Satellite-Based Non-Terrestrial Networks in 5G: Insights and Challenges

ATA SATTARZADEH[1], **YUN LIU**[2], **ABDELRAHIM MOHAMED**[1], **RUILIANG SONG**[2],
PEI XIAO[1], **(Senior Member, IEEE), ZHIQUN SONG**[2], **HAIPENG ZHANG**[2],
RAHIM TAFAZOLLI[1], **(Senior Member, IEEE), AND CHUANFENG NIU**[2]

[1]5GIC and 6GIC, Institute for Communication Systems (ICS), University of Surrey, Guildford GU2 7XH, U.K.
[2]The 54th Research Institute, China Electronics Technology Group Corporation, Shijiazhuang 050000, China

Corresponding author: Yun Liu (yunl_001@163.com)

ABSTRACT Non-terrestrial networks (NTNs) will become an indispensable part of future wireless networks. Integration with terrestrial networks will provide new opportunities for both satellite and terrestrial telecommunication industries and therefore there is a need to harmonize them in a unified technological framework. Among different NTNs, low earth orbit (LEO) satellites have gained increasing attention in recent years and several companies have filed federal communication commission (FCC) proposals to deploy their LEO constellation in space. This is mainly due to several desired features such as large capacity and low latency. In addition, recent successful LEO network deployments such as Starlink have motivated other companies. In the past satellite and terrestrial wireless networks have been evolving separately but now they are joining forces to enhance coverage and connectivity experience in the future wireless networks. The 3rd Generation Partnership Project (3GPP) is one of the dominating standardization bodies that is working on various technical aspects to provide ubiquitous access to the 5G networks with the aid of NTNs. Initial steps have been taken to adopt 5G state of the art technologies and concepts and harmonized them with the conditions met in non-terrestrial networks. In this article, we review some of the important technical considerations in 5G NTNs with emphasis on the radio access network (RAN) part and provide some simulation based results to assess the required modifications and shed light on the design considerations.

INDEX TERMS New radio (NR), non-terrestrial networks (NTNs), 5G.

I. INTRODUCTION

5G networks are evolving to meet the new demands in the telecommunication industry. The increase in demands for better services including high speed, low latency and IoT massive connection, is pushing the communication industry towards a unified network infrastructure in order to provide customized services for various verticals industries. Different use cases have diverse and sometimes conflicting requirements which calls for novel technological solutions. Unlike its predecessors, 5G is an integration of vertical technologies and architectures which is designed to be highly adaptive and can tailor and optimize itself for different services. This is in sharp contrast with previous generations where the service requirements are sacrificed and compromised to fit the network capabilities. Therefore, 5G needs to have a very dynamic network structure to properly manage its resources

The associate editor coordinating the review of this manuscript and approving it for publication was Cesar Vargas-Rosales.

and accommodate heterogeneous wireless connectivities. Despite all the improvements in 5G, coverage is still one of the main barriers that prevents many users around the world from gaining access to wireless networks. According to Global System for Mobile Association (GSMA) reports, roughly 4 billion people still do not have access to the internet to date [1].

Wireless networks coverage is limited to the access points' locations and their footprints. To enhance the coverage the number of access points should be increased and they should be deployed in the locations where the service is requested. Apparently, due to the economic and logistic reasons, the cost for global terrestrial network deployment cannot be justified. This is where the non-terrestrial networks (NTNs) comes in as an effective solution to complement the terrestrial networks and provide coverage and capacity for under-served and isolated areas. Although the coverage extension can be identified as the main reason of interest in NTNs but there are several other benefits to integrate NTNs with terrestrial

networks. The 3rd generation partnership project (3GPP) as the dominating standardization development body, has been working to make satellites viable for coverage extension and service provision [2]. Accordingly, several use cases have been identified for NTNs which in turn promotes their special role in the future network. In addition to providing 5G coverage for isolated, unserved or under-served areas, NTNs will reinforce the 5G service reliability by providing service continuity for machine to machine (M2M) and IoT devices and enhance 5G network scalability by providing efficient multicast/broadcast resources for data delivery towards the network edges or even directly to user terminals. A more comprehensive list of NTN use cases in 5G is available in [3].

Integrating the 5G networks with satellite communications as a unified system is a paradigm with technical challenges. These challenges and potential incompatibilities should be identified, assessed and then proper solutions should be proposed to resolve them. Apart from the 3GPP, several other projects such as Sat5G, VITAL, SATis5 and SANSA have also demonstrated solutions to federate the satellite networks with 5G technologies [4], [5]. These projects cover different aspects from the New Radio (NR) compatibility in satellite communications to the network architecture and methods of implementing a unified hybrid 5G network using technology enablers including software defined radio (SDR) and network function virtualization (NFV).

In this paper we review some of the main challenges and solutions for 5G NTNs. In the next section, we will provide a review on the main components of satellites based non-terrestrial networks including satellites and user equipment and the connectivity structure. In Section III, four fundamental procedures in 5G network are revisited. These include synchronization, paging, random access and handover. We review the impact of NTN on these items and using simulations we give insight into how they are affected and how they should be modified. Finally, Section IV concludes the paper.

II. SATELLITES-BASED NON-TERRESTRIAL NETWORKS

In satellite-based NTNs, satellites play the role of coverage extension for isolated areas and provide edge content delivery services for the terrestrial nodes. Integrated with the 5G terrestrial networks, the service continuity would be obtained and the goals of a unified heterogeneous network can be achieved. In general, satellites are categorized into three different classes, namely Geostationary Earth Orbiting (GEO), Medium Earth Orbiting (MEO) and Low Earth Orbiting (LEO) satellites.

GEO satellites are deployed at an altitude of 35786 Km over the equator and have a circular orbit with zero inclination angle. In an earth-fixed reference frame, their location appears to be fixed in the sky. This is interesting since the transmitting and receiving antennas usually do not need any specific tracking capabilities. MEO satellites are deployed at an altitude between 7000 Km to 25000 km and usually use a circular orbit with different inclination angles. An example

for MEO constellation is O3b network [6]. Leo satellites have the lowest altitude between 300 Km to 1500 km and have circular orbits which can be inclined or polar. The orbit period of the GEO satellites is the same as the orbit period of the earth but in the case of MEO and LEO, it is much shorter and for this reason, they are usually referred to as Non-GEO satellites.

