Architectures, standardisation, and procedures for 5G Satellite Communications: A survey

Architectures, standardisation, and procedures for 5G Satellite Communications: A Survey

A. Guidotti[a,*,1], S. Cioni[b], G. Colavolpe[c], M. Conti[a], T. Foggi[c], A. Mengali[b], G. Montorsi[d], A. Piemontese[c] and A. Vanelli-Coralli[a]

[a]*Department of Electrical, Electronic, and Information Engineering "Guglielmo Marconi," University of Bologna, Italy*
[b]*European Space Agency, Noordwijk, The Netherlands*
[c]*Department of Engineering and Architecture, University of Parma, Italy*
[d]*Department of Electronics and Telecommunications, Politecnico of Turin, Italy*

ARTICLE INFO

Keywords:
Satellite Communications
New Radio
Non-Terrestrial Networks
PHY/MAC procedures
5G
System architecture

ABSTRACT

The 5G ecosystem is radically shifting Information and Communication Technologies (ICT) towards a system directly tailored and optimised to support vertical markets and services. In order to design and implement the ambitious 5G network, Satellite Communications (SatCom) have been recognised as a promising solution to extend and complement terrestrial networks in unserved or under-served areas, as reflected by recent commercial and standardisation endeavours. Within 3GPP, this is reflected in the novel Study Item on Non-Terrestrial Networks to introduce 5G (New Radio, as per 3GPP nomenclature) technologies in SatCom. However, typical satellite channel impairments, as large path losses, delays, and Doppler shifts, pose severe challenges to the realisation of a satellite-based 5G network. In this paper, we thoroughly review the standardisation process performed within 3GPP from Rel. 15 to the present, discussing the different technologies and the related system architectures and interfaces. Based on these discussions, we assess the impact of the satellite channel characteristics on the major physical and medium access control layers, *i.e.*, Downlink Synchronisation, Random Access, and Hybrid Automatic Repeat reQuest. A set of potential solutions to the identified shortcomings is finally discussed.

1. Introduction

Today, Information and Communication Technologies (ICT) are at the very heart of any industrial, social, and economical activity around the globe. New technologies and applications continue to affect the way we perceive and interact with the world around us, but they also continuously fuel the users' expectations and necessities. As a consequence, the last years have been characterised by an unprecedented demand for improved broadband connectivity, near-zero latency services, and ultra-reliable and heterogeneous communications. A trend that is likely to continue, and actually further increase in the next years; for instance, forecasts report 5.3 billions users for Internet by 2023 and 14.7 billions Machine-to-Machine (M2M) connections, [1]. In such a rapidly evolving environment, the devices, services, and technologies that are currently well established in the global market, as, for instance, the 3GPP Long Term Evolution (LTE) standard, need to be significantly enhanced. However, improvement in existing technologies alone is not sufficient to meet the above significantly demanding requirements or to sustain novel market segments. The design of new technologies, leading the definition of new standards as 5G, or New Radio (NR) in 3GPP nomenclature, is thus of paramount importance. In particular, the massive scientific and industrial interest in 5G communications is motivated by the key role that these future systems will play in the global economic and societal processes to support the next generation vertical services, as massive Internet of Things (IoT), automotive, e-Health, Industry 4.0, etc., [2, 3].

5G systems are bringing a radical shift in the way both the access and the core networks are designed. Unlike past general-purpose standards to which applications and services were adapted, the 5G system is directly tailored and optimised to support a plethora of services and applications of the vertical markets. Such a flexible framework brought to the definition of very challenging requirements for the next generation of access technologies, in line with the forecast for future networks previously outlined. This is reflected in the ambitious objectives in terms of key performance indicators as, *e.g.*, peak data rates up to 20 Gbps in the downlink and 10 Gbps in the uplink, latencies on the User Plane (UP) down to 0.5 ms for time-critical applications and down to 4 ms for broadband services, [4]. In addition, 5G systems are characterised by advanced flexibility and re-programmability (*e.g.*, network slicing or Network Function Virtualisation, NFV), seamless global mobility in heterogeneous environments, and the support for multiple radio access technologies. With respect to the latter, 3GPP clearly states that Satellite Communications (SatCom) can provide significant benefits in terms of coverage extension and increased network availability, [5]. Thanks to their inherently large footprint, satellites are an efficient solution to complement and extend terrestrial networks, both in densely populated areas and in rural zones, as well as provide reliable Mission

This work has been supported by European Space Agency (ESA) ARTES Advanced Technology (3C.017) project "Spin-In of 3GPP terrestrial radio access technology for SATCOM." Opinions, interpretations, recommendations and conclusions presented in this paper are those of the authors and are not necessarily endorsed by the European Space Agency.

a.guidotti@unibo.it (A. Guidotti)
ORCID(s): 0000-0003-4480-0010 (A. Guidotti)

Critical services, [6]-[8]. While in the past satellite and terrestrial networks evolved almost independently from each other, making it difficult to *a posteriori* integrate them, the definition of the 5G paradigm offered the unique opportunity for these communities to define a harmonised and fully-fledged architecture from the very beginning.

In this context, 3GPP recognised the added value of SatCom and initiated several Study Items (SI) and succeeding Work Items (WI) related to Non Terrestrial Networks (NTN) for NR. It is expected that the integration of both satellite and aerial access networks in the 5G ecosystem can bring manifold benefits, [9]: i) complement 5G services in under- or un-served areas; ii) improve the 5G service reliability and continuity for M2M or IoT devices, or for Mission Critical services; and iii) enable the 5G network scalability by means of efficient multicast/broadcast resources for data delivery.

In this paper, we provide a detailed discussion and analysis of NTN systems both in terms of standardisation status, architecture solutions, and technology challenges for the Physical (PHY) and Medium Access Control (MAC) layers that shall be coped with in order to fully integrate non-terrestrial components in the 5G framework. In fact, since the NR air interfaces and the related procedures have been specifically designed for terrestrial systems, it is of paramount importance to properly assess the impact of typical satellite channel impairments on the PHY and MAC procedures.

1.1. Literature Review

Since the beginning of the NTN studies, an increasing interest and participation of companies from both the satellite and terrestrial communities has been witnessed. In addition to the standardisation efforts, which are the focus of this paper, also many funded completed and on-going projects are related to SatCom-based 5G systems, as, for instance: i) Sat5G, related to the development and validation of key technologies for enhanced Mobile Broadband (eMBB) scenarios, [10]; ii) SATis5G, in which a testbed to demonstrate the benefits of SatCom components in NR systems has been developed, including both Geostationary Earth Orbit (GEO) and Low Earth Orbit (LEO) systems, [11]; iii) 5G-VINNI, focused on developing an end-to-end 5G facility to be used to demonstrate implementation of the 5G infrastructure so as to support key 5G performance indicators, [12]; iv) 5GENESIS, a project aimed at providing an open set of tools to facilitate 5G experimentations, [13]; v) CloudSat, focused on the review and analysis of virtualisation and softwarisation technologies and the definition of their applicability in the SatCom context; vi) VITAL, considering the combination of terrestrial and satellite systems by bringing NFV into SatCom and by enabling Software Defined Networking (SDN) based resources management in hybrid satellite-terrestrial networks, [15]; and vii) SANSA, aimed at enhancing the performance of mobile wireless backhaul networks, both in terms of capacity and resilience, while assuring an efficient use of the spectrum, [16].

In addition to the above effort in terms of funded projects, also the scientific literature has provided a valuable addition to the global 5G knowledge, [6], [17]-[37]. In [6], the authors provide an overview of the possible 5G-based SatCom architectures and identify some of the most critical issues in terms of PHY and MAC layer procedures, as well as for the waveforms. In [17]-[19], the authors focused on resource allocation algorithms for multicast transmissions, in particular analysing the performance of the TCP protocol in a LTE-based GEO system. In [20]-[21], LEO mega-constellations in Ku-band were proposed to provide LTE broadband services, also taking into account the impact of typical satellite channel impairments as large Doppler shifts and delays on the PHY and MAC layer procedures. Moving to the introduction of 5G in SatCom systems, preliminary analyses have been addressed by the authors in [22]-[23], in which the focus was on the PHY and MAC layer techniques that shall cope with satellite channel impairments; these works were based on several assumptions due to the fact that the 3GPP standardisation for SatCom-based 5G was still in its infancy, which brought to a more detailed analysis in [6]. In [24], an interesting analysis of the adaptability of frequency-localised waveforms based on Orthogonal Frequency Division Multiplexing (OFDM) is provided in the context of non-linear satellite channels. The authors of [25] provide a valuable evaluation of the performance of live 4K video streaming over a 5G core network supported by a live GEO backhaul. Multimedia content delivery is also addressed in [26], in which a novel radio resource management algorithm for efficiently managing multicast multimedia content transmissions over satellite networks is designed, providing robust performance with respect to State-of-the-Art (SoA) solutions. In [27], the authors focus on Narrowband IoT (NB-IoT) and propose an interesting uplink scheduling technique to be used in LEO constellations, which aims at mitigating the level of the differential Doppler down to a value tolerable by the IoT devices. The authors of [28] addressed energy-efficient communications in satellite-based 5G systems, in particular investigating the fundamental relationship between spectral efficiency and energy efficiency. The authors of [29] provide an interesting feasibility analysis on the utilisation of C-band ($3.4 - 3.8$ GHz) spectrum for mobile 5G services, in particular taking into account the impact on existing Fixed Satellite Service (FSS) systems. In [30], a top-down network architecture for the integration of nanosatellites in 5G systems in the millimeter wave domain is described; the system performance is evaluated in terms of Signal-to-Interference-plus-Noise Ratio (SINR) in the presence of fading, shadowing, and interference, and both random access and routing aspects are discussed. The work in [31] is focused on the application of Non Orthogonal Multiple Access (NOMA) solutions to different satellite architectures, aiming at satisfying the demanding 5G requirements in terms of availability, coverage, and efficiency. Aspects related to spectrum assignment and sharing have been addressed in [32] and [33], respectively. In the former paper, an intelligent spectrum assignment algorithm is proposed based on dynamic cooperation between primary and secondary spectrum users for 5G-based SatCom systems. In the latter, the authors intro-

duce an advanced technique for spectrum sharing between satellite and terrestrial systems; in particular, the satellite spectrum is shared with terrestrial in-building small cells and the discussion is more focused on the terrestrial link performance assessment in the presence of a supporting satellite system. The authors of [34] and [35] focus on the design of the Random Access preamble. As for the former paper, a long preamble sequence is built by concatenating multiple Zadoff-Chu sequences and, to improve the detection performance, a joint correlation scheme is exploited based on power-delay profiles. In the latter publication, the authors propose a Random Access preamble based on a pruned Discrete Fourier Transform (DFT) spread Filer Bank Mutlicarrier (FBMC) waveform. In [36], the authors provide a detailed overview of the NTN system-level assumptions, together with a description of the methodology to be used for link budget calibration and the corresponding results. Finally, in [37], an interesting overview of 3GPP activities for NTN is provided; in particular, the authors first provide a quick introduction on the NTN scenarios and system characteristics and then focus on idle and connected mode UEs, providing an overview of the related mobility challenges.

1.2. Paper Contribution and Organisation

In this paper, we provide a detailed analysis of NTN systems in terms of standardisation, architectural solutions, and technology challenges for the PHY and MAC layers that shall be coped with in order to fully integrate non-terrestrial components in the 5G framework. To this aim, we provide a detailed survey discussing the status of standardisation, the architectures, and some of the PHY/MAC procedures of NTN systems. By means of numerical simulations, performed according to the latest 3GPP recommendations, we provide an insight on the most critical PHY/MAC challenges related to the integration of a satellite component in the NR system and a set of possible solutions, together with their pros and cons.

The remainder of this paper is organised as follows: i) in Section 2, we outline the status and time plan for 3GPP NTN standardisation, discuss the different architectural options, together with the related air interfaces proposed within 3GPP, and highlight the main advantages and disadvantages; ii) in Section 3, we describe the main PHY and MAC procedures in NR systems that will be considered throughout the paper, *i.e.*, the downlink synchronisation, Random Access, and Hybrid Automatic Repeat reQuest procedures as defined in the NR standard, also highlighting the most relevant aspects to be addressed; iii) in Section 4, we extensively discuss the challenges related to the introduction of non-terrestrial components in NR systems from the perspective of the considered PHY and MAC procedures, supported by numerical simulations and also proposing and assessing potential solutions; iv) finally, Section 5 concludes this work.

2. NTN Systems

In this Section, we provide a detailed overview of NTN systems in terms of the 3GPP standardisation activities and time plan, applications, architectures, and the related air interfaces. Moreover, we define a set of scenarios that will be addressed in terms of the NR procedures.

2.1. Standardisation

As previously discussed, 3GPP recognised SatCom systems to be a valuable solution to complement and integrate terrestrial NR networks. To this aim, several SI and succeeding WI were initiated within the Radio Access Network (RAN) and the Service and System Aspects (SA) Technical Specification Groups (TSG). More specifically, the first analyses highlighting the importance of satellites in the realisation of the fully-fledged 5G system were performed within Release 15 and summarised in TS 22.261, [5]; in that version of the TS, the main focus for satellites in NR was related to Mission Critical and industrial services.

Moving from the preliminary assessment in Rel. 15, an increasing interest rose within 3GPP for SatCom solutions in NR, also from key players in the terrestrial community, leading to the definition of NTN. These systems cover not only satellites, but also High Altitude Platform Systems (HAPS) and Unmanned Aircraft Systems (UAS), which are generally referred to as aerial systems in NTN. In this context, the following WIs and SIs can be identified based on the reference Release, as also summarised in Fig. 1 (it shall be noticed that the leading TSG is reported in this figure: also other RAN or SA groups contribute to the WI/SI studies apart from the leading one).

