

Received August 26, 2019, accepted September 1, 2019, date of publication September 6, 2019, date of current version September 19, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2939938

5G Evolution: A View on 5G Cellular Technology Beyond 3GPP Release 15

AMITABHA GHOSH¹ (Fellow, IEEE), ANDREAS MAEDER¹, MATTHEW BAKER¹,
AND DEVAKI CHANDRAMOULI¹

Nokia Bell Labs, Naperville, IL 60563, USA

Corresponding author: Amitabha Ghosh (amitava.ghosh@nokia-bell-labs.com)

ABSTRACT The 5G System is being developed and enhanced to provide unparalleled connectivity to connect everyone and everything, everywhere. The first version of the 5G System, based on the Release 15 (“Rel-15”) version of the specifications developed by 3GPP, comprising the 5G Core (5GC) and 5G New Radio (NR) with 5G User Equipment (UE), is currently being deployed commercially throughout the world both at sub-6 GHz and at mmWave frequencies. Concurrently, the second phase of 5G is being standardized by 3GPP in the Release 16 (“Rel-16”) version of the specifications which will be completed by March 2020. While the main focus of Rel-15 was on enhanced mobile broadband services, the focus of Rel-16 is on new features for URLLC (Ultra-Reliable Low Latency Communication) and Industrial IoT, including Time Sensitive Communication (TSC), enhanced Location Services, and support for Non-Public Networks (NPNs). In addition, some crucial new features, such as NR on unlicensed bands (NR-U), Integrated Access & Backhaul (IAB) and NR Vehicle-to-X (V2X), are also being introduced as part of Rel-16, as well as enhancements for massive MIMO, wireless and wireline convergence, the Service Based Architecture (SBA) and Network Slicing. Finally, the number of use cases, types of connectivity and users, and applications running on top of 5G networks, are all expected to increase dramatically, thus motivating additional security features to counter security threats which are expected to increase in number, scale and variety. In this paper, we discuss the Rel-16 features and provide an outlook towards Rel-17 and beyond, covering both new features and enhancements of existing features. 5G Evolution will focus on three main areas: enhancements to features introduced in Rel-15 and Rel-16, features that are needed for operational enhancements, and new features to further expand the applicability of the 5G System to new markets and use cases.

INDEX TERMS 5G new radio, 5G core, 5G system, non-public network, industrial IoT, time sensitive communication, ultra-reliable low-latency communications, integrated access and backhaul, converged edge and core clouds, positioning, NR-unlicensed, non-terrestrial network.

I. INTRODUCTION

5G provides a highly flexible and scalable network technology for connecting everyone and everything, everywhere. It provides a resilient cloud-native core network with end-to-end support for network-slicing. It enables new value creation through support for new services based on three major use case domains, namely enhanced mobile broadband (eMBB), URLLC, and massive machine type communications (mMTC).

The initial commercial deployments of NR are already under way during 2019, focusing on eMBB using the Release 15 (“Rel-15”) version of the 3GPP specifications.

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

The basis for URLLC is inherent in the Rel-15 version of the 5G System, especially in respect of support of low latency.

For the mMTC component, NR is complemented by the machine-type communications technologies known as LTE-M and Narrow Band IoT (NB-IoT) already developed by 3GPP in Rel-13, which provide unrivalled low-power wide-area performance covering a wide range of data rates and deployment scenarios.

Subsequent releases of the 3GPP specifications will build in a backward-compatible manner on the foundation provided by Rel-15, as illustrated in Figure 1. In that spirit, the second phase of 5G is currently being standardized as Rel-16 and is scheduled to be completed by March 2020.

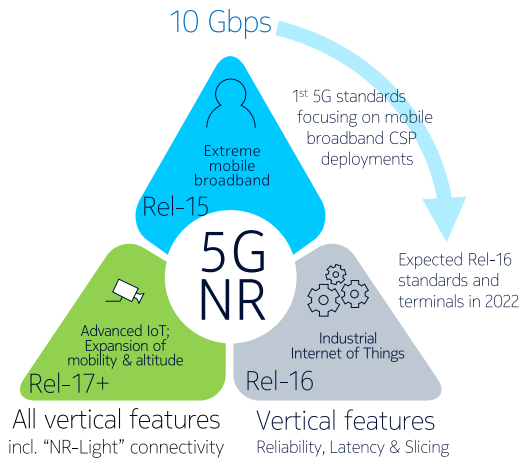


FIGURE 1. Evolution of 5G from Rel-15 to Rel-17.

In addition to enhancing the Rel-15 features, Rel-16 focuses on enabling full support for the Industrial Internet of Things (IIoT) for Industry 4.0, including enhanced URLLC and TSC, introducing support for NPNs, operation in unlicensed spectrum, and deployment enhancements by means of IAB operation mainly geared towards mmWave networks.

With Rel-16 specification work ongoing in 3GPP, planning for the feature content of Rel-17 is already underway, targeting specification availability in mid-2021. From the 5G System Architecture perspective, Rel-17 and beyond is expected to include (but not limited to) enhanced support of IIoT and enhanced support of NPN, enhanced support of wireless and wireline convergence, support for multicast and broadcast architecture, proximity services, enhanced support of multi-access edge computing, and enhanced support of network automation. On the Radio Access Network (RAN) side, in June 2019 the 3GPP community identified the major topics of interest for consideration for Rel-17, including NR-Light which aims to enable lightweight communications for industrial sensors and similar applications, IIoT, MIMO enhancements, sidelink enhancements for both V2X and public safety, support for Non-Terrestrial Networks (NTNs) and coverage enhancement techniques, as well as beginning the work to extend 5G NR to operate in frequencies beyond 52 GHz which is expected to culminate in specifications in Rel-18. The final set of Rel-17 features will be selected in December 2019, and the specification phase is being planned to take 15 months subsequently [4], as shown in Figure 2. It is expected that features which are not in the end included in Rel-17 will be included in later releases. The main enhancements for Releases 16 and 17 are illustrated in Figure 3 in the context of the Nokia Bell Labs Future X architecture [5].

In Section II, we describe the enhancements to and evolution path of the 5G system architecture and security features in Releases 16 and 17, and in Section III we similarly describe the RAN aspects. Finally, conclusions are drawn in Section IV.

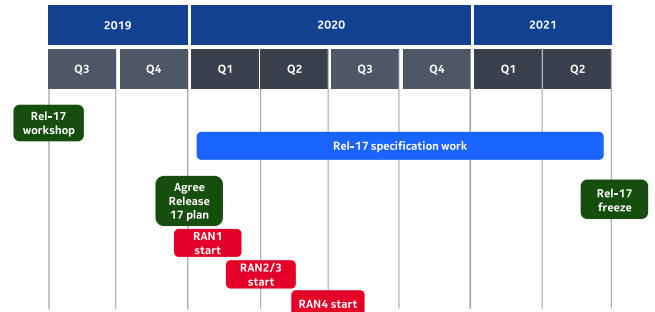


FIGURE 2. Rel-17 timeline.

II. 3GPP REL-16 AND PROPOSED REL-17 5G SYSTEM ARCHITECTURE AND SECURITY FEATURES

The 5G system (5GS) architecture builds on a cloud native foundation by means of the Service Based Architecture (SBA) in the 5G core (5GC), with the objective to provide universal connectivity via all access technologies. 5GC also provides architectural agility by introducing compute-storage split and enablers for 1:N redundancy for control plane resiliency, efficient transaction processing and operational efficiency (enabling reduced CAPEX and OPEX).

As illustrated in Figure 4, Rel-16 focuses on IIoT-related enhancements (URLLC, TSC), NPNs (also referred to as private networks), wireless and wireline convergence and full system resiliency [1]. The Rel-17 architecture will provide:

- further enhancements for analytics-powered networks and enablers for network automation (eNA),
- support for proximity services,
- a multicast and broadcast architecture,
- enhancements to support edge computing,
- enhancements to support the IIoT framework,
- support for NTN and drones (also referred to as unmanned aerial systems - UAS) [1], [3].

A. ANALYTICS-POWERED NETWORKS

Rel-15 and Rel-16 specify the framework for data collection and data analytics in the 5GS by introducing the Network Data Analytics Function (NWDAF) as illustrated in Figure 5. This will pave the way towards utilizing the full potential of artificial intelligence (AI) and machine learning (ML) techniques to realize network automation with as little human interaction as possible. On the one hand, more localized “Cognitive Self-Management” for every radio access point and edge cloud entity is required, i.e. all the communication relationships and operational requirements are learned in-situ. On the other hand, in analytics-powered networks the Network Management moves to higher-level abstractions for clusters of nodes, governed by a continuum of intent- and policy-based management interfaces, thus allowing a tight alignment with service management procedures.