Recently and after a silent period from the 90's, significant attention is being gained by LEO satellites and different companies have shown strong interest in deploying them with large constellations. Since the number of satellites in these constellations is large, they are usually referred to as mega-constellations. The main purpose of these constellations is to provide ubiquitous internet access to the users [7]. In satellite networks with Mega-constellation, hundreds or thousands of LEO satellites are deployed to provide global coverage and user service connectivity. There are several reasons that make LEO satellites appealing. The first is the acceptable delay of data transmission using LEO satellites. One of the main concerns of integrating conventional GEO satellites with terrestrial networks is the large round-trip time which imposes heavy restrictions on the networks and their services. Keeping in mind that 5G aims to minimize the delay in its services, LEO satellites can have round trip times around 25 ms or even less which is an order of magnitude less than traditional GEO satellites. LEO satellites can provide better coverage at high latitudes where GEO satellites would suffer high propagation losses and longer round-trip times. The other interesting property of the LEO satellites is the better spectral efficiency achieved by these networks. In LEO the free space path loss is lower than MEO or GEO. Apart from lower signal energy loss, deploying a large number of satellites with restricted footprints and focused beams on the ground, like a cellular network, can provide better frequency reuse capabilities which enhances the total network average spectral efficiency. Reduction in cost of satellite development and production and improvements in satellite deployment capabilities also contribute to the promotion of using LEO satellites in recent years.

In this paper, our focus is on LEO satellites and we review some design challenges faced when integrating 5G and LEO networks.

A. USER EQUIPMENT

In NTN two different types of user equipment are considered. The first one is handheld devices with Omni-directional antennas and the other one is the Very Small Aperture Terminals (VSAT) terminals with directional antennas [8]. The handheld devices work in lower frequency bands such as S-band and it is envisioned that they would be able to connect to both terrestrial and non-terrestrial 5G networks directly, making it possible to have a ubiquitous 5G network access. Devices supporting dual connectivity will be able to seamlessly handover 5G connections between terrestrial and non-terrestrial networks. Low-end IoT devices also will operate in lower frequency bands and employ

Omni-directional antennas with minimum complexity. These devices do not need to support dual connectivity [10]. On the other hand, VSAT terminals will operate in higher frequency bands such as Ka-band. While bandwidth in handheld devices is usually limited, VSAT terminals can operate with higher bandwidth. They use directional antennas with higher gains and better spectral efficiencies. VSAT terminals that connect to the non-GEO satellites should be able to track LEO or MEO satellites and therefore they are equipped with steerable beam antennas. In [8] typical characteristics for each of these devices are summarized.

B. SATELLITE PAYLOAD

Three different options have been proposed for satellite payloads which are shown schematically in Fig.1. The first one is the transparent payloads where the received signals are forwarded to the ground gateway (GW) without any processing. The received signals are only up-converted or down-converted in frequency and then they are forwarded to the GWs. The gateway will direct the received signal to a gNB (i.e. 5G base station) which is usually deployed nearby. This is a simple architecture that has been adopted in many satellites systems. An example of the modern LEO constellation utilizing transparent payloads is OneWeb [11]. In addition, several conventional broadcasting GEO satellites use transparent payloads. In this architecture, the gNB which is located at the GW, can simultaneously be connected to multiple satellites. It is also possible to have multiple instances of gNBs at a GW, which are coordinated and share the same resources. The ground station will have multiple highly directional tracking antennas deployed at the same site. Here each satellite will act as a mirror which reflects the transmitted signals of the gNBs to cover a larger area. One can consider this as a gNB with a large set of different beams or sectors covering a wide area. The access to these networks are very dependent on the location of the GWs and therefore, their location should be carefully optimized to achieve the required coverage, throughput and delay constraints. Since in transparent payloads the transmitted data is not processed at the satellites, we can not have inter-satellite links (ISLs) to build a space-based network and route data packets directly between satellites.

The second payload type is the regenerative payload which embarks on-board processors to receive and decode exchanged signals. In 5G NTN with direct access to the satellites, the connection link between users and the satellite would use 5G NR, and the link between satellite and gateways which is usually referred to as feeder link is a transport link with arbitrary protocols. Using ISLs, satellites are able to directly route a packet without using the GWs. This will produce a space-based network where in many cases can reduce the packet transmission latency in long-distance communications [12]. Having on-board processing capabilities and using software-defined radios (SDRs) and software-defined networks (SDN) bring flexibility and adaptability to these networks.
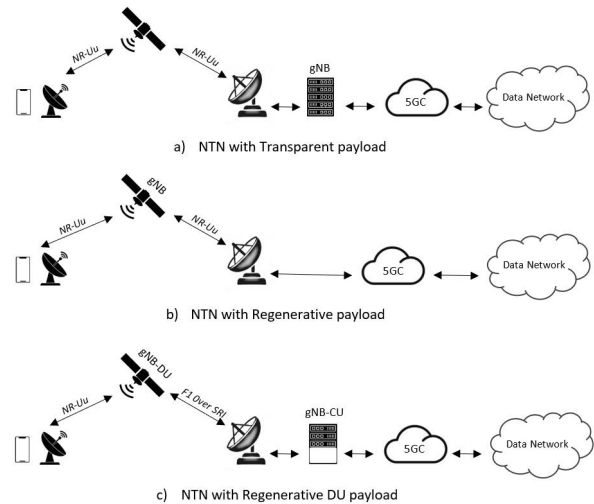

FIGURE 1. Possible network structures in NTN.

The third payload type is a centralized unit-distributed unite (CU-DU) separated structure which has been proposed and used in NG-RAN [13]. The gNB is divided into CU which is deployed at the GW and DU which is deployed on the satellites. Usually, CU will host radio resource control (RRC), service data application protocol (SDAP) and packet data convergence protocol (PDCP) and DU hosts radio link control (RLC), medium access control (MAC), and physical layers (PHY). The F1 link which connects CU to DU is implemented using the satellite radio interface (SRI). Each CU can connect to multiple DU and may perform coordination according to the CU-DU separation options [8].

In any of these three payload types, the link between the user and the satellite can also be relayed. When UE uses a relay to connect to satellites, 5G-NR will be used to communicate with the relay, and the relay can use 5G-NR or other types of interfaces such as DVB-S2X to connect to the satellites and GWs. Here, there is no need to modify the conventional 5G user handsets and it is only the relay node that is responsible for communicating with the satellites.

