Ericsson White Paper
1/28423-FGB1010937
April 2020

ERICSSON

5G wireless access: an overview

Abstract

The capabilities of 5G wireless access extend far beyond previous generations of mobile communications. 5G wireless access provides wireless connectivity for a wide range of new applications and use cases. This includes industries such as automotive, logistics, public safety, media, and manufacturing. Examples of these capabilities include very high data rates, spectral efficiency and mobility requirements, ultra-low latency, ultra-high reliability, and very low device cost and energy consumption. These capabilities are realized by the development of LTE in combination with 5G New Radio (NR) technologies, and include:

— The additional availability of frequency bands in low, mid, and high bands that together provide a combination of coverage, very high bitrates, and ultra-low latency
— An ultra-lean design to enhance network energy performance and reduce interference, interworking, and LTE co-existence
— The compatibility to prepare for future use cases and technologies, and the low latency to improve performance and enable new use cases
— The extensive usage of beamforming and a massive number of antenna elements for data transmission and control plane procedures

What is 5G?

The term "5G" is used to refer to the fifth generation of mobile wireless technologies, which originated with analog mobile telephony in the late 1980s and has progressed to the point where all people and things can be connected to the Internet. Every generation of mobile technology has aimed to provide connectivity anywhere and anytime. The underlying technological objectives and the capabilities of the network, however, have continued to shift into a new generation every 7 to 10 years, with each generation being designed to serve societal needs over a duration of 2 to 3 decades across markets around the globe. The impact of these generations on the way we communicate may be viewed across various dimensions. These include: (1) service offerings, (2) air-interfaces, (3) data rates (4), spectrum ranges, and (5) performance.

Early generations began with only isolated concerns, such as telephone conversations, while later ones evolved to digital data communication and more sophisticated service architectures. Accordingly, the focus in past generations was consistently on operation in wider spectrum ranges and with higher data rates and traffic handling capacities. These objectives continue to be important today, with 5G being able to operate over much wider range of frequency bands than ever before. The added attention to massive IoT support, the development of support for critical services, and the reimagining of the core network functionality to better support distributed cloud computation and service orchestration expands the potential of 5G further.

The combination of cellular technologies such as 5G NR, Long-Term Evolution (LTE), and NB-IoT (Narrowband Internet of Things) address a broader reach for cellular communications, spanning public telecommunications services as well as private industry applications and verticals, like energy, transportation and smart cities. 5G NR offers support for critical services that place extreme demands on latency and link reliability while LTE for Machine Type Communication (MTC) (aka LTE-M) and NB-IoT tackle huge numbers of IoT devices that are deployed as sensors and actuators, often with energy and power limitations. Thus, the overall aim of 5G is to provide ubiquitous connectivity for any kind of device and any kind of application that may benefit from being connected.

The 5G system is built on radio access nodes and distributed and centralized data centers, allowing for a flexible allocation of workloads. These nodes and data centers are connected via programmable transport networks, which are connected via backbone nodes that carry the information from the access nodes to the data centers, where most of the data is stored and the network itself is managed. In addition to this, the management of applications, cloud, transport, and access resources can be allocated centrally in the data center or be flexibly allocated as necessary.

5G systems will have a significant role to play — not just in the evolution of communications but in the evolution of businesses and society as a whole. The 5G system is designed to handle the phenomenal increase in traffic from previous years in a way where value can be captured by stakeholders with minimal impact on consumer net costs. The Ericsson Mobility Report records global traffic levels at 38 exabytes per month at the end of 2019, with a projected four-fold increase to 160 exabytes per month by 2025 [1].

As with previous generations, 5G radio access technologies lie at the heart of 5G systems, providing the wireless connectivity for a wide range of new applications and use cases, most prominently including industries such as automotive, logistics, public safety, media, and manufacturing. As a result, it has been accelerating the development and implementation of IoT.

3GPP standards and 5G initiatives in industries

The access radio and core network of any cellular technology is standardized by the 3GPP (the 3rd Generation Partnership Project) organization. Although 3GPP began work on the specification of 5G in March 2017, they had already standardized LTE-M and NB-IoT features that addressed some of the Machine Type Communications requirements (such as long battery life, low device cost, and coverage extension) in Releases 13 and Release 14, by 2016 and 2017 respectively.