Release 15 Two SIs were initiated in 2017: i) "Study on NR to support Non Terrestrial Networks," with RAN as responsible group with the support of RAN1 (related to Layer 1); and ii) "Study on using Satellite Access in 5G," under the supervision of SA1 (Services). With respect to the former, the objective has been that of (see [9]): defining the NTN deployment scenarios and the related system parameters, identifying and assessing the potential key impact areas on the NR, and identifying the required adaptations of the 3GPP channel models for NTN; following these analyses, also solutions for the identified key impacts on the RAN protocols and architectures have been preliminarily proposed. As for the latter, it shall be noticed that it was initiated in 2017 but then moved to Release 16 and associated to the WI on "Integration of Satellite Access in 5G," discussed in the next paragraph. This SI focused on the definition of a set of use cases for the integration of a satellite component in NR, together with the identification of the potential services, and a categorisation of use cases based on service continuity, ubiquity, and scalability, [38].

Release 16 In the system and architecture framework, the following activities were initiated within Rel. 16: i) a SI on "Study on architecture aspects for using satellite access in 5G," within SA2 (Architecture), [39]; ii) a WI on "Integration of Satellite Access in 5G," under the supervision of SA1, [5]; and iii) a SI on "Study on management and orchestration aspects with integrated satellite components in a 5G network," under SA5 (Management), [40]. As also re-

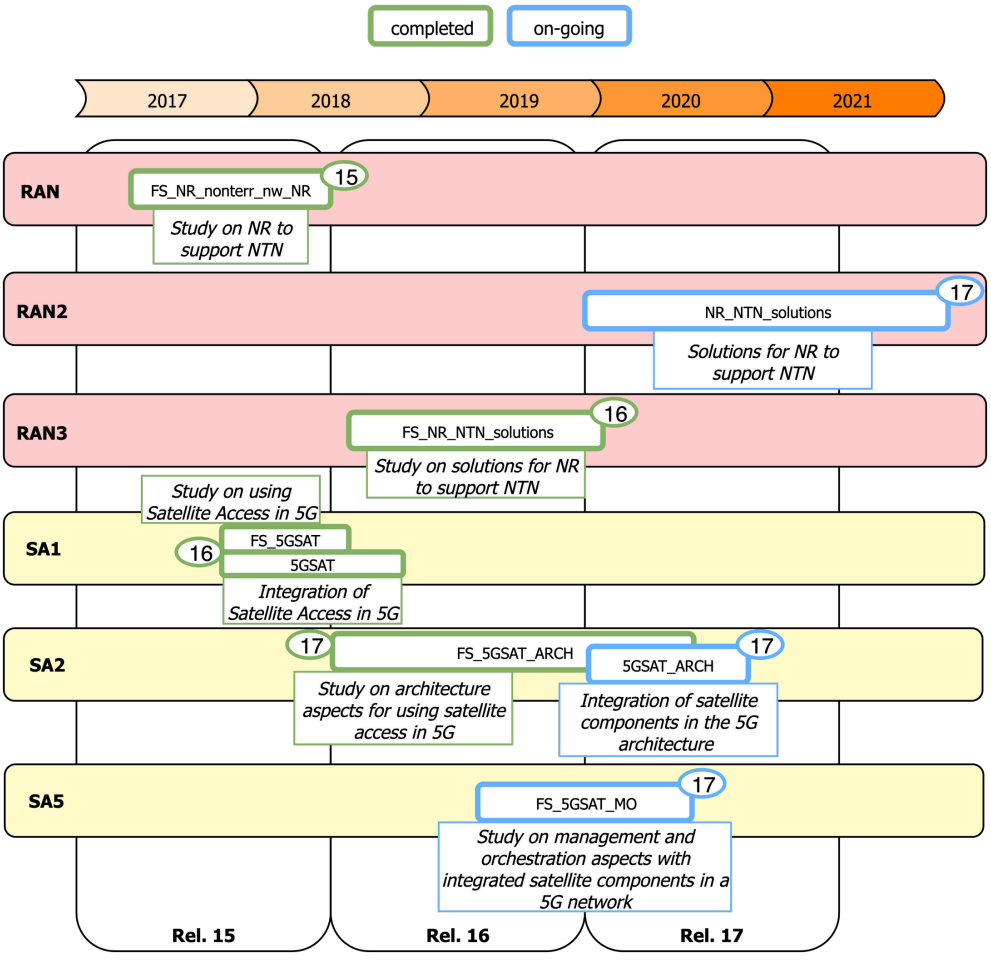

Figure 1: 3GPP timeline for the development of NTN (Mar. 2020, https://www.3gpp.org/ftp/Information/WORK_PLAN/).

ported on the 3GPP website[1] More in detail, the impacted areas related to the integration of SatCom in NR were identified, together with a set of potential solutions for two specific use cases from [38]: roaming between terrestrial and satellite networks, and 5G fixed backhaul between satellite enabled NR-RAN and the 5G Core; moreover, issues related to the interaction between the RAN and the core network were also discussed, [39]. The requirements of NR systems provided in [5] were extended with a satellite component in terms of multiple access technologies and connectivity models; moreover, specific key performance indicators for NTN were also included. Finally, aspects related to management and orchestration for NTN are currently being addressed, in particular identifying the most critical issues (and possible solutions) when including a satellite component in the NR network, [38]. With respect to the access technologies, RAN3 (Interfaces) led the activities on the SI "Study on solutions for NR to support NTN," which was completed at the end of 2019. Within this SI, building on the results obtained within Rel. 15 in [9], a set of required adaptations enabling NR technologies and operations in the NTN context were addressed, covering several issues in RAN1 (Physical layer), RAN2 (Layer 2 and 3), and RAN3 (Interfaces). In particular, the performance assessment of NR in scenarios including GEO and LEO satellites were provided at both system and link level, together with a preliminary set of potential solutions for NR adaptations at Layers 2 and 3. Moreover, some architecture aspects were modified with respect to TR 38.811, which is superseded from this point of view by TR 38.821, [41].

Release 17 Based on the above, at the end of 2019, two Work Items were officially started for NTN: i) "Solutions for NR to support NTN," under RAN2 (Layers 2 and 3) activities, but covering RAN1/2/3 technologies and techniques; and ii) "Integration of satellite components in the 5G architecture," under SA2. The activities for the former are still being agreed within 3GPP; however, the focus should be on the following aspects: i) consolidation of the performance assessment provided in [41] and of the potential impacts at PHY and MAC level; ii) the analysis of aspects related to Layers 2 and 3, *e.g.*, handover and dual connectivity; and iii) identification of the potential requirements for the upper layers based on the considered architectures. The latter WI aims

[1] See https://www.3gpp.org/ftp/Information/WORK_PLAN/ and https://www.3gpp.org/DynaReport/GanttChart-Level-2.htm#bm, both the former and the latter SIs were moved and completed under Rel. 17, as highlighted in the figure.

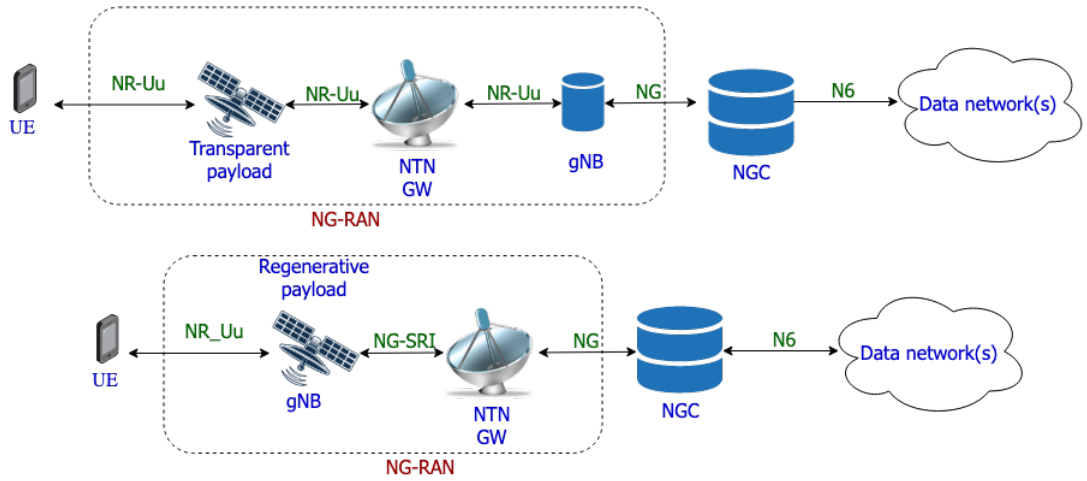

Figure 2: Architectural options with direct user access with transparent (top) and regenerative (bottom) payload. Applicable to: A, C1, and C2 (transparent); B, D1, and D2 (regenerative).

at extending the analysis provided in [39]: i) identification of impact areas of the SatCom integration in NR systems, in particular aiming at minimising such impact; ii) analysis of the issues related to the interaction between the RAN and the core network; and iii) identification of solutions for the two use cases highlighted above (terrestrial/satellite networks roaming and 5G fixed backhaul). In this context, several areas will be considered, including network discovery and selection, UP handling, network slicing, and many others, [42].

2.2. NTN System Architectures

Based on the Radio Access architecture and interfaces described in 3GPP TR 38.801, [43], a set of architecture options has been defined within the NTN Study Item, as per TR 38.811, [9]. The impact of the NTN environment on the technologies developed for terrestrial NR systems is strictly linked to the scenarios and architectures under consideration: the type of "flying" element and its capabilities, the constellation, and the use cases shall be clearly identified. Moreover, this impact does not only depend on the specific implementation of the SatCom user, space, and ground segments, but also on how these are interconnected and mapped to the NR network elements. In terms of the space segment, two macro-categories have been identified: the space-borne, *i.e.*, satellite-based communication platforms, and the airborne, *i.e.*, HAPS, devices. In the following, if not otherwise specified, we focus on space-borne platforms, *i.e.*, GEO or LEO constellations.

In the 3GPP framework, six macro-scenarios have been identified as outlined in Table 1. In all of the proposed scenarios, the user service link can be either in S-band (*e.g.*, 2 GHz) or in Ka-band (*e.g.*, 20 GHz in the downlink and 30 GHz in the uplink). With respect to LEO constellations, a further distinction is based on the type of on-ground coverage that is provided; in particular, on the one hand, the payload can be equipped so as to be able to steer the on-board antenna in order to always cover the same on-ground area,

Table 1
Summary of NTN reference scenarios per system type.

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

which means that the on-ground beams are not moving (as in the GEO scenarios). On the other hand, when no such capability is present, the area covered by each satellite will move accordingly with the satellite movement on its orbit, *i.e.*, the on-ground beams are moving with the satellite since the coverage center is always located at the sub-satellite point. Another important aspect to be highlighted is that scenarios D1 and D2 are the only ones in which it is possible to implement an Inter-Satellite Link (ISL), if needed for the chosen system architecture. These scenarios, which are classified based on the payload type and the presence or absence of an ISL, can then be mapped to different system architectures, as outlined below. From the beginning of NTN standardisation activities, scenarios A, C2, and D2 have been considered with higher priority; however, recently, the possibility to implement steerable beams (scenarios C1 and D1) is receiving an increasing attention. Finally, scenario B is not of interest, for the moment being.

The different architectural options can be broadly categorised based on: i) the type of satellite payload, either *transparent* or *regenerative*, in which the satellite can contain a full gNB(s) (next generation NodeB), part of it in case functional split solutions are implemented, or a Relay Node (RN); and ii) the type of user access link, either *direct* or *relay-based*, in which the User Equipment (UE) is connected to a RN and not directly to the satellite. Before detailing the architectures, it is worth highlighting that, in Figs. 2-4, a single system GW is represented for the sake of clarity; however, it shall be noticed that multiple GWs can be deployed

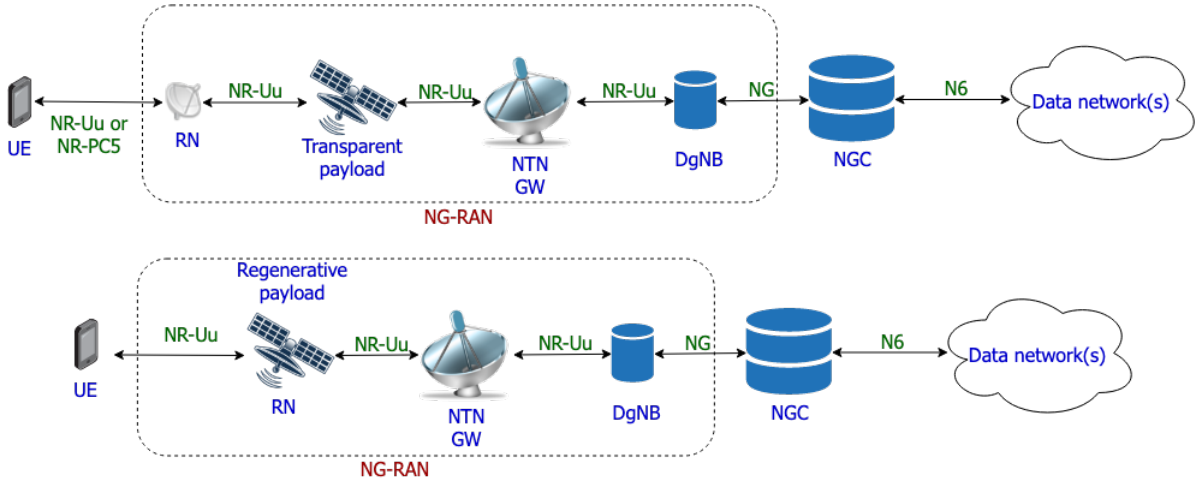

Figure 3: Architectural options with relay-based user access with transparent (top) and regenerative (bottom) payload. Applicable to: A, C1, and C2 (transparent); B, D1, and D2 (regenerative).

to provide connectivity to the data network(s). In this context, it is also worth mentioning that, for GEO systems, an on-ground beam is associated to a specific GW, while for non-GEO solutions specific requirements for service continuity shall be met to ensure that the successive serving GWs can manage the handover due to the satellites' movement.

