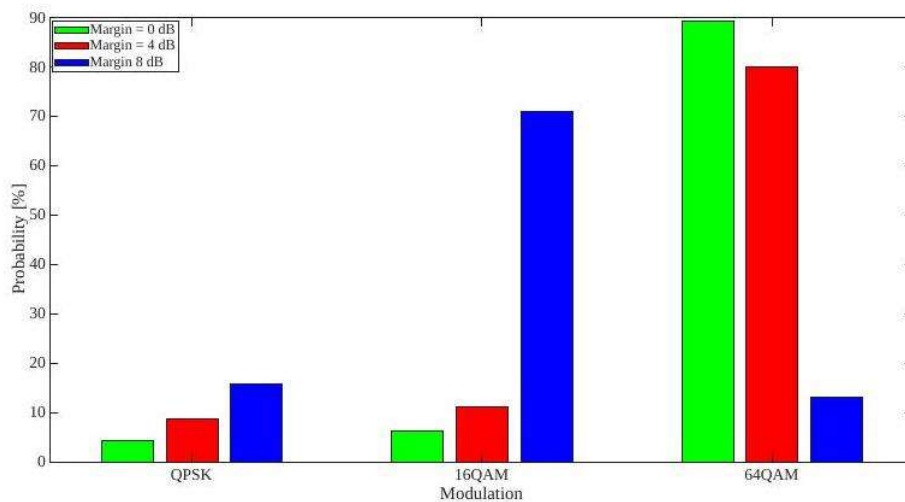


Figure122: Distribution of Modulations a function

proved to provide the best performance for OFDM as well, and, from Figure 125 in particular, it can also be concluded that OTFS would require a more extended set of options than just 1 and 2 in order to optimize the performance for all realizations of SINRs. In fact, from Figure 126 it can be noticed that plausibly a better result can be obtained by connecting many curves computed optimizing over a wider range of values for the noise variance (this was not possible at this point due to the time-consuming simulations).

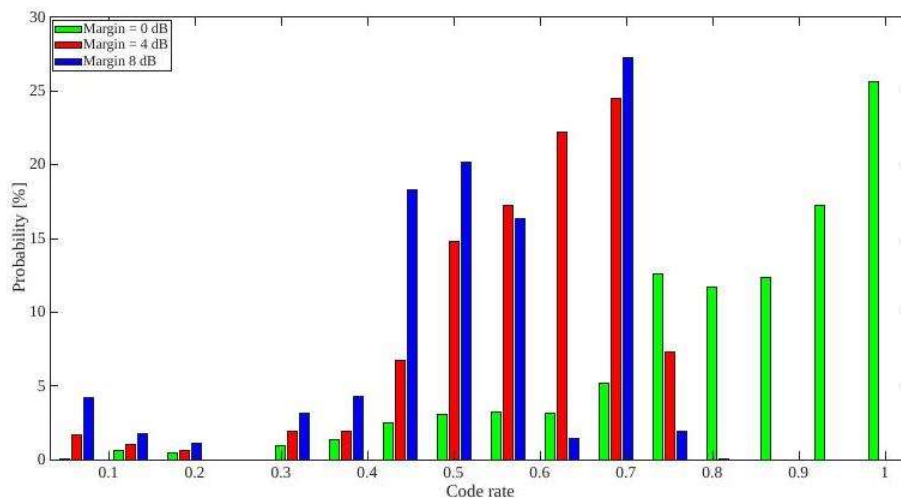


Figure123: Distribution of code rates a function of CSI margin.

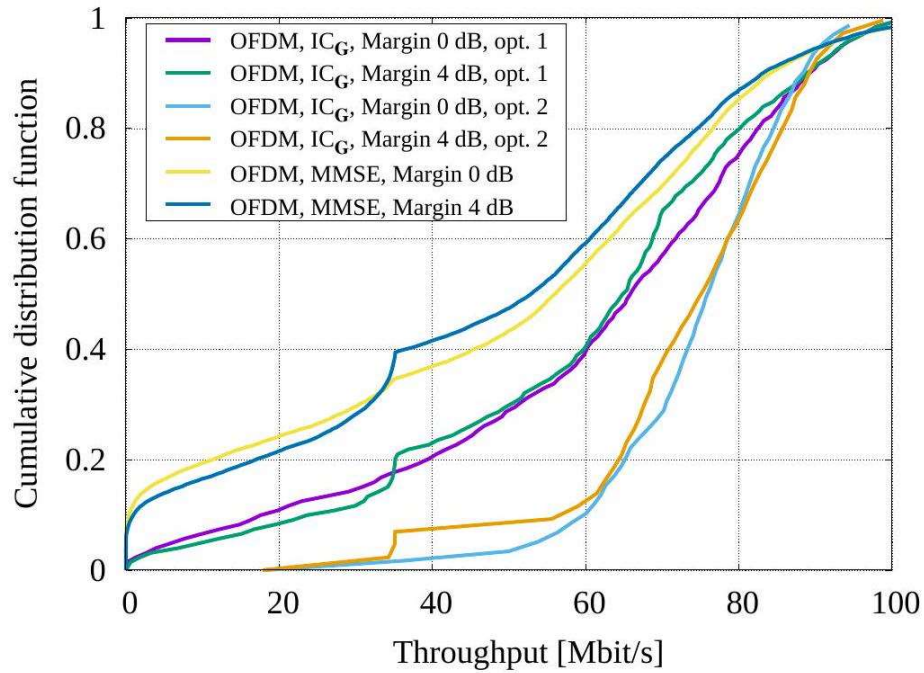


Figure124: Cumulative distribution function of the pragmatic-capacity-based throughput for OFDM.

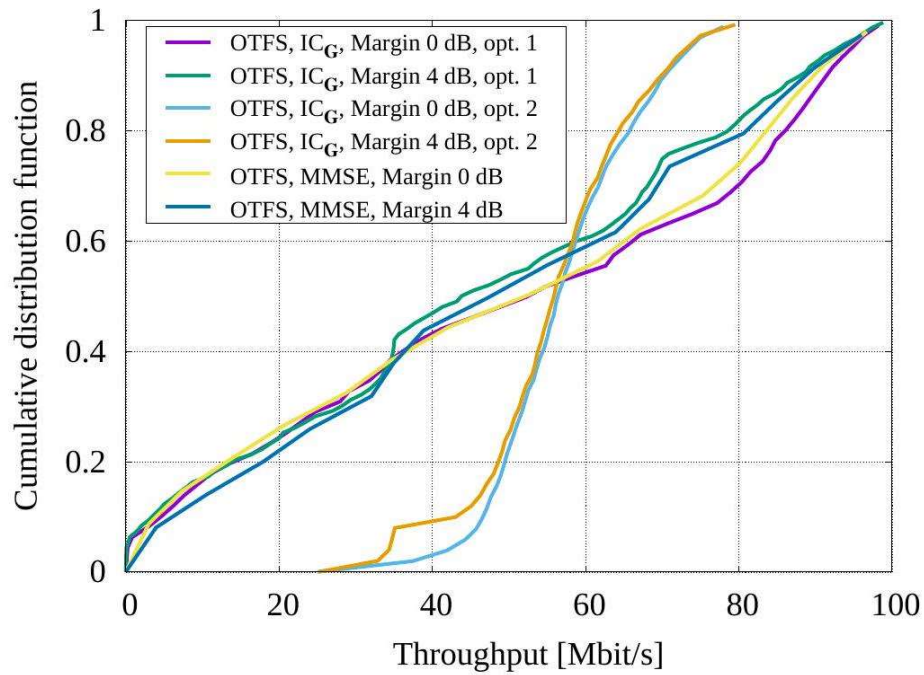


Figure125: Cumulative distribution function of the pragmatic-capacity-based throughput for OTFS.

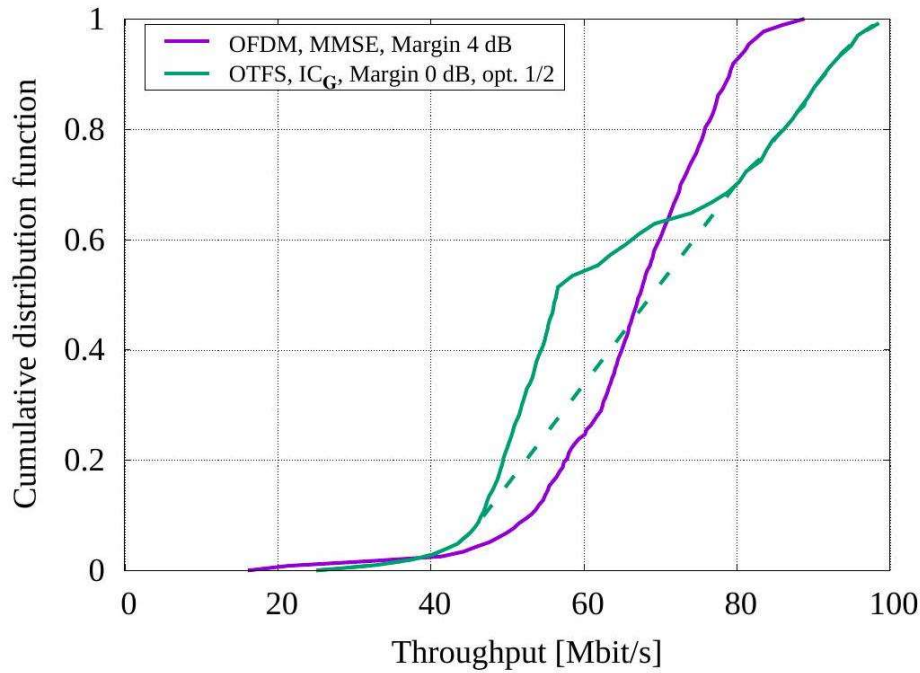


Figure 126: Comparison of the (best) results for OTFS and OFDM.

4.4.5 Advanced receiver processing over fading channel

As mentioned in the previous section, the main advantages of the cell-free paradigm are offered when the channel does not correspond to the optimistic clear-sky condition, but to a more realistic channel with some kind of, at least, large scale fading. Therefore, to show the actual benefits of serving the same users from two different satellites, we employed the model described in [53], and in particular we referred to the highway environment, that includes parameters for 20 GHz carrier frequencies and 30° of inclination. This channel is modelled by the classical two-state semi-Markov chain, where the transitions between two states identified as GOOD and BAD are defined by the set of parameters reported in Table 34 for the specific case at hand. In the table, G,B stand for GOOD and BAD states, parameter (μ, σ) refers to the mean and standard deviation of the log-normal law assumed for events duration, $dur_{minG,B}$ refers to the minimum possible events duration, $(\mu_{MA G,B}, \sigma_{MA G,B})$ to the mean and standard deviation of the average direct path amplitude (M_A) over one event, $(h_{1G,B}, h_{2G,B})$ to the multipath power, $(g_{1G,B}, g_{2G,B})$ to the standard deviation of M_A , $L_{corrG,B}$ to the direct path amplitude correlation distance, $f_1 \Delta M_A + f_2$ to the transition length (being ΔM_A the difference between M_{AG} and M_{AB}). The choice of such a model is motivated by the fact that for transmissions in Ka-band with directive antennas a narrow-band (flat-fading) channel can be assumed. Also, uncorrelated links were assumed as satellites serving the same user are spatially separated both in terms of distance (hundreds or thousands of km) and angles of arrival. Delays and Doppler shifts are computed for each satellite based on its true orbital speed and position.

The first numerical results were then derived by using a standard front-end at the user side, thus still combining the received signals non-coherently; all other parameters are same as previously described. This approach leads to a suboptimal performance as, even if a good timing advance estimation would entail a correct reception at symbol time, the phase uncertainty (also due to the different Doppler shifts), that cannot be avoided, is responsible for the fluctuation of the received signal power.

Table 34: ITU-R P.681-11 channel parameters for highway environment

Parameter	GOOD	BAD
$\mu_{MA\ G,B}, \sigma_{MA\ G,B}$ (m)	1.27, 1.86	-0.31, 1.35
$dur_{min\ G,B}$ (m)	0.01	0.5
$\mu_{MA\ G,B}, \sigma_{MA\ G,B}$ (dB)	-0.16, 0.39	-5.92, 8.20
$h_{1\ G,B}, h_{2\ G,B}$ (dB)	0.0, -29.61	-0.34, -14.39
$g_{1\ G,B}, g_{2\ G,B}$	0.0, 0.39	-0.41, 0.0
$L_{corr\ G,B}$ (m)	0.15, 1.28	
$f_1\Delta M_A + f_2$	[0.001, 0.861]	

4.4.5.1 Non-coherent detection

The following results are related to non-coherent signal combining at the user, as in sect. 4.4.4, but with the narrowband (NB) fading channel. Here, in the single satellite case only one of the transmitted signals was used for the detection, to have a fair comparison. Figure 127 and Figure

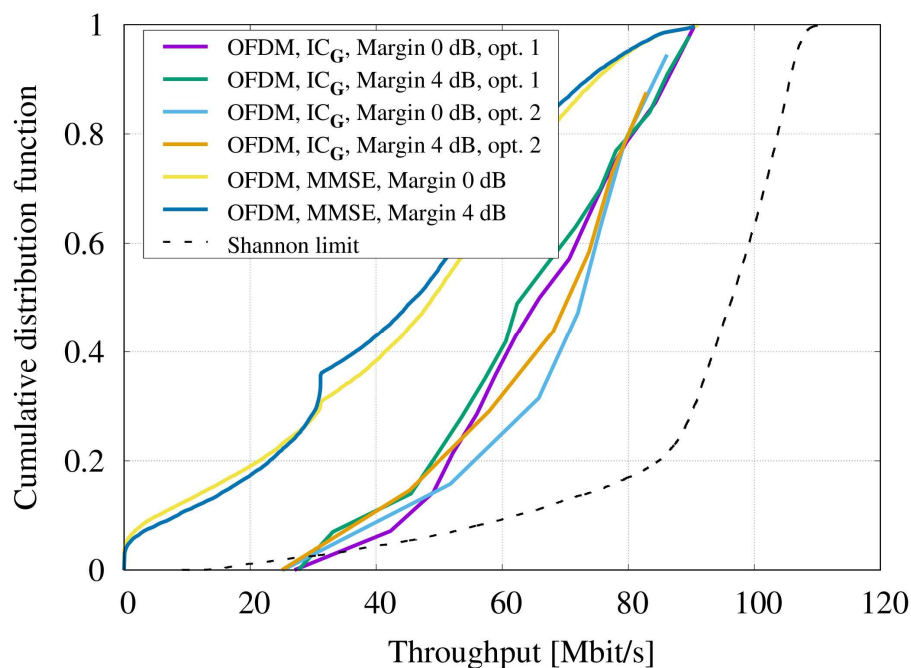


Figure 127: Cumulative distribution function of the pragmatic-capacity-based throughput for OFDM with one transmitting satellite and fading channel.

128 show the CDFs of OFDM and OTFS when only one satellite is transmitting over the fading channel. It can be noticed that the IC_G detector entails the best performance, and that the CSI margin has a limited impact, in the case of OFDM, whereas the performance of the OTFS is more or less the same if parameters are properly optimized, with a slight advantage for the MMSE detector.

Figure 129 and Figure 130 on the other hand show the results when two satellites are serving each user; in this case the results are only marginally better for both OFDM and OTFS, with the IC_G detector still performing better for OFDM and both detectors having similar performance with OTFS.

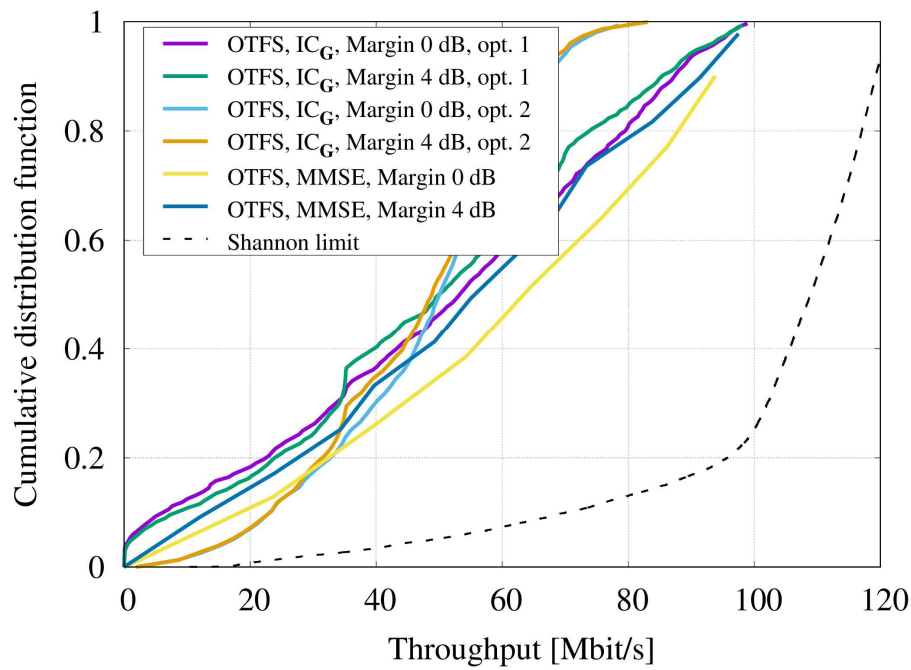


Figure128: Cumulative distribution function of the pragmatic-capacity-based throughput for OTFS with one transmitting satellite and fading channel.

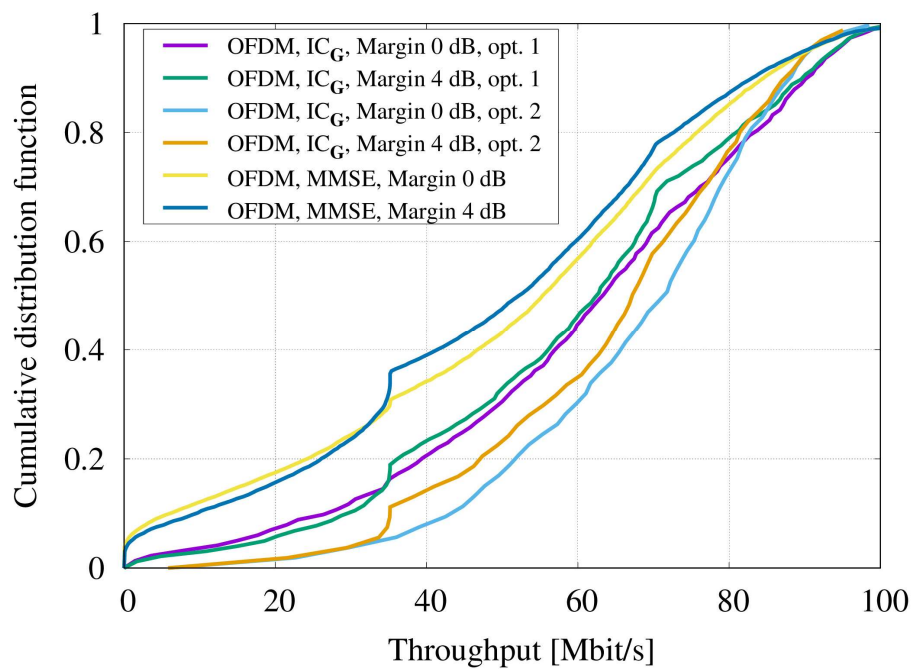


Figure129: Cumulative distribution function of the pragmatic-capacity-based throughput for OFDM with two transmitting satellites and fading channel.