C. SATELLITE BEAMS

The design of the satellite antenna can have a huge impact on the throughput and procedures involved in implementing 5G NTN including handover and paging. There are two main types of antenna beams that are used for satellites namely the fixed-beam antennas and the steerable-beam antennas. In the first category as the name indicates, antenna beams are fixed relative to the satellite body and therefore they will move on the ground as the satellite moves. The speed of the beams on the ground depends on the satellite altitude and can vary from about 6.9 km/s to 0 for LEO with 600 km altitude and GEO satellites respectively. On the other hand, the steerable beams can keep the beam on a target user as long as the minimum elevation angle for the satellite network is not reached. In phased array antennas, by steering the beam from the bore-sight the beam-width will increase and at the same

time, the gain of the beam will begin to decrease. Modern phase array antennas usually are capable of managing these changes and reducing the beam widening and scanning loss effects due to the beam steering. This type of antenna can also potentially reduce the handover rate in the system. With fixed beam antennas, as the satellites move rapidly, we need to perform intra-satellite inter-beam handovers but with steerable beams, this type of handover can be avoided. It is worth noting that due to the limited number of RF chains, in either of these antenna types only a limited number of beams can be active at each time. The number of active beams in practical satellite networks is usually less than 40 [7]. Therefore, utilizing effective beam management and resource allocation techniques are crucial. The application of advanced signal processing techniques like massive MIMO, can help to improve the efficiency of the spectrum utilization [14]. By increasing the number of beams, the spectrum can be more effectively reused in the network and more users can be simultaneously served by the satellite. An example for LEO satellites with fixed beams is OneWeb which has 16 highly elliptical beams and both Starlink and Telesat use steerable and adjustable beams [15], [16].

III. 5G NON-TERRESTRIAL NETWORKS

In this section, we review some of the main challenges in deploying 5G using non-terrestrial satellite networks. Delay, doppler, and short visiting times are the main characteristic of Non-GEO satellite networks. We will focus on the main procedures in the network and discuss the technical challenges. To meet those challenges, either the standard procedures should be slightly updated and changed or smart solutions should be proposed to successfully accommodate available network structures and procedures.

A. SYNCHRONIZATION

Delay and doppler are the two main channel impairments that should be addressed in NTNs. The first step for users to access a 5G network is to receive the synchronization signals transmitted from the satellites and synchronize with a gNB. The synchronization signal blocks (SSBs) are transmitted in periodic bursts with a periodicity of 5 ms to 160 ms in time. Each burst has 240 sub-carriers in frequency and 4 OFDM symbols length [17]. The long delay and large doppler frequencies make synchronization challenging. The doppler will cause (carrier frequency offset) CFO at the receiver which will degrade the performance of the SSB detection. The amount of CFO depends on the relative speed of the UE and satellite. In terrestrial networks, the maximum doppler for a user with 1000 km/h speed in 3 GHz carrier frequency region would be about 2.8 kHz and this would raise up to 28 kHz for 30 GHz carrier frequencies. However, the doppler frequency in NTN can be as high as 50 kHz up to 720 kHz at 3 GHz and 30 GHz frequencies respectively. This is far beyond what is tolerable in conventional 5G receivers which should be usually less than 5ppm. Therefore, in order to keep the receiver structure unchanged, we need some

mechanisms to reduce the carrier frequency offset (CFO) at the receivers and meet the acceptable CFO levels in 5G. One common approach is to perform CFO or doppler pre-compensation at the transmitter. The pre-compensation is performed according to the doppler of a selected point for each beam (i.e. centre of each beam) and can effectively reduce the maximum CFO within acceptable regions [8]. This is possible because of the known location of the beams and ephemeris of the satellite at the network side. The residual doppler depends on the size of the footprint of each beam and smaller footprints will have smaller residual doppler. For example let us consider a satellite orbiting in 600 km height with minimum elevation angle for visibility equal to 40°. Following the guidelines provided in 3GPP reports [8], for the beam layout which obeys the hexagonal tessellation for the terrestrial networks in UV-plane, the footprint of the antenna can be produced as what is shown in Fig. 2 [8]. The radius of the whole coverage area would be approximately 620 Km. Assuming the radius of the footprint of the nadir beam to be ≈50 km, we need 85 beams to fill the coverage area. In such a pattern, without performing any doppler pre-compensation, the maximum carrier frequency offset produced by doppler and experienced at the receiver is about 17.5 ppm. After pre-compensation for each beam and according to center point of the beams, this value is reduced to approximately 2.1 ppm for any individual beam which would be tolerable at conventional 5G receivers.

Timing Advance (TA) is another mechanism in 5G that will be affected due to large delays. The purpose of TA is to adjust the transmission timing of different users so that their received signals at the gNB are aligned in time. Large propagation delays in satellite communications impose serious restrictions on reusing the same terrestrial network procedures and timing adjustments. The propagation delay in NTN depends on the altitude of the satellite, payload type and location of GWs, and ranges from several milliseconds to hundreds of milliseconds. In contrast, propagation delays in terrestrial networks are usually less than 1ms. This calls for redesigning the timers to meet the new circumstances [18]. In terrestrial 5G networks, TA estimation is performed in the random access procedure where the user wants to establish a connection with the network. The gNB estimates the TA by measuring the reception time of the received random access preamble and feedback the TA to the UE. The maximum TA which can be compensated is dependent on the NR numerology and ranges from 2 ms for 15 kHz subcarrier spacing to 0.15 ms for 240 kHz sub-carrier spacing [2]. These values will restrict the NTN beam size.

In NTN, if UE has access to its own location and satellite ephemeris, then it can calculate the TA and also the doppler frequency with a good resolution. Otherwise, TA should be adjusted using the network indication. Here, the common part of the delay can be calculated by the network and broadcasted to the users in each cell. Therefore, the TA adjustment procedure will then only compensate for the remaining TA difference between users in a cell. Consequently, the coverage
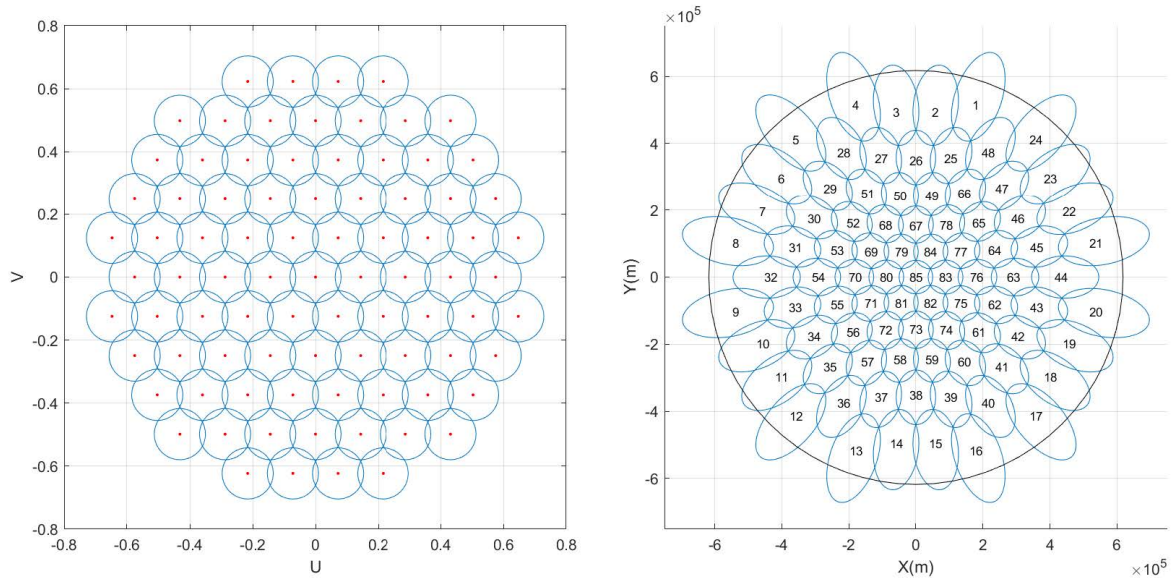
FIGURE 2. Left figure shows that antenna beam layout in UV-plane and the right figure shows the same beam layout on the ground.