The 5G standardization work was, thereafter, divided into two major phases. The first phase included standardization of the fundamental 5G building blocks that had already been completed in March of 2019 (Release 15), to which further enhancements were added (Release 16). These enhancements constituted phase two. The 5G timeplan including the 3GPP releases is shown in Figure 1.

5G combines two radio technologies: a novel radio interface technology — denoted as new radio (NR) — and LTE. In order to accelerate the 5G schedule, 3GPP divided Release 15 into three steps. The first step, consisted of non-standalone (NSA) NR architecture, was completed in December 2017. The second step, tackling standalone (SA) NR, was completed in June 2018. In the operation of NSA networks, LTE is a must for initial access and mobility handling, while its SA counterpart can be deployed independently of LTE. The third and last step for Release 15 was a late drop completed in March 2019, which included more architecture options (for example, the possibility to connect 5G NodeBs (gNBs) to an Evolved Packet Core (EPC) and then operating NR and LTE in multi-connectivity mode, wherein NR is the master node and LTE is the secondary node). Introducing a 5G standard

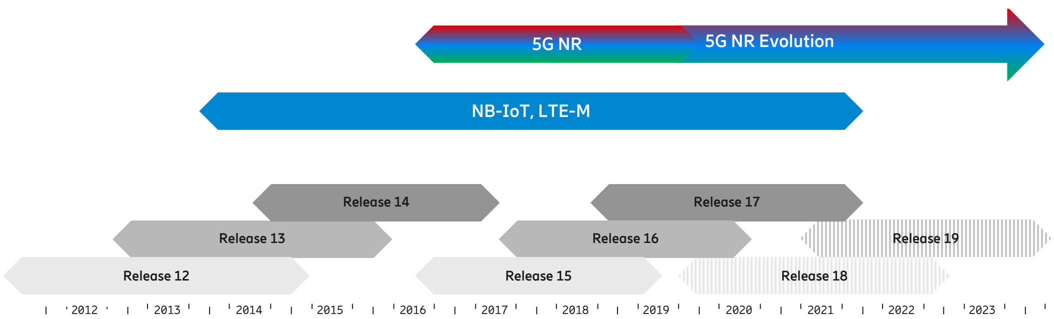


Figure 1 5G Timeplan and 3GPP Releases 12 to Release 19. NB-IoT and LTE-M are being evolved in 3GPP Release 17. The overlaps between the shown 3GPP releases are due to the fact that releases start with the services and system aspect (SA) studies roughly a year before the RAN specifications begin.

is but one step needed to secure a solid 5G ecosystem. The automotive and manufacturing industries, for example — stand out as unique beneficiaries traditionally removed from the cellular ecosystem — have taken concrete steps to embrace the potential of 5G as a complement to their needs.

The 5G Alliance for Connected Industries and Automation (5G-ACIA), launched in April 2018 [2], aims to firstly, establish a common language between OT (operational technology) and ICT (information and communications technology) and secondly, to ensure that the requirements of the industrial domain are considered in 5G standardization, ultimately paving the way for a 5G ecosystem for the industrial domain.

The 5G Automotive Association (5GAA), created in 2016 [3], is a cross-industry organization comprised of companies from the automotive, technology, and telecommunications industries whose main goal is to develop end-to-end solutions for future mobility and transportation services.

A short history of 5G

The journey towards 5G began almost a decade ago with research activities. The EU project METIS set the foundation for 5G between 2012 and 2014 [4], which fueled a number of global discussions and activities, including 5G-PPP (Europe), 5G Forum (Korea), IMT-2020 Promotion Group (China), and 5GMF (Japan), to name a few. These activities set the stage for the 3GPP standard discussions that were to begin in 2015 and which officially started on Radio Access Networks (RANs) in 2016, followed by the completion of the fundamental building blocks in June 2018 [5]. As a result of these endeavors, the first 5G deployments finally arrived in early 2019 in South Korea and the US. Shortly after, the first mobile phones became available, where the largest-scale deployment occurred in Korea, culminating in a total of 5 million users (which accounts for approximately 40 percent of the global 5G penetration) [1].