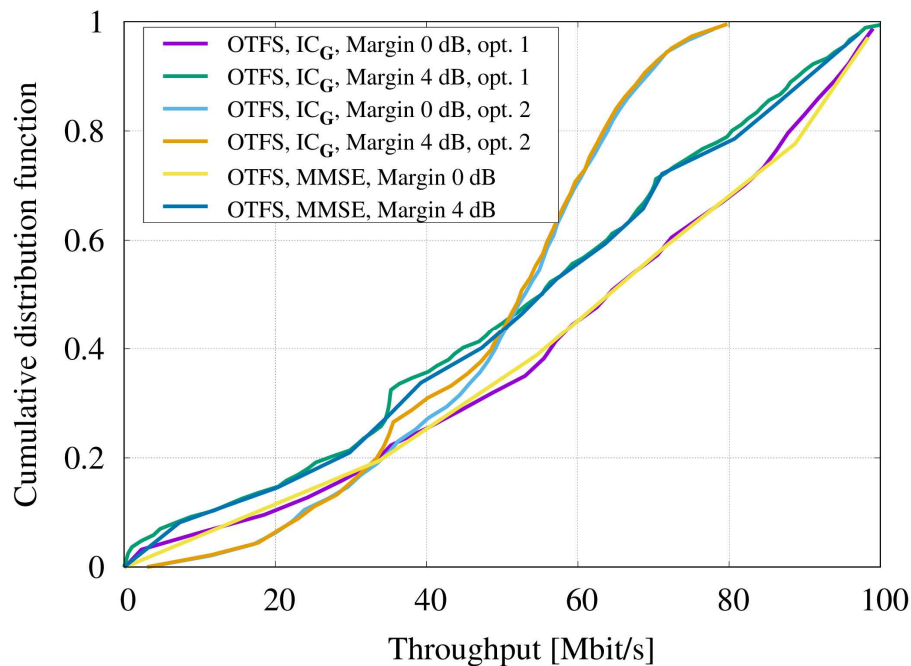


Figure130: Cumulative distribution function of the pragmatic-capacity-based throughput for OTFS with two transmitting satellites and fading channel.

4.4.5.2 Coherent detection

To overcome the problem related to non-coherent signal reception at the user side, which typically happens when the users are equipped with single antennas, directive antennas (i.e., uniform planar array - UPA - antennas) must be employed not only at the transmit side but at the receive side as well, so that the signal processing chain can be replicated for each received signal (RF chains, channel gain, frequency and phase estimation, etc.). This approach can be seen as a sort of maximum ratio combining (MRC), a well-known technique in wireless communications that aims at maximizing the SNR at the receive side when a MIMO transmission or a multipath channel is envisaged. A processing complexity increase must be paid to obtain better results as shown in the following figures, where optimized parameters (most of them) were employed. Also, ideal combining at the receiver is here considered, which means ideal interference rejection at the user side, as the main purpose was to demonstrate the relative advantage of using two satellites instead of one rather than accurate estimates of the performance, that, in any case, would depend on a relevant number of system parameters and further optimizations. It is worth noting that another possible solution would be an equivalent of selection combining, where the power of the signals at the antenna array can be estimated and the array can be set to receive only the signal with the highest sensed SNR. This approach would lead to an intermediate result with respect to the single and double transmitting satellites with MRC scenarios.

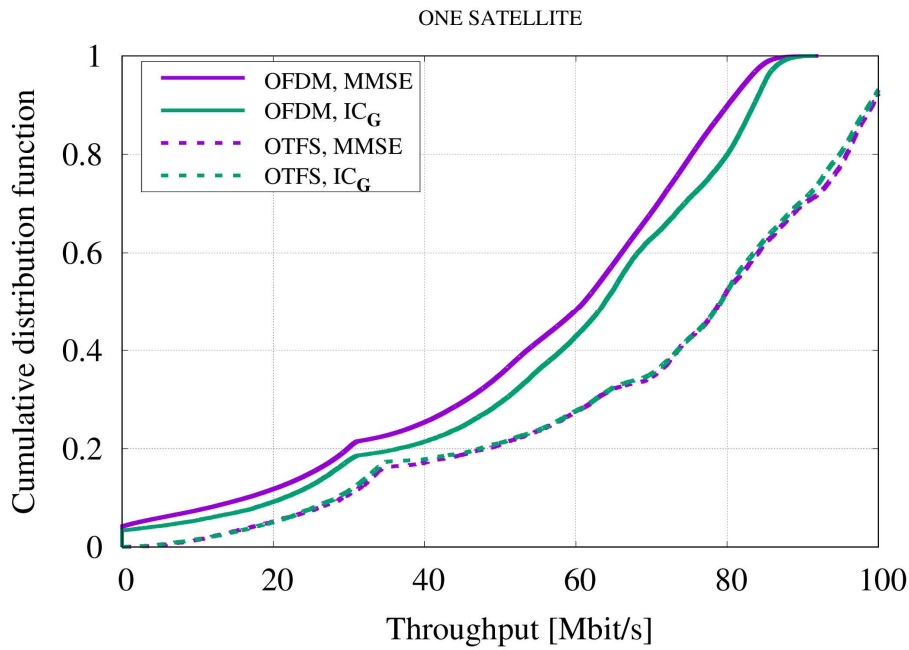


Figure131: Cumulative distribution function of the pragmatic-capacity-based throughput for OFDM and OTFS with one transmitting satellite, fading channel and antenna arrays at the receive side (MRC).

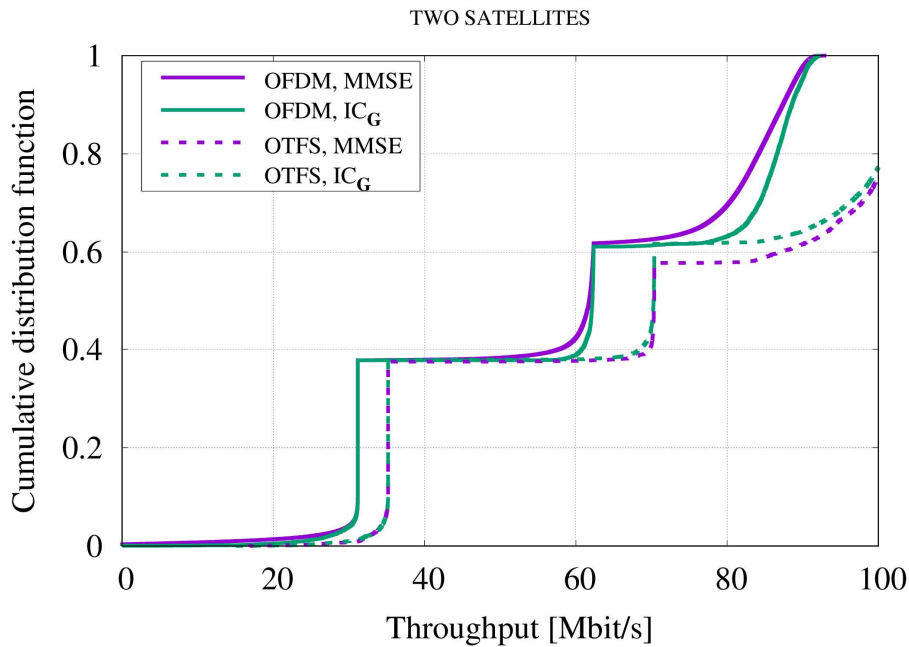


Figure 116: Cumulative distribution function of the pragmatic-capacity-based throughput for OFDM and OTFS with two transmitting satellites, fading channel and antenna arrays at the receive side (MRC).

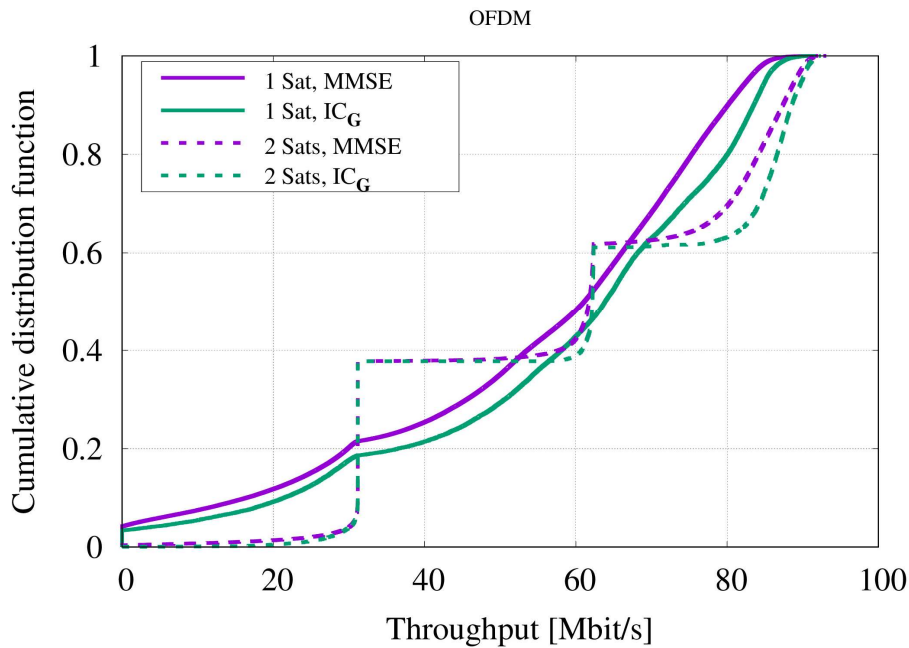


Figure133: Cumulative distribution function of the pragmatic-capacity-based throughput for OFDM with one or two transmitting satellites, fading channel and antenna arrays at the receive side (MRC).

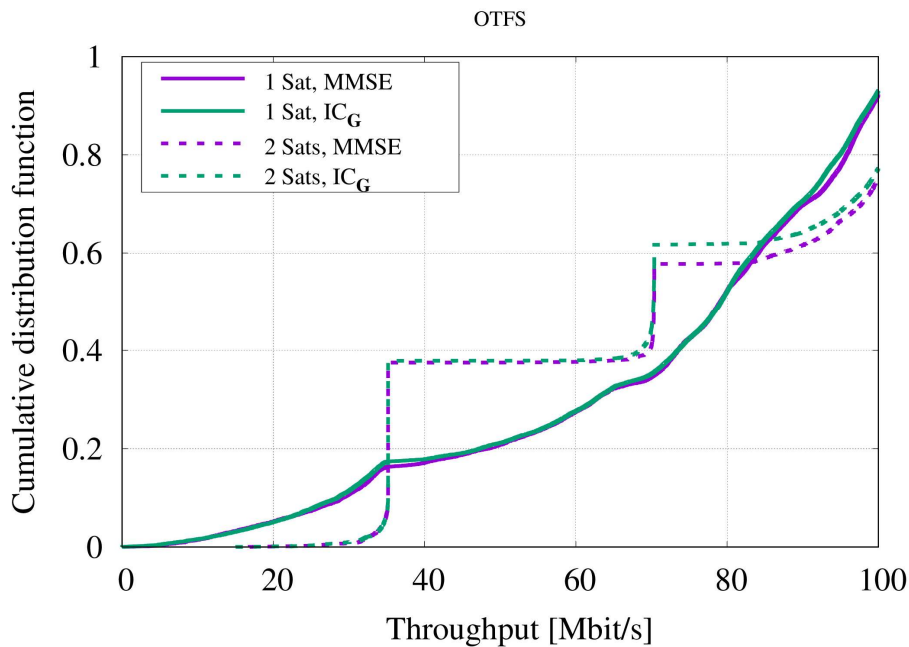


Figure134: Cumulative distribution function of the pragmatic-capacity-based throughput for OTFS with one or two transmitting satellites, fading channel and antenna arrays at the receive side (MRC).

Figure 131 and Figure 116 present results on the CDFs of one and two transmitting satellites, respectively. Receiver parameters were optimized to a large extent (to save computational time), whereas simulation parameters are same as listed in Table 32 except for the channel that is the NB channel previously described. The simulated scenario data (i.e., satellite orbital propagation and snapshots), nevertheless, are not the same as in the previous figures. The curves show that OTFS performs better than OFDM in the single-transmitting satellite case, whereas the gain is reduced in the two-transmitting satellites case. Figure 133 and 134 show the same results but with a comparison over the same modulation formats, to facilitate a deeper comprehension of the outcomes. It can be noticed that the case of two transmitting satellites (i.e., the cell-free case) presents almost staircase shapes of discrete distributions, because the SNR advantage entailed by the coherent sum of the received signals is not fully exploited by the system, and the pragmatic capacity saturates; in fact, the CSI reporting procedure was here set up to work on individual satellites, therefore the modulation and code rate scheme (MCS) selection was set upon a lower SNR than the actual one, which is the sum of the two, even if the shadowing penalty (i.e., the estimation of the channel attenuation that is supposed to be slower than a frame duration) is taken into account. Hence, even better results can be obtained if the information on the estimated SNR over each satellite-user link is shared among nodes (satellites) in the network, or just exploited by the central node, so that the MCS selection can take into account the overall SNR at each receiving user.

The following figures show the CDFs of the estimated throughput when the CSI reporting procedure evaluates this kind of information. Results outline a significant improvement that can be obtained by just selecting proper modulations and code rates based on data that are available at gateways.

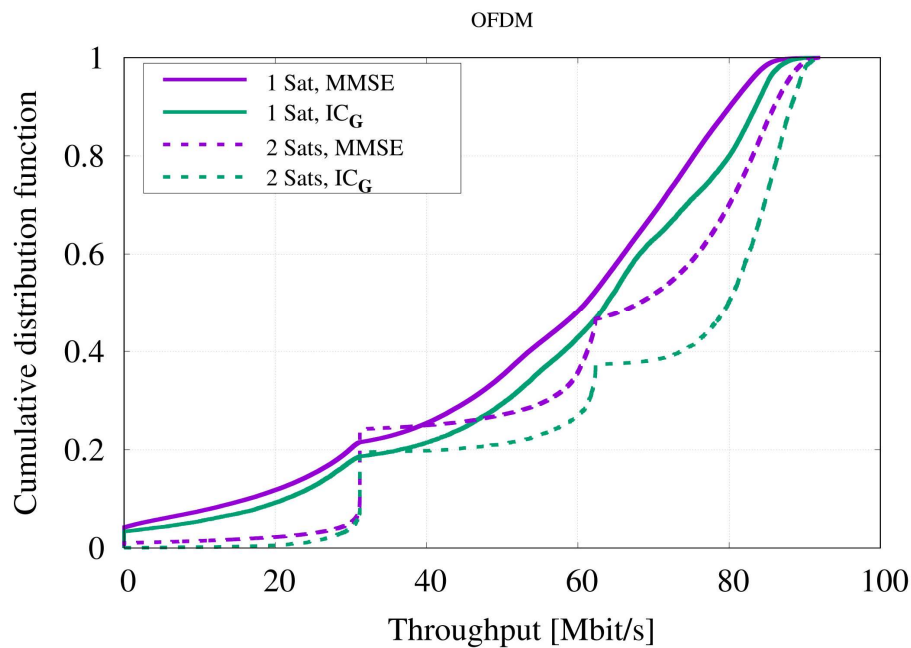


Figure 135: Cumulative distribution function of the pragmatic-capacity-based throughput for OFDM with one or two transmitting satellites, fading channel, antenna arrays at the receive side (MRC), and modified CSI procedure.

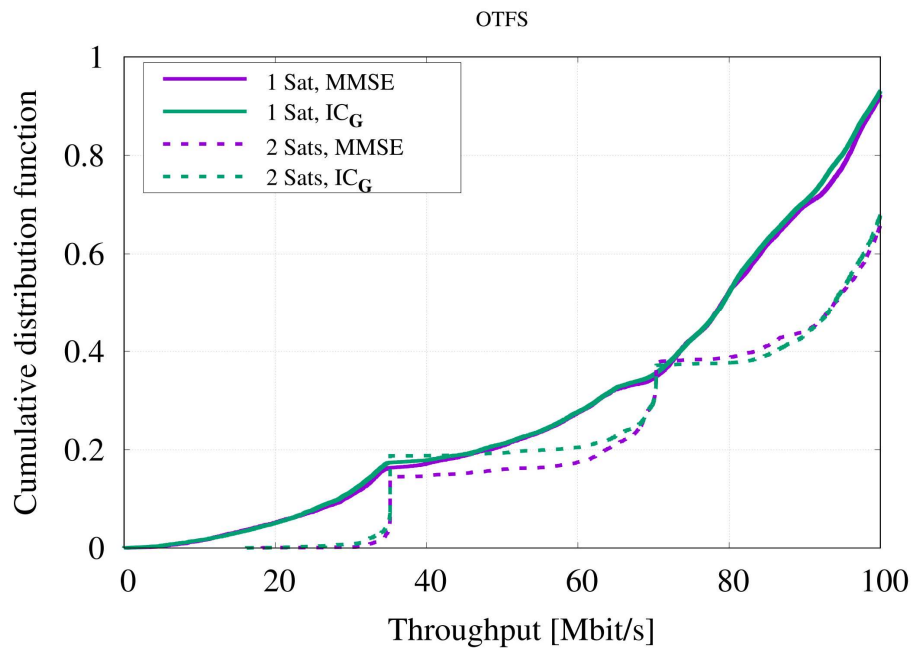


Figure 136: Cumulative distribution function of the pragmatic-capacity-based throughput for OTFS with one or two transmitting satellites, fading channel, antenna arrays at the receive side (MRC), and modified CSI procedure.

Finally, some results were carried out also in terms of block error rate (BLER), as for the single-transmitting satellite in clear-sky conditions, therefore including all the detection processing and the actual code rates from the standard MCS selection procedure, in case of CSI reporting for individual satellites. As for the previous results, a CSI correction margin was considered to emulate the channel penalty estimation which cannot be implemented on our specific scenario environment. These results must be evaluated as preliminary as further optimizations are possible (cfr. results in Table 33).

Table 35: Block error rate performance as a function of CSI margin

scenario	Modulation	Detector	CSI Margin [dB]		
			0	4	8
1 Sat	OFDM	IC _G	0.4152	0.1593	0.099
1 Sat	OFDM	MMSE	0.2961	0.1465	0.1190
1 Sat	OTFS	IC _G	0.1495	0.041	0.034
1 Sat	OTFS	MMSE	0.1566	0.046	0.029
2 Sats	OFDM	IC _G	0.0557	0.009	0.005
2 Sats	OFDM	MMSE	0.0322	0.021	0.015
2 Sats	OTFS	IC _G	0	0	0
2 Sats	OTFS	MMSE	0	0	0

5 SPECTRUM SHARING

Currently, 3GPP does not foresee spectrum sharing between TN and NTN. Available studies detail the adjacent channel leakage ratio (ACLR) and adjacent channel sensitivity (ACS) requirements to ensure that NTN and TN can operate in adjacent bands in the edges of Ka-band (17GHz DL and 27 GHz UL), so that the throughput degradation is limited to 5%, as detailed in [54]. The studies related to spectrum sharing are deferred to a later stage. The growing interest in the FR2 spectrum for both terrestrial and non-terrestrial applications highlight the need to investigate the coexistence of NTNs and TNs within these frequency ranges. The analysis considers that the satellite beams are synthesized to deliver user-specific data. Consequently, maximum EIRP is preserved. The beam widening techniques introduced in Section 3 are not considered.

The coexistence scenario is particularly challenging as the coverage area of satellite nodes contains several terrestrial base stations. In such a case, unless some actions are taken in the spatial and the terrestrial segment, interference events are bound to arise. Without proper spectrum management, frequent network handovers can disrupt services, making it difficult to maintain stable connections. To make progress towards the operation of TN and NTN in shared frequency resources, 5G-STAR DUST conducts an initial analysis. First, the operating bands are identified.

According to the channel arrangements specified in [55], NTN deployments in FR2 are allowed to operate in 3 frequency bands. The duplex mode of all these bands is frequency division duplex (FDD). In alignment with many Ka band links, the downlink band (Satellite to Earth Station) is the lower frequency. The details are provided in Table 36.

Table 36: Operating bands in FR2-NTN

Satellite operating band	Uplink (UL) operating band SAN receive / UE transmit $F_{UL,low} - F_{UL,high}$	Downlink (DL) operating band SAN transmit / UE receive $F_{DL,low} - F_{DL,high}$	Duplex mode
n512	27500 MHz - 30000 MHz	17300 MHz - 20200 MHz	FDD
n511	28350 MHz - 30000 MHz	17300 MHz - 20200 MHz	FDD
n510	27500 MHz - 28350 MHz	17300 MHz - 20200 MHz	FDD

The band n512 is applicable in the countries subject to European Conference of Postal and Telecommunications Administrations (CEPT) and Electronic Communications Committee (ECC) decisions. The bands n511 and n510 are applicable in the US market and the countries subject to Federal Communications Commission (FCC) rules.

The frequency bands in TN are specified in [56]. These bands operate in time division duplex (TDD) mode. The most relevant information is gathered in Table 37.