Direct access Fig. 2 shows the architectures with direct access. In particular, the top diagram refers to a transparent payload and the bottom one to a solution with regenerative payload. The difference between the two architecture options, which also impacts the air interfaces as discussed below, is related to the location of the gNB. With a transparent satellite, the gNB is conceptually located at the GW, while with regenerative payloads it is implemented on-board. In both cases, the user access link between the satellite and the on-ground UE is implemented by means of the traditional NR-Uu air interface; this air interface is specifically designed for terrestrial systems and, thus, it is of paramount importance to properly assess the impact of the satellite channel impairments on the PHY and MAC procedures, as thoroughly discussed in the Sections 3 and 4. As for the feeder link between the satellite and the GW, the solution to be adopted depends on where the NR-Uu interface protocols are terminated, *i.e.*, on the position of the gNB: i) a transparent payload only acts as a radio-frequency repeater, thus not terminating the protocols, which requires the feeder link to be implemented with a NR-Uu solution; ii) in the regenerative payload case, the gNB is located on the satellite and, thus, the feeder link is implemented by means of the NG air interface, *i.e.*, the air interface between a gNB and the Next Generation Core network (NGC). Clearly, this architecture is more expensive and complex; however, this solution also allows to significantly reduce the propagation delays for NR PHY and MAC procedures and, thus, to ease the adaptations that might be needed for NTN (for a detailed discussion, refer to the following sections). Moreover, the NG air interface on the feeder link is a *logical* interface, *i.e.*, can be implemented with any existing Satellite Radio Interface (SRI), as long as specific signalling operations are guaranteed, [44]. Thus, it can be implemented with a modified version of the terrestrial air interface, or even by means of State-of-the-Art (SoA) solutions for SatCom, as the DVB-S2, [45], DVB-S2X, [46], or DVB-RCS, [47], air interfaces.

Relay-based access Solutions based on RN, also known as Integrated Access and Backhaul (IAB) in 3GPP terminology, are currently considered for further study, [41]; however, they are worth to be introduced for the sake of completeness. In this architecture, shown in Fig. 3, the UE is connected to a RN, which is a network entity quite similar to its LTE equivalent. In particular, the following aspects are worth to be highlighted for relay nodes: i) the RN is connected to a Donor gNB (DgNB), which is the entity providing connectivity towards the NGC; ii) the RN can terminate the procedures up to Layer 3; iii) the air interface on the user access link (RN-UE) can be implemented either as the traditional NR-Uu as for direct access or as a NR-PC5 interface, which is a device-to-device (or sidelink) air interface for NR, used, *e.g.*, in enhanced Vehicular-to-Everything (eV2X) applications; and iv) the air interface on the backhaul link (RN-DgNB) is a modified version of the NR-Uu air interface, in which only Radio Frequency characteristics and minimum performance requirements are different. Based on these operational aspects, the RN basically acts as a traditional gNB from the users' perspective, while it is seen as a UE from the DgNB, which motivates the implementation of the NR-Uu interface on both links and is also reflected in the RN attach procedure. The architecture with RNs and transparent payloads (top diagram in Fig. 3) is more complex with respect to the direct access scenarios due to the introduction of a potentially large number of on-ground relays, acting as gNBs; these entities must be managed by several DgNBs and, thus, the overall system cost might increase. However, since the RNs can terminate the protocols up to Layer 3, no modification is needed on the user service link. In this case, the im-

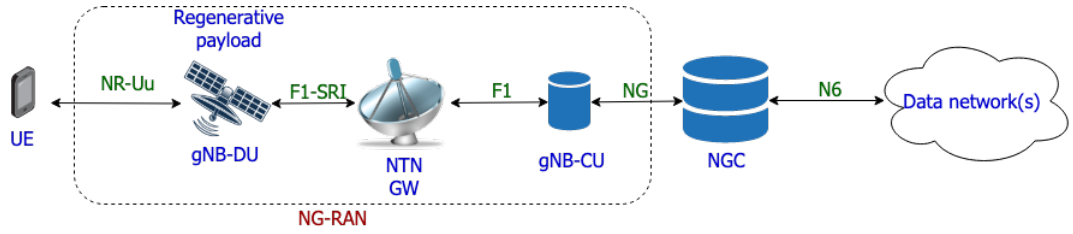

Figure 4: Architectural option with direct user access, regenerative payload and functional split. Applicable to B, D1, and D2.

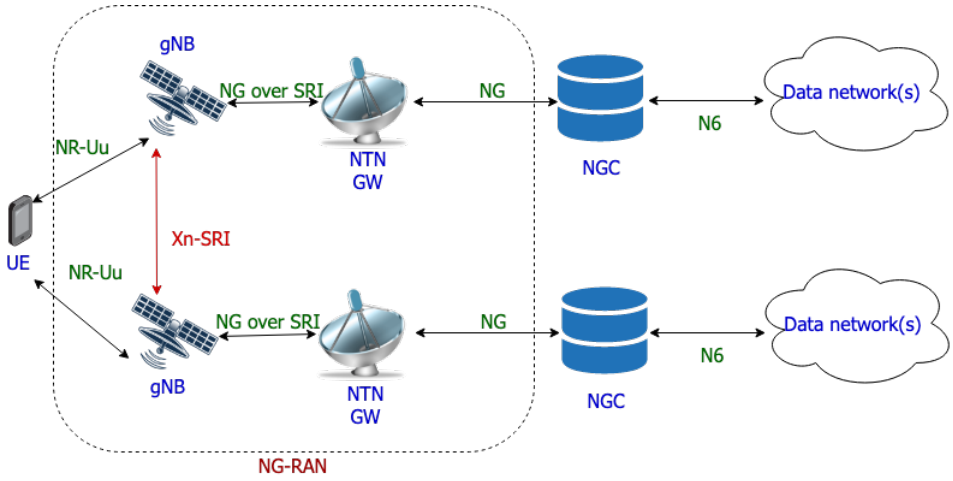

Figure 5: Architectural option with direct user access, regenerative payload and ISL. Applicable to B, D1, and D2.

pact of typical satellite channel impairments thus has to be assessed on the backhaul link only. When an on-board RN is implemented, *i.e.*, the regenerative solution, the payload cost increases but we also have the advantage of terminating the protocols up to Layer 3 on-board the satellite, as in the regenerative direct access architecture. In terms of feeder link, the connection between the RN and the DgNB must be implemented with a NR-Uu air interface, [46]. Finally, it shall be noticed that the NR-PC5 air interface can be implemented only when the RNs are on-ground.

Functional split Fig. 4 shows an architecture option with direct access and regenerative payload, in which functional split concepts are applied. The functional split allows scalable solutions, a significant adaptability to the different use cases and vertical services, and enhanced performance in terms of load and network management; moreover, it is at the basis of NFV and SDN. Clearly, this solution also increases the overall system cost. As provided in TS 38.401, [48]: i) a gNB can be split into a Centralised Unit (gNB-CU) and one or more Distributed Units (gNB-DU); ii) a gNB-DU can be connected to only one gNB-CU; iii) the air interface to be used between the gNB-CU and its gNB-DUs is the F1 air interface; iv) the F1 air interface is logical as the NG, *i.e.*, as long as specific signalling operations are ensured, it can be implemented by means of any existing standard, [49]. The split between CU and DUs can be implemented at different layers, and even within a given layer, as detailed in [43]; however, the most considered option in NR, which is also that considered for NTN for the moment being, is as follows: i) PHY, MAC, and Radio Link Control (RLC) are implemented in the DU; and ii) the Packet Data Convergence Protocol (PDCDP) and Service Data Application Layer (SDAP) for the UP or the Radio Resource Control (RRC) for the Control-Plane are implemented in the CU. Finally, it is worth highlighting that intermediate solutions can also be envisaged: apart from a CU and the controlled DU, there can be an Intermediate Unit (IU) that further splits the gNB in 3 entities and controls several DUs. This option is not yet considered in the current status of NTN systems.

Multi-connectivity In Fig. 5, we show an architecture in which the UE can be simultaneously connected to more than one satellite. The satellites are assumed to be regenerative, but it shall be noticed that also solutions with transparent satellites (and, thus, no ISL) are considered for NTN; in addition, the possibility to have multi-connectivity with one NTN satellite and a terrestrial gNB are also foreseen. In the solution represented in Fig. 5, the regenerative satellites are connected by means of the Xn air interface, which is a logical interface as the NG and F1, [50].

Constellations In the framework of NTN integration in 5G networks, the design of satellite constellations is another key aspect. As a matter of facts, according to the type of scenario and of service that should be provided, and the related requirements to be met, different satellite constellations can be used. For instance, a constellation with global coverage

would be required for a use case in which latency constraints are particularly stringent, while a constellation with a long revisit period could fit an IoT scenario where the device does not need to be constantly connected to the network. Then, the orbital parameters, such as the orbit inclination and altitude, the number of orbital planes, and the number of satellites per orbital plane, must be designed accordingly. As for the possible orbits, a first distinction can be made between elliptical and circular; in terms of NTN standardisation, the current analyses are focusing on the latter, [41]. In particular, aiming at assessing the network performance with different satellite constellations design, two main simulation methodologies have been proposed in TR 38.821, [41]:

- *constellation-based*: this approach is based on the definition of a reference constellation, in which all of the orbital parameters, the antenna pattern, and RF characteristics shall be defined, [51]-[53]. Notably, for a system aiming at achieving global coverage, the satellite altitude, the minimum elevation angle and inclination, as well as the Half Power Beam Width (HPBW) must be carefully chosen. According to the definitions in [41], a set of preliminary constellation parameters has been proposed in the past months in order to define an initial framework for performance evaluation. In particular, both GEO and LEO systems, at an altitude of 600 and 1200 km, are being addressed. As for the latter, a Walker star constellation layout with an inclination of 87.5° has been proposed, in which the reference parameters to compute the antenna patterns are defined in TR 38.821, [41], for both S-band and Ka-band. According to the constellation altitude and carrier frequency, the number of satellites per orbit and the number of beams per satellite change, such to ensure global coverage. The list of detailed parameters can be found in [51]. Another solution has been proposed in [52] based on a Walker Delta constellation, with similar parameters to the Galileo constellation, but at LEO altitude; since the constellation coverage depends on where the users are located, the coverage area to be simulated must be defined.

- *regional-based*: the second approach is based on the definition of a regional beam layout for multiple satellites. Instead of simulating the whole satellite constellation, as for the constellation-based methodology, a regional coverage for a specific geographical area is defined; then, the performance assessment is performed on this area to reduce simulation complexity while achieving accurate results, [55], [54].

2.3. Applications

Back in 2012, ITU-R started the development of the International Mobile Telecommunications (IMT) systems for 2020 and beyond (IMT2020). Under this programme, ITU-R defined the following three categories of service, which paved the way for the definition of 5G applications and requirements, [56]: i) enhanced Mobile Broadband (eMBB), for data-driven services requiring large capacities over wide coverage areas, also on moving platforms (*i.e.*, ships, planes, etc.); ii) massive Machine Type Communications (mMTC), characterised by a large number of devices typically requesting limited capacity without strict time constraints, *i.e.*, IoT applications; and iii) ultra Reliable and Low Latency Communications (uRLLC), for services with strict latency (1 ms) and reliability (packet losses in the order of 10^{-5}) requirements. Within 3GPP NTN, these use cases have been moved to a SatCom framework, thus defining a list of possible roles for NTN systems. In this context, it shall be noticed that NTN systems are currently focusing on space-borne platforms, which are not well suited for uRLLC use cases due to the stringent latency requirements. This aspect is reflected in TR 38.811, in which for NTN only eMBB and mMTC applications are considered, [9]. However, it is worth mentioning that this does not prevent NTN to have a role in uRLLC, *e.g.*, reliable backhauling for eV2X, Tactile Internet, or Public Protection & Disaster Relief (PPDR). Below, we provide an overview of the eMBB and mMTC services; for a detailed discussion on application aspects, the reader can refer to [8], [9], [56] and the references therein.

eMBB For this type of service, several applications can be envisaged for NTN systems to coexist and complement terrestrial networks, which can be broadly classified as follows:

- enhanced connectivity: the NTN system can cooperate with the terrestrial network(s) to provide broadband connectivity to underserved areas (*e.g.*, home, small offices, or for big events in ad-hoc built-up facilities), between the NGC and the cells in unserved areas or on moving platforms, or to ensure high availability to critical networks preventing network outage.

- broadcasting: NTN systems can support terrestrial mobile networks by offloading the traffic of media and entertainment contents or group communications at the 5G network edges; also Direct-to-Home (DTH) services or the provision of multimedia content to handheld terminals or on-board moving platforms or autonomous cars (*i.e.*, ICT on the road for eV2X) belong to this use case.