The NWDAF in the 5GC plays a key role as a functional entity which collects KPIs and other information

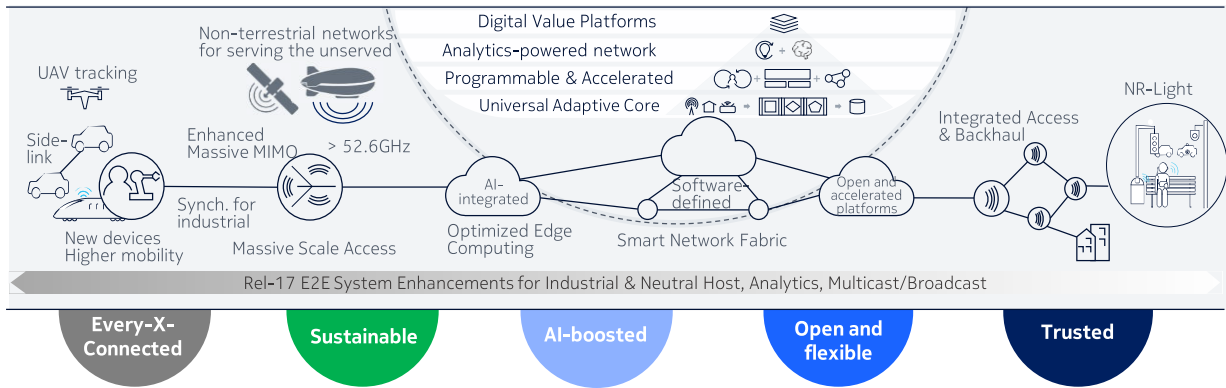


FIGURE 3. E2E Architecture for 5G evolution.

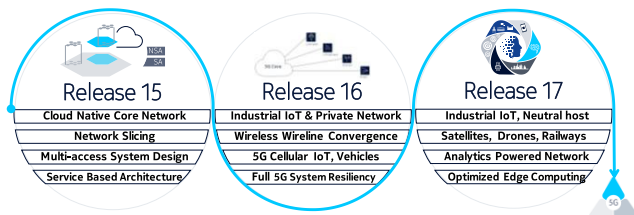


FIGURE 4. 5G Architecture Evolution.

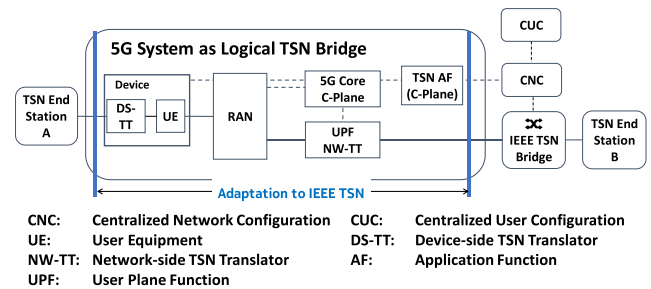


FIGURE 6. Rel-16 architecture support for integration with IEEE TSN.

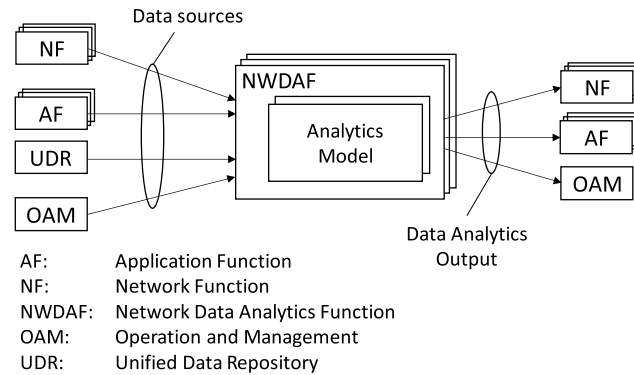


FIGURE 5. NWDAF as network data analytics function in the 5G system.

about different network domains and uses them to provide analytics-based statistics and predictive insights to 5GC network functions, e.g. to the Policy Control Function (PCF). Advanced ML algorithms can utilize the information collected by the NWDAF for tasks such as mobility prediction and optimization, anomaly detection, predictive QoS and data correlation.

The following are some of the objectives for the Rel-17 work [7] on enhanced network automation:

- NWDAF-assisted predictable network performance;
- UE driven analytics;
- NWDAF analytics exposure to applications, for example in Smart City applications such as alleviating urban traffic congestion.

B. ENHANCEMENTS FOR VERTICALS AND INDUSTRIAL IOT

Support for E2E IIoT in the 5GS is based on a set of enabling features for which the foundation is being specified in Rel-16:

- Support for Time Sensitive Communication
- Non-public (i.e. private) networks
- Support for 5G LAN-type services
- Enhanced location services

In Rel-16, **TSC support** is based on integration with the IEEE Time Sensitive Networking (TSN) set of standards developed by the IEEE 802.1 TSN task group [6]. TSC is a communication service that supports deterministic communication and/or isochronous communication with high reliability and availability. This is achieved by hard guarantees for QoS characteristics such as latency bounds, packet loss and reliability, and synchronization down to the nanosecond level [1].

The architectural model considers the 5G System as a bridge for integration of a 5G system within a TSN operating network, as illustrated in Figure 6. A TSN translator function provides interoperation capability on both the device and the network sides, including support for time synchronization, hold-and-forward for de-jittering and link layer discovery and reporting. Furthermore, QoS enhancements were introduced to support efficient scheduling in the RAN and 5G System for deterministic traffic.

It is proposed that the support for IIoT will be further enhanced [8] to support TSC without relying on IEEE TSN

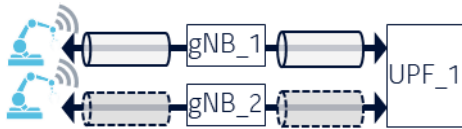


FIGURE 7. Two 5G base stations (gNBs) and one UPF serving the UE(s).

support in the network, thus enabling private WAN (wider area network) type deployments. Use cases for this type of network include logistics, shipping harbors and audio-visual production studios. Feature-wise, this includes optimizations for UE-to-UE TSC communication over a single User Plane Function (UPF) as shown in Figure 7, and exposure of deterministic QoS and synchronization capabilities to external entities. Additionally, 5GS integration with IEEE TSN will be further improved specifically for uplink time synchronization, including support for multiple working clock domains.

Support for NPNs includes enhancements for network deployments which are dedicated to specific use cases like IIoT and are fully isolated to allow only authorized subscribers to camp in the network (hence referred to as “non-public”). Two types of deployments are envisioned with NPN: Stand-alone NPN, and public network integrated NPN. Stand-alone NPN related enhancements include introduction of a Network Identifier (NID) for unique network identification when a shared Public Land Mobile Network Identifier (PLMN ID) is used, enabling network/cell discovery, network (re-)selection using NID, access control and barring to prevent unauthorized users, NPN-specific authentication mechanism (allowing support for UE(s) without USIM), RAN sharing between NPN and PLMN, and access to PLMN services via NPN RAN and vice versa. Public network integrated NPN related enhancements include support for closed access groups (CAGs), access control and cell (re-)selection using CAG.

It is proposed that support for NPN will be further extended in Rel-17 [9] by introducing neutral host models where the network owner and service provider are not necessarily the same. This also includes enablers for accessing stand-alone NPNs using credentials from 3rd party service providers and public network operators. Furthermore, support for emergency services will also be introduced for standalone NPN deployments, as well as support for onboarding and provisioning of non-public networks.

Enhancements to support **5G-LAN services** [10] enable optimized routing and local switching within the 5G System for UE to UE communication (i.e. without routing the traffic to the data network), including functionality for management of 5G-LAN groups which provide service capabilities similar to those of Virtual LANs (VLANs). For example, in the industrial environment, different 5G-LAN groups could be maintained and managed for different device types.

Finally, **location services** for IIoT need to meet stringent requirements on accuracy down to a few centimeters, and on latency both for time-to-first-fix (TTFF) after a location request is initiated and for continuous localization of the

device. This will impact the location and positioning architecture by introducing a RAN-based location management component (LMC) which reduces the latency incurred due to signaling with the Access and Mobility Management Function (AMF), thus being able to process measurement data faster. Necessary enhancements are proposed to be specified as part of the enhanced LCS architecture and 5G System procedures [11] in Rel-17.