region or equivalently beamwidth of antenna patterns should be carefully designed to meet the maximum possible TA adjustment values in the initial access of the users in 5G. Reusing the sample beam layout presented in Fig. 2, the maximum TA difference for the users in the coverage area of each beam is presented in Fig. 3. Accordingly, the results show that the maximum TA difference is experienced in the outer beams where the footprint of the beams expands and in our example, the values are less than 0.7 ms. Therefore, to be able to perform TA adjustments during initial access, only 15 kHz and 30 kHz numerologies can be reused and for higher numerologies, the beam size should be reduced.

Apart from initial TA adjustments, the network attempts to keep the TA value in the correct region by sending the TA command during the UE connection period (RRC-connected). Again the maximum TA adjustment or refinement depends on the Numerology used in the NR and should be kept in acceptable regions [8]. Our simulations for different constellation parameters show that 5G NR will not face restrictions in TA tracking in NTN.

Some designs propose to use different beams for the control-plane and user-plane of the network. A wider beam with large coverage transmits the SSB and broadcasts other necessary information to the users. Then a narrow beam will be directed towards UEs to serve them for data transmission. We should note that in this design, the doppler and TA pre-compensation performance would be limited due to the large footprint of the wide beam. Also, the location of the users should be available at the gNB to perform the beamforming towards the UEs. This design will further impose restrictions on the paging and random access capacity of the network which is discussed next section.

As mentioned before, synchronization begins with SSB block detection at the receiver. In NTN, Each beam of the satellite can either represent a separate cell, transmitting a

unique Physical Cell-ID (PCI) and their own SSB or a group of them can join to represent a cell using one common PCI [8]. In this case each beam will still transmit its own SSB and the situation will resemble a multibeam gNB. Currently, the maximum number of beams that can be supported in each gNB are 4, 8 or 64 in frequency range 1 (FR1), frequency range 2 (FR2) and mm-wave frequency ranges, respectively. So, an antenna capable of generating multiple beams should divide its beams between multiple instances of gNB to meet the 5G designs for multi-beam cases.

B. PAGING AND RANDOM ACCESS PROCEDURES

Paging and random access are two important control plane procedures that should be studied under NTN conditions. Studying their capacity and limitations will also give insight into the supportable user density in NTNs. When there is waiting data for an idle user in the network, the user will be paged. In 5G, a paging message is sent in paging occasions (PO) which occurs in a paging frame. In a non-multibeam scenario at most 4 out of 10 subframes can be used for paging and a paging message can at most include 32 paging records. Each paging record will include the UE identity of the user being paged. Therefore, the maximum number of UE that can be paged in a second is calculated as:

$$\text{Paging capacity} = N_{PF} \times N_{POperPF} \times N_{UEperPO}. \quad (1)$$

where N_{PF} is the number of paging frames, $N_{POperPF}$ is the number of paging occasions in each frame and $N_{UEperPO}$ is the number of paging records in each PO. The maximum for N_{PF} is 100 per second and this implies that the maximum paging capacity would be $100 \times 4 \times 32 = 12800$ per second. This capacity is calculated for a single beam. Usually, the paging message is transmitted in the whole tracking area where a user is camping and this information is obtained according to the
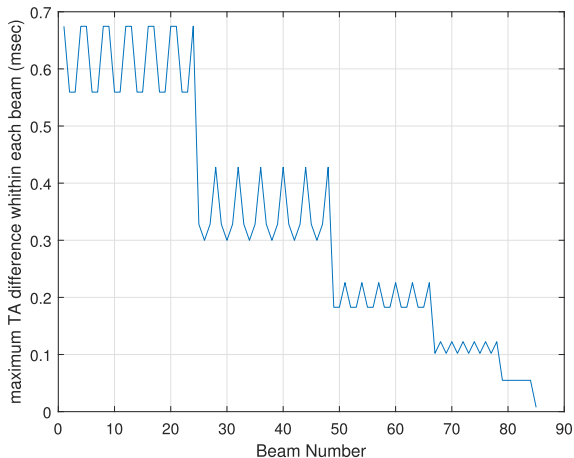
FIGURE 3. Maximum TA difference between users in each beam.

FIGURE 4. Average number of required beams to cover a location area.

location area information available in Access and Mobility Management Function (AMF) in 5G core. Therefore, the real paging capacity of each satellite depends on the number of beams which is required to fill a tracking area region. In a fixed beam scenario, the number of beams required to cover a tracking area is not a fixed value and changes as the location and coverage area of each satellite change. Fig. 4 shows the average number of beams required to cover a tracking area with different radius. To generate this figure, we have used the same settings described in the synchronization section and we have considered a circular tracking area with different radius. Then we have calculated the average number of required beams over time to cover the tracking area. By increasing the nadir beam radius, the average number of beams required for paging will decrease but at the same time, the number of cells or beams of each satellite will also decrease which will reduce the total paging capacity. According to this figure, for a tracking area with an average radius of 50 Km, on average we need 4 beams to perform the paging for each user. Clearly there would be a success probability for each round of paging that will reduce the final paging capacity of the network.