- public safety: both wide and local area public safety services can be significantly supported by NTN. The former, by means of direct communications to and between the emergency responders' terminals, achieving the continuity of service in mission critical scenarios anywhere; the latter, by guaranteeing a reliable communication between the NGC and the tactical cells, which can be connected by the satellite(s) allowing to exchange data, audio, and video signals between the responders and the coordination center.

mMTC In this category, the terrestrial networks can benefit from the traffic offloading offered by a space-borne platform or the guarantee of service continuity in under-/un-served areas. More specifically, NTN systems can bring significant

advantages for both wide and local area IoT applications: i) in the former case, there can be many IoT devices scattered over a large area (*e.g.*, vehicle software updates, smart agriculture, surveillance of critical infrastructures, asset tracking, etc.) that need a reliable connection to the central server for reporting; ii) in the latter, there can be groups of sensors collecting local information (*e.g.*, smart grid subsystems, monitoring of containers on moving platforms, etc.), eventually connected to each other, that need to report to the central server in order to define the orders to be sent to the actuators. In this context, for mMTC services we focus on LEO constellations.

For eMBB and mMTC applications, we consider multi-beam direct access solutions with both transparent and regenerative payloads. However, taking into account the current prioritisation of 3GPP NTN activities, the main focus will be on the former. Thus, the following assumptions hold throughout this work: i) the user access link (UE-satellite) is implemented with the NR-Uu air interface; and ii) the feeder link (satellite-GW) is implemented with the NR-Uu interface for transparent payloads and with the NG-SRI interface for the regenerative case. As for LEO constellations, we focus on scenarios C2 and D2, *i.e.*, the on-ground covered beams move together with the satellite along its orbit.

3. NR Procedures

In this section, we provide an overview of some of the most critical procedures at PHY and MAC level for NR systems, which might be deeply impacted by typical satellite channel impairments when considering the NTN framework.

3.1. Downlink Synchronisation

The first step in the implementation of a connection is the detection of the synchronisation signal block at the user terminal side. The procedure of downlink synchronisation relies on the frame and symbol synchronisation that the UE has to perform in order to obtain timing and frequency references, which enable the detection of the Physical Broadcast CHannel (PBCH), so that the information conveyed by the broadcast channel allows the UE to start the Random Access (RA) procedure. The NR procedure is similar to that in LTE, but the structure of the block, represented in Fig. 6 is more complex and flexible (in LTE it is not explicitly referred to as SSB, Synchronisation Signal Block). The SSB has a constant time-frequency structure, that is 240 subcarriers are occupied over four OFDM symbols. The first OFDM symbol transmits one out the three possible Primary Synchronisation Signals (PSS), Binary Phase Shift Keying (BPSK) modulated, which occupy the central 127 subcarriers among the 240 available; the 3 different sequences are obtained by shifting the same m-Sequence (better performance in case of timing and frequency offset ambiguity than the LTE Zadoff-Chu sequences). Thus, the UEs can run just three parallel correlators to detect the PSS. After the PSS detection, symbol synchronisation is obtained in two steps. First a coarse estimation of the frequency offset allows the detection of the correlation peak. Once the PSS has been detected, a fine estimation can be carried out by exploiting a proper frequency compensation technique. The UE can now search for the Secondary SS (SSS), which occupies the same 127 subcarriers as the PSS, but in the third OFDM symbol of the SSB. The SSS is one out of 336 possible Gold sequences, still BPSK modulated, and the detection relies on correlations that can now exploit the already acquired symbol timing. The combination of the detected PSS and SSS indexes allows to identify the cell identity (ID) index, which is in fact one out of 1008, given by the simple formula $3 \times \mathrm{SSS}id + \mathrm{PSS}id$. It is worth noting that some subcarriers of the SSB do not carry information, specifically 8 before and after the SSS, and the $240 - 127$ remaining subcarriers of the first OFDM symbol not occupied by the PSS. All the rest of the SSB subcarriers are used to transmit the PBCH, which consists of 432 payload resource elements (RE) and of 144 demodulation reference signals (DMRS) REs, for a total of 576 REs. The PBCH payload is given by 56 bits - 32 message bits of the master information block (MIB) and 24 CRC bits - which are polar encoded to 512 bits, which then undergo rate matching so that the final payload size is 864 bits, scrambled and mapped to 432 REs QPSK modulated. The detection of the DMRS, which is performed by correlating among $L = 4$, 8, or 64 possible sequences, depending on the frequency range at hand, allows the identification of 2 (for $L = 4$) or 3 bits (for $L = 8, 64$) of the SSB index (3 more from PBCH are needed in frequency range 2, when $L = 64$, where the SSB index is labeled with 6 bits), whereas the position of the DMRS is determined by the cell ID already acquired. The subcarrier spacing (SCS) of the SSB may be 15, 30, 120, or 240 kHz, depending on the type of SSB, as explained in the following, and on the frequency range, therefore the SSB bandwidth can be up to 60 MHz. The information transmitted in the PBCH are partly constant over 80 ms (8 frames), and are associated with the MIB, whereas other information is not constant over this time range (*e.g.*, the SSB index). The MIB carries several information needed by the UE to establish a connection, *i.e.*, the cell accessibility, the position of the system information block type 1 (SIB1), the location of

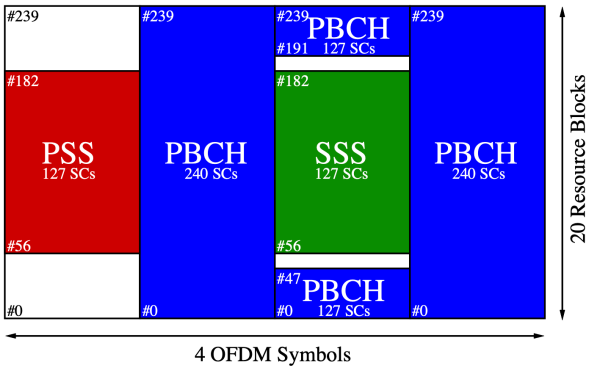

Figure 6: Synchronisation signal block, including primary synchronization signal, secondary synchronization signal, and physical broadcast channel.

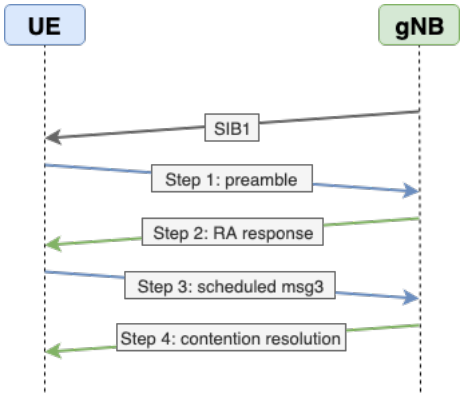

Figure 7: Random Access: contention-based procedure.

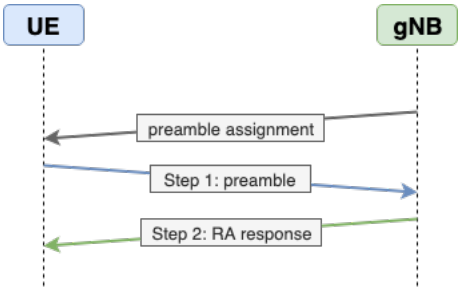

Figure 8: Random Access: contention-free procedure.

the common research grid, the system frame number (partly carried out by the encoding of MIB), etc. The SSB index is associated to different occurrences of the SSB within one half-frame (5 ms, either the first or second half of the frame), which form the SSB burst, as to implement beamforming. Some occurrences of the SSB may not be present, so that some flexibility is available to the gNB for the transmission of the SSB burst, in order to reduce the energy consumption and adapt to current cell conditions (the SSB is the only always-on signal in NR). For the same reason, the SSB burst periodicity may 5, 10, 20 (assumed as default), 40, 80, 160 ms. The maximum number of SSB occurrences in the burst is 4, for carrier frequency below 3 GHz, is 8 for carrier frequency within 3 and 6 GHz (these scenarios are referred to as frequency range 1), and is 64 for carrier frequency above 6 GHz (frequency range 2). The SSB types are five, and which one is transmitted depends on the SCS and the frequency range: type A (SCS 15 kHz), type B (30 kHz), or C (30 kHz) for Frequency Range 1 (FR1), type D (120 kHz) or E (240 kHz) for FR2. For NTN there is clearly no advantage in transmitting SSB bursts with more than one SSB, since no beamforming is possible.

3.2. Random Access

The RA procedure provides the UL synchronisation and it allows the UE to identify the exact timing for data transmission on the Physical Uplink Shared CHannel (PUSCH) and Physical Uplink Control CHannel (PUCCH). It follows the initial access, in which the UE identifies the serving cell based on one of the two following methodologies: i) scanning all RF channels in the NR bands to find a suitable (*i.e.*, best-serving) cell; or ii) leveraging on previously stored information of carrier frequencies and cell parameters. Once a cell has been identified, the UE can decode the SSB, which is transmitted according to a periodicity defined by the RRC, [57]. From the SSB, several information blocks can be decoded; for RA, critical information is carried by the System Information Block 1 (SIB1), which provides configuration parameters to all of the cell UEs.

The information provided by the SIB1 for RA depends on the type of RA procedure that is implemented. As shown in Figs. 7 and 8, both approaches are based on the transmission of a RA preamble, [58]: i) contention-based RA (CBRA) is implemented when the UE is trying to access the NR network for the first time or in case it lost the connection; ii) contention-free RA (CFRA) is the procedure applied when the UE was previously synchronised, *e.g.*, during a handover process. Focusing on the former, the SIB1 provides the RA configuration, as, *e.g.*, time and frequency locations of the RA channel resources (RA occasions) and all of the the parameters needed to build the RA preamble, extensively discussed in [57]-[59]. In the following, we focus on the preamble structure and on the procedure timers; in fact, these are two of the most challenging aspects when aiming at limiting the required modifications to RA in NTN compared to terrestrial systems.

Preamble The RA preamble structure is shown in Fig. 9 and it includes: i) a Cyclic Prefix (CP) with duration T_{CP}, aimed at reducing the Inter Symbol Interference (ISI) in fading environments; ii) a Zadoff-Chu random sequence with duration T_{SEQ}, sent by the UE to the gNB to request its connection to the NGC; and iii) a Guard Time (GT) with duration T_{GT}. The duration of the GT is directly related to the maximum cell radius, since it provides the maximum amount of differential delay that can be experienced by any two UEs in the cell, [23], [60]-[63]. As detailed in Section 4.2, the maximum differential delay shall not exceed the guard time T_{GT}. While the CP duration is directly provided in 3GPP documents, T_{GT} is not specified, but it is a consequence of the slotted nature of the RA channel; in particular, since the RA preamble shall fit an integer number of time slots, the GT can be computed as follows:

$$T_{GT}(\mu, \mathrm{f}) = N_{slot}(\mu, \mathrm{f})\, T_{slot}(\mu, \mathrm{f}) - T_{CP}(\mu, \mathrm{f}) - T_{SEQ}(\mu, \mathrm{f}), \tag{1}$$

where $\mu = 0, 1, 2, 3$ is the SCS parameter, f is the preamble format, $T_{slot}(\mu, \mathrm{f})$ is the time slot duration, and

$$N_{slot}(\mu, \mathrm{f}) = \left\lceil \frac{T_{CP}(\mu, \mathrm{f}) + T_{SEQ}(\mu, \mathrm{f})}{T_{slot}(\mu, \mathrm{f})} \right\rceil \tag{2}$$

is the condition ensuring that $T_{CP} + T_{SEQ} + T_{GT}$ is an integer multiple of T_{slot}. The values of T_{CP} and T_{SEQ} are provided in [59] as a function of the preamble format and of the SCS

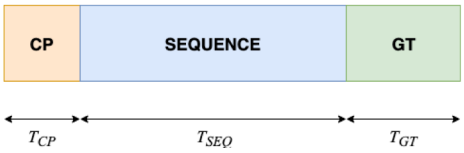

Figure 9: Random Access preamble structure.

used in the RA channel, Δf^{RA}: i) with formats 0, 1, 2, 3 (corresponding to long sequences in the time domain), Δf^{RA} can be either 1.25 or 5 kHz and these values can be considered for Frequency Range 1, *i.e.*, below 6 GHz; ii) formats $A1, A2, A3, B1, B2, B3, B4, C0, C2$ (short sequences) can be implemented with $\Delta f^{RA} = 15 \cdot 2^{\mu}$ kHz, with $\mu = 0, 1$ for FR1 and $\mu = 2, 3$ for FR2, *i.e.*, above 6 GHz.

Procedure timers Focusing on Fig. 7 for the CBRA procedure, the following operations are performed:

- Step 1: the UE randomly choses a RA preamble belonging to the format defined in SIB1 and also based on preliminary information on the expected amount of resources to be used in step 3. The preamble is sent to the gNB along with a temporary network identifier, which is computed based on the RA preamble.
- Step 2: the gNB responds to the request with a RA Response (RAR) message, which shall be received by the UE within a RA time window starting after the transmission of the last preamble symbol. This timer is the RRC parameter *ra-ResponseWindow* and it can be as large as 180 slots, [57], with an upper limit of 10 ms, [64]. If this counter expires, the UE can attempt a new RA procedure, up to 200 tentatives.
- Steps 3 and 4: the NR UE is assigned a final network identifier, subject to the successful resolution of any possible contention. In particular, Hybrid Automatic Repeat reQuest (HARQ) is implemented in step 3, with a contention timer up to 64 subframes, *i.e.*, 64 ms, *ra-ContentionResolutionTimer* in [57]. If the UE receives a correct response in Step 4, the procedure is successful.

As for the CFRA procedure, from Fig. 8 it can be noticed that the preamble is directly assigned by the gNB; this approach greatly simplifies the following steps, since only the preamble transmission and the response from the gNB are required because Step 3 contentions are avoided *a priori*.

3.3. HARQ

HARQ (Hybrid Automatic Repeat reQuest) is one of the key technologies introduced in NR to improve the link adaption capability of the air interface. In the following, we provide a review of the link adaptation techniques and then focus on the HARQ implementation in the NR standard.