C. EXPANSION OF MOBILITY AND ALTITUDE

One of the main drivers for 5G Evolution is to expand the reach of mobile connectivity beyond current boundaries. Several features are proposed to be introduced which will address this objective:

- Enhancements for NTN [12], including both satellites and High Altitude Platform Stations (HAPS) (see Section III for further details);
- Connectivity and control of Unmanned Aerial Vehicles (UAVs) [13] including identification, tracking and authentication.

D. PROXIMITY SERVICES

Support for proximity services (ProSe) is beneficial both for public safety and commercial services. The main objective of the proposed Rel-17 work [14] in this area is to define a common architecture for public safety and commercial ProSe services. It is expected to introduce support for both ProSe discovery (direct discovery) and communication (one-to-one direct communication and one-to-many direct communication). In case of public safety, it is essential to support ProSe discovery and communication when the UE is out of coverage, e.g. in case of disaster relief in remote areas without network coverage.

E. MULTIMEDIA BROADCAST AND MULTICAST SERVICES (MBMS)

Support for MBMS will be introduced for NR mainly for public safety use cases, V2X applications and railways [15]. These use cases require broadcasting/multicasting over an area potentially wider than a single cell. A content distribution mechanism and architecture for content distribution towards multiple base stations (known as “gNBs” in 5G NR) within a Multi-Cell Point-to-Multipoint (MC-PTM) area will be introduced as an enabling feature. In the radio layers, this feature needs to be kept simple in order to ensure widespread adoption in the User Equipment (UE), and therefore it is expected to employ independent broadcasting in each cell, rather than requiring explicit support for wide-area single-frequency-network (SFN) operation.

F. EDGE COMPUTING ENHANCEMENTS

The 5G System in Rel-15 introduces a strong foundation to enable edge computing [16] by means of support for local UPF (re-)selection, concurrent access to local and central data networks, application influence for traffic routing, etc.

Rel-16 provides further enhancements to improve coordination of mobility procedures with the application. Later releases will introduce further enhancements to support edge computing for an extended set of vertical use cases such as URLLC, V2X and Augmented/eXtended Reality (AR/XR). This includes improvements to support for discovery and seamless change of application server, and traffic steering from local to central data networks after processing. In a complementary manner, application enablers will be studied for interactions between UE, application server and network (e.g. for edge discovery at the application layer).

G. SECURITY EVOLUTION

5G networks ensure privacy of their users, confidentiality protection, integrity of the traffic they transport and protection against attacks that can affect availability, integrity of the network and confidentiality of stored data. Rel-15 introduced a broad range of security features aligned with the general 5G architecture evolution principles.

Building upon the Rel-15 security framework, new security features in Rel-16 help multiple industry segments. Support for Non-Public Networks with new authentication schemes will drive 5G adoption in industrial environments. Security features for IoT communications will help massive IoT adoption. With slice-specific authentication in addition to primary authentication, slice tenants gain increased access control and security isolation between slices. Network Slice Selection Assistance Information (NSSAI) for slice access can be protected, if necessary. Security for duplicated transmissions (for URLLC) is expected to provide support for new applications such as medical imaging. Integrity protection in the user plane will prevent packet injection and manipulation of user packets.

The following are some of the key security features specified in Rel-15 and Rel-16:

- Unified authentication framework and access-agnostic authentication.
- Primary and secondary authentication in public and non-public networks (including support for slice-specific authentication).
- Increased home control, e.g. for authentication and steering of roaming.
- Enhanced subscriber privacy
- Enhanced Security for Radio Resource Control (RRC) and Non-Access Stratum (NAS) signaling
- Support for user plane integrity protection (covering all three use case domains, i.e. eMBB, URLLC and massive IoT).
- Secure service-based architecture and inter-PLMN inter-connection
- Security for interworking between the 5GS and the Evolved Packet System (EPS) of 4G.

We expect that in the course of the further evolution of 5G security, new features will be introduced including security support for NR-Light (see Section III), enhanced NPN,

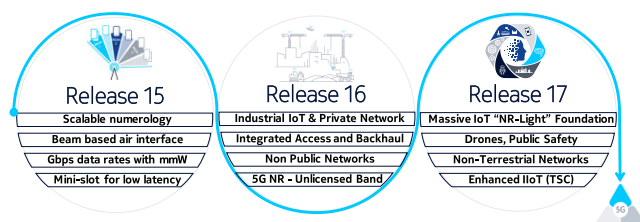


FIGURE 8. 5G NR Evolution.

proximity services (considering commercial services such as V2X and public safety applications), MBMS, NTN and UAVs (e.g. for identification, control and tracking), and new features to increase security of Virtual Network Functions (VNF). Furthermore, it is expected that UEs will be able to verify the legitimacy of base stations before attempting to connect, which should make the deployment of fake base stations more difficult. With security assurance specifications for all 5G nodes, operators and regulators can be assured of the security compliance of the 5G system.

III. 3GPP REL-16 AND PROPOSED REL-17 RAN FEATURES

The key themes of the evolution of 5G NR are summarized in Figure 8, building on the ultra-flexible radio interface design specified in Rel-15. In this section, we review the key NR enhancement features being introduced in Rel-16, and those expected to be introduced in Rel-17 or beyond. Rel-16 commences the focus on IIoT which will continue in Rel-17, as well as introducing IAB, unlicensed band operation, MIMO enhancements for further improvement of eMBB, positioning enhancements, and sidelink operation for direct device-to-device communication. Looking ahead to Rel-17, as also shown in Figure 8, we focus in addition on NR-Light and support for UAVs and NTN. Study of technologies for operation at frequencies above 52.6 GHz is also expected to commence in earnest in the Rel-17 timeframe, with normative specifications following in Rel-18.

A. IIOT ENHANCEMENTS

The foundation for URLLC in NR was laid in Rel-15, with support for ultra-low latency to address mission-critical and time-sensitive communications, mainly in support of IIoT. In Rel-16, URLLC is being enhanced to address new use cases with more stringent reliability requirements (e.g. factory automation requiring reliabilities as good as 10^{-6} packet error rate). Key components of the Rel-16 enhanced URLLC feature include:

- a) Control channel enhancements including new compact downlink control information formats with improved reliability, increased downlink control information monitoring capability to minimize delay and service blocking due to the inability to schedule a user, and support for more than one uplink HARQ-ACK within a slot to reduce latency and provide separate feedback for different services.

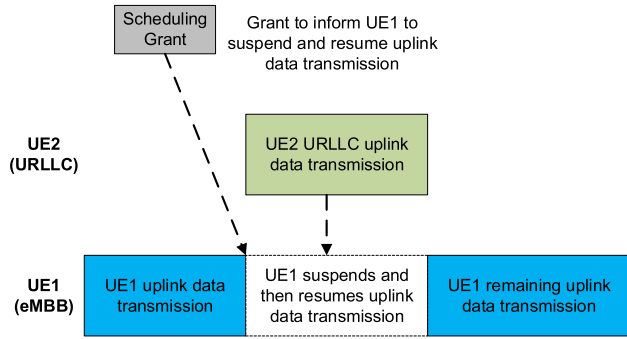


FIGURE 9. Uplink traffic type pre-emption in Rel-16 URLLC.

- b) Scheduling and HARQ enhancements to allow for out-of-order uplink scheduling and HARQ-ACK. This reduces latency by allowing a UE to be scheduled with a second uplink packet while still receiving the first uplink packet.
- c) Support for multiplexing and pre-empting different traffic types in the uplink as illustrated in Figure 9. This allows the gNB to interrupt (i.e. pre-empt) data transmission from one user to accommodate higher-priority data from another user.
- d) Support for multiple active grant-free uplink transmission configurations to accommodate different service flows (e.g. one configuration for a robot arm and one configuration for a pressure sensor).
- e) Support for the ability to schedule two or more uplink data repetitions that can be in one slot or across a slot boundary for improved reliability.

Further enhancements of IIoT in Rel-17 are expected to focus primarily on enhanced support for TSC, as outlined in Section II.