Table 37: Operating bands in FR2-TN

Satellite operating band	Uplink (UL) operating band BS receive / UE transmit $F_{UL,low} - F_{UL,high}$	Downlink (DL) operating band BS transmit / UE receive $F_{DL,low} - F_{DL,high}$	Duplex mode
n257	26500 MHz – 29500 MHz	26500 MHz – 29500 MHz	TDD
n258	24250 MHz -27500 MHz	24250 MHz -27500 MHz	TDD

n259	39500 MHz – 43500 MHz	39500 MHz – 43500 MHz	TDD
n260	37000 MHz – 40000 MHz	37000 MHz – 40000 MHz	TDD
n261	27500 MHz – 28350 MHz	27500 MHz – 28350 MHz	TDD
n262	47200 MHz – 48200 MHz	47200 MHz – 48200 MHz	TDD
n263	57000 MHz – 71000 MHz	57000 MHz – 71000 MHz	TDD

Notice that the overlapping occurs between n257 and n261 and the UL of n512, n511 and n510. In such a case, we can identify different coexistence scenarios in Table 38.

Table 38: Coexistence scenarios

	Interfering system	Victim system	Direction	Frequency	Bandwidth
Scenario 1	NTN UE	TN gNB	UL to UL	28 GHz	200 MHz
Scenario 2	TN UE	NTN gNB	UL to UL	28 GHz	200 MHz
Scenario 3	NTN UE	TN UE	UL to DL	28 GHz	200 MHz
Scenario 4	TN gNB	NTN gNB	DL to UL	28 GHz	200 MHz

The most detrimental situation is experienced when the unwanted signals transmitted from the interfering system leak through the main beam of the victim system. This event is known as in-line interference. It shall be noted that in-line interference is unlikely to happen in Scenario 3 and Scenario 4. Further details are provided hereinafter upon defining the parameters.

To model the terrestrial deployments, it is assumed that base stations are in urban environments.

5.1 PARAMETERS

SATELLITE PARAMETERS

This deliverable assumes the payload antenna model introduced in [2], featuring four apertures (TX-LHCP, TX-RHCP, RX-LHCP, RX-RHCP), each consisting of 512 radiating elements in an approximated circle shape with a Hexa-Deca structure, and inter-element spacing of 0.65λ . The resulting beam footprint at Nadir has a radius of approximately 41 km. The gain of each element is 4.7 dBi, their pattern can be modelled by a $\cos^2()$ function and each have an associated power amplifier of 65 mW. Moreover, we assume digital beamforming, so beamforming weights can be updated at OFDM symbol level. The resulting direct radiating array (DRA) has several degrees of freedom in the generation of beam layouts.

Table 39: Satellite parameters

Parameters	Value
UL carrier Frequency	28 GHz
Bandwidth	200 MHz
G/T (dB K ⁻¹)	6.58

NTN UE PARAMETERS

In FR2, as recommended by 3GPP for system level evaluations, the normalized antenna pattern of a Very Small Aperture Terminal (VSAT) UE is

$$G_{UE}(\theta) = \begin{cases} 4 \left| \frac{J_1(2\pi f_c d_A \sin\theta/c)}{2\pi f_c d_A \sin\theta/c} \right|^2 & \theta \neq 0 \\ 1 & \theta = 0, \end{cases} \quad (124)$$

where $J_1(\cdot)$ is the Bessel function of the first kind and first order with the argument, d_A is the antenna aperture radius, θ is the angle measured from the antenna boresight and c is the light propagation speed. The main parameters are listed in Table 40 according to [18] and [54].

Table 40: NTN UE parameters

Parameters	Value
UL carrier Frequency (GHz)	28
Bandwidth (MHz)	200
Tx Antenna Gain (dBi)	39.7
Rx Antenna Gain (dBi)	43.2
Transmit power (dBm)	33
Antenna altitude (m)	22

TN GNB PARAMETERS

The beamforming antenna is based on an antenna array and consists of a number of identical radiating elements with a fixed inter-antenna separation distance. We adopt the 3GPP antenna field pattern defined in [57]. Accordingly, the element pattern is given by

$$G_E(\theta, \phi) = G_{MAX} - \min\{-[G_V(\phi) + G_H(\phi)], A_m\}, \quad (125)$$

where θ, ϕ are the elevation and azimuth angles, respectively. The reference coordinate system is defined in Figure 137.

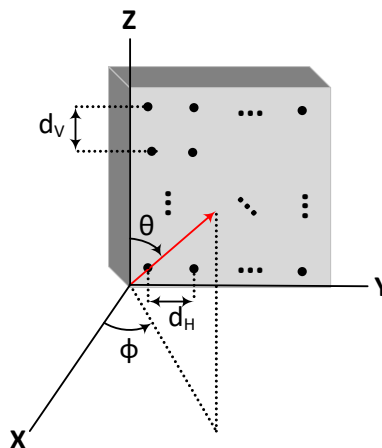


Figure137: Definition of θ, ϕ in the TN gNB coordinate system

The element parameters are defined in Table 41.

Table 41: Base station element pattern

Parameters	Value
Horizontal Radiation Pattern (dB)	$G_H(\phi) = -\min\left\{12\left[\frac{\phi}{\phi_{3dB}}\right]^2, A_m\right\}$
Horizontal and vertical 3 dB beamwidths (°)	$\phi_{3dB} = 90, \theta_{3dB} = 90$
Front-to-back ratio (dB)	$A_m = 30$

Vertical Radiation Pattern (dB)	$G_V(\theta) = -\min \left\{ 12 \left[\frac{(\theta - 90^\circ)^2}{\theta_{3dB}} \right], SLA_v \right\}$
Side Lobe lower level (dB)	$SLA_v = 30$
Element Gain (dBi)	$G_{MAX} = 5.5$

The composite array pattern is formulated as

$$G(\theta, \phi) = G_E(\theta, \phi) + 20 \log_{10} \left(\sum_{n=1}^{N_V} \sum_{m=1}^{N_H} v_{m,n} w_{m,n} \right). \quad (126)$$

The components of the steering vector are given by

$$v_{m,n} = e^{j2\pi \left((n-1) \frac{d_V}{\lambda} \cos(\theta) - (m-1) \frac{d_H}{\lambda} \sin(\theta) \sin(\phi) \right)}. \quad (127)$$

for $n = 0, \dots, N_V - 1$ and $m = 0, \dots, N_H - 1$. Hence, the total number of antenna elements is $N_V N_H$. The variables d_V and d_H represent the vertical and horizontal inter-antenna spacing, respectively, and λ denotes the wavelength.

To synthesize a beam in the target direction (θ_t, ϕ_t) , the antenna coefficients are computed as

$$w_{m,n} = \frac{1}{\sqrt{N_H N_V}} e^{-j2\pi \left((n-1) \frac{d_V}{\lambda} \sin(\theta_t) - (m-1) \frac{d_H}{\lambda} \cos(\theta_t) \sin(\phi_t) \right)}. \quad (128)$$

where θ_t is the electrical down tilt steering and ϕ_t is the electrical horizontal steering. The rest of the parameters are listed in Table 42 according to [58] and [54].

Table 42: TN gNB parameters

Parameters	Value
Cellular layout	Hexagonal, 3 sectors/cell (19 cell wrap-around), uncoordinated
Inter-site distance (m)	300
Antenna array configuration	$N_H = 16, N_V = 8$
Bandwidth (MHz)	200
Horizontal/Vertical radiating element spacing	$d_V = 0.5\lambda, d_H = 0.5\lambda$
Antenna height (m)	25
Down tilt ($^\circ$)	10 (urban environment)
UL and DL carrier Frequency (GHz)	28
Noise figure (dB)	10
Array Ohmic loss (dB)	2
Conducted power (before Ohmic loss) per antenna element (dBm)	22
Vertical coverage range ($^\circ$)	90-120

TN UE PARAMETERS

TN UE antenna adopts the same model as the TN gNB antenna. The radiating elements are characterized by the values provided in Table 43. The rest of the parameters are listed in Table 44.

Table 43: UE element pattern

Parameters	Value
------------	-------

Horizontal Radiation Pattern (dB)	$G_H(\phi) = -\min\left\{12\left[\frac{\phi}{\phi_{3dB}}\right]^2, A_m\right\}$
Horizontal and vertical 3 dB beamwidths (°)	$\phi_{3dB} = 90, \theta_{3dB} = 90$
Front-to-back ratio (dB)	$A_m = 25$
Vertical Radiation Pattern (dB)	$G_V(\theta) = -\min\left\{12\left[\frac{(\theta - 90^\circ)}{\theta_{3dB}}\right]^2, SLA_v\right\}$
Side Lobe lower level (dB)	$SLA_v = 25$
Element Gain (dBi)	$G_{MAX} = 5$

Table 44: TN UE parameters

Parameters	Value
UE distribution	Average 10 UEs per sector. UEs on flat ground. 100% outdoors
Antenna array configuration	$N_H = 2, N_V = 2$
Bandwidth	200 MHz
Horizontal/Vertical radiating element spacing	$d_V = 0.5\lambda, d_H = 0.5\lambda$
Transmit power (dBm)	23
Antenna height (m)	1.5
UL and DL carrier Frequency (GHz)	28
Noise figure (dB)	10

5.2 INTERFERENCE ANALYSIS

This section evaluates the interference power received by the victim system, which is dictated by the coexistence scenario. To conduct the analysis, it is assumed that

- The TN gNB and the TN UE shall be in the field of view of the serving satellite;
- The NTN UE points to the serving satellite accurately;
- 5G-STAR DUST constellation is used to provision service to NTN users;
- 250m as maximum distance assumed between NTN UE and TN gNB.

The performance metric is the interference-to-noise ratio (INR), which is expressed in the logarithmic scale as

$$INR[dB] = P_T[dBW] + G_T(\Delta\theta_T, \Delta\phi_T)[dB] + \frac{G_R(\Delta\theta_R, \Delta\phi_R)}{T} [dB/K] - 10 \log_{10}(K_B)[dBW/K/Hz] - 10 \log_{10}(B_W)[dBHz] - P_L[dB]. \quad (129)$$

In notation terms, $P_T, G_T(\theta_T, \phi_T), G_R(\theta_R, \phi_R), P_L$ are respectively the interfering system transmit power, the transmit antenna gain of the interfering system at the boresight offset $(\Delta\theta_T, \Delta\phi_T)$, the receive antenna gain of the victim system at the boresight offset $(\Delta\theta_R, \Delta\phi_R)$ and the propagation loss between the interfering system and victim system. The noise on the victim system is formulated as function of the Boltzmann constant $K_B(-228 dBW/K/Hz)$, the temperature T and the bandwidth B_W .

For the computation of the system temperature, the following formula applies

$$T = T_0(10^{0.1NF} - 1) \quad (130)$$

where T_0 is the ambient temperature (usually 290 K) and NF is the noise figure in dB.

INTERFERENCE ANALYSIS FOR COEXISTENCE SCENARIOS 1 AND 3

This section evaluates the interference power that single TN gNB or TN UE receive from the NTN UE. The interference level received by the TN gNB and the TN UE depends on the position of the satellite and the location of the NTN UE, which dictates the user elevation angle θ_{EL} . According to information reported in [1], the range of this angle is defined by $30^\circ \leq \theta_{EL} \leq 90^\circ$.

In the most detrimental situation, the main lobe of the NTN UE antenna is directed towards the TN gNB. Since the terrestrial base station points to the ground, this in-line interference event depends on the down tilt of the terrestrial antenna and the minimum user elevation angle. Thus, the higher the height of terrestrial base stations, the higher the likelihood that any of the beams synthesized by the TN gNB could be along the path that connects the NTN UE and the satellite. This situation is only encountered if the NTN UE and the TN gNB are in close proximity, e.g., less than 1km.

Nevertheless, assuming that terminal antennas are oriented toward the base station at a given elevation angle, in-line interference events between NTN UEs and TN UEs are unlikely to occur. As a result, the interference level will be significantly higher in Scenario 1 compared to Scenario 3. Based on that, the subsequent analysis will just focus on Scenario 1.

The situations discussed above are illustrated in Figure 138.

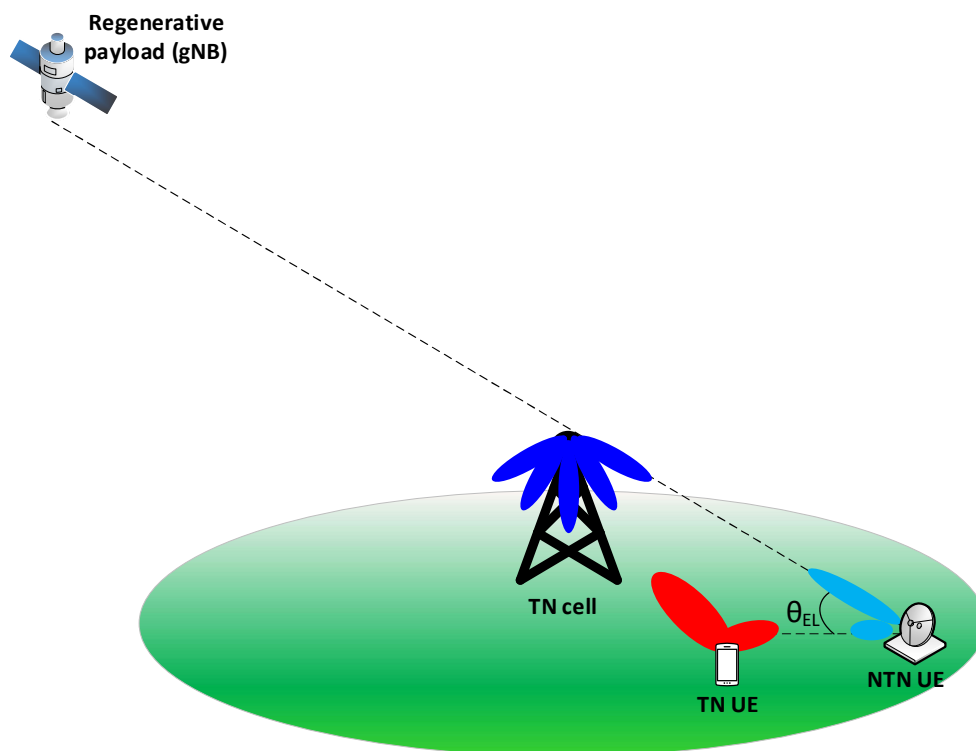


Figure138: System model for coexistence scenarios 1 and 3.

The relevant information to define the path loss model is obtained from [58]. The deployment scenario that is considered corresponds to the urban macro scenario. The model differentiates between LoS and NLoS. Nevertheless, to analyse the coexistence between NTN and NT, only

LoS will be considered. Essentially, because it corresponds to the most harmful situation in terms of interference.

Based on the model defined in [58], the signal path between the interfering NTN UE and the TN gNB experiences several stages of attenuation, as outlined below:

$$P_L = \begin{cases} PL_1 & 10m \leq d_{2D} \leq d_{BP} \\ PL_2 & d_{BP} \leq d_{2D} \leq 5km \end{cases} \quad (131)$$

where

$$PL_1 = 28.0 + 22 \log_{10}(d_{3D}) + 20 \log_{10}(f_c)$$

$$PL_2 = 28.0 + 40 \log_{10}(d_{3D}) + 20 \log_{10}(f_c) - 9 \log_{10}((d_{BP})^2 + (h_{BS} - h_{UT})^2).$$

The expressions depend on the average building height, which is set to $h = 5m$, and the distance breakpoint

$$d_{BP} = \frac{4h_{BS}h_{UT}f_c \times 10^9}{c}.$$

In notation terms, f_c is the carrier frequency normalized by 1GHz, $c = 3 \times 10^8 m/s$ is the propagation velocity in free space. The antenna heights are respectively h_{BS} and h_{UT} . The default parameters are provided in Table 41 and Table 42. The distance definitions are indicated in Figure 139. The distance that is used to computing the path loss is obtained as

$$d_{3D} = \sqrt{(d_{2D})^2 + (h_{BS} - h_{UT})^2}.$$

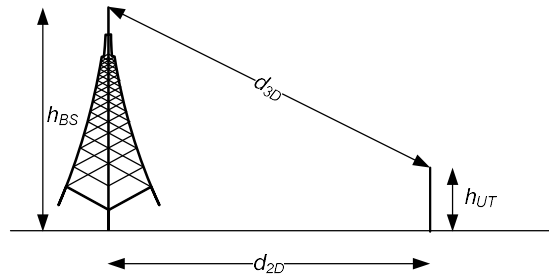


Figure139: Definition of d_{2D} and d_{3D}

INTERFERENCE ANALYSIS FOR COEXISTENCE SCENARIOS 2 AND 4

This section evaluates the interference power that single NTN gNB receives from the TN gNB or the TN UE. The interference level received by the NTN gNB depends on the position of the satellite and the pointing direction of TN UE and TN gNB antennas.

Since the antenna employed by the TN gNB is expected to be directive and pointing towards the ground, the corresponding leakage observed by the satellite may not significantly contribute to the performance degradation. Based on that, the subsequent analysis will just focus on Scenario 2.

Although the satellite cannot be positioned along the path that connects the TN gNB and the TN UE, it still could receive some interference in Scenario 2. To keep at a reasonable level

complexity and power consumption, the array that is employed by the TN UE typically has a low number of antenna elements. The resulting beamwidth may radiate power within an angular sector where the satellite is visible. In this scenario, the aggregated interference from multiple TN UEs might degrade the NTN link.

The scenario under study is illustrated in Figure 140.

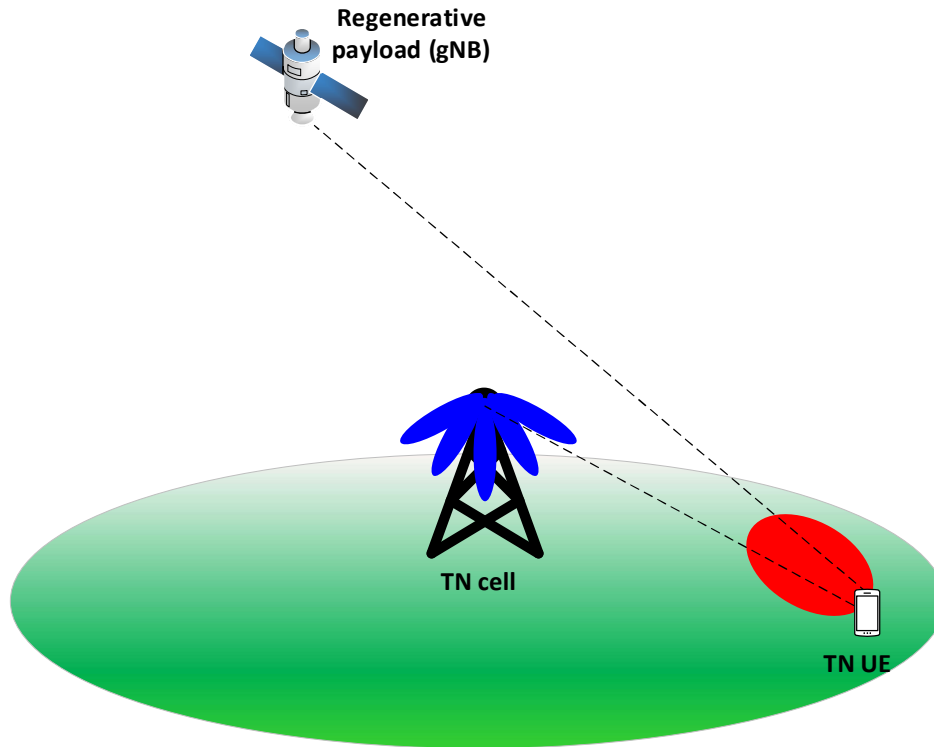


Figure140: System model for coexistence scenarios 2 and 4.

The signal path between the NTN gNB and the interfering terrestrial solely experiences the free space loss, namely,

$$P_L = L_{fs},$$

It is worthwhile mentioning that this is the basic path loss model. The free space loss at a distance d in meters and frequency f in GHz is given by

$$L_{fsl} = 32.45 + 20 \log_{10}(f) + 20 \log_{10}(d).$$

5.3 NUMERICAL RESULTS

This section provides some numerical results to understand the detrimental effects when TN and NTN operate on shared frequency resources. The INR metric is provided in Table 45 using the system parameters specified in Section 5.1. Results are provided for the coexistence scenario 1 and 2, as they represent the most harmful situations.