Link adaptation techniques HARQ is a technology complementary to Adaptive Coding and Modulation (ACM) and Power Control (PC). In Adaptive Coding and Modulation, the user throughput is adapted to the time varying channel conditions by properly changing the coding and modulation scheme to ensure reliable communication. With power control, on the other hand, link adaptation is obtained by changing the power level and, consequently, the channel conditions, according to the desired user throughput. Both techniques strongly rely on the availability of reliable channel state information for the next transmission opportunity at the transmitter (CSIT). The presence of unreliable or outdated CSIT strongly affects the performance of ACM and PC link adaptation techniques. In this case, the only possible countermeasure is to adopt a proper margin, defined based on the quality of the available CSIT, to optimally trade between capacity and frame error rate. HARQ is particularly effective in this case as it allows to adjust the coding rate depending on the actual channel conditions, with a control loop driven by the receiver.

HARQ differs from the more traditional ARQ technique because the retransmission process (HARQ process) is controlled at the first stages of the physical layer, namely by the channel decoder. In Type II HARQ, the soft-information available at the decoder is preserved during the HARQ process from one transmission to the following one by inserting memory also at the receiver side. The availability of soft-information considerably improves the performance of the decoder and, consequently, provides higher throughput. The additional advantage of controlling the process at the very first receiver stages is that of minimising the control loop latency, a fundamental parameter for its efficiency. Type II HARQ is further classified in two main categories: i) Chase Combining HARQ (CC-HARQ); and ii) Incremental Redundancy HARQ (IR-HARQ). With CC-HARQ, coded bits, or a subset of them, are retransmitted in following transmissions; at the receiver, the corresponding soft-information, in the form of bit log-likelihood ratio (LLR), is simply accumulated before next decoding attempt. This approach has the effect of increasing the effective Signal-to-Noise-Ratio (SNR) of each retransmitted bit sequence. With IR-HARQ a new *redundancy version* of the same information transport block is transmitted at each new transmission attempt. The effective code rate at the receiver is then truly lowered at each retransmission and the effect is that of accumulating the mutual information of the transport block instead of increasing the SNR. IR-HARQ is always better than CC-HARQ in terms of performances, although this difference may be negligible at low SNR.

IR-HARQ process In NR, the channel encoding scheme is based on Low Density Parity Check (LDPC) codes and it has been designed to efficiently support IR-HARQ. The ACM procedure selects the coding rate and modulation scheme (MCS), depending on the CSIT, and consequently determines the transport block size (TBS) K for the set of allocated resources n in the current slot. The initial TBS and rate deter-

mine the base graph of the LDPC used for encoding. Four redundancy versions (RVs) are available, corresponding to the reading of the generated LDPC codeword starting from 4 different positions in a circular buffer containing it. The NR standard allows to select the sequence of RVs indexes that are sequentially used during the HARQ process and the maximum number of retransmissions.

When HARQ is active, *i.e.*, $N_{TX} > 1$, after transmitting the first RV, the process waits for a NACK/ACK feedback from the receiver. The NACK signal triggers the transmission of the new redundancy version, while the ACK signal (or the last allowed retransmission) release the HARQ process. In order to deal with the latency involved in the HARQ control loop, a number of HARQ processes are managed in parallel at both the transmitter and receiver side. Each process stores and updates the information relative to the transmission of a single transport block; the processes are allocated at the first transmission and released at the last transmission or after the correct reception of the packet. Each process requires, at the transmitter side, a buffer to store the codeword associated to the base LDPC graph and, at the receiver side, a buffer to store the corresponding soft information. The number of required HARQ processes to manage a control loop with a latency of D slots and a maximum of N_{TX} transmissions is $D(N_{TX} - 1)$; in NR, the maximum number of HARQ processes is bounded to 16, corresponding to a maximum latency of 5 slots with 3 retransmissions or 8 slots for 2 retransmissions.

4. Technical Challenges and Solutions

Depending on the selected system configuration, in particular in terms of scenarios and architecture options as outlined in Sections 2.2 and 2.3, respectively, different technical challenges will arise in order to implement the NR procedures discussed in the previous section. In particular, the impact of typical satellite channel impairments shall be taken into account, since these can be extremely different with respect to terrestrial systems. In the following, we focus on the most critical issues in the procedures identified in the previous Section when taking into account the large delays, Doppler shifts, and path losses of SatCom systems.

4.1. Downlink Synchronisation

The main problem is due to Doppler effects in LEO links [41]. The satellite orbital velocity is in the order of 4-7 km/s for altitude in the range 600-1200 km, and induced Doppler shift and rate can be visualized in Fig. 10. For uncompensated systems, the value of Doppler shift can be as high as 20 or 24 ppm, for altitude of 1200 and 600 km respectively, whereas the Doppler rate can reach 0.09 and 0.27 ppm/s for the same altitudes, as shown in Fig. 10. The amount of Doppler shift can be reduced to a few ppm in case of pre/post-compensation (with the exception of very large beam footprint size, in fact the residual Doppler shift in case of 1000-km diameter can still be around 15 ppm). In addition, 10 ppm of frequency offset must be accounted for due to UE crystal accuracy, and an almost negligible Doppler shift

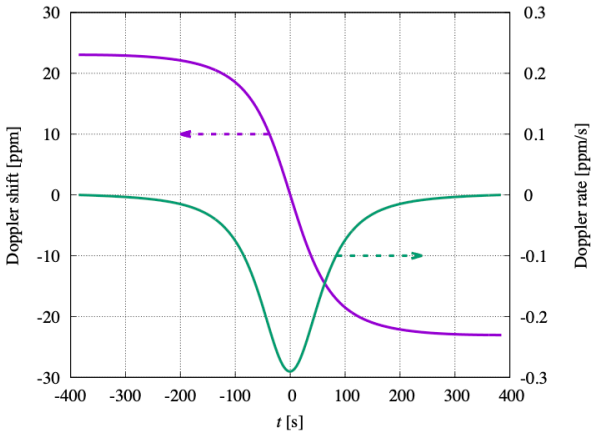

Figure 10: Doppler shift and rate in ppm and ppm/s respectively for satellite altitude of 600 km, as a function of time.

due to the UE movement (which also entails Doppler spread) and satellite oscillator accuracy. Such high uncompensated values may represent an obstacle for UEs synchronisation. Moreover, another impairment caused by the Doppler is the timing drift [65]. The compensation can be implemented effectively by knowing the satellite ephemeris, so that a pre-compensation can be performed at the receive side or at the transmit side (with respect to the beam center). In the former case, the UEs need to also know their own position, whereas in the latter case the pre-compensation can be made with respect to the beam center, hence a significant residual Doppler shift can affect the signal received by UEs at the edge of the beam footprint, which can be up to several hundreds of kms. In this latter case, the UEs need to also know their own position. The performance of the current synchronisation algorithms allows to detect the SSB as required with the residual frequency offset due to Doppler effects, but must be further investigated without pre-compensation, [66]. Indeed, the possibility to synchronise low-class UEs, lacking of advanced processing features like the ephemeris knowledge, *e.g.*, in the mMTC use case, would represent an interesting field of research, which would rely on the assessment of the current synchronisation algorithms in worst case scenarios, and on the investigation of simpler but more performing algorithms. Pre-/post-compensation methods basically reduce the residual frequency offset to values due to UE crystal accuracy. Preliminary studies highlight that better performance in the timing drift compensation can be expected for higher SCS (see [65], where both open - self-adjustment by UE - and closed-loop - UE exploits information by network - algorithms have been considered), whereas the performance of the SSB detection in presence of normalised residual frequency offset is almost independent from the SCS. Better performance was observed by using PSS/SSS-based compensation techniques with respect to DMRS-based ones. Then, it was observed that the UE needs to compensate the timing drift in order to avoid ISI and phase errors, and to reduce the equaliser complexity.

4.2. Random Access

As previously mentioned, one of the most impacting impairments on the RA procedure is the propagation delay, [6], [20]-[23]. In particular, two different requirements arise when focusing on the GT period and on the procedure timers: i) the maximum differential delay between any two users in a given beam shall be totally absorbed by the GT period; and ii) the Round Trip Time (RTT) between each UE and the network element implementing the RRC (which is a function of the implemented functional split option) shall be at most equal to the procedure timers. To assess these aspects, an analysis at beam level and at single UE level, respectively, is required; in fact, the preamble GT period is impacted by the maximum differential delay within a beam, while the procedure timers by the overall RTT on the link of a single UE.

Let us consider a single multi-beam satellite; this assumption does not affect the generality of our analysis, since with multiple satellites the same approach and considerations apply to each one, both at beam and at single link level. In terms of beam layout and payload characteristics, two sets of satellite parameters are proposed in TR 38.821, [41]. These two sets provide the payload configuration for both the UL and the DL and are to be used as a baseline for system-level calibrations depending on the considered use case. In the following, we consider the NTN system with the parameters from Set-1 ([41, Table 6.1.1.1-1]), summarised in Tables 2 and 3 for the sake of completeness and for which an extensive discussion can be found in [36]. It is worth highlighting that, in case Set-2 parameters are of interest, the same procedure as that discussed in the following can be applied by simply using different configuration parameters. It is worth highlighting that we consider $N_{tiers} = 6$ concentric tiers around the central beam in the coverage, *i.e.*, the maximum coverage for NTN, [41]. Depending on the satellite orbit, a different elevation angle at the coverage center ε_C, is assumed, [41].

It is worth highlighting that, for NTN, the current assumption is that each UE has GNSS capabilities and has also knowledge of the satellite orbit; in this case, unless significant errors are introduced by the GNSS estimates, no issue arises since each UE can pre-compensate the uplink transmission. However, this solution poses additional challenges as, *e.g.*, how the UEs can have updated information on the satellite orbit (they can be deployed in an area with no terrestrial connection from which it can be downloaded), how much battery consumption is introduced for GNSS and the related pre-compensation, etc. Thus, we focus on the case in which this possibility is not available and assess the impact of the differential delays on the RA procedure.

Preamble For the preamble analysis, let us focus on a the generic b-th beam. We need to ensure that the maximum differential delay between any two UEs in the beam does not exceed the delay that can be absorbed by the GT, *i.e.*:

$$\max_{i,j\in\mathcal{B}_b} \Delta T\left(\tau_i, \tau_j\right) \leq T_{GT}\left(\mu, \mathrm{f}\right), \tag{3}$$

Table 2
Beam layout parameters for GEO systems, [41].

Band	Parameter	Value	Units
Ka	Equiv. antenna radius	2.5	m
	3 dB aperture	≈ 0.0882	deg
	Beam radius	55	km
	ε_C	45	deg
S	Equiv. antenna radius	11	m
	3 dB aperture	≈ 0.2006	deg
	Beam radius	125	km
	ε_C	45	deg

Table 3
Beam layout parameters for LEO systems, [41].

Band	h_{sat}	Parameter	Value	Units
Ka	600 km	Equiv. antenna radius	0.25	m
		3 dB aperture	≈ 0.8823	deg
		Beam radius	10	km
		ε_C	90	deg
	1200 km	Equiv. antenna radius	0.25	m
		3 dB aperture	≈ 0.8823	deg
		Beam radius	20	km
		ε_C	90	deg
S	600 km	Equiv. antenna radius	1	m
		3 dB aperture	≈ 2.2064	deg
		Beam radius	25	km
		ε_C	90	deg
	1200 km	Equiv. antenna radius	1	m
		3 dB aperture	≈ 2.2064	deg
		Beam radius	45	km
		ε_C	90	deg

where $\mathcal{B}_b$ is the set indexes of the UEs belonging to the b-th beam. We denoted by τ_i, with $i \in \mathcal{B}_b$, the delay between the satellite and the i-th UE, which can be computed as:

$$\tau_i = \frac{d_i}{c} = \frac{R_E\left(\sqrt{\frac{\left(R_E+h_{sat}\right)^2}{R_E^2} - \cos^2\varepsilon_i} - \sin\varepsilon_i\right)}{c}, \tag{4}$$

where d_i is the i-th UE slant range, $R_E = 6371$ km being the Earth's mean radius, h_{sat} the satellite altitude, ε_i the elevation angle of the i-th user, and c the speed of light. Figs. 11 and 12 provide an example of differential delay. In particular, they are obtained with LEO satellites at $h_{sat} = 600$ km and Ka- and S- band, respectively, with the latter showing an increased coverage due to the lower carrier frequency. In this case, since the coverage center is at the Sub-Satellite Point (SSP), a symmetric behaviour of the differential delay arises and beams that are farther away from the coverage center experience increased differences. The differential delays are significantly larger due to the increased beam size.

In order to assess the impact of the differential delay on the RA preamble, we consider the preamble formats from $A1$ to $C2$ with: i) $\mu = 0, 1$ in S-band; and ii) $\mu = 2, 3$ in Ka-

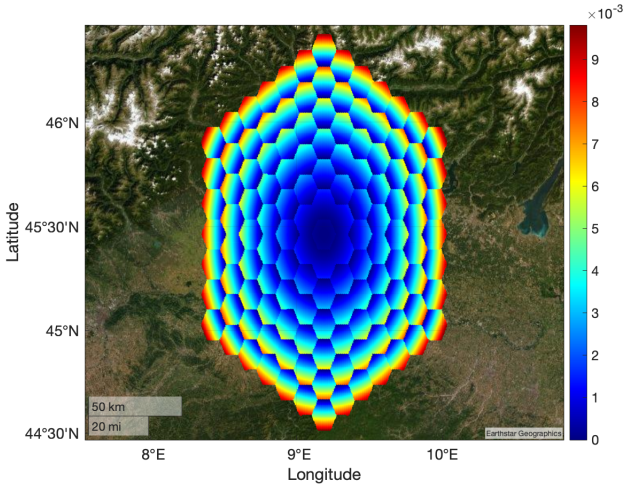

Figure 11: Differential delay per user in [ms], LEO system in Ka-band, $h_{sat} = 600$ km.