Each user that is paged needs to go through the random access procedure to initiate any connection. In 5G, the physical random access channel (PRACH) is based on a slotted Aloha random access scheme. Therefore, the probability of collision between contending users trying to access the network is calculated as:

$$P_{collision} = 1 - e^{-\frac{\gamma}{M}}. \tag{2}$$

where M is the number of access opportunities per second and γ is the random-access arrival rate per second. To maintain a high quality of service usually $P_{collision}$ is kept as low as 1%. The value of γ is dependent on some of the other parameters in the system. The number of users that have been paged and are trying to access the network and the number of users that try to initiate a connection are the main contributors to γ. In a terrestrial 5G networks M depends on the network parameters including the frequency range, PRACH configuration index, format of preamble, frequency
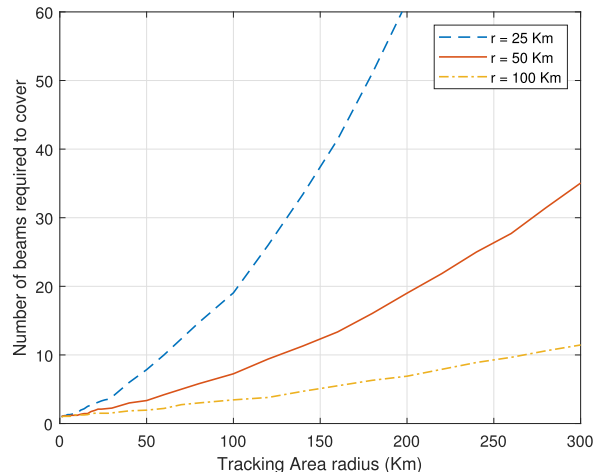
multiplexing for the PRACH occasions, and number of PRACH opportunities per second [8]. The maximum for M is achieved with 56 possible preambles, setting frequency multiplexing equal to 8 and 1000 PRACH opportunities per second which leads to 448000 available PRACH opportunities per second per cell. Using (2) and setting $P_{collision} = 0.01$ we can calculate the maximum random-access arrival rate equal to 4500 requests/sec which is less than the calculated maximum paging capacity of each cell. Therefore, in our example, PRACH capacity is the dominant factor that determines the maximum capacity of each cell.

The gNB needs to wait until it receives the transmitted preambles of all users before processing them. Due to the large delay differences between users in an NTN cell, this may take longer than usual. This is shown in Fig. 5 where the time difference between the reception of the nearest user and further users is twice the delay difference of the users and spans the reception window of the gNB. This imposes a restriction on the number of available PRACH opportunities in the network. Referring to our example and beam layout presented in Fig. 2 and TA difference values presented in Fig. 3, we have 85 beams (cells) in the antenna coverage region where the available paging opportunities in the outer ring which has 24 beams, is at most 500 rather than the maximum of 1000, which is available in the remaining cells. This reduces the maximum random-access arrival rate to 2500 requests/s in these cells which compared to 4500 requests/s, is 44% less.

To give a rough estimate of the maximum paging opportunities provided by each satellite in our example, we can have 61 cells with a maximum paging capacity of 4500 UE/s and 24 cells with the capacity of 2500 UE/s which sum up to 334500 paging/s. Dividing by the coverage area of each satellite, the maximum user density that can be supported by each satellite would be ≈ 0.3 user/km². Also considering the effect of location area size with a radius of 50 km, the supported density will further decrease to ≈ 0.1 user/km².
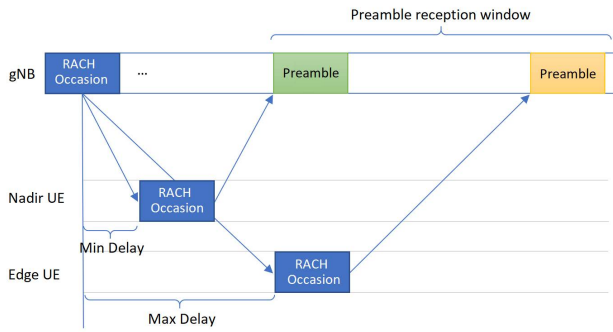
FIGURE 5. PRACH preamble reception timing.

FIGURE 6. Total signaling load for paging and TAU procedures for different scenarios.

C. PAGING AND TRACKING AREA MANAGEMENT

In terrestrial networks, the tracking area is a set of adjacent cells that defines the perception of the core network from the possible location of the UE. When there is a need for paging a UE e.g. network has data waiting to be transmitted to the UE, the paging signal will be sent out to all the cells comprising the tracking area. To keep the network update about the location of each UE, when the tracking area of a user is changed due to the movement of the user, there is a need to update the tracking area. There is a trade-off between the size of the location area and the number of cells that paging signals should be transmitted. Smaller tracking areas will reduce the paging load, while larger tracking areas will reduce the need for frequent Tracking Area Update (TAU) signaling which is more signaling-intensive compared to the paging procedure.

The tracking areas can be defined as fixed or moving with the beams. In non-terrestrial networks with moving cells and moving tracking areas, even fixed users need to update their tracking area information frequently which will lead to extreme signaling loads that is difficult to be handled by the network. A more practical solution is to consider fixed tracking areas on the ground. In this case, the Tracking Area Code (TAC) transmitted by the gNB will be dependent on the location of the satellite. For example in satellites with regenerative payloads, the TAC should be changed according to their coverage region. The UE will listen to the TAC broadcast and in case this code is changed will inform the network for the TAU. As in terrestrial networks, there is a trade-off between the tracking area size and the paging load in the system. Large tracking areas will reduce the TAU request loads but at the same time will reduce the paging capacity of the network, since the paging signal should be transmitted through multiple beams covering the paging area.

In a fixed tracking area scenario, the TAU signaling load is dependent on the velocity distribution of the users. Users with higher speeds will generate more TAU signaling load since they tend to change their tracking area more frequently. On the other hand, the paging load for each region is dependent on the user density and their call-rates. Call-rate refers to the number of times that the network needs to page a UE for a specific purpose e.g. data waiting to be transmitted to the UE or even TAU request triggered by the network.
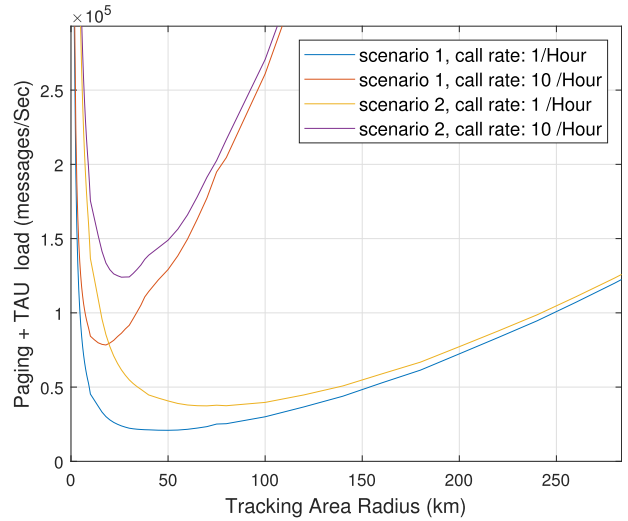
To give an estimate of optimized tracking area sizes in NTN, we have evaluated the TAU and paging signaling loads in a network. In this simulation, the satellites orbits are polar with an inclination angle of 86° and the height is set to 800 Km. There are 18 planes in the network and 36 satellites equally spaced in each plane. The antenna settings are the same as in Fig. 2. The compromise between TAU signaling load and paging signaling load has been shown in Fig. 6.