5G applications

Conforming to the ITU nomenclature for IMT-2020 [6], 5G targets three main use case families with distinct connectivity requirements: enhanced mobile broadband (eMBB), massive machine type communications (mMTC) and ultra-reliable low latency communications (URLLC) (also called critical machine type communications (cMTC)), shown in Figure 2.

Mobile broadband addresses human-centric use cases such as mobile telephony and media delivery. eMBB enables large volumes of data transfer and extreme data rates. Typical usage is in mobile phones and mobile PCs/tablets.

mMTC and URLLC target machine-centric use cases. The focus of mMTC is on providing connectivity to a massive number of low complexity narrow-bandwidth devices that infrequently send or receive small volumes of data. These devices can be in challenging radio conditions requiring coverage extension capabilities and may solely rely on battery power supply. Common use cases include low-cost sensors, meters, actuators, trackers, and wearables.

URLLC is for use cases with stringent requirements on reliability and latency such as AR/VR, advanced wearables, autonomous vehicles, real-time human machine collaboration, cloud robotics and real-time coordination and control of machines and processes. The reliability is defined as probability of successful data delivery within a specified time duration.

In reality, the division among the three categories has been made to particularly ease the understanding of the requirements; however, these do not fit several use cases, including those where some MTC-type communications might have mobile broadband requirements, as well as others. In essence, for many use cases, they can have hybrid requirements stemming one, two, or three categories.

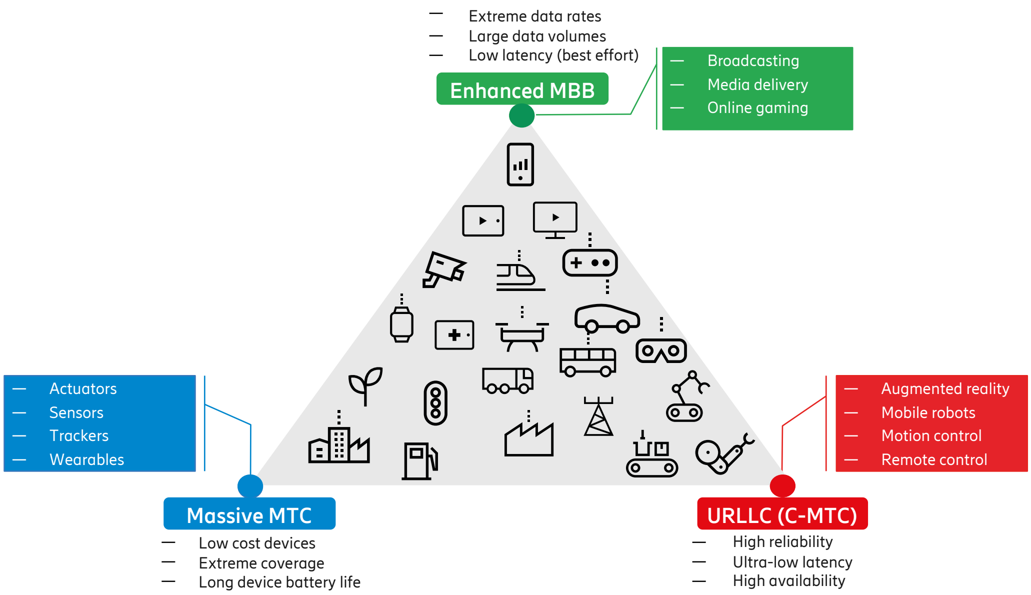


Figure 2 5G application categories.

5G for Internet of Things

Many industries are already experiencing the benefits of cellular connectivity, for example in the consumer electronics, automotive, railways, mining, utilities, healthcare, agriculture, manufacturing and transportation sectors. There are over one billion cellular Internet-of-Things (IoT) connections in 2020 and Ericsson forecasts around 5 billion connections by 2025 [1]. Almost every industry is exploring the potential of 5G for fundamentally transforming their businesses.