B. INTEGRATED ACCESS AND BACKHAUL

IAB is being introduced in Rel-16 as a key enabler for fast and cost-efficient deployments, mainly targeting dense mmWave deployments outdoors [17]. IAB nodes may be deployed for four fundamental purposes: (i) to remediate isolated coverage gaps, (ii) to provide backhaul where fiber deployment is sparse, (iii) to enhance system capacity and (iv) to bridge coverage from outdoor to indoor. IAB nodes use the same spectrum and air-interface for access and backhaul and create a hierarchical wireless multi-hop network between sites. The hops eventually terminate at a donor node which is connected by means of a conventional fixed backhaul to the core network.

Figure 10 depicts the architecture of IAB with a central IAB donor node connected to the core network, parent links to the next node, child links to the node downstream from the core network, and access links to devices served directly by the IAB node. The IAB architecture leverages the gNB logical split architecture with a centralized Unit (CU) at the IAB donor node and Distributed Units (DUs) at IAB nodes.

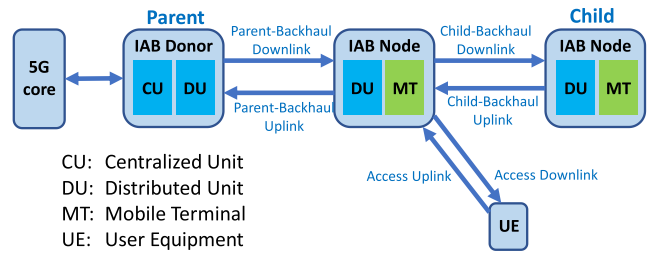


FIGURE 10. IAB architecture.

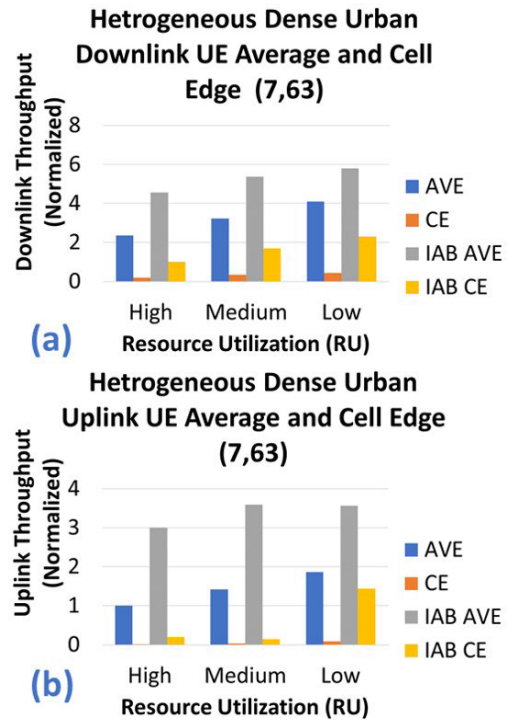


FIGURE 11. IAB relative performance.

An IAB node contains a Mobile Terminal (MT) part which behaves as a UE towards the parent node (either a donor node or another IAB node); this means it monitors the Synchronization Signal Blocks (SSBs) transmitted by the parent node for beam discovery, it monitors Downlink Control Information (DCI) for scheduling, and it performs random access towards the parent node. On the child links, the DU part of an IAB node behaves as a gNB, transmitting SSBs, scheduling resources using DCI messages and responding to random access attempts from the next-hop IAB node. Similarly, on the access links the IAB node behaves exactly as a normal gNB, providing the NR radio interface for UEs in its coverage area.

Figure 11 shows the relative downlink and uplink IAB performance in a heterogenous dense urban scenario—having 7 fiber-connected macro donors and 63 wirelessly connected micro IAB nodes under the conditions of high, medium and low loads.

The system simulated in Figure 11 uses a 30 GHz carrier frequency with 400 MHz aggregated system bandwidth

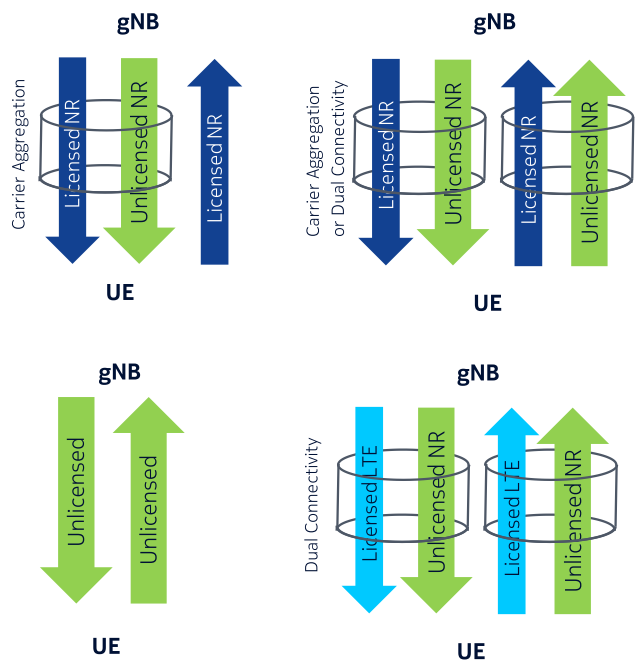


FIGURE 12. Main deployment scenarios for NR-U.

(access+backhaul) and time-division duplexing (TDD). The macro node inter-site distance is 200 meters and the micro IAB nodes are dropped randomly. The transmit powers of macro and micro nodes are 40 dBm and 33dBm respectively, while the UE power is 23dBm. Clear gains are shown when IAB nodes are added to the network, improving both the UE average throughput and cell edge throughput. Larger gains are seen at lower loading, demonstrating that IAB is more effective at coverage enhancement than capacity enhancement.

Enhancements of IAB in a subsequent release may include multi-connectivity for improved network resilience, and Space-Division Multiplexing (SDM) and/or Frequency-Division Multiplexing (FDM) of the different links in order to improve deployment flexibility.

C. NR FOR UNLICENSED SPECTRUM (NR-U)

Access to unlicensed spectrum provides an important tool to increase capacity for both service providers and private networks. For service providers, NR-U enables access to additional spectrum to improve the cellular network operation by offloading traffic in hot-spots. NR-U also enables the operation of standalone networks in unlicensed spectrum without any access to licensed spectrum.

NR-U is being introduced in Rel-16, with the focus on eMBB services in the 5 GHz and 6 GHz frequency bands [18]. Rel-16 NR-U supports a variety of deployment scenarios, as illustrated in Figure 12. It uses the same flexible frame and slot structure and fundamental physical layer design and protocol stack as NR Rel-15, hence limiting the magnitude of changes to the UEs compared to licensed band operation. NR-U adds channel access procedures to enable fair

coexistence with other systems such as IEEE 802.11 variants or LTE Licensed-Assisted Access (LAA).

In Rel-17, NR-U enhancements are expected to expand NR-U use cases beyond eMBB services. For example, improvements to the communication latency and reliability may be introduced to better facilitate industrial use cases.

In due course, there is also an interest to support NR in the unlicensed spectrum in the 57-71 GHz frequency range, where up to 14 GHz of spectrum is globally available. The narrower beams and higher attenuation at these frequencies will provide higher spatial isolation than is seen in lower frequency bands, which will change the coexistence mechanisms required. However, in common with possible future licensed band operation above the current upper operating frequency of NR (namely 52.6 GHz), the power amplifier characteristics need to be carefully considered and a waveform needs to be designed to allow efficient operation at these frequencies. The necessary analysis for this is expected to be carried out during the Rel-17 timeframe, leading to specifications for a common waveform for operation in both licensed and unlicensed spectrum in Rel-18.

D. MIMO ENHANCEMENTS

The 5G NR air interface provides extensive support for large scale antenna arrays often referred to as Massive MIMO [19]. The main reasons for deploying Massive MIMO are coverage and capacity enhancement, which are critical given the expected year-on-year increases in demand for mobile broadband data services [22]. The NR MIMO framework is scalable and flexible and supports arbitrarily large antenna arrays with arbitrary antenna configurations for both FDD and TDD deployments [29]. The NR MIMO framework also supports arbitrary array architectures such as fully digital, hybrid, or analog array architectures (see [23]) for carrier frequencies up to 52.6 GHz.

The air interface of NR is “beam-based”, which means that all the channels in NR can be beamformed for range extension and coverage enhancement. To support beam management, Channel State Information (CSI) can be acquired by means of codebook-style feedback as well as through leveraging TDD channel reciprocity. Two kinds of codebook-based CSI feedback are provided, one of which, known as Type II CSI feedback, is particularly aimed at low-rank multi-user MIMO (MU-MIMO). Type II CSI feedback provides higher precision angular-domain feedback than Type I, and recent studies (e.g. [20]–[22]) have shown that the gains from the Type II CSI over the best LTE codebooks are in the order of 25-30% in full buffer traffic depending on the scenario and array configuration.