To generate the results we have assumed that there are 4 different categories of user speeds i.e. fixed, low-speed, medium-speed and high-speed users. For each category, a Normal distribution for the speed of users is considered. The low speed velocities are modeled as $\mathcal{N}(5, 3)$, the medium speed with $\mathcal{N}(70, 30)$ and high-speed with $\mathcal{N}(250, 100)$. The user density has been set to 20 users per km^2. For scenario 1, the users' combination is set to be 50% fixed users, 30% low-speed, 15% medium-speed and 5% high-speed. For scenario 2, these ratios are changed to 30%, 30%, 30%, and 10% respectively. For each scenario the effect of two different call-rates has been shown. The results imply that for each scenario there is an optimal tracking area size that minimizes the total signaling load of the network. With higher call-rates, the choice of the proper tracking area size becomes more important and deviating from the optimal values will significantly affect the total signalling load of the network.

D. HANDOVER CHALLENGES

Providing a reliable and seamless connection for the users in non-terrestrial networks, especially when we are dealing with fast-moving LEO satellites is another challenge in the NTN. The signaling load, service interruption time, and handover success rate should be carefully analyzed. Depending on the payload type, antenna beam type, and how we define a cell in the network, we can have different types of handover. In a network with steerable antennas, mostly the handover is only needed when UE has to switch to another satellite

and this is due to the limited coverage of the serving satellite. This type of handover is usually referred to as inter-satellite handover. When using satellites with fixed beam antennas, entering every new beam, the UE may need to perform a handover. This type of handover is called intra-satellite handover. We can also have another type of handover called inter-network handover which occurs when switching between two different networks where at least one of them is NTN.

Before diving into the handover details, it is worth taking a look at the handover procedure in 5G [19]. The conventional method of handover in 5G is a hard handover where the connection with the source gNB is broken before making the new connection with the Target gNB. In terrestrial networks due to the short round trip times (RTT), this will not impose any significant issue to the network. However, in NTNs the RTT is not negligible, thus the procedure should be assessed carefully. The user will send the report of its received power from different satellites to its serving (source) gNB and the gNB will decide when to initiate the handover for each UE. To perform the handover, the source gNB will first send a request to the target gNB. The target gNB will perform the admission and respond to the source gNB. The source gNB will inform the UE to initiate the handover. As the gNB receives the handover acknowledgment from the target gNB, the downlink data transmission to the UE will be terminated and after the UE receives the handover initiation command from the source gNB it will stop uplink transmission. Then the UE will go through a random access procedure to connect to the target gNB using a 4-step RACH procedure which will take at least 2 RTTs. During the time that the source gNB has stopped the downlink transmission, it will forward the downlink data of the user to the target gNB where the data will be buffered until the connection with the UE has been established. Upon completing the handover procedure, the target gNB will send a request to the core network to update the UPF anchor point.

In a network with transparent satellites, gNBs are implemented at the gateways. Therefore, regardless of whether there is one gNB responsible for all the beams of the satellite or multiple gNBs, they are co-located and handover can be performed quite fast. The data transfer between possible gNB will not occupy any network capacity. In the case where one gNB is controlling multiple beams, the handover would be local and there is no need to send any request to the AMF and core network. Note that in the transparent case, beams that are connected to the gateway can belong to multiple satellites.

In a network with regenerative satellites, gNBs can be implemented on the satellites. Assuming a fixed beam scenario, intra-satellite handover is easy and does not need to go through the gateways but the inter-satellite handover is more complicated. To perform the handover, the source and target gNBs should exchange control data. The control signal exchange will go through the ISL if available, otherwise they should go through a common gateway which will increase the delay. Therefore, the handover usually takes

place between satellites that have ISL in between. For example in a constellation with inclined orbit satellites, the handover often happens between satellites that are all ascending or descending (because usually there is no ISL between ascending or descending satellites). Also note that to transfer UE data from the source to destination gNBs, ISL capacity should be utilized. However, the RTT in regenerative payloads is shorter compared to the transparent payloads.

From the network point of view, the number of handovers in the network that should be handled is important. The handover rate not only depends on the constellation and antenna parameters, but also depends on the connection duration. For example, we do not expect that IoT devices will ever need to perform a handover because their connection duration is usually very short in time. On the other hand, for any long connection, we should perform handovers every few minutes. The handover can be triggered by signal strength measurements. A conventional method is to use a threshold and compare the received signals from different satellites. In NTNs we may use location information to initiate handovers. Other parameters such as angle of the users and TA values or doppler values may also be exploited as indicators for handover time. For example, if the TA value exceeds a predefined value, the handover can be triggered. The location of GWs and the need for feeder link handover also have their own impact on the handover rate.

In situations where satellites need to perform a feeder link handover, some of the connections may need to perform the handover prior to any handover requests triggered by other methods. Note that satellites are usually able to connect to multiple GWs at the same time and therefore, only a portion of users may be forced to handover in the case of satellite feeder link handover. Also, it is worth mentioning that there is no fundamental difference between inter-satellite handover rate for transparent and regenerative payloads and the handover rate will be the same for both.

For the UE with directional antennas, the handover requires more considerations. The user should be able to monitor the received signal from other satellites while being connected to the serving satellite. If the antenna can have only one beam i.e. dish antennas, then usually the UE can not have information about the received signal strength from other visible satellites. As a result, user terminals with array antennas and capable of generating multiple beams are more applicable in these scenarios. Knowing the ephemeris of the satellites in the network will help to shape the beam towards the best directions and listen to other satellites more intelligently.