5.3.1 Scenario 1

As expected, in scenario 1 the magnitude of the interference increases, as the satellite elevation angle decreases. In this situation, the direction of the interference is gradually aligning with the boresight of the victim system. Consequently, at low satellite elevation angles, the interference leaks through the main beam of the victim system. In Scenario 1, some additional measures shall be considered to reduce the magnitude of the interference, specially at low satellite elevation angles.

Table 45: INR in coexistence scenarios 1 for different satellite elevation angles

Satellite elevation angle	60°	45°	30°
INR[dB] (Scenario 1)	-12.4	-4.9	4.6

5.3.2 Scenario 2

To conduct the performance assessment in Scenario 2, it is considered that the satellite reception is interfered from the uplink transmission of a 19-cell terrestrial cluster.

The simulation parameters are summarized in Table 46 where a carrier frequency of 28 GHz is considered with a LEO satellite at an altitude of 1300 Km and antenna pattern as defined in [2].

Table 46: simulation parameters

Category	Parameter	Value
Physical Parameters	Carrier frequency	28 GHz
	Satellite altitude	1300 km
UE and Transmit Parameters	Transmit power	-7 dBW
	System bandwidth	200 MHz
	UE antenna elements	2 × 2
Cellular Layout	Inter-site distance	300 m
	Number of cells	19
	Cell radius	173.2 m
UE Distribution	UEs per cell	10
	UE distance from gNB	10–150 m (uniform)
	UE azimuth angle	0–360° (uniform)
Satellite Geometry	Elevation angles	30°, 40°, 50°, 60°, 70°, 80°, 90°
Antenna and channel Models	Path loss model	Free-space path loss
	Satellite and UE antenna	Section 5.1
Monte Carlo Setup	Monte Carlo trials	100,000

A 19-cell hexagonal cellular layout is generated using an inter-site distance of 300 m and a two-ring configuration. Within this network, UEs are randomly distributed for each Monte Carlo trial. Specifically, 10 UEs per cell are placed with uniformly distributed angles between 0° and 360° and radial distances uniformly drawn between 10m and 150 m from their serving gNB. Figure 141 illustrates a snapshot from one of the trials, showing the 19-cell hexagonal cluster with both gNBs and UEs.

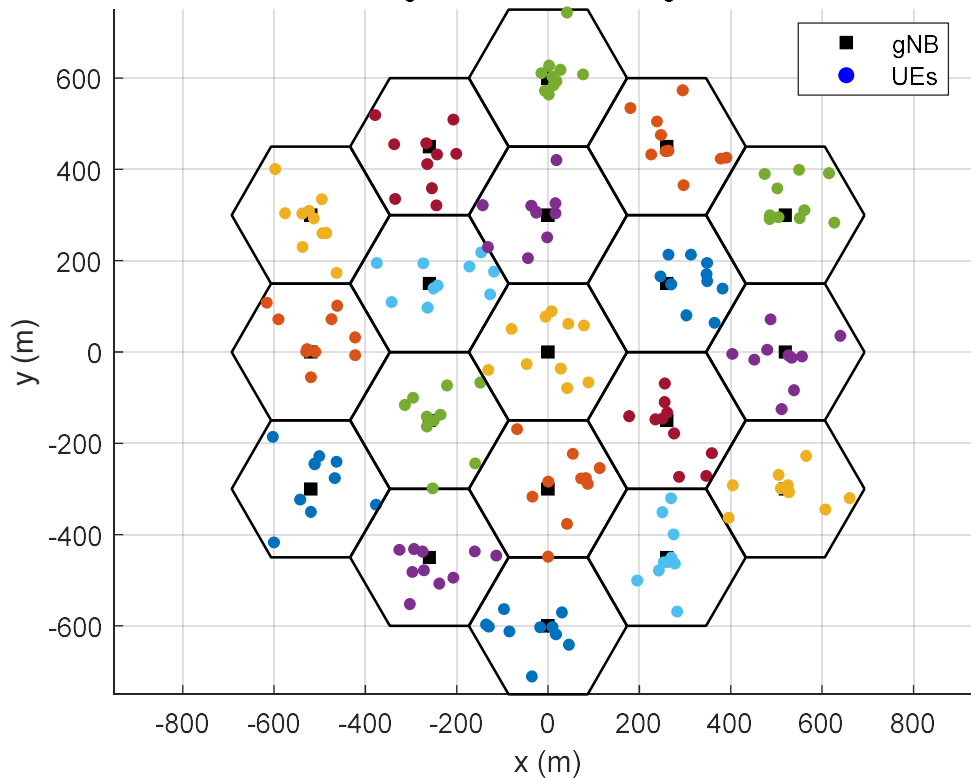


Figure 141: Spatial map of the 19-cell cluster: cell hexagons, gNB locations (black squares) and the UE positions.

For the considered satellite elevation angle, the satellite position is determined relative to the different gNBs. The slant range between each UE and the satellite is then computed, along with the corresponding elevation angle from the UE to the satellite. A free-space model path loss model is employed to calculate the received interference at the satellite side. The INR contributions from all UEs across all cells are aggregated to obtain the total INR for each Monte Carlo trial. This process is repeated over a sufficient number of trials to ensure statistical reliability.

Figure 142 illustrates the CDF of the accumulated INR at the satellite for different satellite elevation angles. It can be observed that the CDF curves shift progressively towards higher INR values as the satellite elevation increases, indicating a higher probability of stronger interference at higher elevations. This behaviour is primarily attributed to the reduced slant range between the UEs and the satellite, which leads to lower pathloss and consequently higher received interference power.

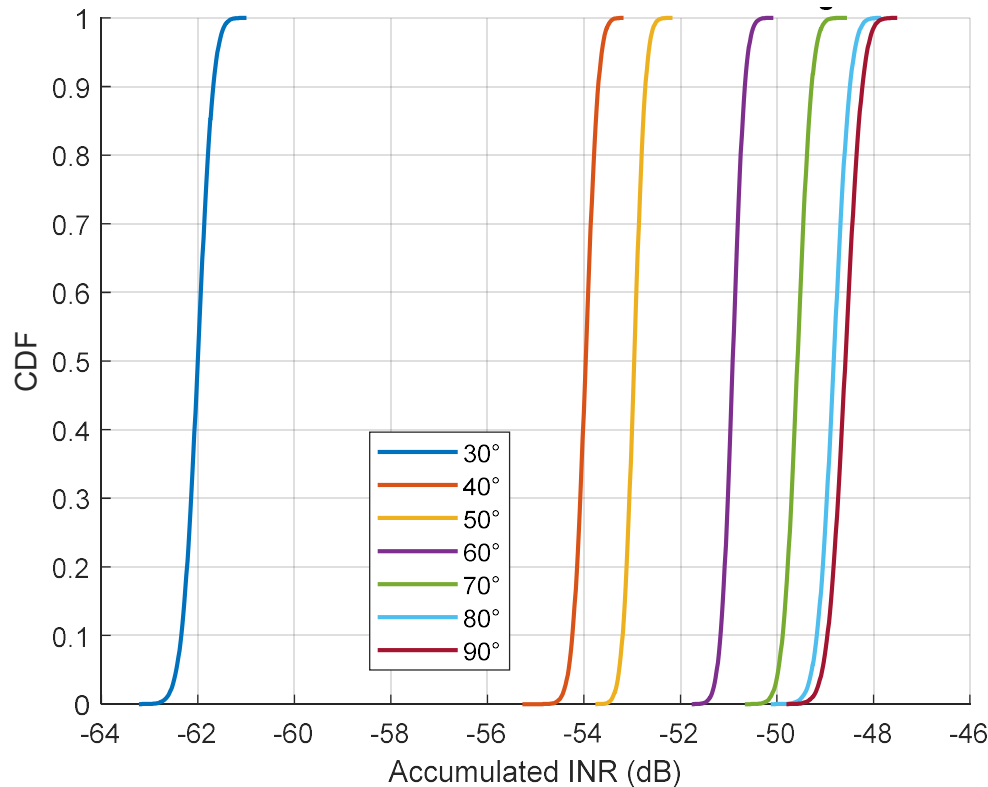


Figure 142: CDFs of accumulated INR (dB) at the satellite for elevation angles ranging from 30° to 90°, based on 100,000 Monte Carlo trials per elevation angle.

Figure 143 presents the mean aggregated INR at the satellite receiver as a function of the satellite elevation angle. The curve shows a clear increasing trend, which is consistent with the result of Figure 142. That is, higher elevation angles correspond to shorter propagation distances and reduced free-space path loss. Specifically, the mean INR increases from approximately -62 dB at 30° to about -49.5 dB at 90°. The resulting interference in the simulation is generally considered non-critical from a system performance perspective as the thermal noise remains the dominant impairment. The results highlight the strong dependence of the INR on network parameters. Nevertheless, a noticeable impact on satellite reception would only occur if a very large number of clusters were simultaneously deployed within the beam. Considering that the radius of satellite beams typically ranges between 50 and 100 km together with the results presented in this section, it appears unrealistic that terrestrial cellular users could significantly degrade the quality of the NTN link. This is primarily due to the high propagation losses experienced in the Ka band. Another important aspect is that terrestrial UEs synthesize directional beams to connect to the base station. As a result, the power that is radiated towards the satellite depends on the antenna pointing direction. This means that only a small subset of interfering signals leaks through the main lobes, resulting in a limited impact on satellite reception. Under these conditions, uplink power control or scheduling constraints are not essential for protecting the satellite link.

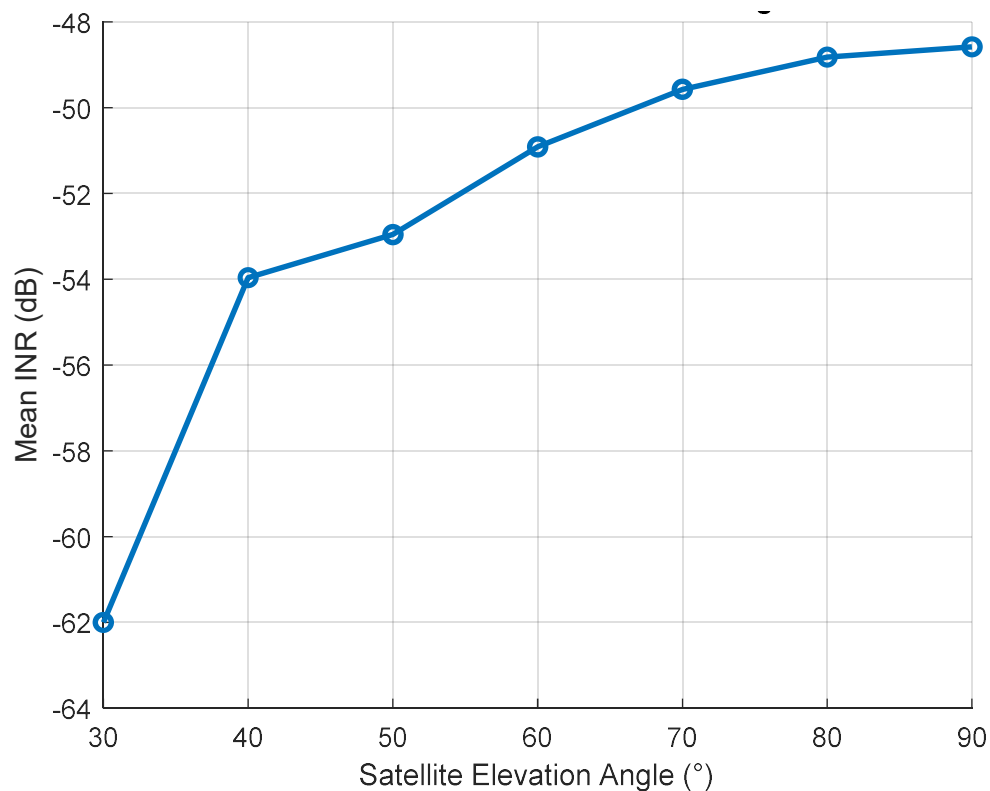


Figure 143 Mean accumulated INR (dB) at the satellite vs satellite elevation angle, computed over 100,000 Monte Carlo trials per elevation.

6 CONCLUSIONS

This concluding section provides the main findings and draws the conclusion.

Mapping schemes between NTN beams and NR cells

The impact of two different mapping schemes on intra-satellite mobility aspects has been analysed. The one-to-one mapping, where each NTN beam is associated with a NR cell handles mobility across beams using the HO mechanism. It is important to remark that the HO procedure uses L3 signalling. To reduce signalling overhead and latency, the many-to-one mapping scheme is more attractive. In this case, multiple NTN beams are associated with a single NR cell, allowing mobility to be managed through beam management rather than frequent HOs. This significantly reduces signalling and latency, making many-to-one mapping the preferred option for NTN deployments. Furthermore, the multi-beam NTN scenario provides the basis for designing efficient beam hopping schemes.

Beam hopping schemes

Beam management forms the foundation to implement periodic beam hopping. Within this scope, the deliverable focuses on two key aspects: the scheduling of control information and the design of beam footprints and layouts.

With respect to the beam lattice, the preferred priority is given to have a fixed grid on Earth. This is interesting from a system point of view, as a stationary user that is inside the FoV of the satellite will be served by the same beam in the visibility period. Based on the beam shapes and the grids, different options have been compared. One possibility is to deploy a homogenous grid according to the footprint radius at Nadir. The consequence is that the FoV is divided into a large number of beams and, thus, large overlapping will be experienced at the edge due to the beam deformation. The alternative is to reduce the number of beams. This is achieved by increasing the crossover level beyond 3 dB. This approach outperforms other strategies, such as beam widening through phase tapering, by lowering both the number of beams and edge overlapping. Building on this method, a strategy has been developed to select the set of beams illuminated simultaneously. The approach maps consecutive beam indices into rectangular blocks, repeating each index across latitude and longitude. By selecting beams from different blocks, it is ensured that active beams do not overlap.

Building upon the proposed mapping as well as the beam footprint and layout designs, frame structures for both UL and DL have been investigated with the objective of reducing signaling overhead. The flexibility provided by the standard makes it possible to schedule SSBs, PDCCCH and PDSCH for SIB1, SIB19, msg2, msg4 and paging messages within the same half frame. This scheduling approach enables the system to sweep the coverage area with a periodicity of 20 ms. As a result, the default SSB periodicity can be reused without affecting the compatibility with legacy UEs.

In the UL, FoV coverage is mainly required for RACH signaling. To avoid complex solutions, the UL beam illumination plan is designed using the same beam pattern as in the DL. Consequently, the coverage area is divided into the same number of beams for both uplink and downlink. For the allocation of random access resources in UL frames, multiple beams are served within the same frame. Based on the beam-hopping evaluation, the RACH periodicity defined in the standard (up to 160 ms) is sufficient to ensure 100% UL coverage.

It is important to remark that Earth-fixed beam layout becomes inefficient in terms of coverage. The main reason is that a fixed regular grid on Earth does not match well with the beam footprint widening caused by the Earth curvature. Depending on the strategy, it may result in

large beam overlapping, coverage holes or in reduced beam gains. Given these limitations, the beam illumination pattern has been also investigated for Earth-moving beams. Earth-moving beams offer two key advantages over Earth-fixed beams. First, fixing a regular grid in the angular domain of the satellite antenna results in a non-regular grid of beam footprints on Earth, where the inter-beam distance naturally increases as the beams widen due to curvature. This enables full coverage of the region of interest with fewer beams and with a consistent overlap across the entire field of view. Second, the beamforming weights can remain fixed and do not require frequent updates as the satellite passes over the service area. These benefits translate into a more efficient scheme for broadcasting the common signalling. In particular, the number of slots allocated to common signalling within an SSB period can be reduced by up to 24% when using Earth-moving beams compared with Earth-fixed beams. On the negative side, the signalling required to manage mobility can increase significantly in the Earth-moving design, since the beams move along with the satellite trajectory.

System-level beamforming

With respect to system-level beamforming aspects, we aimed at: i) defining the system-level procedures to support CSI-based and location-based beamforming and estimating the corresponding reporting overheads; ii) designing a methodology to derive databases for non-uniform traffic distributions; iii) design scheduling algorithms with flexible priority metrics; iv) assess the performance with ideal and non-ideal ancillary information; and v) design and assess the performance of federated beamforming from multiple satellites.

The main outcomes and future activities are reported hereafter:

- The system-level procedures designed in [1] for standalone algorithms have been applied without relevant modifications also to federated beamforming. In terms of overhead, it was shown that CSI-based techniques are extremely challenging, requiring up to 32 Mbit every 80 ms assuming 1000 users in the service area and 32 bits per complex channel coefficient with the 512 radiating elements antenna designed in 5G-STARLUST. In addition, the availability of channel estimates at radiating element level also calls for proper modifications to the CSI-RS or specifications in general. As for location-based techniques, these require much less overhead, but they might pose an issue from the privacy point of view (users' locations are not available at RAN level, typically). However, it was shown that the performance is not particularly impacted by non-ideal location estimation; this aspect suggests that approximate locations, not posing privacy issues, might be exploited without affecting the achieved capacity.
- A simple and flexible methodology for the generation of non-uniform traffic distributions over land areas has been defined by exploiting the EC JRC population dataset. This tool allows to easily define traffic ranges in terms of the minimum and maximum capacity request and also identify the specific area of interest (rural, sub-urban, urban, dense-urban). In addition, a simple methodology for air traffic scenarios has also been defined.
- Two scheduling algorithms have been designed: i) PQS, inspired by priority queueing principles, in which multiple mutually exclusive classes can be introduced to classify the inbound packets; and ii) PRS, in which the definition of the classes is not needed and the users are classified in a single look-up table taking into account the residual unmet capacity and visibility time. The two algorithms show a good performance, with the PRS outperforming the PQS in terms of flexibility (no priority class has to be defined) and achieved capacity (it is better tailored to the current users' conditions); it was also shown that the two algorithms have the same big-o computational complexity. For both algorithms, two scheduling constraints (CSI-based and location-based) have

been introduced to avoid ill-conditioned channel matrices to have a better performance.

- The performance assessment showed that: i) as mentioned above, PRS provides a better achieved capacity and lower unmet capacity, thanks to its inherent flexibility compared to the need for PCs in PQS; ii) in the standalone NTN node scenario, the system is capable of providing 25-28 Gbps over a visibility period of approximately 7 minutes for a service area that comprises most of the European land areas. However, in dense traffic scenarios (in particular due to an increased minimum capacity request), a significant amount of unmet capacity is still present. Since the scheduling constraints and algorithms have been numerically optimised, this outcome suggests that the performance is close to its optimum and system-level modifications would be needed to improve it, such as: i) increasing the number of satellites while increasing the minimum elevation angle (which reduces the path losses thus improving the overall SINR), increasing the available on-board power, reducing the number of users (we extensively discussed the impact of the number of allocations per user in the study), or using multiple satellites in a federated approach.
- Comparing CSI-based and location-based MMSE algorithms, since we are in clear-sky conditions no difference arises with ideal ancillary information. Notably, in LOS or NLOS channels, the location-based approach might suffer from the additional misalignment between the estimated and the actual channel matrices due to the lack of knowledge on the stochastic factors in the channel coefficients. When considering non-ideal ancillary information, the location-based approach is almost unaffected by location errors even in the order of 10 km; this is almost expected, since an error of a few km when considering a satellite orbiting the Earth at 1000 km altitude yields to a negligible difference in the slant range. However, for LOS/NLOS channels, these observations shall be further verified. The performance loss for CSI-based solutions with non-ideal estimation is significant, in particular in low SNR regimes and/or with less symbols for the data-aided estimation. Notably, the performance might be improved by designing novel and more advanced estimation algorithms or by using more symbols for the estimation. Finally, an extensive analysis has been performed related to the CSI or geographical scheduling constraints to be used to avoid ill-conditioned channel matrices, providing valuable insights for the current study and also future activities.
- The performance in the federated beamforming case follows the same trends discussed above. The numerical assessment was performed comparing the standalone case with two different implementations of federated MIMO: i) equal power, in which the NTN nodes in the swarm collectively have the same on-board power of the standalone node; and ii) maximum power, in which each NTN node in the swarm has the power of a single node. Notably, the power boost in the latter case yields to a significantly better performance. This is in line with the previous observation on the need for system-level modifications to improve the performance in the standalone beamforming architecture.