The wireless connectivity across various industries can be grouped into four distinct sets of requirements. To address these requirements, Ericsson has defined four IoT connectivity segments [7]: Massive IoT, Broadband IoT, Critical IoT, and Industrial Automation IoT*, as illustrated in Figure 3. Each IoT connectivity segment is multi-purpose, addressing multiple use cases in multiple industries.

Massive IoT addresses the ITU-R mMTC requirement with NB-IoT and Category-M or Cat-M devices (low-complexity device series defined as part of LTE-M). LTE-M /NB-IoT can efficiently co-exist with 5G NR in the same spectrum and fulfill all 5G massive MTC requirements, as set out in the IMT-2020 and 3GPP standards, in terms of coverage, latency, data rate, battery life, and connection density [4][5]. LTE-M and NB-IoT have a smooth and future-proof evolution in 5G networks [4].

Broadband IoT connectivity adopts the capabilities of eMBB for IoT to provide large volumes of data transfer, much higher data rates and lower latencies than Massive IoT, while enabling additional capabilities for IoT such as extended device battery life, extended coverage, and uplink heavy data rates.

* "Industrial Automation IoT" segment refers to the scenarios where the 5G system is integrated into the Ethernet based deterministic networks used for the most stringent use cases in industrial automation. Otherwise, industrial automation use cases are diverse and the other three IoT connectivity segments are also relevant for industrial automation use cases.

rates. There are more than 500 million Broadband IoT users as of 2020 [1]. Commercial usage today is dominated by personal cars, commercial vehicles, trains, wearables, gadgets, cameras, sensors, actuators and trackers.

Critical IoT connectivity is for time-critical communications. It enables ultra-reliable and/or ultra-low latency communication at a variety of data rates. In contrast to Broadband IoT which achieves low latency on average and best effort basis, Critical IoT can deliver data within specified latency bounds with required guarantee levels, even in heavily loaded networks. Such time-critical use cases are in almost every industry.

Industrial Automation IoT aims at enabling seamless integration of cellular connectivity into the wired industrial infrastructure used for real-time advanced automation. It includes capabilities for integrating 5G system with Ethernet based industrial protocols and Time Sensitive Networking (TSN) [7][8][9].

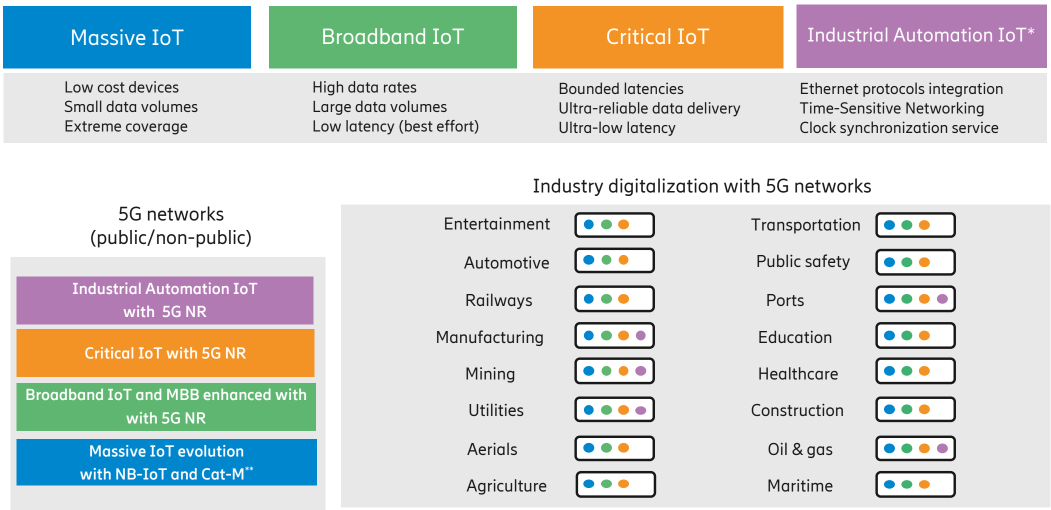


Figure 3 Ericsson's IoT connectivity segments for industries using 5G NR, NB-IoT and Cat-M**.