In Fig. 7 the average handover rate for the users in a NTN is shown. The payload type in this simulation is transparent and the satellites' orbits are polar with an inclination angle of 86° and the height is set to 800 km. There are 18 planes in the constellation and there are 36 satellites equally spaced in each plane. The handover rate is calculated using simulation for different connection periods and for different users speeds. The first observation is that the users' speed can be ignored
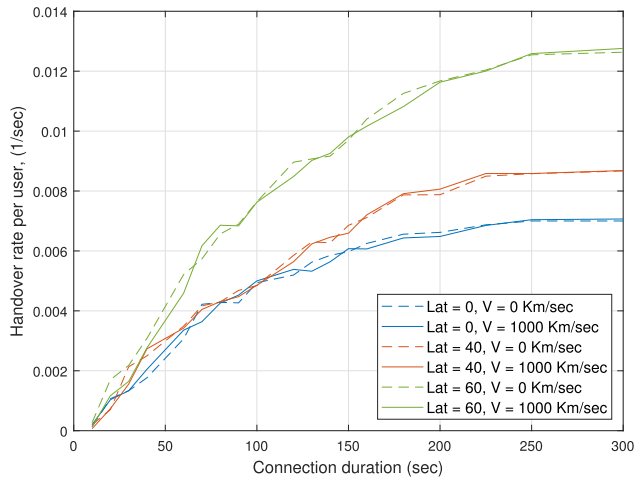
FIGURE 7. Empirical handover rate for users with different speed and locations.

mainly due to the high speed of the satellites and there is no much difference between handover rates for users with 1000 km and 0 km speed. The second observation is that at higher latitudes, the hand over rate is slightly increased mainly due to the fact that satellites get closer to each other in polar orbits and there are more overlaps between footprints of adjacent satellites. The third observation is that as the connection duration goes beyond 300 s, the average handover rate reaches its maximum. It can be observed from this figure that for example for the users located in $+40°$ latitude, the average handover rate for long connection duration is approximately 0.0085 handovers per second per user or equivalently the average time between each handover is about 117.6 seconds.

IV. CONCLUSION

Non-terrestrial networks will be an indispensable part of the future networks and several use cases for NTNs have been envisioned. However, adopting the existing 5G technology to be used in NTNs and especially in satellite-based NTNs is challenging and there are several system-level and link-level issues to be addressed. In this paper, we have reviewed some of these challenges and some limitations have been demonstrated in practical 5G satellite networks through our simulation-based assessments. In particular, we have shown that each beam's coverage area should be deliberately confined to meet the CFO requirements in standard 5G devices. We have shown the impact of NTNs on the paging and random access procedures in 5G and we have analysed how these procedures interact to limit the user capacity of the network as well as how they interact with tracking area sizes to keep the signaling load minimum. Also, handover procedure in NTN is reviewed and practical handover rates in LEO constellations and their dependency on the location of the users are evaluated.

REFERENCES

[1] K. Bahia and S. Suardi. (Jul. 2019). *The State of Mobile Internet Connectivity 2019*. [Online]. Available: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2019/07/GSMA-State-of-Mobile-Internet-Connectivity-Report-2019.pdf

[2] *Study Using Satell. Access 5G*, document TR 22.822, Release 16, 3GPP, Jun. 2018. [Online]. Available: https://www.3gpp.org/

[3] *Study New Radio (NR) to Support Non-Terrestrial Netw. (NTN)*, document TR 38.811, Release 15, 3GPP, Oct. 2019. [Online]. Available: https://www.3gpp.org/

[4] *System Architecture: Final Report*, document H2020-ICT-2014-1, Project VITAL, Deliverable D2.3 Jun. 2016.

[5] H.-I.-2.-1. P. SANSA and D. Deliverable. (Jan. 2016). *Definition of Reference Scenarios, Overall System Architectures, Research Challenges, Requirements and KPIs*. [Online]. Available: http://www.sansa-h2020.eu/deliverables

[6] S. Min, *Engineering Design and Application of Satellite Communication System*. Beijing, China: House of Electronics Industry, 2015.

[7] I. D. Portillo, B. G. Cameron, and E. F. Crawley, "A technical comparison of three low earth orbit satellite constellation systems to provide global broadband," *Acta Astronautica*, vol. 159, pp. 123–135, Jun. 2019.

[8] *Solutions for NR to Support Non-Terrestrial Netw. (NTN)*, document TR 38.821, Release 16, 3GPP, Jan. 2020. [Online]. Available: https://www.3gpp.org/

[9] SATis5. (2018). *Demonstrator for Satellite-Terrestrial Integration in the 5G Context*. Accessed: Feb. 20, 2019. [Online]. Available: https://satis5.eurescom.eu/

[10] A. Guidotti, A. Vanelli-Coralli, M. Conti, S. Andrenacci, S. Chatzinotas, N. Maturo, B. Evans, A. Awoseyila, A. Ugolini, T. Foggi, L. Gaudio, N. Alagha, and S. Cioni, "Architectures and key technical challenges for 5G systems incorporating satellites," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 2624–2639, Mar. 2019.

[11] WorldVu Satellites Limited. *Fixed Satellite Service by WorldVu Satellites Limited LOI-20160428-00041 FCC Filing*. Accessed: 2017. [Online]. Available: https://fcc.report/IBFS/SAT-LOI-20160428-00041

[12] M. Handley, "Using ground relays for low-latency wide-area routing in megaconstellations," in *Proc. 18th ACM Workshop Hot Topics Netw.*, Nov. 2019, pp. 125–132.

[13] *NG-RAN*, document TS 38.401, Release 15, 3GPP, Jul. 2019. [Online]. Available: https://www.3gpp.org/

[14] L. You, K.-X. Li, J. Wang, X. Gao, X.-G. Xia, and B. Ottersten, "Massive MIMO transmission for LEO satellite communications," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1851–1865, Aug. 2020, doi: 10.1109/JSAC.2020.3000803.