In Rel-16, the NR MIMO framework is being enhanced in a number of ways. One drawback of Type II CSI feedback is its high overhead, and Rel-16 is therefore introducing a new frequency-domain compression technique, but also extending its applicability to higher rank transmissions (up to rank 4). Transmission and reception at multiple Transmission and Reception Points (TRPs) is also being introduced, enabling

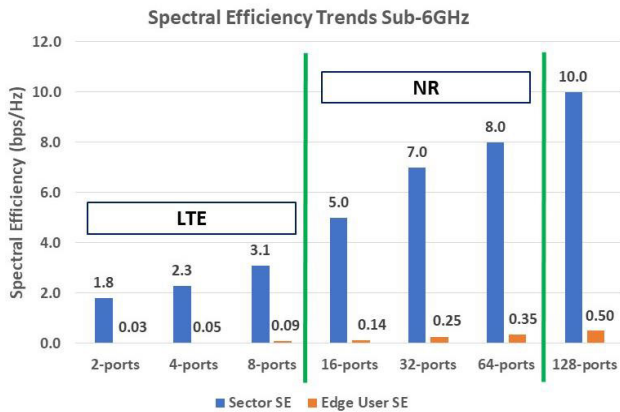


FIGURE 13. MIMO SE trends at 2 GHz [22].

a kind of coordinated multipoint (CoMP) operation based mainly on non-coherent joint transmission; this is especially relevant for ensuring reliability for URLLC services. Support for multi-beam operation is being enhanced, which is particularly useful for mmWave frequencies. Finally, enhancements are being introduced to support full power uplink transmission in the case of UE implementations with multiple power amplifiers.

For Rel-17, the CSI feedback overhead is expected to be further reduced, for example by extending the new frequency-domain compression mechanism of Rel-16 into the layer and time domains, as well as investigating how reciprocity may be exploited to a greater extent even in FDD operation. The beam management framework is also expected to be improved for overhead and latency reduction. Multi-TRP operation is expected to be enhanced, for example in the area of CSI acquisition, especially for FR2, where blockage effects may be problematic. Rel-17 is also likely to address UL-MIMO operation with additional support for non-codebook-based operation and enhancements to codebook-based operation.

The NR air interface in Rel-15 started out from the beginning with support for a maximum of 32 ports in a CSI-RS resource, but with a scalable and flexible framework for supporting arbitrary arrays operating both below 6 GHz (typically with fully digital architectures) and above 6 GHz all the way up to 52.6 GHz (typically with hybrid or analog architectures). Figure 13 shows the progression in spectral efficiency (SE) from 2 ports to 128 ports comparing LTE with NR on a 2 GHz carrier (see [22] for additional details).

As the operating frequency increases, the path loss characteristics become steadily worse, but the physical size of the antennas becomes smaller, thereby enabling more antennas to fit in the same physical space. As a result, arrays with larger numbers of antennas (e.g. 512 antenna elements at 28 GHz) can be used to provide higher beamforming gains to overcome the poor path loss conditions. In line-of-sight propagation, with a constant antenna aperture (i.e. overall array size), there is therefore no degradation in range with carrier frequency; it should be noted, however, that penetration loss (e.g. from outdoor to indoor) does increase and

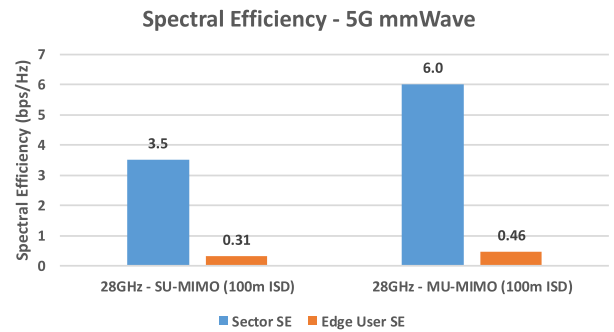


FIGURE 14. Spectral Efficiency at mmWave frequencies.

is dependent on the materials. Figure 14 shows the SE at mmWave frequencies.

E. NR POSITIONING

Precise and up-to-date location knowledge is an essential requirement for emergency calls as well as new services like IIoT.

The Rel-15 version of NR already provides positioning protocol support for RAT-independent positioning techniques, as well as the possibility of measurement gaps on the NR carrier to enable the UE to perform time difference of arrival (TDOA) measurements on an LTE carrier. Like the later releases of LTE, the NR positioning protocol also provides support for Real-Time Kinematic (RTK) positioning using the carrier phase of positioning satellite systems, which can give centimeter-level positioning accuracy in some outdoor scenarios.

In Rel-16, native support for positioning on the NR carrier is being introduced, which will be applicable both for low frequencies (FR1), defined as 410-7125MHz) and for mmWave frequencies (FR2, defined as 24-52.6 GHz). Some methods which are available in LTE are again being introduced in NR, namely downlink-based Observed Time Difference of Arrival (OTDOA), Uplink Time Difference of Arrival (UTDOA), and Enhanced Cell ID (E-CID). Some new positioning methods are also being specified for NR: Multicell Round Trip Time (Multi-RTT), Uplink Angle of Arrival (UL-AoA) and Downlink Angle of Departure (DL-AoD).

As part of this work, a positioning reference signal (PRS) is being added to NR, and some enhancements of the uplink sounding reference signal (SRS) are being made. Various measurements are being specified to support these RAT-dependent solutions. Figure 15 shows how Multi-RTT positioning is expected to operate: multiple gNBs make measurements of the time difference between transmitting a PRS and receiving an SRS from a UE, and the UE measures the time difference between receiving the PRS and transmitting the SRS. The gNBs and UE send their measurements to the location server, for which purpose the LTE Positioning Protocol (LPP) is reused for the UE measurements and a new NR Positioning Protocol “a” (NRPPa), based on LPPa, is used for the gNB measurements. This enables the location

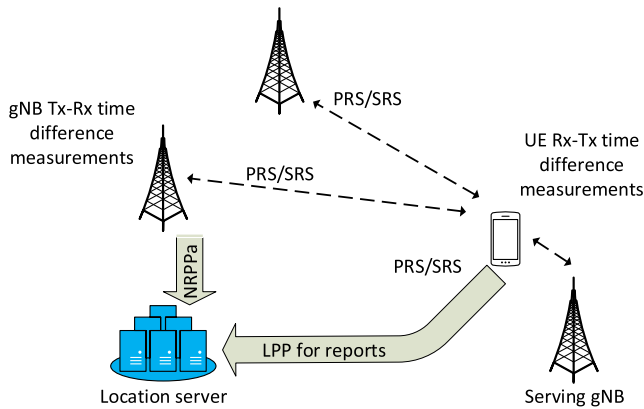


FIGURE 15. Example of NR positioning using Multi-RTT.

TABLE 1. Expected NR location accuracies.

Positioning methods, frequency range, bandwidth	50% location accuracy
OTDOA, FR1, 5 MHz	3.3m
OTDOA, FR1, 100 MHz	0.5m
OTDOA, FR2, 400 MHz	0.3m

server to determine the RTT to multiple gNBs, from which multi-lateration can be used to locate the UE. Multi-RTT may be used in conjunction with OTDOA and/or UTDOA, to overcome the impact of synchronization uncertainties between the gNBs.

On the architecture side, Rel-16 is also working on the provision of a local LMC in the RAN, which can reduce the latency of positioning in local network scenarios such as IIoT.

Whether further positioning enhancements are needed in Rel-17 will depend on the eventual analysis of the positioning accuracy achievable with the Rel-16 techniques. One area which may be considered is the inclusion of relative positioning between UEs using direct sidelink transmissions.

Table 1 summarizes the expected 50% location accuracies for both FR1 and FR2 based on simulation.

F. SIDELINK COMMUNICATION

The Rel-16 version of NR is introducing a sidelink feature to enable direct communication between terminals [30].

Unlike LTE, the NR sidelink is being specifically designed first for the requirements of Cellular V2X (C-V2X), including vehicle-to-vehicle, vehicle-to-pedestrian and vehicle-to-roadside unit (RSU) communication, as illustrated in Figure 16.

While C-V2X in LTE is primarily designed for broadcasting of basic safety messages, NR will also support more advanced use cases with lower latency, larger payloads and higher data rates, as well as both groupcast and unicast modes. These capabilities will provide the foundations for operations such as platooning and remote driving.