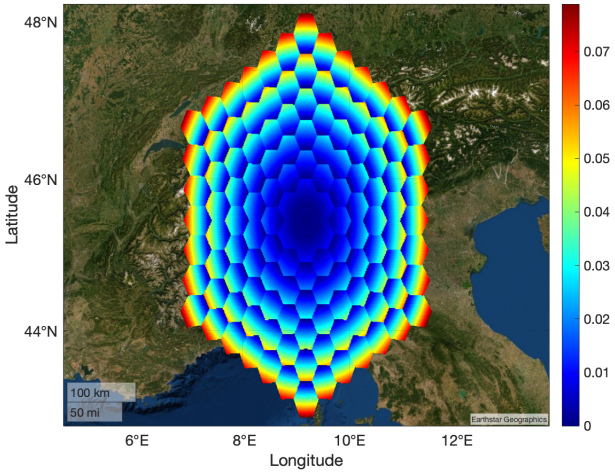

Figure 12: Differential delay per user in [ms], LEO system in S-band, $h_{sat} = 600$ km.

band. This choice is motivated by observing that formats $0, 1, 2, 3$ are allowed with subcarrier spacings on the Physical RA CHannel (PRACH) of 1.25 and 5 kHz, which are not foreseen for NTN, at the present time. As per (3), the differential delays in each scenario are compared to the T_{GT} for each format in S-band ($\mu = 0, 1$) and Ka-band ($\mu = 2, 3$), to assess which combinations are feasible without modifications. Figs. 13 and 14 show the percentage of users for which the differential delay is above the GT duration.

- GEO systems pose the most difficult challenges, due to both the satellite altitude and the elevation angle at the coverage center, which introduce large differential delays. Consequently, even by exploiting the preamble providing the largest margin (format C0, with $T_{GT} = 0.89, 0.45, 0.22, 0.11$ ms for $\mu = 0, 1, 2, 3$), we have: i) in S-band, 16% and 53% of users that cannot perform the RA procedure for $\mu = 0$ and 1, respectively; and ii) in Ka-band, these values are 38% and 76% for $\mu = 2, 3$. With other formats, the percentages can be as large as 95% and 98% with format B4 in Ka-band.

- LEO systems do not exhibit significant issues, thanks to the reduced altitude and the coverage area centered at the SSP; this can be noticed in the figures by observing that often no bars are present for this case, corresponding to a null percentage of users with differential delay above the GT duration. The only case in which a non-null percentage of users will experience an outage related to the preamble is with format B4 and $h_{sat} = 1200$ km: i) in S-band, we have 0.15% and 22% for $\mu = 0, 1$; and ii) in Ka-band, the outage is 0.05% with $\mu = 3$, while it is null for $\mu = 2$.

Based on the above analysis, in general, GEO systems are those for which particular attention shall be devoted to the RA procedure in terms of the RA preamble. In particular, in order to allow the UEs to send the RA preamble and at least be sure that the differential delay constraint is respected, two possible solutions might be envisaged. First, novel format options can be defined with increased guard times; however, this approach would require modifications to the current NR RA procedure. Moreover, for GEO systems very large guard times or cyclic prefixes are needed and, thus, the RA procedure would be significantly slowed down; this can also have an impact on the system throughput and, thus, it is not recommended. The second possibility is to have a system architecture that ensures lower differential delays; this can be performed by identifying the minimum elevation angle at the coverage center ensuring that condition (3) is met. With this approach, no modification would be needed to the RA procedure, since any challenge would be resolved geometrically. The other scenario in which the condition in (3) is not met is the LEO case at 1200 km with format B4; in particular, in S-band the percentage of users experiencing larger differential delays is non-negligible and shall be properly taken care of. The solutions proposed for GEO scenarios can, of course, still be taken into account. However, with respect to the definition of novel formats allowing for larger differential delays, the additional delay introduced in the RA procedure would be detrimental; in fact, since the satellite is now moving, the RA procedure shall be completed as fast as possible so as to avoid that the UE is granted resources that can be provided by the connected satellite for a too limited period (before an handover procedure is needed). Thus, a geometry solution seems to be the best option also for LEO cases. Both for GEO and LEO scenarios, in case neither of the above options are feasible, a third solution to the differential delay issue is clearly the reduction the beam size. However, in this case more beams are required to cover the same on-ground area; moreover, in the LEO case, and in particular when considering the fixed beams scenarios, this would also increase the rate at which satellite handovers are needed, thus signif-

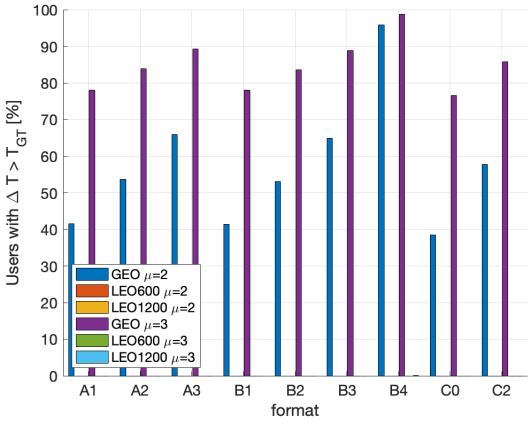

Figure 13: Percentage of users for which condition (3) is not met, Ka-band.

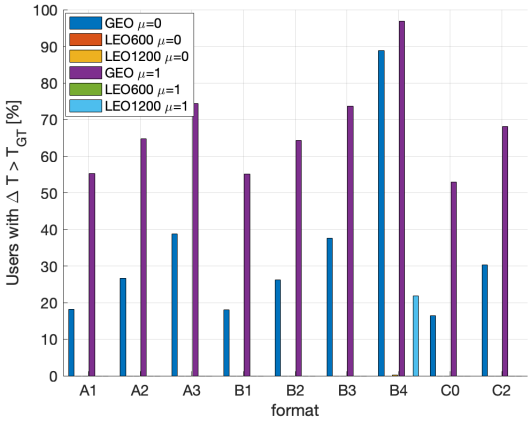

Figure 14: Percentage of users for which condition (3) is not met, S-band.

icantly increasing the system complexity.

Finally, it shall be noticed that solutions based on the modification of the RA preamble structure are being discussed in the 3GPP context. In particular, in [41, Section 6.3.3], the following modifications have been initially proposed: i) single Zadoff-Chu sequences with larger subcarrier spacings; ii) multiple Zadoff-Chu sequences with different roots; iii) Gold or m-sequences with additional modulations and precoding; and iv) a single Zadoff-Chu sequence combined with different scrambling sequences. These solutions, differentl from a simple increase in T_{CP} or T_{GT}, are based on deeper adjustments of the preamble structure. Moreover, in [41, Section 7.2.1], a discussion on the improvement of the RACH capacity is provided, which shall be taken into account in case solutions as the above proposed extension of T_{CP} or T_{GT} are exploited to assess the system performance.

Procedure timers In this case, we focus on a single link between a UE in the generic b-th beam and the network entity which terminates the RA protocol. Depending on the type of satellite payload, the termination can happen either onboard or at the system GW. In the latter case, the following condition shall be met in order to ensure that the generic i-th user can perform the RA without modifications:

$$2\tau_i + 2\tau_{GW} \leq T^{(x)}, \forall i \in \mathcal{B}, \tag{5}$$

where $\mathcal{B} = \bigcup_{b=1}^{N_B} \mathcal{B}_b$ is the set of indexes of all users (N_B is the total number of beams) and $t_{GW} = d_{GW}/c$ is the propagation delay between the GW and the satellite. $T^{(x)}$ is the procedure timer, with $x = RAR$ and $x = CR$ to denote the RA response and the Contention Resolution timers, respectively. When a regenerative payload is implemented, this condition can be relaxed to the following:

$$2\tau_i \leq T^{(x)}. \tag{6}$$

In the transparent payload case, it is assumed that the GW sees the satellite at an elevation angle $\varepsilon_{GW} = 10°$, which is the minimum value as per TR 38.821, [41]. The following delays can thus be computed for the GW-satellite link: i) 135.38 ms for GEO; ii) 10.45 ms for LEO at $h_{sat} = 1200$ km; and iii) 6.45 ms for LEO at $h_{sat} = 600$ km. Based on the above and on the delay computation in (4):

- GEO systems are critical, since all of the users in the coverage area will experience an RTT larger than the timers $T^{(RAR)}$ and $T^{(CR)}$. This happens for both transparent and regenerative payloads, due to the large link delay between the satellite and each user, in the order of approximately 125 ms;
- for LEO systems, the propagation delay between a user and the satellite is in the order of a few ms (approximately 2 for $h_{sat} = 600$ km and 4 at $h_{sat} = 1200$ km). Thus, even with transparent payloads there is no issue for the Contention Resolution timer, since $T^{(CR)} = 64$ ms. As for the RAR timer of 10 ms, on the one hand, when regenerative payloads are implemented no issue arises; on the other hand, with transparent payloads the additional delay of $2\tau_{GW}$ (20 ms at 1200 km and 13 ms at 600) becomes critical and no user is able to complete the procedure.

When a regenerative payload is considered, only GEO satellites experience issues in terms of the procedure timers. Even assuming a coverage centered at the SSP, the overall RTT values would be greater than the timers and, thus, this solution cannot be adopted. The only possibility is that of increasing these timers for NTN RA procedures in GEO scenarios, which would also impact the system performance and, thus, needs to be further analysed. For LEO satellites, only the case at $h_{sat} = 1200$ km can be challenging, in terms of the RA response timer and similar approaches as those for GEO systems can be envisaged. Finally, it shall be noticed that these aspects are related to procedure timers impacted by the single UE RTT, and not to differential delays as in the preamble GT analysis; as a consequence, no precompensation can be implemented by exploiting GNSS capabilities and MAC level modifications are thus required.

Table 4
RTT values for HARQ analysis.

Scenario	System	RTT [ms]
A	GEO, service/feeder	545
B	GEO, service	271
C	LEO, 600 km	28
C	LEO, 1500 km	52
D	LEO, 600 km	13
D	LEO, 1500 km	24

4.3. HARQ

As pointed out in [41], one of the main challenges introduced by the satellite scenario is the considerable increase of the RTT with respect to that considered for terrestrial links. In Table 4, we report the RTT from Tables 5.3.2.1-1 and 5.3.4.1-1 in [9] for the scenarios provided in Table 1. The increased RTT impacts all closed loop procedures devised in NR. In particular, the large RTTs have the consequence of increasing the latency and memory requirement of the HARQ process; this, in turn, impacts on both the performances and the system requirements. The maximum number of managed parallel HARQ processes can be evaluated as:

$$N_{HARQ} = \mathrm{RTT}(N_{TX} - 1)2^{\mu}, \tag{7}$$

where RTT is expressed in [ms] and the slot length is $2^{-\mu}$ ms. Assuming the the smallest value of SCS numerology $\mu = 0$, this corresponds to up to 1635 HARQ processes for the worst case (GEO scenarios). One important question pointed out in [41] is the opportunity and usefulness of keeping this technology in satellite scenarios. It should be noticed however that a large RTT also reduces the effectiveness of the other link adaptation techniques (ACM and PC), as the Channel State Information (CSI) is available at the transmitter with a latency that is at least RTT/2.

As discussed below, on the one hand the use of HARQ further increases the latency and memory requirements of the satellite system; on the other hand, its effectiveness is particularly important in systems in which the CSIT quality is poor or outdated and other link adaptation techniques become highly suboptimal. In order to assess the performance of both HARQ and ACM vs the RTT we will use the procedure outlined in Algorithm 1.

The algorithm keep track of several counters and assumes a constant number of resources n per slot. The number of used resources N, the total number of transmitted information bits K_T, the number of correctly received bits K_R, the number of transmitted information blocks B_T, and the number of correctly received information blocks B_R. In Algorithm 1 , FB_t represents the feedback information sequence and CSI_t the sequence of channel state information. At each slot (index t) the transmitter check the (outdated) value of FB_{t-D} to decide weather to allocate a new transport block (line 7), or transmit a new RV of a previously transmitted block (line 11). The transport block size K is computed according to the function $f(\cdot)$ which have access to the outdated sequence of CSI ($\mathrm{CSI}_{-\infty}^{t-D/2}$ line 8). The emulation of

Algorithm 1 IR-HARQ process description. D is the RTT delay measured in number of slots ($D = \mathrm{RTT} \times 2^{\mu}$). U is a random variable uniformly distributed in $[0, 1]$.