To further improve the system performance, solutions based on AI/ML can clearly be designed and implemented. This activity has been performed in Task 4.3 and reported in D4.7.

Further activities in this area will cover, *e.g.*, the performance assessment in LOS and NLOS channel conditions, the design of more advanced scheduling algorithms (to improve the capacity performance and/or reduce the computational complexity), and the optimisation of the constellation design targeting the application of federated beamforming (*i.e.*, understanding if there are optimal constellation sizing options to further improve the performance).

A performance comparison has also been performed to compare the user-centric PRS designed in Task 4.2 with a beam-centric variant, BC-PRS, which takes into account the beam lattices optimised in Task 4.1. As expected, the performance of the user-centric solution is better in terms of both achieved and unmet capacity; however, the performed analyses highlight a potential direction for future research: the joint optimisation of the beam lattice in terms of signalling overhead and beamforming performance.

OTFS

In this deliverable the feasibility of OTFS as a possible waveform for satellite communications based on the NR standard was demonstrated. Several simplified detectors were designed, among which one was implemented and simulated, showing a good performance/complexity trade-off. Then, a specific architecture to map the OTFS delay-Doppler grid to the NR OFDM standard time-frequency grid was proposed, and numerical results demonstrated convincing results. A first preliminary scenario with two transmitting satellites was also investigated, to demonstrate the feasibility of such waveform in a user-centric system. Results in clear-sky condition do not highlight evident benefits, as expected, although with a more realistic channel (i.e., in presence of at least shadowing) the proposed system would likely take advantage of the OTFS intrinsic robustness and capability to exploit multiple path signalling.

When the realistic fading channel was considered, the expected advantage of the cell-free paradigm with two serving satellites per user could be measured and proved. Results showed the gain that can be obtained especially when coherent combining of the received signals is enabled by the complexity increase required to separately process the signals (i.e., by using directive antennas and duplicated RF chains and channel estimation at the receive side). Further gains can be obtained if the CSI reporting procedure includes the overall SNR estimated by the user so that proper modulation and coding schemes can be selected based on a more accurate estimation. Numerical results were provided mainly in terms of pragmatic capacity and block error rates, therefore accounting for both actual modulations and code rates, and the novel detector was demonstrated to perform likewise the more complex standard MMSE.

Spectrum sharing

The growing interest in the FR2 spectrum for both terrestrial and non-terrestrial applications has motivated the study on TN-NTN coexistence within these frequency ranges. This deliverable analyses different scenarios, focusing on interference in both directions: the power received by terrestrial gNBs and UEs from satellite terminals, and the interference induced by the terrestrial segment on the satellite. The analysis reveals that the most severe degradation occurs when NTN UEs interfere with TN gNBs, and when TN UEs disrupt the reception of the NTN gNB hosted on the satellite. Numerical evaluations indicate that even a single NTN UE can significantly impair the uplink performance of a terrestrial deployment. By contrast, a single TN UE has negligible impact on satellite reception, while the aggregate interference generated by multiple UEs could severely affect the quality of the satellite link. However, in the scenario under study where satellites operate at the Ka band, the numerical results suggest that the impact on satellite reception is minimal, primarily due to the high propagation losses characteristic of this frequency range. Another relevant aspect is that terrestrial UEs synthesize directional beams to communicate with the base station. As a result, the power that is radiated towards the satellite depends on the antenna pointing direction. This means that only a small subset of interfering signals leaks through the main lobes, resulting in a limited impact on satellite reception. In light of these findings, mitigation measures should prioritise the TN-NTN coexistence scenario where NTN UEs interfere TN gNBs. In this case, to prevent service disruption and reduce the interference, different solutions can be envisaged, ranging from

complete coordination up to minimal information exchange between networks. The most relevant strategies include:

- Inter-cell interference coordination scheme;
- Spatial processing;
- Exclusion zones.

Inter-cell interference coordination scheme. The principle of this coordination scheme is to allocate orthogonal PRBs to neighbouring cells. This can be accomplished by implementing a centralized multi-cell scheduling algorithm that involves a non-negligible signalling overhead. However, this coordination scheme demands for extensions in NTN-TN coexistence scenarios, as it requires time and frequency synchronization between gNBs. Furthermore, this mechanism requires an Xn interface between the gNB that belongs to the cooperative set. Without Xn signalling, an alternative mechanism shall be implemented to exchange resource scheduling strategies between terrestrial and satellite gNBs.

Spatial processing. Spatial diversity is used to mitigate interference. This approach leverages the increased number of antennas enabled by high frequency bands. In this context, if the NTN UEs are provisioned with the geographical location of terrestrial deployments, dynamic interfering nulling methods can be applied. This solution can be used at the transmitter to protect terrestrial base stations. Remarkably, the same strategy could be applied to terrestrial elements, i.e., TN UEs and gNBs, if they are provisioned with the satellite ephemeris. This solution could be used at the transmitter to protect satellite uplinks without compromising the quality of terrestrial links. At the receiver, spatial processing should be performed to suppress the interference in the cellular links. The key idea is motivated by the power predominance of LOS paths with respect to NLOS paths. The potential issue is the nulling direction, which is the location of the satellite, it is highly likely to vary over time. This is the case with 5G-STARDUST constellation that is composed of LEO satellites. Hence, the angular change must be taken into account to recalculate the beamforming weights. However, as the numerical assessment indicates, the impact on satellite reception is minimal and priority should be given to protect TN gNBs.

Exclusion zones. The idea is to separate NTN and TN cell centres by defining an area where certain network elements are restricted from operating. This technique is especially relevant for the coexistence scenarios 1 and 3, which are defined in Table 38, where the NTN UE interferes TN gNBs and TN UEs. In this situation, a radius of exclusion is defined to keep base stations and terrestrial users out of high interference zones. In practice, this can be realized by setting a minimum distance between satellite and terrestrial cells. Leveraging the link budget analysis, the minimum distance can be computed to guarantee that a pre-defined INR is not exceeded.

7 REFERENCES

- [1] SNS JU Project 5G-STAR DUST, D4.3 "Preliminary Report on the Unified Air Interface, User-Centric and Beamforming Solutions," May 2024.
- [2] SNS JU Project 5G-STAR DUST, D3.2 "End-to-end ground and space integrated architecture", February 2024.
- [3] 3GPP TS 38.300, "NR and NG-RAN overall description; stage 2 (release 18)".
- [4] 3GPP TR 36.881, «Study on latency reduction techniques for LTE (release 14)».
- [5] 3GPP TR 38.802, "Study on new radio access technology physical layer aspects (Release 14)".
- [6] M. Giordani, M. Polese, A. Roy, D. Castor y M. Zorzi, «A Tutorial on Beam Management for 3GPP NR at mmWave Frequencies,» *IEEE Communications Surveys & Tutorials*, vol. 21, nº 1, pp. 173-196, 2019.
- [7] 3GPP TSG-RAN WG1 Meeting #118-bis, «R1-2408426 - On NR-NTN downlink coverage enhancement».
- [8] J. O. Coleman, D. P. Scholnik y K. R. McPhail, «Phase-only tapers for regular planar arrays, a heuristic nonlinear-FM approach,» de *IEEE International Symposium on Phased Array Systems and Technology*, 2010.
- [9] D. Orlando y A. Farina, «Phase-only array transmit beamforming without iterative/numerical optimization methods,» *Frontiers in Signal Processing*, vol. 3, 2023.
- [10] 3GPP TS 38.104, «NR; Base Station (BS) radio transmission and reception (Release 19)».
- [11] 3GPP TS 38.213, «NR; Physical layer procedures for control (Release 19)».
- [12] 3GPP TS 38.331, «NR; Radio Resource Control (RRC); Protocol specification (Release 19)».
- [13] 3GPP TSG RAN WG1 #117, «Session notes for 9.11 Non-Terrestrial Networks (NTN) for NR Phase 3 and Internet of Things (IoT) Phase 3,» R1-2405699.
- [14] R. M. D. G. Y. B. E. Kovalchukov, «An Accurate Approximation of Resource Request Distributions in Millimeter Wave 3GPP New Radio Systems,» de *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*, 2019.
- [15] «3GPP TS 38.214 version 19.1.0 Release 19, "Physical layer procedures for data," Oct. 2025».
- [16] 3GPP TS 38.211, «NR; Physical channels and modulation (Release 19),» 2025.
- [17] X. A. a. M. Á. Vázquez, «Evaluation of New Radio Beam Management Framework for LEO Satellites,» de *2023 IEEE Conference on Standards for Communications and Networking (CSCN), Munich, Germany, 2023*, pp. 241-246, doi: 10.1109/CSCN60443.2023.10453212.
- [18] 3GPP, TR 38.821 "Solutions for NR to support non-terrestrial networks (NTN) (Release 16)", March 2023.
- [19] A. Guidotti, C. Amatetti y A. Vanelli-Coralli, Federated Cell-Free MIMO in Nonterrestrial Networks: Architectures and Performance, *IEEE Transactions on Aerospace and Electronic Systems*, June 2024.
- [20] EC Project DYNASAT, D3.3 "Bandwidth Efficient Techniques Design – Version 2", November 2022.
- [21] EC Project DYNASAT, D3.6 "Bandwidth efficient techniques evaluation (V3)", February 2023.
- [22] M. Schiavina, S. Freire y K. MacManus, GHS-POP R2022A - GHS population grid multitemporal (1975-2030), European Commission, Joint Research Centre (JRC), 2024.
- [23] S. Freire, K. MacManus, M. Pesaresi, E. Doxsey-Whitfield y J. Mills, Development of new open and free multi-temporal global population grids at 250 m resolution. *Geospatial Data in a Changing World*, Association of Geographic Information Laboratories (AGILE), 2016.

- [24] European Commission, Archive: Urban-rural typology, https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Archive:Urban-rural_typology, March 2018.
- [25] L. Dijkstra y H. Poelman, Regional Working Paper 2014 "A harmonised definition of cities and rural areas: the new degree of urbanisation", Regional and Urban Policy, 2014.
- [26] SNS JU Project 6G-NTN, D2.4 "Market and business model analysis", January 2025.
- [27] 3GPP, TR 38.811 "Study on New Radio (NR) to support non-terrestrial networks (Release 15)", September 2020.
- [28] 3GPP, TR 38.803 "Study on new radio access technology: Radio Frequency (RF) and co-existence aspects (Release 14)", September 2019.
- [29] A. Guidotti, C. Sacchi y A. Vanelli-Coralli, Feeder Link Precoding for Future Broadcasting Services, IEEE Transactions on Aerospace and Electronic Systems, August 2022.
- [30] U. Mengali y A. D'Andrea, Synchronisation Techniques for Digital Receivers, New York: Springer New York, 1997.
- [31] S. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Englewood Cliffs, NJ: Prentice Hall, 1993.
- [32] 5G-STARUST, «D4.3; Report on the Unified Air Interface, User-Centric and Beamforming Solutions,» May 2024.
- [33] Z. Wei, S. Li, W. Yuan, R. Schober and G. Caire, "Orthogonal Time Frequency Space Modulation Part I: Fundamentals and Challenges Ahead," *IEEE Communications Letters*, vol. 27, no. 1, pp. 4-8, 2023.
- [34] Z. Wei, W. Yuan, S. Li, J. Yuan and D. W. K. Ng, "Transmitter and receiver window designs for orthogonal time-frequency space modulation," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2207-2223, 2021.
- [35] S. K. Mohammed and V. Khammammetti, "OTFS-based multiple-access in high Doppler and delay spread wireless channels," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 528-531, 2019.
- [36] P. Raviteja, K. T. Phan, Y. Hong and E. Viterbo, "Interference cancellation and iterative detection for orthogonal time frequency space modulation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6501-6515, 2018.
- [37] F. Lampel, A. Alvarado and F. M. Willems, "Orthogonal time frequency space modulation: A discrete Zak transform approach," , *arXiv preprint arXiv:2106.12828*, 2021.
- [38] L. Gaudio, M. Kobayashi, G. Caire and G. Colavolpe, "On the effectiveness of OTFS for joint radar parameter estimation and communication," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 5951-5965, 2020.
- [39] G. Ungerboeck, «Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems,» *IEEE Transactions on Communications*, vol. 22, n° 5, pp. 624-636, 1974.
- [40] G. Colavolpe, D. Fertonani and A. Piemontese, "SISO detection over linear channels with linear complexity in the number of interferers," *IEEE J. Sel. Topics in Signal Proc.*, vol. 5, pp. 1475-1485, 2011.
- [41] D. G. C. Bandeira, D. L. Ruyet, M. Pischella and J. C. M. Mota, "Performance evaluation of low-complexity algorithms for orthogonal time-frequency space modulation," *Journal of Communication and Information Systems*, vol. 35, no. 1, pp. 138-149, 2020.
- [42] E. Conti, A. Piemontese, T. Foggi, G. Colavolpe and A. Vannucci, "Efficient message-passing detection for multi-satellite systems using OTFS modulation," in *IEEE Aerospace Conference*, Big Sky, MT, USA, 2025.

- [43] A. Vehtari, A. Gelman, T. Sivula, P. Jylänki, D. Tran, S. Sahai, P. Blomstedt, J. P. Cunningham, D. Schiminovich and C. Robert, "Expectation propagation as a way of life: A framework for Bayesian inference on partitioned data," *arxiv*, 2014.
- [44] M. Seeger, «Bayesian Gaussian process models: PAC-bayesian generalisation error bounds and sparse approximations,» PhD thesis, University of Edinburgh, 2003.
- [45] S. Wu, L. Kuang, J. L. Z. Ni, D. Huang and Q. Guo, "Low-complexity iterative detection for large-scale multiuser MIMO-OFDM systems using approximate message passing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 902-915, 2014.
- [46] T. P. Minka, «Divergence measures and message passing,» Microsoft Research Cambridge, Tech. Rep., 2005, mSR-TR-2005-173, 2005.
- [47] T. Minka, «Expectation propagation for approximate Bayesian inference,» de *Proc. UAI*, 2021.
- [48] A. Vannucci, G. Colavolpe, R. Pecori and L. Veltri, "Estimation of a Gaussian source with memory in bursty impulsive noise," in *IEEE ISPLC*, 2019.
- [49] A. Mirbadin, E. Kiani, A. Vannucci and G. Colavolpe, "Estimation of Gaussian processes in Markov-Middleton impulsive noise," in *1st Global Power, Energy and Communication Conference (GPECOM)*, 2019.
- [50] A. Vannucci, G. Colavolpe and L. Veltri, "Estimation of correlated Gaussian samples in impulsive noise," *IEEE Commun. Letters*, vol. 24, no. 1, pp. 103-107, 2020.
- [51] F. R. Kschischang, B. J. Frey and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498-519, 2001.
- [52] 3GPP Technical Specification Group Radio Access Network, «Study on New Radio (NR) access technology: Radio Frequency (RF) and co-existence aspects,» de *3rd Generation Partnership Project (3GPP), Technical Report (TR) 38.803*, 2022.
- [53] R. I.-R. P.681-11, «Propagation data required for the design systems in the land mobile-satellite service,» 2019.
- [54] 3GPP TR 38.863, «Solutions for NR to support non-terrestrial networks (NTN): Non-terrestrial networks (NTN) related RF and co-existence aspects (Release 18)».
- [55] 3GPP TS 38.101-5, «NR; User Equipment (UE) radio transmission and reception; Part 5: Satellite access Radio Frequency (RF) and performance requirements (Release 18)».
- [56] 3GPP TS 38.101-2, «NR; User Equipment (UE) radio transmission and reception; Part 2: Range 2 Standalone (Release 18)».
- [57] 3GPP TR 37.840, «Study of Radio Frequency (RF) and Electromagnetic Compatibility (EMC) requirements for Active Antenna Array System (AAS) base station».
- [58] 3GPP TR 38.901, «Study on channel model for frequencies from 0.5 to 100 GHz».
- [59] 3GPP TS 37.340, «NR; Multi-connectivity; Overall description; Stage-2 (Release 18),» March 2024.
- [60] 3GPP TS 24.193, «5G System;Access Traffic Steering, Switching and Splitting (ATSSS); Stage 3 (Release 18),» April 2024.
- [61] 3GPP TS 36.300, «Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 18),» Feb. 2024.

ANNEX A: EXTENDED PERFORMANCE ASSESSMENT WITH UNIFORM DISTRIBUTIONS

7.1 OVERVIEW OF THE SYSTEM MODEL

In this Annex, we extend the performance assessment of the random scheduling algorithms described in [1]. For the sake of completeness, we report in Algorithm 5 (RS) and Algorithm 6 (RS-MD) the random scheduling algorithms that were considered, for which the details are available in [1]. Please note that the nomenclature has been slightly modified to be aligned with the terminology used throughout the present document.

Algorithm 5: Random Scheduling (RS) algorithm, [1].

Input	
• $N_{UE} \times 1$ column flag vector $\mathbf{v}_{vis}$ • $N_{UE} \times 1$ allocation flag vector $\mathbf{f}$ • $N_{UE} \times 1$ allocation number vector $\mathbf{a}$	
Output	
• $N_B \times N_{sched}^{(slot)}$ scheduling matrix $\mathbf{S}^{(m)}$, $m = 1, \dots, N_{sched}$	
for $t = 1, \dots, N_{sched}^{(slot)}$	
1	Identify the visible users without allocations: $\mathbf{v}_{vis}^{(m)} = \mathbf{f} \odot \mathbf{v}_{vis}$ Identify the users with at least one allocation: $\mathbf{v}_{vis}^{(m,1)} = (\mathbf{f} + 1)_2 \odot \mathbf{v}_{vis}$
	if $\text{Card}(\mathbf{v}_{vis}^{(0)}) \geq N_B$: randomly select N_B indexes from $\mathbf{v}_{vis}^{(m)}$ and assign them to $\mathbf{S}^{(m)}$: $\mathbf{s}_{t,:}^{(m)} = \omega(\mathcal{I}(\mathbf{v}_{vis}^{(m)}), N_B)$
2	update the allocation flag vector $\mathbf{f}$: $\mathbf{f}_k = 1, \forall k \in \mathcal{I}(\mathbf{S}^{(m)})$ update the allocation number vector $\mathbf{a}$: $\mathbf{a}_k = \mathbf{a}_k + 1, \forall k \in \mathcal{I}(\mathbf{S}^{(m)})$
	else:
	assign the available users to $\mathbf{S}^{(m)}$: $\mathbf{s}_{t,1:\text{Card}(\mathbf{v}_{vis}^{(0)})}^{(m)} = \mathbf{v}_{vis}^{(m)}$
3	randomly select $N_B - \text{Card}(\mathbf{v}_{vis}^{(m)})$ indexes from $\mathbf{v}_{vis}^{(1)}$ and assign them to $\mathbf{S}^{(m)}$: $\mathbf{s}_{t,1+\text{Card}(\mathbf{v}_{vis}^{(0)}):N_B}^{(m)} = \omega(\mathcal{I}(\mathbf{v}_{vis}^{(m,1)}), N_B - \text{Card}(\mathbf{v}_{vis}^{(m)}))$
	update the allocation flag vector $\mathbf{f}$: $\mathbf{f}_k = 1, \forall k \in \mathcal{I}(\mathbf{S}^{(m)})$ update the allocation number vector $\mathbf{a}$: $\mathbf{a}_k = \mathbf{a}_k + 1, \forall k \in \mathcal{I}(\mathbf{S}^{(m)})$
end for	

As already remarked in [1], it can be noticed that RS is a special case of RS-MD when there is no constraint on the distance among users in the same time slot, i.e., when $\rho_{GEO} = 0$ in step 3 of the RS-MD algorithm. The basic principle of RS-MD is to randomly allocate users in the slot as long as there are available beams and the distance from the users already allocated is

not below a predefined constraint that is defined based on the parameter ρ_{GEO} (an integer value representing the multiples of the beam radius to be considered as a guard distance).

Algorithm 6: Random Scheduling with Minimum Distance (RS-MD) algorithm, [[1]].