It should be noted that wherever network coverage is available, V2X can be supported via the base stations using the

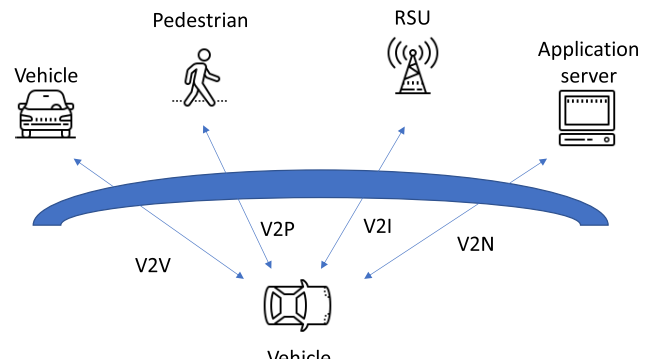


FIGURE 16. Vehicle-to-Everything.

URLLC functionality of NR together with edge computing to deliver low latencies. If the sidelink is being used, the physical layer resources can either be scheduled by the gNB (known as mode 1) or autonomously selected by the UE (mode 2) in a contention-based manner.

Three physical channels are specified for NR V2X operation, namely i) physical sidelink control channel (PSCCH) containing sidelink control information (SCI), ii) physical sidelink shared channel (PSSCH) and iii) physical sidelink feedback channel (PSFCH). The PSCCH and PSSCH are time-division-multiplexed to enable low latency and energy efficiency.

Interoperability between NR V2X and LTE V2X is supported by means of gNB scheduling of LTE V2X sidelink resources; NR side-link scheduling by LTE Uu is also supported. The Rel-16 sidelink will also support public safety applications, where emergency workers can benefit from direct device-to-device communication. Further enhancements under discussion for Rel-17 include extensions to multi-carrier operation to support higher data rates, enhancements for mmWave operation, and more advanced CSI feedback.

G. NR-LIGHT

A version of NR known as NR-Light is planned to be introduced in Rel-17, aiming to address use cases that cannot be met by NR eMBB, URLLC or mMTC.

The feature is intended to provide low-complexity solutions to address IIoT and other verticals in both FR1 and, as one of the distinguishing features compared to NB-IoT and LTE-M, also in mmWave frequencies (FR2). NR-Light is not intended to replace the mMTC low-power wide-area use cases currently supported via LTE-M/NB-IoT but will target the following requirements: i) higher data rate and reliability, and lower latency, than LTE-M & NB-IoT; ii) lower cost/complexity and longer battery life than eMBB and iii) wider coverage than eMBB.

Figure 17 provides an illustrative comparison of NR-Light to other 5G features. Specifically, NR-Light will address the following objectives and use cases [28]:

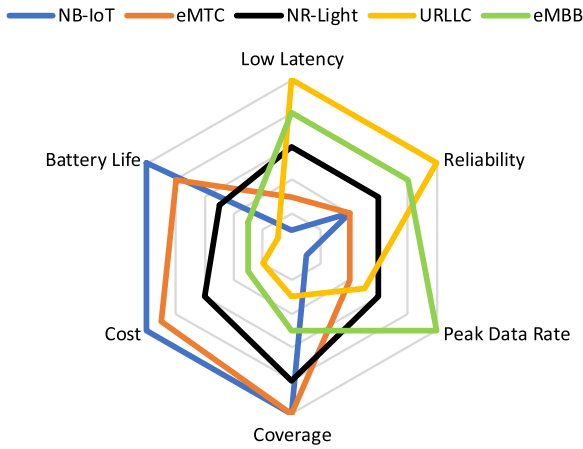


FIGURE 17. Comparison between NR-Light and existing 5G features.

- a) Moderate data rates up to 100 Mbps to support, for example, live video feed, visual production control and process automation.
- b) Moderate latency of around 10-30 ms to support, for example, remote drone operation, cooperative farm machinery, time-critical sensing and feedback, and remote vehicle operation.
- c) Low-complexity devices with module cost comparable to LTE.
- d) Coverage enhancement compared to eMBB.
- e) Low power consumption with longer battery life than eMBB.

The feature will specify new low-complexity UEs that will support NR’s existing Bandwidth Part (BWP) definition within a carrier in order to minimize divergence from legacy NR UEs.

H. NON-TERRESTRIAL NETWORKS (NTN)

NTNs are the key to supplementing the coverage of terrestrial networks and extending service provision to remote areas of the Earth, for example to support global IoT and asset-tracking. The most interesting deployment modes for NTN are expected to use Low Earth Orbit (LEO) satellites and HAPS such as balloons and airplanes. Geostationary Earth Orbit (GEO) satellites have the advantage that full global coverage can be achieved with only 3 satellites, but their RTT is long and they are expensive to deploy. The typical beam footprints and delay budgets for GEO and LEO satellites and HAPS are shown in Table 2.

Figure 18 shows the expected NTN architecture with base stations mounted in satellites or HAPS. It comprises of a service link (a.k.a. access link) and a feeder link (a.k.a. backhaul link). The feeder link terminates in the ground gateway which is connected to the 5GC. There may also be HAPS-to-HAPS links or satellite-to-satellite links to provide resilience in case of failure in one feeder link.

NR NTN may exploit the fact that the 5G RAN defines an interface between the CU of the base station and the lower

TABLE 2. Beam footprint and delay for various NTN cases.

	GEO	LEO	HAPS
	36,000 km	500-1500 km	18-25 km
Beam footprint diameter (km)	200-1000	100-500	5-200
Round trip delay (ms)	270	25	3

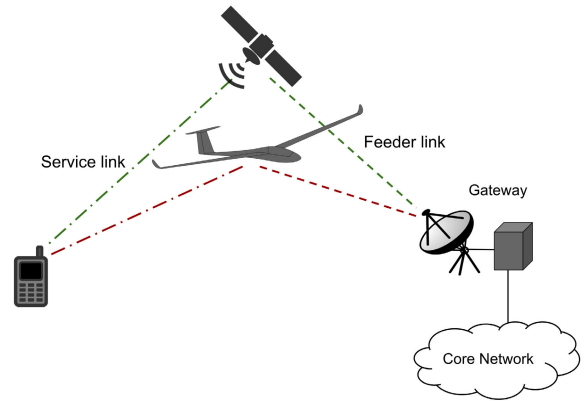


FIGURE 18. NTN architecture.

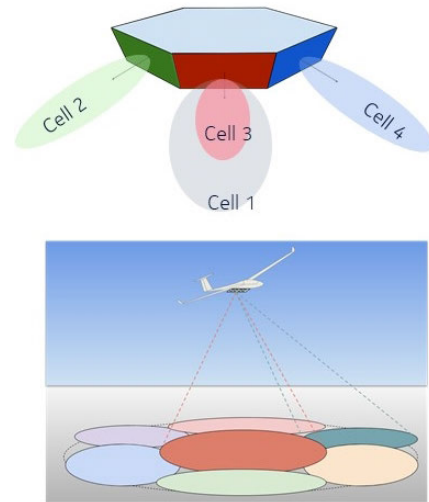


FIGURE 19. HAPS coverage with 7 sectors.

layer functions in DUs. The lighter and less-complex DU may be mounted on the satellite or HAPS, while the CU may be in the ground station; in this case, the feeder link carries the CU-DU interface.

Figure 19 shows an example of a service link configuration with antenna arrays under the HAPS forming sector beams to cover 7 cells in a service area of 100 km radius. The antenna array consists of a downward facing 2 × 2 panel serving the center cell and six 4 × 2 side panels each serving an outer cell. A side panel covers an outer cell in the azimuth domain ±φ° from its boresight, where φ is typically up to 30°. Achievable

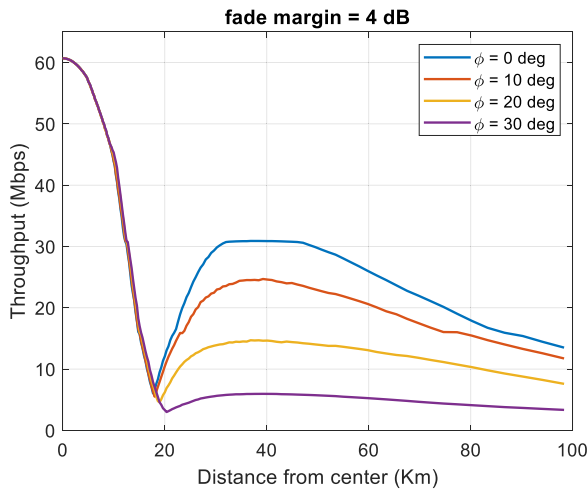


FIGURE 20. Downlink throughput vs. distance from the center.

throughput for an outer cell therefore depends on the azimuth angle with respect to the side panel’s boresight.