Require: The number of resources per slot n
Require: The maximum number of transmissions N_{TX}
1: Set counters N, K_T, K_R, B_T, B_R to zero. $\forall t < 0$, set $\mathrm{FB}_t = \mathrm{CSI}_t = \emptyset$
2: **for** $t = 0$ to $t_{\max}$. **do** ▷ t is the slot index
3: **if** FB_{t-D} =ACK **then**
4: Release the HARQ process
5: **end if**
6: **if** $\mathrm{FB}_{t-D} \in \{\emptyset, \mathrm{ACK}\}$ **then**
7: Allocate a new HARQ process.
8: Set $K = \mathrm{f}\left(\mathrm{CSI}_{-\infty}^{t-D/2}\right)$ ▷ Function for MCS selection
9: Compute the first RV (set $k = 1$)
10: **else** ▷ NACK
11: Compute a new RV (set $k := k + 1$)
12: **end if**
13: **if** $k = N_{TX}$ **then**
14: Release the HARQ process.
15: **end if**
16: Transmit the encoded packet over n channel resources
17: Evaluate FER
18: Evaluate CSI_t
19: **if** U<FER. **then** ▷ Packet not received
20: **if** $k = N_{TX}$ **then**
21: $N := N + k \cdot n$, $K_T := K_T + K$, $B_T := B_T + 1$
22: $\mathrm{FB}_t = \emptyset$.
23: **else**
24: FB_t=NACK
25: **end if**
26: **else** ▷ Packet received
27: $N := N + k \cdot n$, $K_T := K_T + K$, $B_T := B_T + 1$
28: $B_R := B_R + 1$, $K_R := K_R + K$
29: **if** $k = N_{TX}$ **then**
30: $\mathrm{FB}_t = \emptyset$.
31: **else**
32: FB_t=ACK
33: **end if**
34: **end if**
35: **end for**
36: Return N, K_T, K_R, B_T, B_R

the transmission of the packet and FER evaluation is then performed (lines 16-17) . We then produce a uniform random variable $U \in [0, 1]$ and compare it to the evaluated FER. If the variable is smaller than FER (line 19) we declare a failure, request a new transmission if possible, or update the counters. Otherwise (line 26) the HARQ process ends successfully and counters are updated accordingly. The Key Performance Indicators (KPI) for the system under investigation can be evaluated from the counters as follows:

BER $= \frac{K_T - K_R}{K_T}$: ratio of not RX bits over TX bits

FER $= \frac{B_T - B_R}{B_T}$: ratio of not RX frames over TX frames

Goodput $T = \frac{K_R}{N}$: ratio of RX bits over employed resources

Throughput efficiency $T_{\mathrm{e}} = K_R / \sum I_t$: ratio of delivered bits over Shannon bits

Average latency $\bar{L} = D\frac{N}{n \cdot B_T}$ slots

For the evaluation of the lines 16 and 17 in Algorithm 1, we considered two complementary approaches. In the first one,

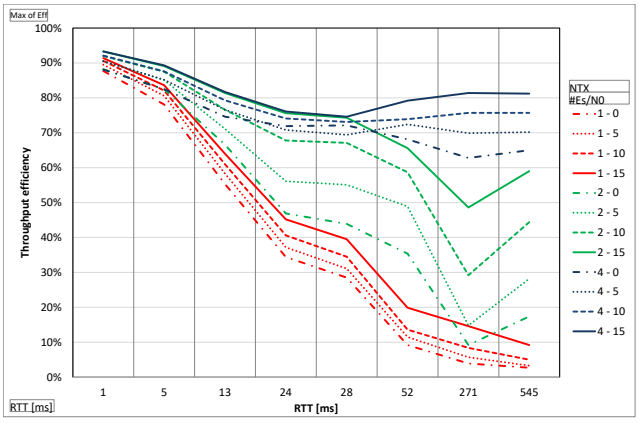

Figure 15: Throughput efficiency vs RTT. NTN-CDL NLOS channel, speed 10 km/h, FER $\leq 10^{-2}$.

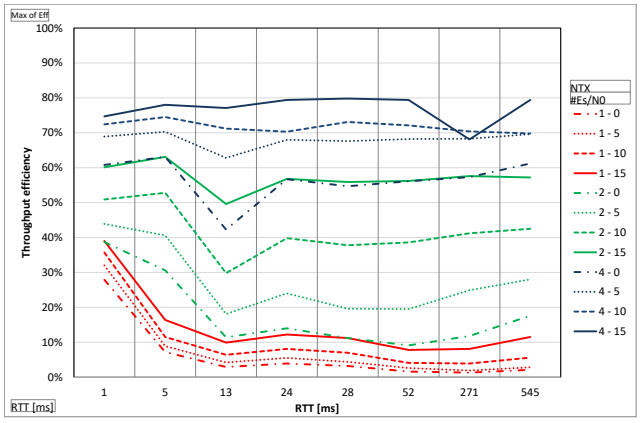

Figure 16: Throughput efficiency vs RTT. NTN-CDL NLOS channel, speed 100 km/h, FER $\leq 10^{-2}$.

we used an approximation of the Frame Error Rate (FER) with information theory expressions for finite block size [67] (see Annex I). This approach allows to rapidly evaluate possible IR-HARQ strategies and to optimise the required procedure parameters, as well as to assess the impact of channel related parameters. In the second one, the FER is evaluated by performing an actual simulation at the symbol rate of the NR PHY. This second much slower approach allows to evaluate additional impairments like suboptimal encoding system (LDPC and finite constellation sets), and decoding system (BICM approach and iterative LDPC decoder).

A crucial feature for the performance of the ACM and HARQ system in the presence of large delay D is the adopted prediction function $\mathrm{f}\left(\mathrm{I}_{-\infty}^{t-D/2}\right)$ for setting the transport block size K in line 8 of Algorithm 1[2] . Optimal linear or non linear ML channel prediction functions can be used and their design depends on the availability of second order channel statistics. In our settings for the results reported below, we simply evaluated the function $\mathrm{f}\left(\mathrm{I}_{-\infty}^{t-D/2}\right)$ in line 11 of Algorithm 1 as follows:

$$\hat{I}_t = \hat{I}_{t-1} + \frac{2}{D}\left(I_{t-D/2} - \hat{I}_{t-1}\right) \tag{8}$$

$$K = \lfloor \hat{I}_t + n \cdot M \rfloor. \tag{9}$$

The simple and suboptimal one-pole filtering in (8), with bandwidth related to the prediction lag $D/2$, is used to improve the prediction capability by removing high frequency components from the process I_t. This filter does not require knowledge of channel second order statistic. The transport block size K is then computed in (9) by adding a suitable margin M to trade the HARQ process latency with the residual FER.

Simulation Results The HARQ procedure described in Algorithm 1, with one of the two proposed approaches, can be used to assess the KPI versus a set of relevant satellite scenarios. Here, we would like to assess the HARQ and ACM performance versus the channel predictability, so that we consider only two reference LOS and NLOS scenarios. For the NLOS scenario, we use the NTN-CDL model, Type A, Suburban NLOS in S-band (2GHz) as defined in Table 6.7.2-6a of [9]. For the LOS scenario, we consider the NTN-CDL model, Type D, Suburban LOS in S-band as defined in Table 6.7.2-5a. In both cases the elevation angle was set to 40°.

The PDSCH was configured with $\mu = 0$ ($\Delta f = 15$ kHz), 52 resource blocks, and 14 symbols per slot, corresponding to a total of $n = 7488$ resources per slot. Channel realizations lasted 10 seconds (10000 slots). The margin M in (9) was optimised for each point of the reported results to maximise the system goodput, while keeping the residual FER below 10^{-2}. In Fig. 19 we provide the throughput efficiency versus a relevant RTT set in the NLOS scenario, for a set of SNR values (0, 5 10, and 15), for $N_{TX} = 1$ (red), 2 (green), and 4 (blue), and a user speed of 1 km/h. Similarly, in Figs. 15-16 we report the results obtained with a user speed of 10 and 100 km/h. Some comments are in order:

- In the NLOS channel, the performance with only ACM (red curves, $N_{TX} = 1$) are unsatisfactory even for moderate speeds of the user. The efficiency rapidly drops below 10% and even below 1% at 100 km/h. The degradation of performance with RTT further increases in Ka-band (20 GHz) where the coherence time is smaller.
- The adoption of HARQ ($N_{TX} = 2$ or 4) provides a significant throughput gain, allowing to reduce the link margin.
- The throughput efficiency degrades more rapidly for low SNR (see the green curves in Fig. 15).
- With ($N_{TX} = 4$), the efficiency reaches a floor that is in the range of 60% to 80% even for an RTT much larger than the channel coherence time.

Fluctuations of the curves are due to the adoption of the non optimal predictor (8), which performs worse for some partic-

[2] As explained in the annex, a suitable CSI proxy that we use here is the channel accumulated information in the slot I_t

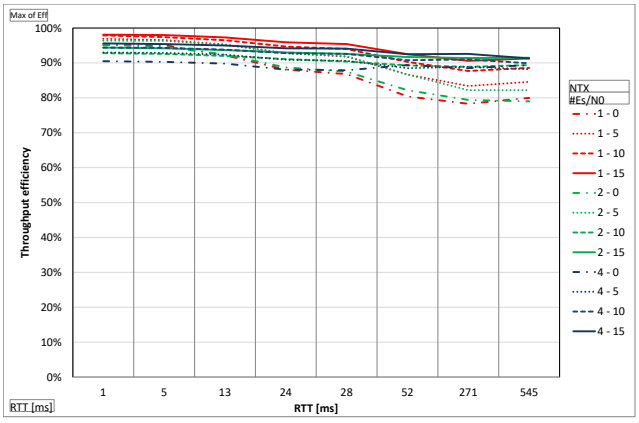

Figure 17: Throughput efficiency vs RTT. NTN-CDL LOS channel, speed 10 km/h, FER $\leq 10^{-2}$.

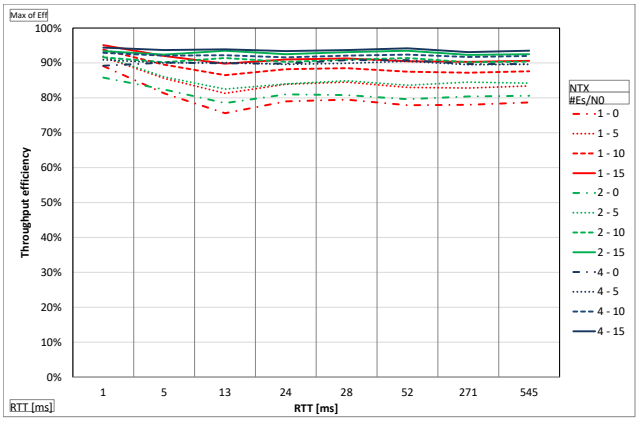

Figure 18: Throughput efficiency vs RTT. NTN-CDL LOS channel, speed 100 km/h, FER $\leq 10^{-2}$.

ular pairs of RTT and channel coherence time. In Figs. 17-18, we report the throughput efficiency versus the RTT for the LOS scenario, for a user speed of 10 and 100 km/h:

- In strong LOS channels, the performance with only ACM (red curves, $N_{TX} = 1$) are in general satisfactory. The efficiency stays in the range above 90% in almost all the cases and it drops to 75% for large RTT and low SNR.
- The adoption of HARQ ($N_{TX} = 2$ or 4) provides some throughput gain, with a maximum at around 15%.

As a general final comment, based on further extensive simulation results, we can observe a strong sensitivity of the ACM system alone to channel selectivity (both in frequency and time). This translates into a very small robustness of the satellite link in some very common scenarios. Despite the increased memory requirements and latency HARQ can actually provide a significant throughput gain and increase link robustness to mobile speed and channel conditions. Dropping the HARQ technology seems to be not advisable if one want to propose the NTN networks as a reliable support and complement to terrestrial cellular networks.

5. Conclusions

In this paper, we provided a thorough and extensive discussion on the introduction of New Radio technologies in Satellite Communication systems, in particular referring to the 3GPP Non-Terrestrial Networks Study Item. A detailed vision of the considered system architectures covering different orbits, payload types, protocol architecture, and radio interfaces has been discussed, highlighting the major benefits and shortcomings of the different solutions. Moreover, an overview of the standardisation process started with Rel. 15 and still on-going towards Rel. 17 has been introduced, together with an analysis of the potential applications of SatCom to 5G. Based on this detailed framework, we then first detailed the procedures involved in Downlink Synchronisation, Random Access, and Hybrid Automatic Repeat reQuest procedures, and then assessed the impact of typical satellite channel impairments on their performance. As for the downlink synchronisation, the main issue is related to Doppler shifts and timing drifts; these can be efficiently coped with by implementing pre-compensation techniques, which, however, require an accurate GNSS capability and knowledge of the satellite ephemeris at the UE. With respect to RA, the impact of the large RTT has been assessed for both the RA preamble format and the procedure timers; similarly to the downlink synchronisation, in case efficient GNSS techniques and the satellite ephemeris are available at the UE, proper pre-compensation techniques can be implemented. In case they are not available, not all of the preamble formats allow the RA preamble detection at the gNB; to overcome this issue, either new formats are introduced (which, however, require modifications to the terrestrial NR standard), or proper system architectures can be defined based on a minimum UE elevation angle (and, thus, maximum differential delay within a beam). As for the RA timers, in particular with transparent payloads, no pre-compensation is possible and, thus, the extension of their maximum values is needed (this is a software modification, which can be easily implemented). Finally, with respect to the HARQ, the main challenge is related to the RTT, since this has a direct impact on the latency and memory requirements in order to implement it; a performance assessment has been provided for both LOS and NLOS scenarios, highlighting that, even though the latency and memory are significantly increased, there is a significant added value in terms of throughput and gain in keeping this procedure also in NTN scenarios.