Input	
$N_{UE} \times 1$ column flag vector $\mathbf{v}_{vis}$ $N_{UE} \times 1$ allocation flag vector $\mathbf{f}$ $N_{UE} \times 1$ allocation number vector $\mathbf{a}$ - Geographic constraint coefficient ρ_{GEO} - Radius r_{uv}	
Output	
$N_B \times N_{sched}^{(slot)}$ scheduling matrix $\mathbf{S}^{(m)}$, $m = 1, \dots, N_{sched}$	
for $t = 1, \dots, N_{sched}^{(slot)}$	
	Identify the visible users without allocations: $\mathbf{v}_{vis}^{(0)} = \mathbf{f} \odot \mathbf{v}_{vis}$
1	Identify the users with at least one allocation: $\mathbf{v}_{vis}^{(1)} = (\mathbf{f} + 1)_2 \odot \mathbf{u}_{vis}$
	Initialise the set of barred users $\delta_{vis}^{(m)} = \mathbf{0}$
	Initialise beam counter: $b = 1$
2	Randomly select 1 index from $\mathbf{v}_{vis}^{(0)}$ and assign it to $\mathbf{S}^{(m)}$: $\mathbf{s}_{t,b}^{(m)} = \omega(\mathcal{I}(\mathbf{v}_{vis}^{(0)}), 1)$
	while $b \leq N_B$:
	if $\text{Card}(\mathbf{v}_{vis}^{(0)}) > 1$:
	randomly select 1 index from $\mathbf{v}_{vis}^{(0)}$: $k = \omega(\mathcal{I}(\mathbf{v}_{vis}^{(0)}) \setminus \mathcal{I}(\delta_{vis}^{(m)}), 1)$
3	compute $\Gamma_{GEO,i}^{(t)}$
	if $\Gamma_{GEO,i}^{(t)}$ is met for all elements, assign the user:
	$b = b + 1$
	$\mathbf{s}_{t,b}^{(m)} = k$
	else insert user k in the set of barred users for this time slot: $\delta_{vis,i}^{(m)} = 1$
	If $\text{Card}(\mathbf{v}_{vis}^{(0)}) = 0$
4	randomly select $N_B - \text{count}$ index from $\mathbf{v}_{vis}^{(1)}$ and assign them to $\mathbf{S}_m$: $\mathbf{S}_m = \omega(\mathbf{v}_{vis}^{(1)}, N_B - \text{count})$ $\text{count} = \text{count} + N_B - \text{count}$
end while	
	update the allocation flag vector $\mathbf{f}$:
5	$\mathbf{f}_k = 1, \forall k \in \mathcal{I}(\mathbf{S}^{(m)})$
	update the allocation number vector $\mathbf{a}$:
	$\mathbf{a}_k = \mathbf{a}_k + 1, \forall k \in \mathcal{I}(\mathbf{S}^{(m)})$
end for	

Table 47 provides the simulation parameters. Compared to the initial assessment reported in [1], the following aspects are worth to be highlighted:

- The user density is fixed to a single value, as the impact of variable densities was already discussed and detailed in [1].

- The channel model is always clear-sky, as the impact of LoS and NLoS channels was already discussed in [1]. Thus, under ideal ancillary information conditions, both MMSE and LB-MMSE provide the same performance (there is no stochastic term in the channel coefficient and, thus, the CSI estimate is the same in both approaches).
- The service elevation angle, *i.e.*, the elevation angle that is used to compute the extension of the service area, is variable between $\varepsilon_S = 30^\circ$ and $\varepsilon_S = 60^\circ$. The objective of the variability of this parameter is to assess the performance when more limited service areas are considered, which leads to: i) reduced path losses, as the minimum slant range in the FoV of the NTN node is increased; and ii) a reduced number of users to be served. Both aspects were observed to be critical in [1] due to the system design, with a very large service area leading to large losses (lowering the overall SINR) and many users to be served (leading to a reduction in the achievable capacity or spectral efficiency, as the allocated time slots per user are reduced). Please note that we are only modifying the dimension of the service area: the users are considered visible as long as their elevation angle is above 30° and, in comparing these scenarios, the satellite orbit is the same.
- We introduce a new parameter ρ_{ML} indicating the maritime to land users ratio. This is a parameter introduced to start moving into the direction of non-uniform users distributions. Basically, the users are distributed according to the density indicated in Table 47 over land areas, while their density is scaled down by the factor ρ_{ML} on maritime areas. Examples of these distributions are provided later in this Annex.
- We consider a standalone NTN node from both the polar constellation (1000 km altitude, 99° inclination) and the inclined constellation (1300 km altitude, 50° inclination).
- We compare the performance with user-centric beamforming with legacy Frequency Reuse 3 (FR3) and FR 4 (FR4) systems.

Table 47: Parameters for system-level simulations with a single NTN node and random scheduling.

Parameter	Value	Comments
Sub-Carrier Spacing (SCS)	120 kHz	In this case, 132 PRBs are available
Operating Frequency Range	FR2 (Ka-band)	See 3GPP TR 38.821
Carrier frequency	20 GHz	See 3GPP TR 38.821
User bandwidth B	190.08 MHz	$132 \cdot 120 \cdot 12$ kHz
User density	$2 \cdot 10^{-3}$ UE/km ²	
Number of radiating elements N_R	512	See Section 2.3.1 of [1]
Power per radiating element $P_{av,el}$	65 mW	See Section 2.3.1 of [1]
Total power on-board P_{av}	15.2218 dBW	$10\log_{10}(0.0065 \cdot 512)$
Geographic constraint coefficient ρ_{GEO}	0 (RS) 1, ..., 7 (RS-MD)	See Section 4.2.4 of [1]
Minimum user elevation angle ε_t	30°	See Section 4.2.3 of [1]
Service area elevation angle ε_S	$30^\circ, 40^\circ, 50^\circ, 60^\circ$	See Section 4.2.3 of [1]
Channel type	clear-sky	See Section 2.4.2.1 of [1] and 3GPP TR 38.811
Scheduling periodicity T_{sched}	2 s (200 slots)	See Section 4.2.1 of [1]
Reporting periodicity T_{rep}	80 ms (8 slots)	See Section 4.2.1 of [1]
Transmission time slot T_{slot}	10 ms	See Section 4.2.1 of [1]
NTN node altitude	1000, 1300 km	See Section 2.1 of [1]
Service area centre	($45^\circ\text{N}, 5^\circ\text{E}$)	See Section 4.2.3 of [1]

UE terminal	VSAT	See Section 2.3.2 of [1] and 3GPP TR 38.811
-------------	------	---

7.2 PERFORMANCE ASSESSMENT

For the performance comparison of the different algorithms and configurations, we focus on the experienced capacity per user. In fact, since the RRM algorithms are random in this case, the objectives of this analysis are to: i) identify the best beamforming algorithm and power distribution strategy; ii) understand the impact of a slightly non-uniform distribution of the users in the area, based on the distinction between maritime and land zones; iii) understand the impact of the constellation (*i.e.*, satellite altitude); and iv) understand the impact of the service area dimensions. All these considerations can be related to the experienced capacity per user.

7.2.1 Polar constellation

Figure 144 shows the service area dimensions considered in the polar orbit scenario. As already highlighted above, the satellite is always moving along all locations indicated by the single ground track and we only vary the dimension of the service area. The ground track corresponds to the service area with 30° elevation, meaning that users are always considered as visible as long as they see the satellite with a minimum elevation angle of 30°.

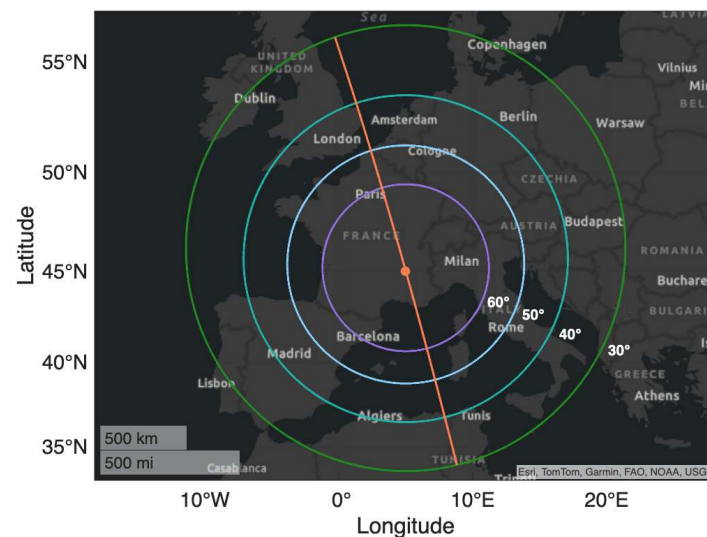


Figure 144: Variable service area dimensions for the polar constellation assessment (standalone node, random scheduling).

7.2.1.1 Uniform distribution

Figure 48 shows an example of user deployment assuming a completely uniform distribution, *i.e.*, $\rho_{ML} = 1$, and considering the variable service area dimensions introduced in the above section and represented in Figure 144.

For each value of the elevation angle defining the dimensions of the service area, we evaluate the average experienced capacity per user. The following solutions have been assessed:

- Beamforming algorithms: **MMSE and CBF beamforming**. Please note again that, in clear-sky conditions, the performance of LB-MMSE is identical and, thus, we only represent one of the algorithms for the sake of clarity and conciseness.

- Power distribution strategy: notably, one of the most relevant configurations to be implemented on top of the selected beamforming algorithm, since it defines how the power is redistributed across the radiating elements and users. In [1], we introduced three solutions: **Sum Power Constraint** (SPC, the theoretical maximum which guarantees the total on-board power is used, but does not control the power per High Power Amplifier on-board), **Maximum Power Constraint** (MPC, which controls the power per HPA, but does not guarantee the total on-board power is exploited), and **Per Antenna Constraint** (which guarantees both the HPA and total power conditions are met, but it disrupts the orthogonality in the beamforming columns). Extensive details are available in [1].

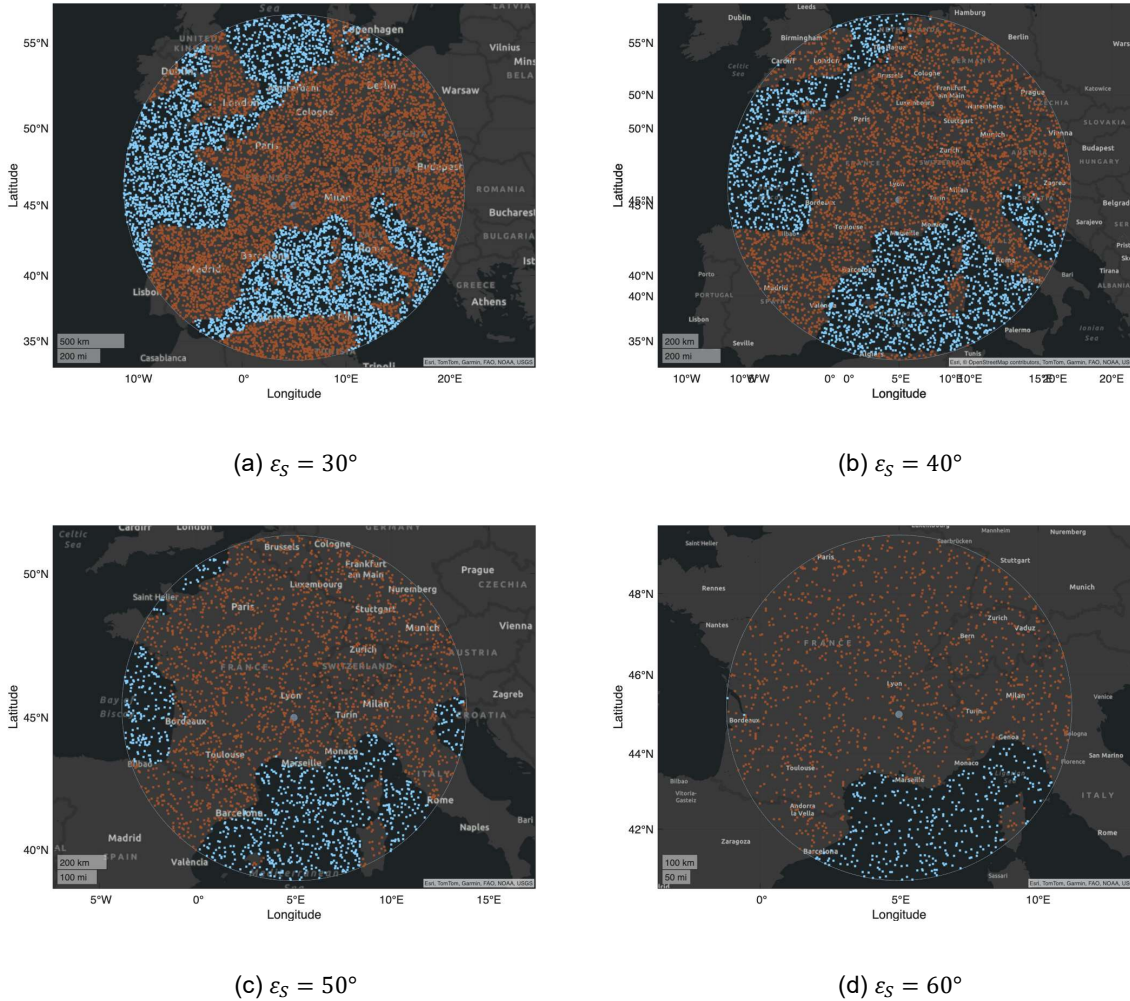
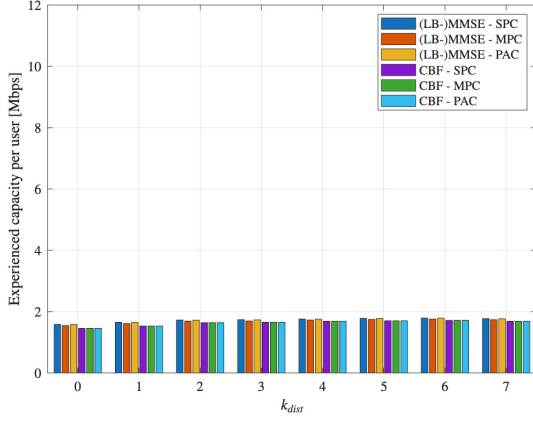


Figure 145: Example of uniform user deployment for the polar constellation and variable service area dimensions.

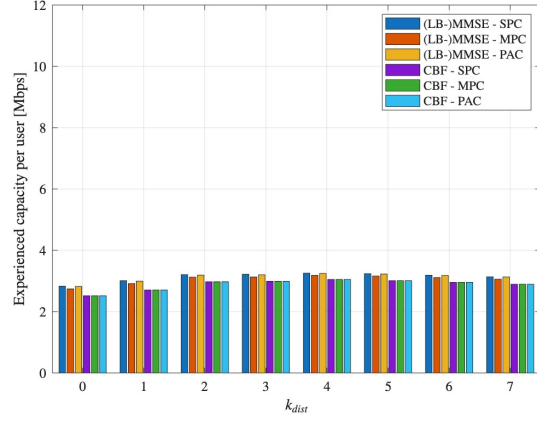
Figure 146 shows the numerical results for the polar constellation, without maritime/land user differentiation (*i.e.*, $\rho_{ML} = 1$) and as a function of the service area dimensions. We can observe the following trends:

- Power distribution strategy: as expected, SPC provides the best performance. Unfortunately, it is also the strategy in which there is no control on the power per radiating element and this might lead to HPAs working in the non-linear regime. Thus, it provides a theoretically optimal performance. MPC is the worst power normalisation, because we are in a scenario with relatively low SNR values due to the large slant ranges and limited

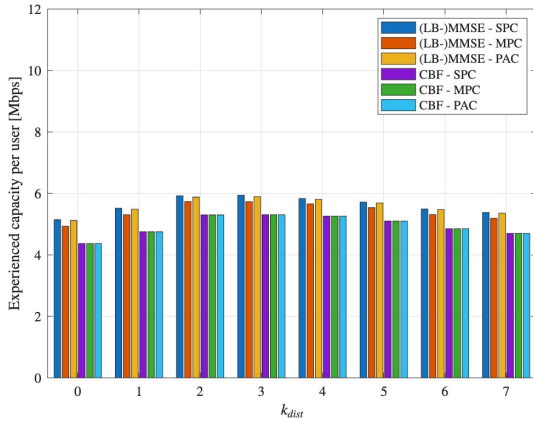
on-board power; thus, not exploiting the entire on-board available power results in a sub-optimal strategy. As such, PAC is the best solution taking into account technical feasibility and performance.



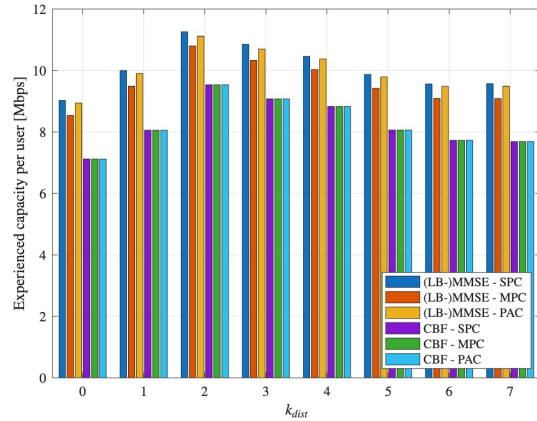
(a) $\varepsilon_S = 30^\circ$ (optimal $\rho_{GEO} = 6$)



(b) $\varepsilon_S = 40^\circ$ (optimal $\rho_{GEO} = 4$)



(c) $\varepsilon_S = 50^\circ$ (optimal $\rho_{GEO} = 3$)



(d) $\varepsilon_S = 60^\circ$ (optimal $\rho_{GEO} = 2$)

Figure 146: Average experienced capacity per user with the polar constellation (1000 km) as a function of the service area dimensions.

- Beamforming algorithm: in line with the initial assessment in [1], MMSE solutions provide a better performance compared to digital beam steering (CBF). This advantage becomes more evident in smaller service areas, when interference is harsher. In fact, MMSE tries to maximise the intended power while limiting interference, instead of CBF in which the signals are only steered in the desired directions. Thus, in scenarios where interference is less critical, *i.e.*, large service area possibly combined with large distances among the users (set by ρ_{GEO}), MMSE and CBF are closer.
- The performance significantly improves in smaller service areas. This is pretty straightforward, as the number of slot allocations per user significantly improves, while the interference, despite increasing, is less impactful. To highlight this trend, we can observe Figure 147, in which we show the CDF of the allocations per user with $\rho_{GEO} = 3, 7$. With smaller service area, clearly, the number of allocations per user increases and, thus, the experienced capacity is significantly improved. It is also interesting to notice that

the larger the service area, the more similar the number of allocations tend to be for a varying value of ρ_{GEO} . This is also reflected in the performance discussed above, as with larger service areas the performance in terms of experienced capacity is similar; this indicates that, in large areas as those that we have in 5G-STARDUST, the most impacting factor (at least with random scheduling and uniform traffic requests) is the number of allocations per user rather than the beamforming algorithm or the RS-MD configuration.