Figure 20 illustrates the simulated downlink throughput at various azimuth angles assuming 20 W transmission power per panel and 4 dB fading margin. It is observed that a cell edge throughput of 5 Mbps can be achieved with this configuration.

A study of NTN design for NR is ongoing in the Rel-16 timeframe [27], and this is expected to lead to specifications in Rel-17.

I. 5G BEYOND 52.6 GHZ

The NR Rel-15 physical layer channels were designed and optimized for frequencies up to 52.6 GHz. At higher frequencies, very large spectrum allocations in both licensed and unlicensed bands are expected to become available, which will enable extremely high capacity and data rates [26]. 3GPP has already nearly completed a study of use cases, deployment scenarios and requirements for the full frequency range 52.6-114.25 GHz, and this is expected to lead to a detailed study of technical considerations in Rel-17 and specifications in Rel-18.

A number of technical challenges exist at these frequencies. Since power amplifier (PA) efficiency and linearity decreases with frequency, a more efficient modulation scheme than OFDM is beneficial, such as single carrier modulation. The key to achieving reasonable efficiency is minimizing the PA back-off necessary to operate within the linear region, and this requires analysis with realistic PA models. It is also well known that local oscillator (LO) phase noise increases with carrier frequency, and to mitigate this it is necessary to consider very large sub-carrier spacings (SCS) (e.g. 960 kHz) and a re-design of the phase tracking reference symbols (PTRS). Figure 21 illustrates that 64QAM and CP-OFDM require a very high SCS (3.84 MHz) to combat phase noise, while single carrier can support 64QAM with a SCS of only 960 kHz.

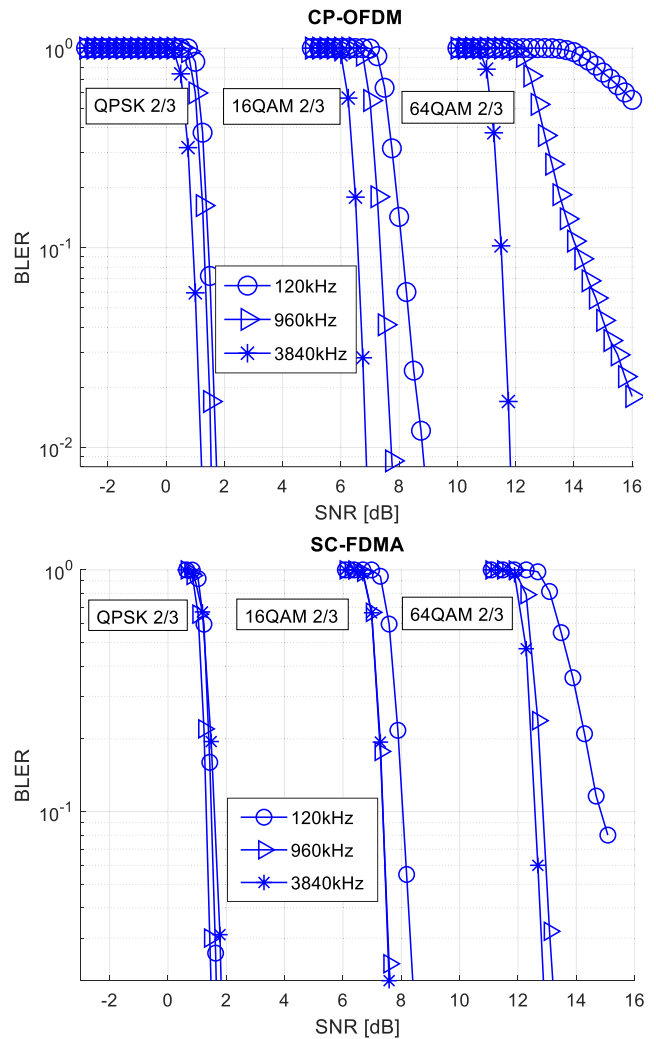


FIGURE 21. BLER vs. Eb/No for CP-OFDM and SC-FDMA for various SCSs.

Finally, at high mmWave bands a greater number of antenna elements are required to compensate for the pathloss, and this results in narrower beams, higher Equivalent Isotropic Radiated Power (EIRP), and path diversity solutions to increase the availability of Line-of-Sight (LoS) channels. Therefore, enhancements to beam management and handling of path diversity are also expected to be studied in the design of NR for this frequency range. It is expected that 3GPP will adopt a harmonized waveform design covering the full frequency range from 52.6 to 114.25 GHz (i.e. including the whole of the W-band).

J. OTHER ENHANCEMENTS

This section has described the most significant new features foreseen for NR over the coming releases, but other smaller features are also under consideration and may include, enhanced support for UAVs, Higher order modulation (1024 QAM) in the downlink, MBMS support as mentioned in Section II, further UE Power Saving, and specific

TABLE 3. 5G evolution features.

Features	Comments
Industrial IoT related features	TSC, NPN, NR Light, Positioning
NR Unlicensed	Initial focus on 5 & 6 GHz bands; 60 GHz later.
MIMO enhancements	Enhanced CSI feedback; multi-TRP operation
Integrated Access & Backhaul	To facilitate network deployment, especially at mmWave frequencies.
Non-Terrestrial Networks	Including primarily LEO satellites and HAPs
NR at high mmWave frequencies	52.6 GHz to 114.25 GHz
Sidelink communication	V2X, Public Safety
Multicast and broadcast	V2X, Public Safety
Edge computing enhancements	Low latency applications
Data collection and data analytics	Analytics powered networks

support for UEs with multiple Subscriber Identity Module (SIM) cards, to name but a few.

IV. CONCLUSION

The 5G standard developed by 3GPP not only offers unprecedented performance in its first release (Rel-15), but also provides scope for extensive enhancements in a backward-compatible manner in the subsequent releases over the coming years. The first sets of new features in Rel-16 and under discussion for Rel-17 are described in this paper, covering both continued enhancements of existing features and introduction of new features to bring the benefits of ultra-high data rate, low-latency, time-bound and highly reliable communications to an ever-wider ecosystem. Table 3 summarizes the key enhancements in the foreseeable future.

ACRONYMS

3GPP	3rd Generation Partnership Project
5GC	5G Core
5GS	5G System
AMF	Access and Mobility Management Function
CU	Centralized Unit
CUC	Centralized User Configuration
DU	Distributed Unit
eMBB	enhanced Mobile Broadband
eNA	enhanced Network Automation
FDM	Frequency-Division Multiplexing
gNB	5G NR base station
HAPS	High Altitude Platform Station
IAB	Integrated Access and Backhaul
IIoT	Industrial Internet of Things
IoT	Internet of Things
LMC	Location Management Component
MBMS	Multimedia Broadcast and Multicast Services

mMTC	massive Machine Type Communications
NB-IoT	Narrow Band IoT
NF	Network Function
NPN	Non-Public Network
NR	5G New Radio
NR-U	NR Unlicensed
NTN	Non-Terrestrial Network
NWDAF	Network Data Analytics Function
OAM	Operations and Management
ProSe	Proximity Services
RAN	Radio Access Network
RRC	Radio Resource Control
RTK	Real-Time Kinematic
SBA	Service Based Architecture
TDD	Time-Division Duplexing
TSC	Time-Sensitive Communication
TSN	Time Sensitive Networking
UAV	Unmanned Aerial Vehicle
UE	User Equipment
UPF	User Plane Function
URLLC	Ultra-Reliable Low Latency Communication
V2X	Vehicle-to-X communication
VLAN	Virtual LAN
VNF	Virtual Network Function

ACKNOWLEDGMENT

The authors would like to thank Rapeepat Ratasuk, Mark Cudak, Fred Vook, Peter Merz, Kari Hooli, Jari Hulkkonen, Jun Tan, Jie Chen, Ryan Keating, Antti Toskala, Miikka Poikselka, Frank Hsieh, Oskari Tervo, Costas Samdanis, Ulrich Dropmann, and Suresh Nair for their extensive help in preparing the article.