5G NR supports eMBB and Broadband IoT performance while also enabling Critical IoT and Industrial Automation IoT segments [7]. The four IoT connectivity segments thus co-exist in one 5G network, whether deployed for public or non-public access. Some devices may need multiple IoT connectivity segments for executing one or multiple use cases, for example, an autonomous vehicle with rich requirements [10].

** The term Cat-M in the figure is used to denote the LTE-M features for low complexity/cost Cat-M type devices.

5G—requirements and capabilities

A new radio interface technology in 3GPP RAN, such as 5G, is specified after an exercise to collect and agree on requirements for the system. These requirements were largely derived from a holistic 5G vision. In this case, the vision was to enable connectivity for a very wide range of applications with diverse characteristics and requirements that bring value in a variety of dimensions, such as latency, data rate, reliability, and security.

Traffic demands for mobile communication systems are predicted to increase dramatically [1], and to support this traffic in an affordable way, 5G networks aim to deliver data at much lower costs per bit compared to the networks of today. The anticipated increase in data consumption will undoubtedly increase the energy footprint of networks. Therefore, 5G is designed to consume significantly lower energy per delivered bit than current cellular networks.

The exponential increase in numbers of connected devices over time, owed to future deployment of billions of wirelessly connected sensors, actuators, and similar devices for mMTC, will demand that the network support new paradigms in device and connectivity management that improve provisioning and do not compromise security. Devices will typically generate or consume very small amounts of data and will individually, or even jointly, have limited impact on the overall traffic volume. However, the sheer number of connected devices seriously challenges the ability of networks to provision signaling and to manage connections.

All-in-all, the capabilities of 5G wireless access extends far beyond those of previous generations of mobile communications. These capabilities include very high data rates and spectral efficiency, very low latency, ultra-high reliability, very low device cost and energy consumption, and energy-efficient networks. All these requirements are based on ITU (International Telecommunication Union) requirements as well as commercial requirements identified by 3GPP and organizations such as 5G-ACIA [11].

The sections below give an overview of the 3GPP/ITU-2020 requirements, including information as well as a highlight regarding to what extent 5G fulfills them. A summary of the key requirements is provided in the tables below, with Table 1 showing a summary of the eMBB requirements and Table 2 showing a summary of the URLLC and mMTC requirements.

Table 1 Requirements for eMBB.

Requirement	DL	UL	Enabling feature
Peak data rate	20 Gbps	10 Gbps	Carrier aggregation, higher modulation, Massive SU MIMO
Peak spectral efficiency	30 bps/Hz	15 bps/Hz	Higher modulation, Massive MIMO
User-experienced data rate (fifth percentile user data rate)	100 Mbps	50 Mbps	Multi-antenna
Cell average spectral efficiency	Scenario dependent, see [12]		Multi-antenna
Fifth percentile user spectral efficiency	Scenario dependent, see [12]		Multi-antenna
Area traffic capacity	10 Mbps/m^2 (indoor hotspot scenario)		Multi-antenna
User plane latency	4 ms, one way for both downlink and uplink		Mini-slot, scalable numerology, fast HARQ, pre-emption

User plane latency	4 ms, one way for both downlink and uplink	Mini-slot, scalable numerology, fast HARQ, pre-emption
Control plane latency	10 ms	Mini-slot, scalable numerology, fast HARQ, pre-emption, RRC inactive
Energy efficiency	High sleep ratio and long sleep duration under low load -	Ultra-lean design
Speed	Up to 500 km/hour for train communication	Multi-antenna (demodulation reference signals)

Table 2 Requirements for URLLC and mMTC.