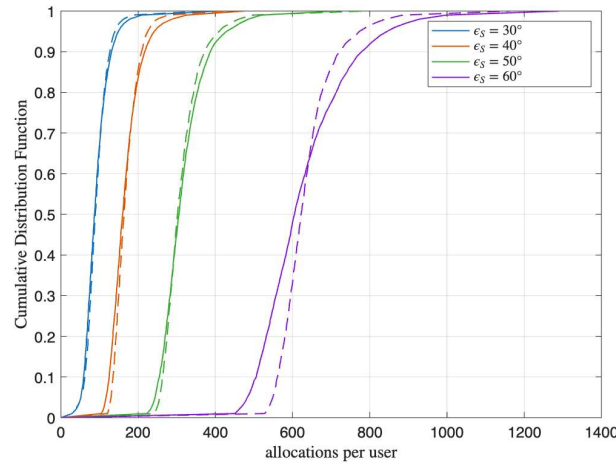
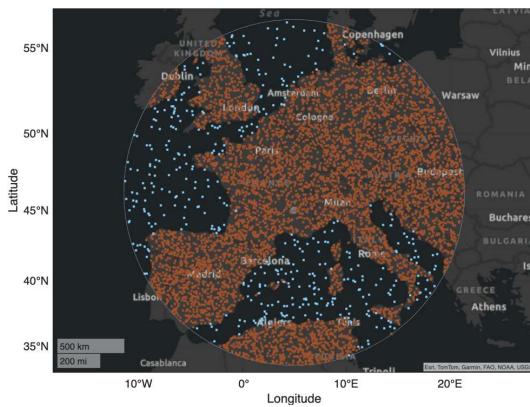


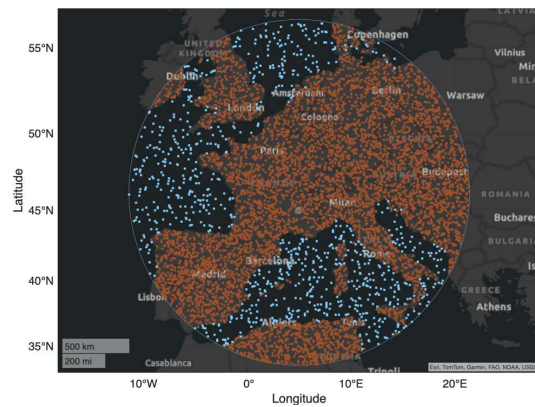
Figure 147: CDF of the number of allocations per user in the polar constellation as a function of the service area dimension with $\rho_{GEO} = 3$ (solid line) and $\rho_{GEO} = 7$ (dashed line).

7.2.1.2 Maritime/land distribution

Below, we provide the performance assessment of the polar constellation when considering a different user density for maritime and land zones of the service area. Figure 148 shows an example of this deployment with $\rho_{ML} = 0.1, 0.2$ and the service area defined with 30° elevations.



(a) $\rho_{ML} = 0.1$

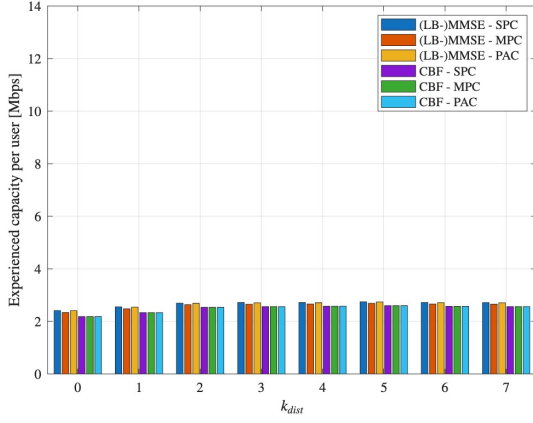


(b) $\rho_{ML} = 0.2$

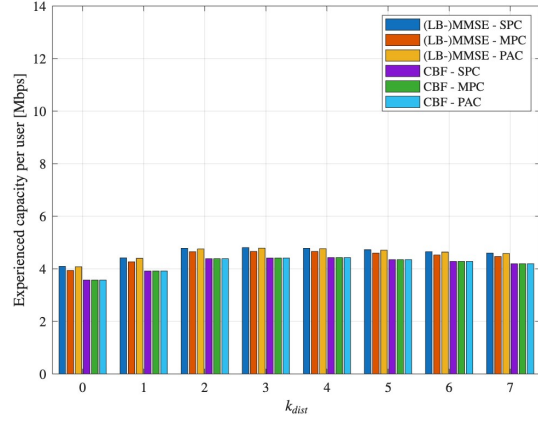
Figure 148: Example of non-uniform user deployment with maritime and land areas for the polar constellation and $\varepsilon_S = 30^\circ$.

Figure 149 and Figure 150 show the experienced user capacity with $\rho_{ML} = 0.1$ and $\rho_{ML} = 0.2$, respectively. In general, the same trends as those discussed above can be observed. Compared to the performance in uniform conditions, the number of users is reduced and, thus,

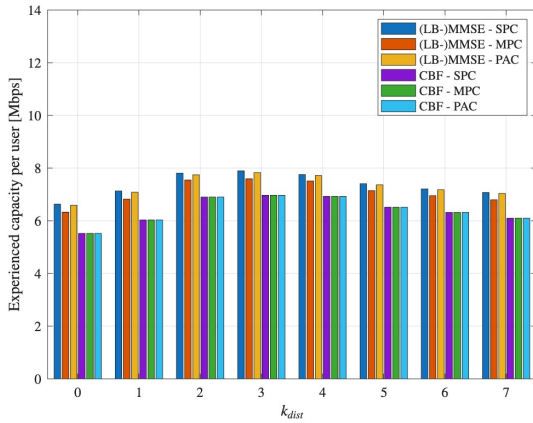
we have a two-fold effect: i) the number of allocations per user increase; and ii) in some cases, a less stringent (lower ρ_{GEO}) condition on the intra-slot distance can be implemented. Ultimately, the performance is slightly increased in these scenarios.



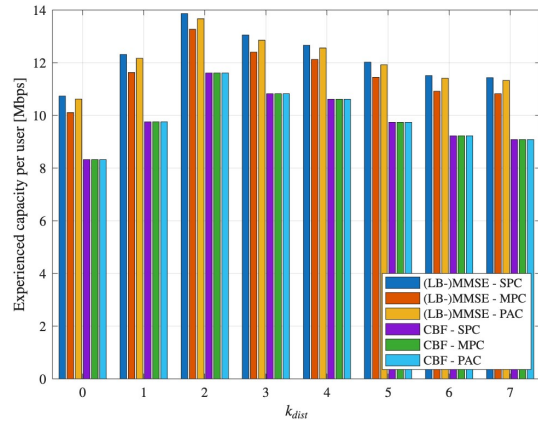
(a) $\varepsilon_S = 30^\circ$ (optimal $\rho_{GEO} = 5$)



(b) $\varepsilon_S = 40^\circ$ (optimal $\rho_{GEO} = 3$)

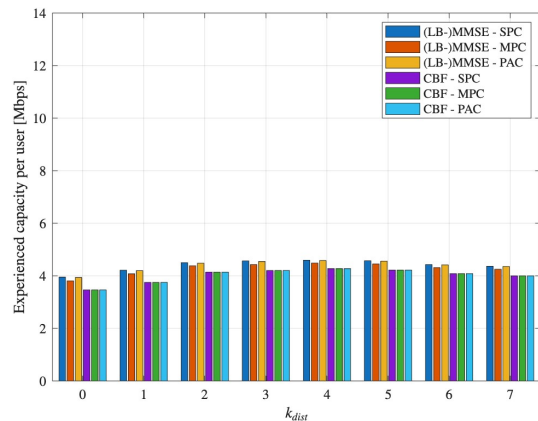
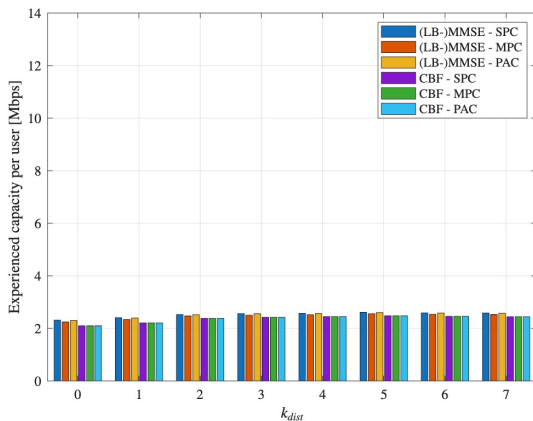


(c) $\varepsilon_S = 50^\circ$ (optimal $\rho_{GEO} = 3$)



(d) $\varepsilon_S = 60^\circ$ (optimal $\rho_{GEO} = 2$)

Figure149: Average experienced capacity per user with the polar constellation (1000 km) as a function of the service area dimensions and with non-uniform distribution ($\rho_{ML} = 0.1$).



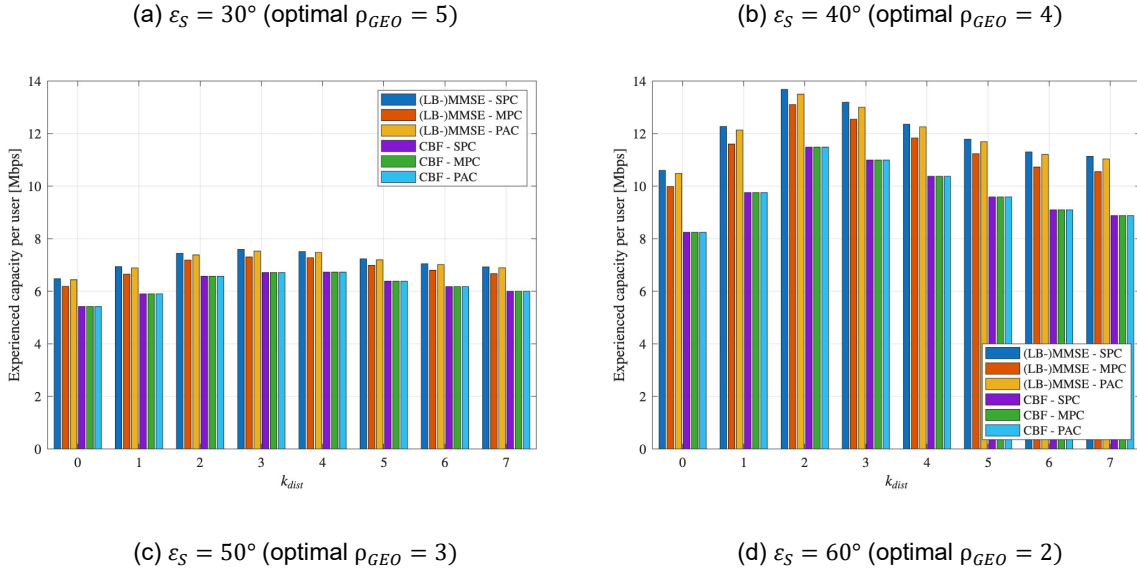


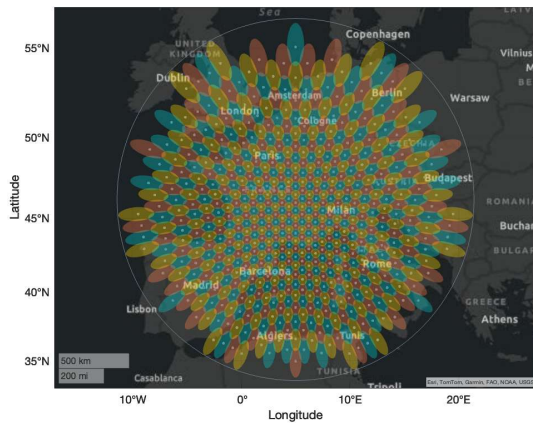
Figure 150: Average experienced capacity per user with the polar constellation (1000 km) as a function of the service area dimensions and with non-uniform distribution ($\rho_{ML} = 0.2$).

7.2.1.3 Comparison with Frequency Reuse schemes

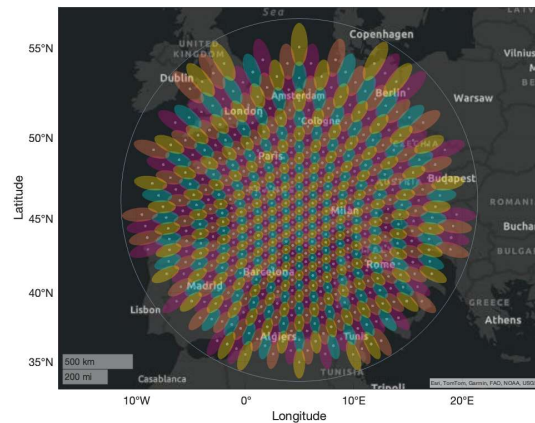
In this section, we compare the performance of user-centric beamforming with a legacy system in which FR3 and FR4 schemes are implemented. In this case, a beam lattice is defined on-ground and 3 or 4 colours are exploited for the resource assignment, as shown in Figure 151. Since we have way more than 24 beams per polarisation in FR3/FR4, in the random scheduler for the performance evaluation a random beam selection has been implemented; in particular, during a generic time slot, a subset of beams is selected to be active during the time slot and, then, one user for each of these beams is scheduled for transmission.

In this context, it shall be noticed that, as described in D3.2, [2], the NTN payload is equipped with an antenna per polarisation and each antenna is capable of generating 24 beams. First, this means that in FR3 and FR4 we consider 24 beams per polarisation, meaning 8 beams per colour in FR3 and 12 beams per colour in FR4 (see Figure 151c). Moreover, in the user-centric (FFR) assessment, as well as with FR3, we are considering a single polarisation; neglecting cross-polarisation interference, this is basically equivalent to assuming half the number of users in the system or obtaining a performance that is almost half the one with two polarisations (doubling the resources would lead to almost doubling the allocations per user). Observing Figure 151, this might lead to a comparison between FR4 and FR3/FFR solutions not entirely fair, as FR4 allocations always assume two polarisations. To circumvent this, and in line with the previous observations, the performance assessment with FR4 is obtained with twice the user density, *i.e.*, $4 \cdot 10^{-3}$ UEs/km².

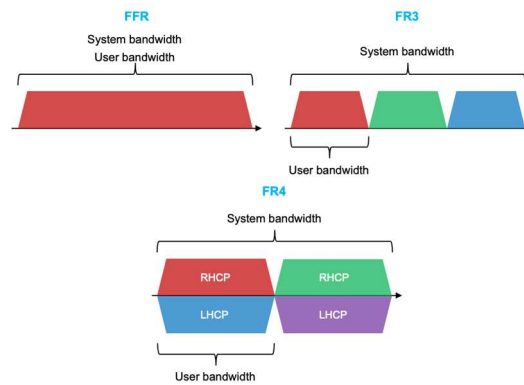
Table 48 summarises the performance of the three resource allocation schemes that we considered. The optimal performance for (LB-)MMSE was obtained with PAC as power distribution scheme and with $\rho_{GEO} = 6$, based on the above extensive performance assessment. It can be noticed, as expected, that the asymptotic rate with FR3 and FR4 is larger, in line with the performance in terms of SINR. In fact, FR3 and FR4 have a significantly better SINR thanks to the separation of the users in the frequency and polarisation domains; however, with FFR and user-centric beamforming we have a much larger user bandwidth, which completely compensates this effect leading to a larger asymptotic capacity. The performance in terms of the experienced capacity per user is aligned with these considerations.



(a) FR3



(b) FR4



(c) Frequency and polarisation allocations with FFR, FR3, and FR4

Figure151: Beam lattice with FR3 (left) and FR4 (right) for the polar constellation.

Table 48: Performance comparison with FR3 and FR4 (relative gains/losses in square brackets).

Parameter	MMSE PAC (optimal)	FR3	FR4
Asymptotic rate [bps/Hz]	4.05	6.61 [+63.12%]	5.70 [+40.74%]
Asymptotic capacity [Mbps]	769.18	419.01 [-45.53%]	541.88 [-29.55%]
Experienced capacity per UE [Mbps]	1.79	1.19 [-33.52%]	1.55 [-13.41%]
Average SINR [dB]	11.82	19.83	17.03

7.2.2 Inclined constellation

Figure 152 shows the service area dimensions considered in the inclined orbit scenario, which are obtained with the same parameters as the polar case with the only difference being the satellite orbit and inclination. Also in this case, the satellite is moving along all locations indicated by the single ground track and we only vary the dimension of the service area; moreover, the users are always considered as visible as long as they see the satellite with a minimum elevation angle of 30° .

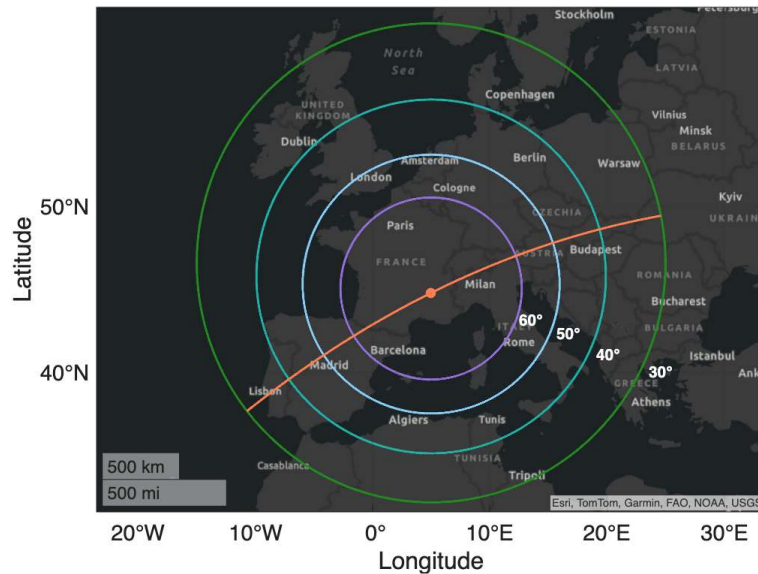


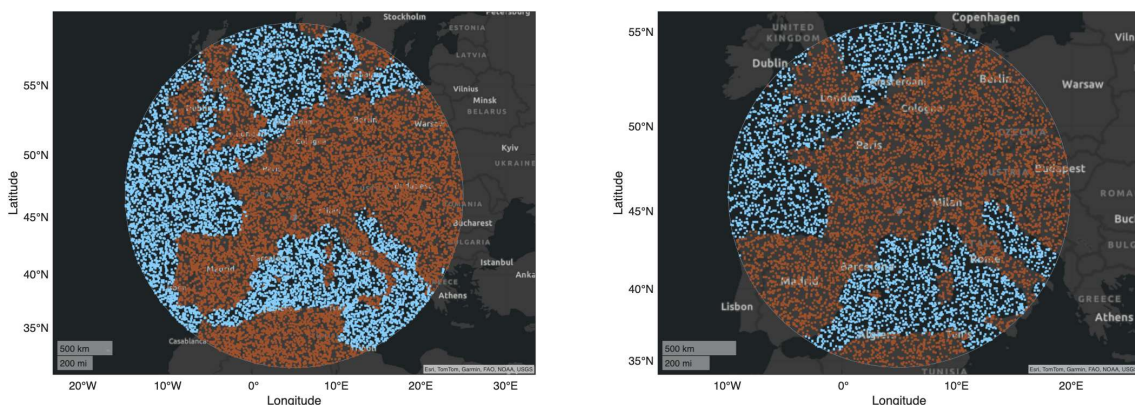
Figure 152: Variable service area dimensions for the inclined constellation assessment (standalone node, random scheduling).

7.2.2.1 Uniform distribution

Similarly to the polar scenario, Figure 153 shows an example of user deployment assuming a completely uniform distribution, *i.e.*, $\rho_{ML} = 1$, and considering the variable service area dimensions represented in Figure 152.

The same algorithms and configurations as those discussed for the polar constellation have been also considered in this case. Observing the performance in Figure 154, the considerations that we made in the polar orbit scenario still hold with the inclined orbit:

- Power distribution strategy: PAC is the best solution taking into account technical feasibility and performance.
- Beamforming algorithm: MMSE solutions provide a better performance compared to digital beam steering (CBF), in with smaller service areas when interference is larger.
- The performance significantly improves in smaller service areas and the optimal value of k_{dist} is almost always the same as that obtained with polar constellations.