Annex I

The FER associated to the transmission of an information packet of size K over n independent Gaussian channels can be well approximated as follows, [67]:

$$\text{FER} \approx Q\left(\frac{I-K}{\sqrt{V}}\right), \tag{10}$$

where I is the accumulated mutual information, V the accumulated dispersion over the n channels, and $Q(\cdot)$ the Gaus-

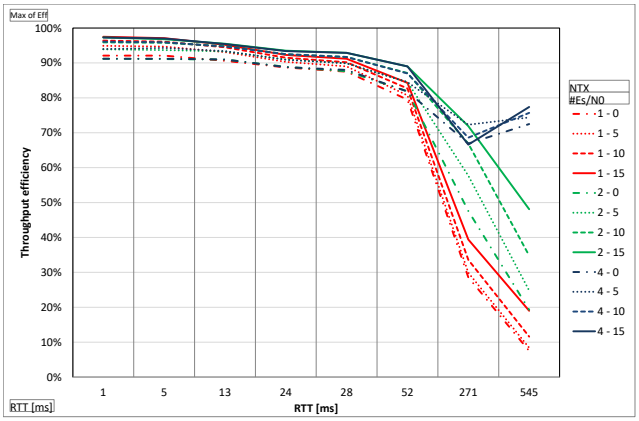

Figure 19: Throughput efficiency vs RTT. NTN-CDL NLOS channel, speed 1 km/h, FER $\leq 10^{-2}$.

sian Q function.

$$I \triangleq \sum_{i=1}^{n} c(x_i), \quad V \triangleq \sum_{i=1}^{n} v(x_i). \tag{11}$$

In (11), x_i is the SNR affecting the i-th resource $x_i \triangleq |\alpha_i|^2/2\sigma_i^2$ and

$$c(x) \triangleq \log_2(1+x) \tag{12}$$

$$v(x) \triangleq \log^2 e \cdot \left(1 - \frac{1}{(1+x)^2}\right). \tag{13}$$

Eq. (10) is obtained by assuming ideal Gaussian input distribution and, consequently, it neglects the additional degradation of using finite size constellations and suboptimal BICM type receivers. Eq. (10) holds also for IR-HARQ systems, [68], where multiple transmission occurs. In this case, we can express the FER at the k transmission as

$$\mathrm{FER}_k \approx Q\left(\frac{I_k - K}{\sqrt{V_k}}\right), \tag{14}$$

where I_k and V_k are the *accumulated* information and dispersions over the resources used so far. The use of (14) for FER evaluation considerably simplifies the evaluation of the core lines 16 and 7 in Algorithm 1. A single simulation of the NR resource allocation and channel model at symbol level allows to generate the sequence of accumulated mutual information and dispersion at slot level

$$I_t = \sum_{i=1}^{n} c(x_{t,i}), \quad V_t = \sum_{i=1}^{n} v(x_{t,i}), \tag{15}$$

where i is the index on resources used on slot t. The core lines 16 and 7 in Algorithm 1 can be subsequently evaluated using (14), where I_k and V_k are the accumulated MI and dispersions over the k slots used for the active HARQ process

$$I_k = \sum_{j=1}^{k} I_{t_j}, \quad V_k = \sum_{j=1}^{k} V_{t_j}. \tag{16}$$

Eq. (14) also suggests that a suitable CSI for the prediction of the system performance is the sequence $\mathrm{CSI}_t = (I_t, V_t)$, rather than other more usual channel quality indicators like average SNR.

References

[1] Cisco White Paper, "Cisco Annual Internet Report (2018-2023)," Mar. 2020. [Online]. Available: https://www.cisco.com/c/en/us/index.html

[2] 5GPPP, "5G empowering vertical industries," Feb. 2016. [Online]. Available: https://5g-ppp.eu/roadmaps/

[3] 5GPPP, "5G innovations for new business opportunities," Mar. 2017. [Online]. Available: https://5g-ppp.eu/roadmaps/

[4] 3GPP TR 38.913 v15.0.0, "Study on scenarios and requirements for next generation access technologies," Jun. 2018.

[5] 3GPP TS 22.261 v17.2.0, "Service requirements for the 5G system," Mar. 2020.

[6] A. Guidotti et al., "Architectures and key technical challenges for 5G systems incorporating satellites ," in *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2624-2639, Mar. 2019.

[7] B. G. Evans, "The role of satellites in 5G," *2014 7th Advanced Satellite Multimedia Systems Conference and the 13th Signal Processing for Space Communications Workshop (ASMS/SPSC)*, pp. 197-202, Livorno, Sep. 2014.

[8] O. Kodheli et al., "Satellite Communications in the New Space Era: A Survey and Future Challenges," submitted to *IEEE COMMUNICATIONS SURVEYS & TUTORIALS*, Feb. 2020.

[9] 3GPP TR 38.811 v15.2.0, "Study on New Radio (NR) to support non-terrestrial networks," Oct. 2019.

[10] SaT5G (Satellite and Terrestrial Network for 5G), EU H2020 5GPPP Phase II project, website: https://www.sat5g-project.eu/

[11] SATis5G (demonstrator for satellite-terrestrial integration in the 5G context), ESA ARTES project, website: https://artes.esa.int/projects/satis5-0

[12] 5G-VINNI (5G Verticals INNovation Infrastructure), EU H2020 5G PPP Phase III project, website: https://www.5g-vinni.eu/

[13] 5GENESIS (5th Generation End-to-end Network, Experimentation, System Integration, and Showcasing), EU H2020 5G PPP Phase III project, website: https://5genesis.eu/

[14] CloudSat (scenarios for integration of satellite components in future networks), ESA ARTES project, website: https://artes.esa.int/projects/cloudsat

[15] VITAL (VIrtualized hybrid satellite-TerrestriAl systems for resilient and fLexible future networks), EU H2020, website: https://cordis.europa.eu/project/id/644843/it

[16] SANSA (Shared Access Terrestrial-Satellite Backhaul Network enabled by Smart Antennas), EU H2020 project, website: http://www.sansa-h2020.eu/

[17] F. Bastia et al., "LTE adaptation for mobile broadband satellite networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2009, art. no. 12, Nov. 2009.

[18] G. Araniti, M. Condoluci, and A. Petrolino, "Efficient resource allocation for multicast transmissions in satellite-LTE network," in *Proc. IEEE Global Commun. Conf.*, pp. 3023-3028, Atlanta, GA, USA, Dec. 2013.

[19] M. Amadeo et al., "A satellite-LTE network with delay-tolerant capabilities: Design and performance evaluation," in *Proc. IEEE Veh. Technol. Conf.*, pp. 1-5, San Francisco, CA, USA, Sep. 2011.

[20] A. Guidotti et al., "Satellite-enabled LTE systems in LEO constellations," in *Proc. IEEE Int. Conf. Commun.*, pp. 876-881 Paris, France, May 2017.

[21] A. Guidotti et al., "LTE-based satellite communications in LEO mega- constellations," *Int. J. Satell. Commun. Netw.*, special issue paper, pp. 1-15, Jun. 2018.

[22] O. Kodheli, A. Guidotti, and A. Vanelli-Coralli, "Integration of satellites in 5G through LEO constellations," in *Proc. IEEE Global Commun. Conf.*, pp. 1-6, Singapore, Dec. 2017.

[23] A. Guidotti, "Beam Size Design for New Radio Satellite Communications Systems," in *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 11379-11383, Nov. 2019.

[24] A. Jayaprakash et al., "Analysis of Candidate Waveforms for Integrated Satellite-Terrestrial 5G Systems," *2019 IEEE 2nd 5G World Forum (5GWF)*, pp. 636-641, Dresden, Germany, Oct. 2019.

[25] C. Ge et al., "QoE-Assured Live Streaming via Satellite Backhaul in 5G Networks," in *IEEE Transactions on Broadcasting*, vol. 65, no. 2, pp. 381-391, Jun. 2019.

[26] G. Araniti et al., "Multimedia Content Delivery for Emerging 5G-Satellite Networks," in *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 10-23, Mar. 2016.

[27] O. Kodheli et al., "An Uplink UE Group-Based Scheduling Technique for 5G mMTC Systems Over LEO Satellite," in *IEEE Access*, vol. 7, pp. 67413-67427, May 2019.

[28] J. Zhang et al., "Energy efficient hybrid satellite terrestrial 5G networks with software defined features," in *Journal of Communications and Networks*, vol. 19, no. 2, pp. 147-161, Apr. 2017.

[29] E. Lagunas et al., "5G Cellular and Fixed Satellite Service Spectrum Coexistence in C-Band," *IEEE Access*, vol. 8, pp. 72078-72094, Apr. 2020.

[30] F. Babich et al., "Nanosatellite-5G Integration in the Millimeter Wave Domain: A Full Top-Down Approach," *IEEE Trans. On Mob. Comp.*, vol. 19, no. 2, pp. 390-404, Feb. 2020.

[31] X. Yan et al., "The Application of Power-Domain Non-Orthogonal Multiple Access in Satellite Communication Networks," *IEEE Access*, vol. 7, pp. 63531-63539, May 2019.

[32] F. Tang et al., "Intelligent Spectrum Assignment Based on Dynamical Cooperation for 5G-Satellite Integrated Networks," *IEEE Trans. On Cog. Commun. and Netw.*, vol. 6, no. 2, pp. 523-533, Jun. 2020.

[33] R. K. Saha, "Spectrum Sharing in Satellite-Mobile Multisystem Using 3D In-Building Small Cells for High Spectral and Energy Efficiencies in 5G and Beyond Era," *IEEE Access*, vol. 7, pp. 43846-43868, Mar. 2019.

[34] L. Zhen et al., "Preamble Design and Detection for 5G Enabled Satellite Random Access," *IEEE Access*, vol. 8, pp. 49873-49884, Mar. 2020.

[35] M. Caus et al., "New Satellite Random Access Preamble Design Based on Pruned DFT-Spread FBMC," *IEEE Trans. On Commun.*, vol. 68, no. 7, pp. 4592-4604, Apr. 2020.

[36] A. Guidotti et al., "Non-Terrestrial Networks: Link Budget Analysis," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, pp. 1-7, Jun. 2020.

[37] H.-L. Maattanen et al., "5G NR Communication over GEO or LEO Satellite Systems: 3GPP RAN Higher Layer Standardization Aspects," in *Proc. IEEE Global Commun. Conf.*, 1-6, Dec. 2019.

[38] 3GPP TS 22.822 v16.0.0, "Study on using Satellite Access in 5G; Stage 1," Jun. 2018.

[39] 3GPP TR 23.737 v17.0.0, "Study on architecture aspects for using satellite access in 5G," Dec. 2019.

[40] 3GPP TR 28.808 0.4.0, "Study on management and orchestration aspects of integrated satellite components in a 5G network," Jan. 2020.

[41] 3GPP TR 38.821 v16.0.0, "Solutions for NR to support non-terrestrial networks (NTN)," Jan. 2020.

[42] 3GPP SA WG2, "Update to FS_5GSAT_ARCH: Study on architecture aspects for using satellite access in 5G," Dec. 2018.

[43] 3GPP TR 38.801 v14.0.0, "Study on new radio access technology: Radio access architecture and interfaces," Apr. 2017.

[44] 3GPP TS 38.410 v16.1.0, "NG-RAN; NG general aspects and principles," Mar. 2020.

[45] ETSI EN 302 307-1, "Digital Video Broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for Broadcasting, Interactive Services, News Gathering and other broadband satellite applications; Part 1: DVB-S2," Jul. 2014.

[46] ETSI EN 302 307-2, "Digital video broadcasting (DVB); Second generation framing structure, channel coding and modulation systems for broad- casting, interactive services, news gathering and other broadband satellite applications; Part 2: DVB-S2 extensions," May 2009.

[47] ETSI EN 301 790, "Digital Video Broadcasting (DVB); Interaction channel for satellite distribution systems," Feb. 2015.

[48] 3GPP TS 38.401 v16.1.0, "NG-RAN; Architecture description," Mar. 2020.

[49] 3GPP TS 38.470 v16.1.0, "NG-RAN; F1 general aspects and principles," Mar. 2020.

[50] 3GPP TS 38.420 v15.2.0, "NG-RAN; Xn general aspects and principles," Jan. 2019.

[51] 3GPP R1-1911858, "Discussion on performance evaluation for NTN," RAN1#99, Nov. 2019.

[52] 3GPP R2-1814877, "Considerations on NTN deployment scenarios," RAN2#103bis, Oct. 2018.

[53] 3GPP R1-1904765, "Considerations on the simulation assumption and methodology for NTN," RAN1#96bis, Apr. 2019.

[54] 3GPP R1-1909693, "Simulation Assumptions for Multi-Satellite Evaluation," RAN1#98, Aug. 2019.

[55] 3GPP R1-1912610, "System simulation results and link budget for NTN," RAN1#99, Nov. 2019.

[56] ITU-R M.2083, IMT Vision - "Framework and overall objectives of the future development of IMT for 2020 and beyond," Sep. 2015.

[57] 3GPP TS 38.331 v15.7.0, "Radio Resource Control (RRC) protocol specification," Sep. 2019.

[58] 3GPP TS 38.804 v14.0.0, "Study on New Radio Access Technology; Radio Interface Protocol Aspects," Mar. 2017.

[59] 3GPP TS 38.211 v16.0.0, "Physical channels and modulation," Dec. 2019

[60] S.Sesia, I.Toufik, M.Baker, *LTE - The UMTS Long Term Evolution*, 2nd ed., Wiley, 2011.

[61] 3GPP TS 38.213 v16.0.0, "Physical layer procedures for control," Dec. 2019

[62] 3GPP R1-1913312, "Summary of 7.2.5.3 on UL timing and PRACH for NTN," RAN1#99, Nov. 2019.

[63] 3GPP R1-1912124, "PRACH design for NTN scenario," RAN1#99, Nov. 2019.

[64] 3GPP TS 38.321 v15.8.0, "Medium Access Control (MAC) protocol specification," Dec. 2019.

[65] R1-1908049, "Discussion on Doppler compensation, timing advance and RACH for NTN," Huawei, HiSilicon, CAICT, 3GPP RAN1 #98, Prague, CZ, August 26th-30th, 2019.

[66] R1-1909479, "Summary of 7.2.5.3 on UL timing and PRACH for NTN," ZTE, 3GPP RAN1 #98, Prague, CZ, August 26th-30th, 2019.

[67] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel Coding Rate in the Finite Blocklength Regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307-2359, May 2010.

[68] B. Makki, T. Scensson, and M. Zorzi, "Finite Block-Length Analysis of the Incremental Redundancy HARQ," *IEEE Wireless Communications Letters*, vol. 3, no. 5, pp. 529-532, Oct. 2014.