REFERENCES

- [1] *System Architecture for the 5G System*, document TS 23.501 V16.1.0, 3GPP, Jun. 2019.
- [2] A. Ghosh, T. Kolding, A. Maeder, S. Redana, and M. Uusitalo, "5G evolution: A view on cellular technology beyond 5G," presented at the Microw. Workshops Exhib. (MWE), Yokohama, Japan, Nov. 2018. Accessed: Sep. 10, 2019. [Online]. Available: https://apmc-mwe.org/mwe2018/pdf/WS_01/FR2A-2_1.pdf
- [3] P. Mogensen, A. Ghosh, A. Maeder, M. Uusitalo, and S. Redana, "5G evolution: A view on cellular technology beyond 5G," presented at the 6G Conf., Levi, Finland, Mar. 2019. Accessed: Sep. 2019. [Online]. Available: <http://www.6gsummit.com/speakers/preben-mogensen/>
- [4] B. Bertenyi. (Jul. 2019). 5G Standardization Update. 3GPP Webinar. Accessed: Aug. 3, 2019. [Online]. Available: <https://www.3gpp.org/news-events/2058-ran-rel-16-progress-and-rel-17-potential-work-areas>
- [5] M. K. Weldon, *The Future X Network: A Bell Labs Perspective*. Boca Raton, FL, USA: CRC Press, 2016.
- [6] *IEEE Standard for Local and Metropolitan Area Networks—Bridges and Bridged Networks—Amendment 31: Stream Reservation Protocol (SRP) Enhancements and Performance Improvements*, IEEE Standard 802.1Qcc-2018, Oct. 2018.
- [7] *Study on Enablers for Network Automation for 5G—Phase 2*, document SP-190557, 3GPP, Sapporo, Japan, Jun. 2019.
- [8] *Study on Enhanced Support of Industrial IoT*, document S2-1908585, 3GPP, Sapporo, Japan, Jun. 2019.
- [9] *Study on Enhanced Support of Non-Public Networks*, document SP-190453, 3GPP, May 2019.
- [10] *Study on Enhancement of Support for 5G LAN-Type Service*, document SP-190454, Reno, NV, USA, May 2019.

- [11] *Study on Enhancement to the 5GC Location Services-Phase 2*, document SP-190452, 3GPP, Reno, NV, USA, May 2019.
- [12] *Study on Architecture Aspects for Using Satellite Access in 5G*, document SP-181253, 3GPP, Reno, NV, USA, May 2019.
- [13] *Study on Supporting Unmanned Aerial Systems Connectivity, Identification, and Tracking*, document SP-181114, 3GPP, Reno, NV, USA, May 2019.
- [14] *Study on System Enhancement for Proximity Based Services in 5 GS*, document SP-190443, 3GPP, Reno, NV, USA, May 2019.
- [15] *Study on Architectural Enhancements for 5G Multicast-Broadcast Services*, document SP-190442, 3GPP, Reno, NV, USA, May 2019.
- [16] *Study on Enhancement of Support for Edge Computing in 5GC*, document SP-190185, 3GPP, Reno, NV, USA, May 2019.
- [17] *Revised WID: Integrated Access and Backhaul for NR*, document RP-190712, 3GPP, Shenzhen, China, Mar. 2019.
- [18] *Revised WID on NR-based Access to Unlicensed Spectrum*, document RP-190706, 3GPP, Shenzhen, China, Mar. 2019.
- [19] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [20] E. Onggosanusi, M. S. Rahman, L. Guo, Y. Kwak, H. Noh, Y. Kim, S. Faxer, M. Harrison, M. Frenne, S. Grant, R. Chen, R. Tamrakar, and Q. Gao, "Modular and high-resolution channel state information and beam management for 5G new radio," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 48–55, Mar. 2018.
- [21] F. W. Vook, W. J. Hillery, E. Visotsky, J. Tan, X. Shao, and M. Enescu, "System level performance characteristics of sub-6 GHz massive MIMO deployments with the 3GPP new radio," in *Proc. IEEE VTC-Fall*, Aug. 2018, pp. 1–5.
- [22] F. W. Vook, A. Ghosh, E. Diarte, and M. Murphy, "5G new radio: Overview and performance," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 1247–1251.
- [23] F. W. Vook, A. Ghosh, and T. A. Thomas, "MIMO and beamforming solutions for 5G technology," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2014, pp. 1–4.
- [24] *Report of 3GPP TSG RAN Meeting #80*, document RP-181999, 3GPP, ETSI MCC, RAN#81, Gold Coast, QLD, Australia, Sep. 2018.
- [25] *New SID on NR Industrial IoT*, document RP-181479, 3GPP, La Jolla, CA, USA, Jun. 2018.
- [26] *Study on NR beyond 52.6 GHz*, document RP-182861, 3GPP, Sorrento, Italy, Dec. 2018.
- [27] *Study on Solutions Evaluation for NR to Support Non Terrestrial Network*, document RP-181370, 3GPP, La Jolla, CA, USA, Jun. 2018.
- [28] *Key Directions for Release 17*, document RP-190831, 3GPP, Newport Beach, CA, USA, Jun. 2019.
- [29] A. Ghosh, "The 5G mmWave radio revolution," *Microw. J.*, vol. 59, no. 9, pp. 22–36, 2016.
- [30] *Revised WID on 5G V2X With NR Sidelink*, document RP-190984, 3GPP, Newport Beach, CA, USA, Jun. 2019.



ANDREAS MAEDER received the Ph.D. degree from the University of Wuerzburg, Germany, in 2008. He is currently a Senior Project Manager and a RAN Expert with Nokia Bell Labs. Since 2015 with Nokia, he has been coordinating the technical aspects of Nokia's standardization and research work on 5G radio and system. He has more than 10 years of experience in the telecommunication industry in various roles in research and standardization, including public funded projects, IEEE, and 3GPP. He was actively contributed as a delegate to the standardization of 3GPP RAN and system architecture, and IEEE 802.16. He is the author of numerous standard contributions and publications, including books and book chapters. He holds numerous patents in the field of mobile communications.



MATTHEW BAKER received the degrees in engineering and electrical and information sciences from the University of Cambridge, U.K. He has contributed to the standardization of UMTS/HSPA, LTE, and 5G in 3GPP, and held the posts of Chairman and subsequently Vice-Chairman of 3GPP TSG RAN WG1, from 2009 to 2017. He is a Chartered Engineer and a Member of the Institution of Engineering and Technology, and has been a Visiting Professor with the University of Reading, U.K. He is currently the Head of radio physical layer and co-existence standardization with Nokia Bell Labs, and a Distinguished Member of the Technical Staff. He has authored many articles and holds numerous patents in the field of mobile communications. He is a Co-Editor of the book entitled *LTE—The UMTS Long Term Evolution: From Theory to Practice* (Wiley, Second Edition 2012).



AMITABHA (AMITAVA) GHOSH (F'15) received the Ph.D. degree in electrical engineering from Southern Methodist University, Dallas, TX, USA. He joined Motorola, in 1990, involved in multiple wireless technologies starting from IS-95, cdma-2000, 1xEV-DV/1XTREME, 1xEV-DO, UMTS, HSPA, 802.16e/WiMAX, and 3GPP LTE. He is currently a Nokia Fellow and the Head of the Radio Interface Group, Nokia Bell Labs. He has 60 issued patents, and has authored multiple book chapters and numerous external and internal technical articles. He is currently involved in 3GPP LTE-advanced and 5G and 5G evolution technologies. His current research interests include digital communications, signal processing, and wireless communications. He was a member of IEEE ACCESS Editorial Board and a coauthor of the book entitled *Essentials of LTE and LTE-A*. He was a recipient of the 2016 IEEE Stephen O. Rice and 2017 Neal Shephard Prizes.



DEVAKI CHANDRAMOULI received the B.E. degree in computer science from Madras University, India, and the M.S degree in computer science from the University of Texas at Arlington, USA. She is currently the Head of North American Standardization with Nokia Bell Labs. She is the rapporteur and a lead of 5G system architecture specification in 3GPP. She is also the rapporteur of 3GPP work related to Industrial 5G, Vertical_LAN for Rel-16, and FS_IIoT in Rel-17 in 3GPP SA2. She has coauthored articles on 5G published by the IEEE and a book on *LTE for public safety* (Wiley, 2015), and coedited the book on *5G for the connected world* (Wiley, 2019). She has (co)authored over 130 patents in wireless communications.