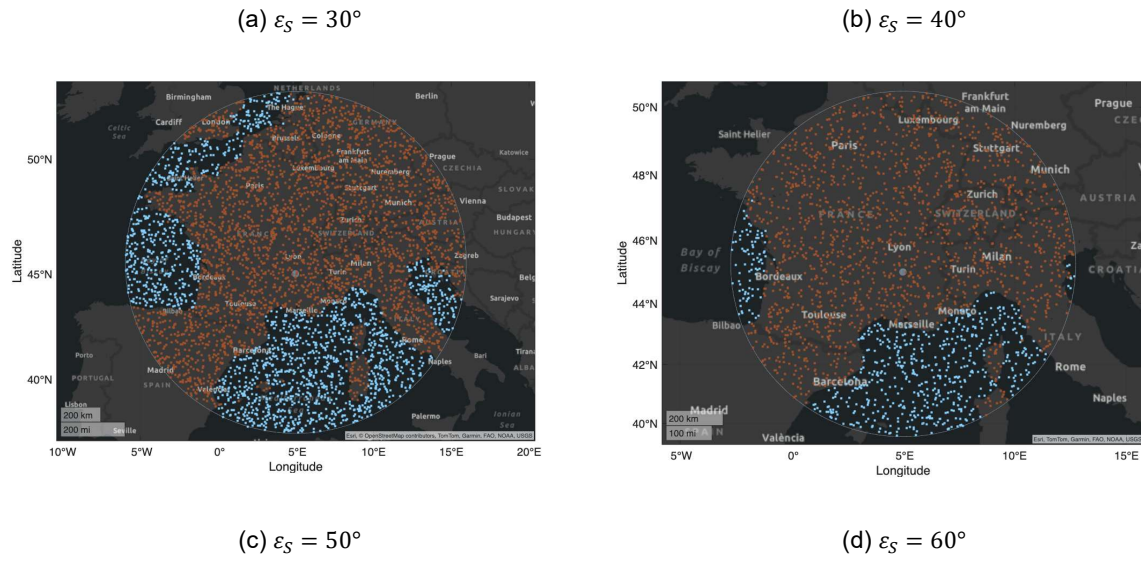


Figure 153: Example of uniform user deployment for the inclined constellation and variable service area dimensions.

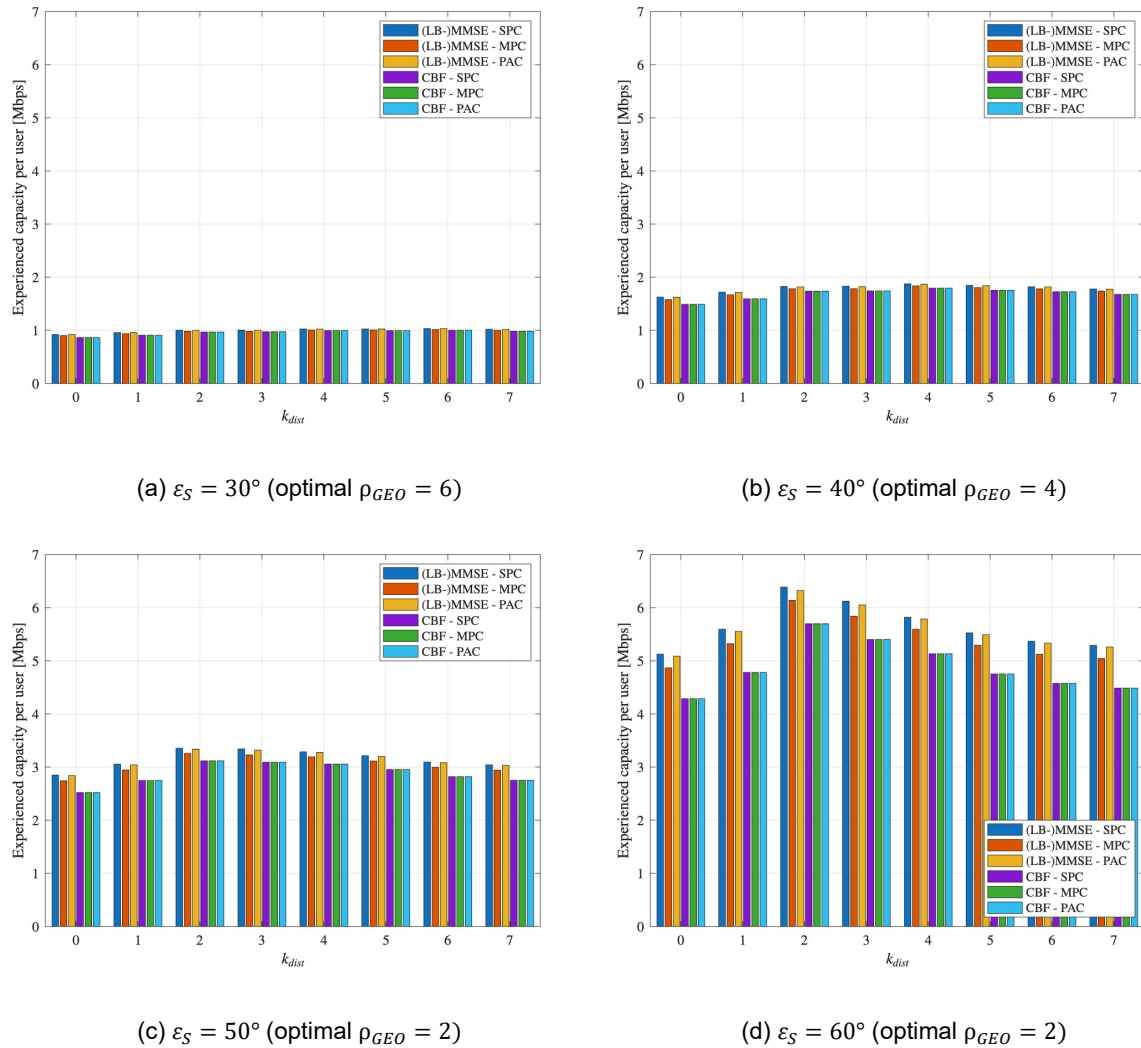


Figure154: Average experienced capacity per user with the inclined constellation (1300 km) as a function of the service area dimensions.

It is worthwhile highlighting that, with the inclined orbit, we are increasing the satellite altitude by 300 km but the transmission power is not adjusted, based on the system architecture defined in [2]. As such, the performance is worse compared to the polar orbit because the slant range is increased; it shall also be mentioned that the increase in the slant range is particularly impactful as the service area at a higher altitude is also larger and, thus, we have much larger path losses for users close to the service area edges.

7.2.2.2 Maritime/land distribution

Below, we provide the performance assessment of the inclined constellation when considering a different user density for maritime and land zones of the service area, as in the polar scenario. Figure 155 shows an example of this deployment with $\rho_{ML} = 0.1, 0.2$ and the service area defined with 30° elevations.

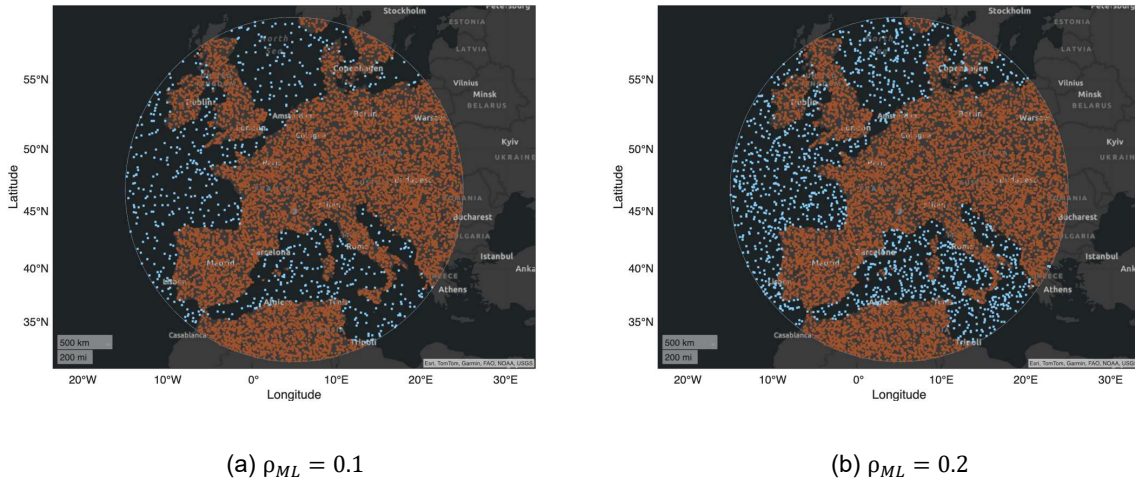
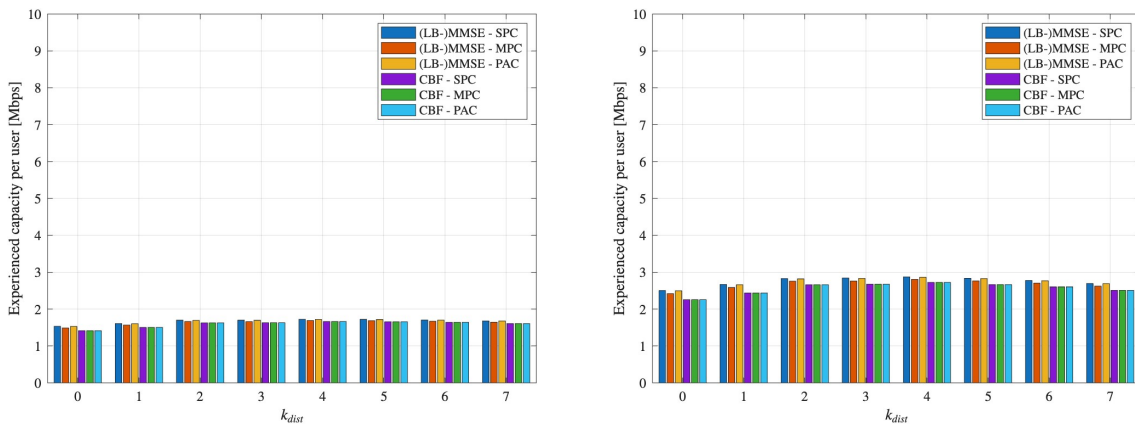


Figure155: Example of non-uniform user deployment with maritime and land areas for the inclined constellation and $\varepsilon_S = 30^\circ$.

Figure 156 and Figure 157 show the experienced user capacity with $\rho_{ML} = 0.1$ and $\rho_{ML} = 0.2$, respectively. The same considerations made in the polar case hold also in this case.



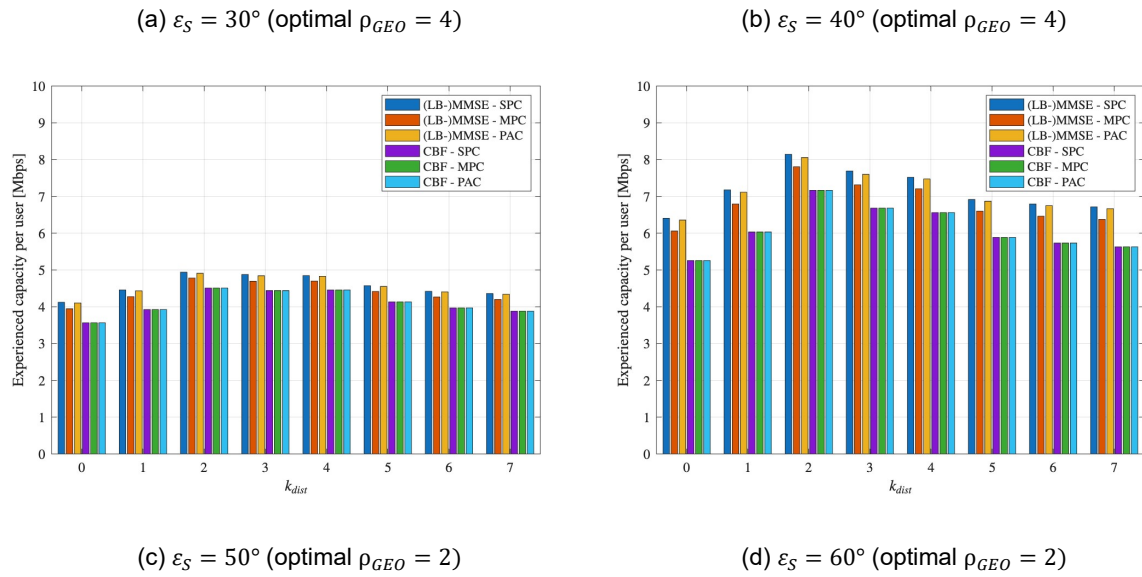


Figure156: Average experienced capacity per user with the inclined constellation (1300 km) as a function of the service area dimensions and with non-uniform distribution ($\rho_{ML} = 0.1$).

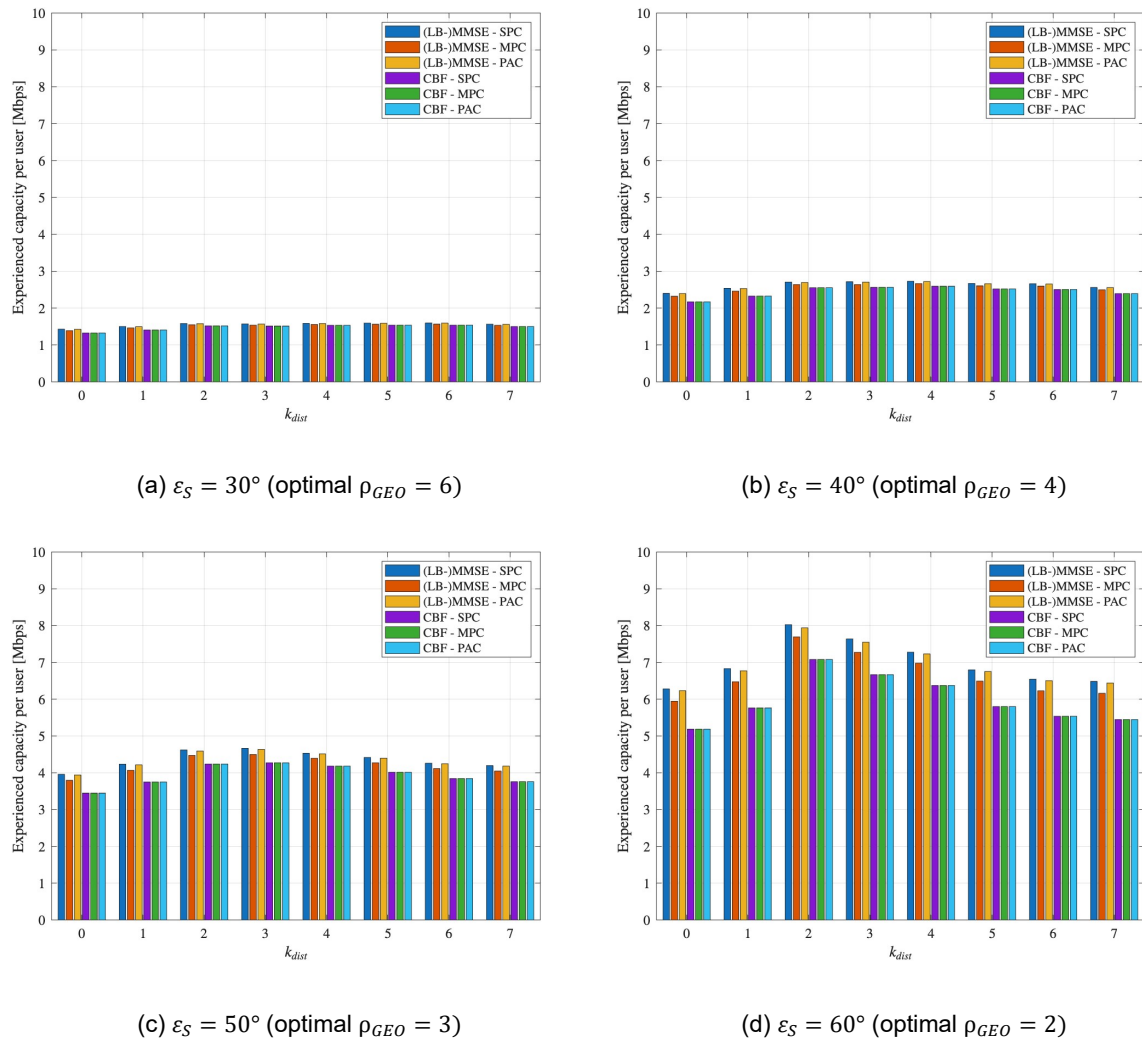


Figure157: Average experienced capacity per user with the inclined constellation (1300 km) as a function of the service area dimensions and with non-uniform distribution ($\rho_{ML} = 0.2$).

7.2.2.3 Comparison with Frequency Reuse schemes

Figure 158 shows the beam lattice and Table 49 summarises the performance comparison of user-centric beamforming with FR3 and FR4 for the inclined orbit; the same assumptions and observations made for the polar constellation case hold also in this scenario. Also in the inclined orbit case the optimal configuration of (LB-)MMSE-PAC is obtained with $\rho_{GEO} = 6$.

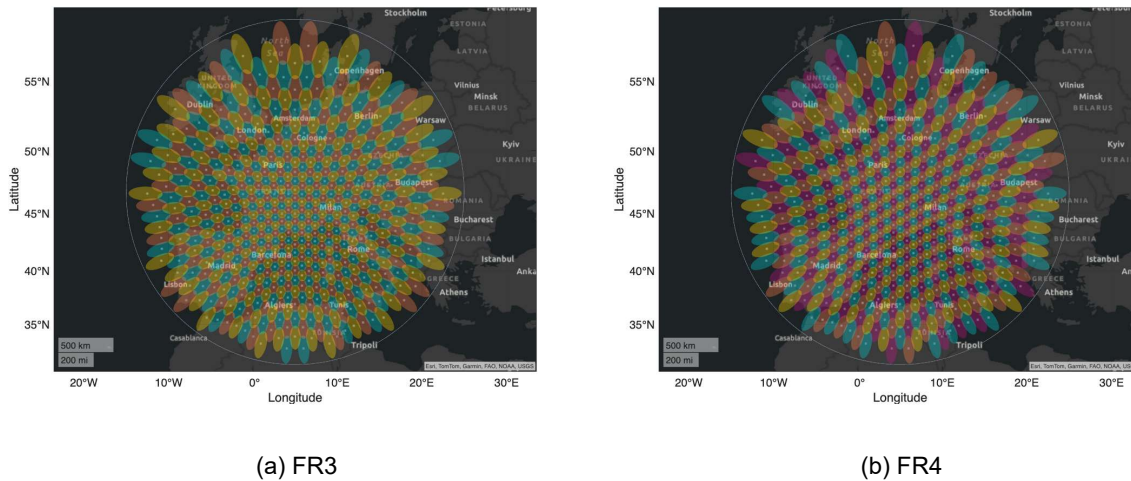


Figure158: Beam lattice with FR3 (left) and FR4 (right) for the inclined constellation.

Table 49: Performance comparison with FR3 and FR4 (relative gains/losses in square brackets) for the inclined constellation.

Parameter	MMSE PAC (optimal)	FR3	FR4
Asymptotic rate [bps/Hz]	3.44	6.29 [+82.45%]	5.38 [+56.39%]
Asymptotic capacity [Mbps]	653.47	398.91 [-38.95%]	511.45 [-21.73%]
Experienced capacity per UE [Mbps]	1.03	0.74 [-28.15%]	0.95 [-7.77%]
Average SINR [dB]	9.79	18.86	16.05

7.3 LESSONS LEARNT

Based on the above extensive performance assessment, we can highlight the most relevant trends that allowed us to identify the most feasible and interesting paths to be explored when introducing non-random scheduling and non-uniform traffic distributions:

- In general, the system is significantly impacted by the large slant ranges, which arise in particular at the edge of the service area due to the combination of the altitudes (1000 km and 1300 km) and the extension of the service area. This aspect, combined with the relatively limited on-board power per radiating element, leads to a system quite limited in terms of the SNR. Considering the large service area, the user experienced capacity is limited because they would need more allocations (so a reduced number of users) to be able to achieve larger capacities, considering a fixed and relatively low SINR. Ultimately,

this indicates that we need to find **solutions to better exploit the limited resources** (power and time slots) to serve the users.

- The performance of user-centric beamforming is improved by properly spacing users in the geographical domain, by avoiding as much as possible to schedule in the same time slot users that are close to each other. This is aimed at limiting situations in which the users' CSI vectors are too similar, which leads to ill-conditioned channel matrices. Both **geographical and CSI-based distancing techniques to improve the average SINR** per time slot can thus be explored.
- With respect to the specific algorithm, MMSE solutions provide a better performance compared to CBF. This advantage is more evident in scenarios with a particularly rich interference. **Both algorithms need to be assessed** with non-uniform traffic requests.
- The best **power distribution strategy** is the **PAC**, which is technically practical since it limits the power per antenna (avoiding HAPS operating in non-linear regime) while entirely exploiting the total available power on-board.

These considerations were at the basis of the development and design of the scheduling algorithms introduced in this report.