[15] Space Exploration Holdings. *SpaceX Ka-Band NGSO Constellation FCC Filing SAT-LOA-20161115-00118*. Accessed: Sep. 12, 2018. [Online]. Available: http://licensing.fcc.gov/myibfs/forwardtopublictabaction.do?file_number=SATLOA2016111500118

[16] Telesat Canada. *Fixed Satellite Service by Telesat Canada SAT-PDR-20161115-00108 FCC Filing*. Accessed: 2016. [Online]. Available: https://fcc.report/IBFS/SAT-PDR-20161115-00108

[17] *NR: Physics Channels Modulation*, document TS 38.211, Release 15, 3GPP, Sep. 2018. [Online]. Available: https://www.3gpp.org/

[18] H.-L. Maattanen, B. Hofstrom, S. Euler, J. Sedin, X. Lin, O. Liberg, G. Masini, and M. Israelsson, "5G NR communication over GEO or LEO satellite systems: 3GPP RAN higher layer standardization aspects," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[19] *NR: NR NG-RAN Overall Description: Stage2*, document TS 38.300, Release 15, 3GPP, Jun. 2019. [Online]. Available: https://www.3gpp.org/

ATA SATTARZADEH received the B.S. degree in electrical engineering from the K. N. T. University of Technology, the M.S. degree in telecommunications from the IUST, and the Ph.D. degree in telecommunications from the University of Tehran, Iran, in 2013. He is currently a Research Fellow with the 5G and 6G Innovation Center (6GIC), Institute for Communication Systems, University of Surrey. His research interests include 5G networks, satellite communications, multiuser MIMO, and signal processing.

YUN LIU was born in Heze, Shandong, China, in 1983. She received the Ph.D. degree in communication and information engineering from the University of Electronic Science and Technology of China (UESTC), in 2013. She is currently a Senior Engineer of communication and networking at the Beijing Research and Development Center, the 54th Research institute of China Electronics Technology Group (CETC). Her research interests include 5G and 6G wireless communication theory and satellite mobile communication, including OFDM technologies, Massive MIMO technology, cognitive radio spectrum sensing technology, resource allocation, and 5G non terrestrial networks technology and applications.

ABDELRAHIM MOHAMED received the B.Sc. degree (Hons.) in electrical and electronics engineering from the University of Khartoum, Sudan, in 2011, and the M.Sc. degree (Hons.) in mobile and satellite communications and the Ph.D. degree in electronics engineering from the University of Surrey, U.K., in 2013 and 2016, respectively. He is currently a Postdoctoral Research Fellow with the 5G Innovation Center (5GIC), U.K. He is also involved in the RRM, MAC, and RAN management work area, and the new physical layer work area with the 5GIC. He worked on multiple projects, including Energy Proportional eNB for LTE-Advanced and Beyond, U.K. DCMS Planning Tool for 5G Network in mm-wave Band Project, EPSRC Stepping Toward the Industrial Sixth Sense, and 5G Non-Terrestrial Networks in parallel to working on the Development of the 5G System Level Simulator. His research interests include radio access network design, system-level simulations and analysis, mobility management, energy efficiency, and cognitive radio. He was a recipient of the Sentinels of Science 2016 Award. He secured the first place and top ranked in the Electrical and Electronic Engineering Department, University of Surrey, U.K., during his M.Sc. studies.

RUILIANG SONG received the B.E. and Ph.D. degrees in microelectronics and solid state electronics from Tianjin University, Tianjin, China, in 2006 and 2009, respectively. From 2009 to 2011, he worked as a Postdoctoral Research Fellow with the Laboratory of CIGS Thin Film Solar Cells, Institute of Optoelectronic Thin Film Devices and Technology, Nankai University, Tianjin. Since 2011, he has been a Senior Engineer with the Beijing Research and Development Center, the 54th Research Institute of China Electronics Technology Group (CETC), Beijing, China. His current research interests include space terahertz communication systems, terahertz devices, and high data rate transmission in 6G application.

PEI XIAO (Senior Member, IEEE) is currently a Professor of wireless communications at the Institute for Communication Systems, home of 5G Innovation Centre (5GIC), University of Surrey. He is the Technical Manager of 5GIC, leading the research team in the new physical layer work area, and coordinating/supervising research activities across all the work areas within 5GIC (www.surrey.ac.uk/5gic/research). Prior to this, he worked at Newcastle University and Queen's University Belfast. He also held positions at Nokia Networks, Finland. He has published extensively in the fields of communication theory, RF and antenna design, signal processing for wireless communications, and is an Inventor on over ten recent 5GIC patents addressing bottleneck problems in 5G systems.

ZHIQUN SONG received the M.S. degree in communication and information system from the Former Research Institute of Communication Measurement and Control Technology under the Ministry of Electronic Industry, in 1985. He is currently the Chief Scientist in the field of communication and transmission of China Electronics Technology Group Company Ltd., and the Chief Engineer of China Electronics Network Communication Group Company Ltd. He has made in-depth research in special communication system, anti-jamming communication, mobile communication, and cognitive radio. His current research interests include 5G private network communication and 6G communication. He is a member of the IMT-2020 (5G) Promotion Group under the Ministry of Industry and Information Technology of China. He is a member of the IMT-2020 (5G) Promotion Group under the Ministry of Industry and Information Technology of China. He has won seven national, provincial, and ministerial science and technology awards, including the National Science and Technology Progress Award. He has published over 20 papers in journals and refereed conference proceedings in those areas and has published a monograph on the cognitive radio technology and applications.

HAIPENG ZHANG received the B.Eng. degree in automation from the University of Science and Technology of China, China, in 2009, and the Ph.D. degree from the School of Information Science and Technology, University of Science and Technology of China, in 2014. He is currently a Senior Engineer of communication and networking at the Beijing Research and Development Center, the 54th Research Institute of China Electronics Technology Group (CETC). His research interests include ambient energy harvesting, the Internet of Things, and satellite networking.

RAHIM TAFAZOLLI (Senior Member, IEEE) is currently a Regius Professor of electronic engineering, a Professor of mobile and satellite communications, and the Founder and the Director of 5GIC, 6GIC, and the Institute for Communication System (ICS), University of Surrey. He has over 30 years of experience in digital communications research and teaching. He was an Advisor at the Mayor of London in regard to the London Infrastructure Investment 2050 Plan. He has given many interviews to international media in the form of television, radio interviews, and articles in international press. He has authored or coauthored more than 1000 research publications. He is regularly invited to deliver keynote talks and distinguished lectures to international conferences and workshops. He is regularly invited by many governments for advice on 5G technologies. He is a fellow of the Royal Academy of Engineering, (F.R.Eng.), FIET, and WWRF.

CHUANFENG NIU was born in Heze, Shandong, China. He received the B.S. degree in electromagnetic and microwave technology from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1998, and the M.S. degree in electronic engineering from Xidian University, Xi'an, China, in 2005. Since 1998, he has been working with the 54th Research Institute, China Electronics Technology Group (CETC), Shijiazhuang of China, where he is currently a Research Professor. His work is focused on the array antenna systems for satellite communication. He published more than 20 contributory papers in refereed journals and international conference proceedings. His research interests include phased array antenna systems and wideband feed.

· · ·