Requirement	Value	Enabling feature
User plane latency	1 ms, one-way for both downlink and uplink	Mini-slot, flexible numerology, fast HARQ, pre-emption
Control plane latency	10 ms	Mini-slot, scalable numerology, fast HARQ, pre-emption, RRC inactive
Connection density	Support for 1,000,000 devices per sq km^2	
Reliability	99.999% success rate	Multi-antenna, robust control and data design (low MCS/CQI), multi-connectivity (including duplication), retransmissions (HARQ)
Mobility interruption time	0 ms	multi-connectivity (dual connectivity)
Battery life	10 years	RRC inactive, power saving mode, enhanced DRX

Very high data rates, spectral efficiency, and mobility requirements

Every generation of mobile communications has been associated with higher data rates compared to its previous generation. The requirements on peak data rate have been defined for 5G as 20 Gbps for downlink and 10 Gbps for uplink. For the downlink, it is seen that the peak data rate is 17.5 Gbps for a 400 MHz-wide component carrier. Aggregating two such component carriers with a total bandwidth of 800 MHz exceeds the ITU-R requirement with a peak data rate of 35 Gbps. For the uplink, the peak data rate on one 400 MHz component carrier is 9.4 Gbps. Aggregating two such component carriers likewise gives a peak data rate well beyond 10 Gbps.

The requirements on peak spectral efficiency are 30 bps/Hz for downlink and 15 bps/Hz for uplink. Efficiencies of up to 44 bps/Hz are achieved on the downlink. Similarly, uplink peak spectral efficiencies of up to 24 bps/Hz are achieved. These metrics are met by using higher modulation schemes, such as 256QAM (both in downlink and uplink), and several data transmission streams or layers (eight layers in downlink and four in the uplink).

When it comes to the requirements on experienced data rates, they are defined to be 100 Mbps for downlink and 50 Mbps for uplink. These are to be obtained in the fifth percentile in the dense urban scenario. The evaluations show that the requirements can be reached assuming a bandwidth of 317 MHz.

The requirements in fifth percentile user spectral efficiency and average spectral efficiency are scenario dependent. Results for all scenarios — fifth percentile and average, downlink and uplink — are above the requirements [12].

The requirement on area traffic capacity is 10 Mbps/m^2 for downlink in the in the indoor hotspot scenario. Results indicate that 10 Mbps/m^2 is reached with a spectrum allocation of 219 MHz.

It is seen that several NR configurations can support the requirement on bandwidths of 100 MHz and above, with the largest component carrier having a bandwidth of 400 MHz. NR can support carrier aggregation (CA) of up to 16 component carriers, offering support for signals in excess of 1 GHz.

To support mobility in NR, there is a requirement of supporting 1.5 bps/Hz, which should be supported for combinations of normalized data rates and terminal speeds at the median uplink signal-to-interference-plus-noise ratio (SINR). It is seen that for all scenarios [13], the median SINR is higher than what is needed to reach the required normalized data rates. Furthermore, there is also a requirement on the mobility interruption time, defined as the shortest time duration, supported by the system during which a user terminal cannot exchange user plane packets with any base station during transitions. The requirement on mobility interruption time is 0 ms and, for NR, the requirement is fulfilled with intra-cell beam mobility or with carrier aggregation.

Ultra-low latency

Ultra-low latency capability is driven by the demands from new applications with extreme requirements. Some envisioned 5G use cases, such as traffic safety and control of critical infrastructure and industry processes, require much lower latency compared to what is possible with the mobile communications systems of today.

The requirements on user plane latency are 4 ms for eMBB and 1 ms for URLLC, one way for both downlink and uplink. It is seen that NR FDD and NR TDD can fulfil the 1 ms uplink latency target.

The IMT-2020 requirement on control plane latency is 10 ms, just as has been adopted by 3GPP. The control plane latency refers to the transition time from a battery efficient state to the start of continuous data transfer. Evaluating NR Release 15 makes it clear that both NR frequency division duplexes (FDDs) and NR time division duplexes (TDDs) can reach the ITU and 3GPP 5G targets on control plane latency.

Ultra-high reliability

Many services will distribute computational capacity and storage close to air interfaces. This offers new capabilities for real-time communication and will enable ultra-high service reliability in a variety of scenarios, ranging from entertainment to industrial process control (meaning that 5G should also enable connectivity with ultra-high reliability and ultra-high availability).

The requirements state that NR should be able to deliver a 32-byte message within 1 ms with a success probability of 99.999 percent at a link quality defined by the fifth percentile SINR in the urban macro scenario. In some cases, loss of connectivity and deviation from quality of service (QoS) requirements would be extremely rare (for example, some industrial applications might need to guarantee availability up to 99.999999 percent).

Very low device cost and energy consumption

Low-cost, low-energy mobile devices have been a key market requirement since the early days of mobile communications; however, to enable the vision of billions of wirelessly connected sensors, actuators, and similar devices, a further step must be taken in terms of device cost and energy consumption. It should be possible for 5G devices to be available at very low costs and with a battery life of several years without recharging. Many of the features needed to extend battery life and reduce costs were introduced in 3GPP Releases 12 and 13 (see chapter 4 in [4]). These include power saving mode (PSM), enhanced discontinuous reception (DRX), reduced peak rate, operation in half duplex FDD, and the introduction of single receive antennas at the user equipment (UE) side.

From a requirement point of view, the connection density in 5G is defined to provide a minimum quality of service support for 1,000,000 devices per square kilometer. This is specified as transferring messages of 32 bytes with an interarrival time of two hours within a delay of 10 s. It can be shown that these requirements are met for NR, LTE-M, and NB-IoT.

The requirement placed on energy efficiency is support for efficient data transmissions in loaded cases as well as low energy consumption when there is no data. It is seen that NR indeed transmits data efficiently, which is also verified by the spectral efficiency requirements, and that NR, through its ultra-lean design, enables low energy consumption when there is no data.

The importance of these factors will increase further in the 5G era, and energy efficiency will, therefore, be an important requirement in the design of 5G wireless access.

Spectrum for 5G

One objective of 5G is to transition the cellular industry to gigabit rates and sub-millisecond latencies for enhanced mobile broadband (eMBB) and critical services. The target data rates, as mentioned earlier, are in order of magnitude over the current peak rates possible with 4G and will achieve support of increased traffic capacity and very high data rates for wide area coverage.

5G is specified to operate over a much wider range of frequencies than before, from below 1 GHz to 100 GHz. Low-band spectrum (below 2.5 GHz) provides excellent coverage, mid- band spectrum (2.5–10 GHz) provides a combination of good coverage and very high bitrates, and high band-spectrum (10–100 GHz) provides the bandwidths needed for the highest bitrates (up to 20 Gb/s) and lowest latencies envisioned for 5G.

Spectrum for 5G continues to be a topic much examined by ITU as well as by regional regulatory bodies and individual countries. The recent World Radio Conference (WRC) 2019, which ended on November 22, identified several bands for IMT. These include 24.25–27.5 GHz globally, 37–43.5 GHz globally, 45.5–47 GHz in more than 60 countries, 66–71 GHz globally, and 47.2–48.2 GHz in Region 2 (that is, North and South America) and additional countries (nearly 100 countries in total). In addition, agenda items for WRC 2023 incorporate further work on potential International Mobile Telecommunications (IMT) identifications or mobile primary allocations within the frequency range of 470 MHz –10.5 GHz

Identification of suitable spectrum is an important harmonization tool, providing a message to the industry that there is an intention to use the frequency band in question for IMT. As a result of the work in and outside of ITU, consensus is emerging around the 3.3–4.2 GHz and 24.25–29.5 GHz frequency bands as preferred choices for the first phase of deployments, with additional activities in 4.4–5.0 GHz, 37–43.5 GHz, and additional bands from WRC-19 to accommodate regional differences and a second stage of 5G deployment. As an example, the US has settled on the use of parts of the 24.25–27.5 GHz, the 27.5–28.35 GHz, the 37–40 GHz, and the 47.2–48.2 GHz bands, while service providers are also aggressively transitioning existing 4G bands ranging from 600 MHz to 3.5 GHz to 5G NR, with possible additional mid-band spectrum in part of 3.7–4.2 GHz. The harmonization process for these bands is also reflected in the definitions of 3GPP frequency bands, with 3GPP NR bands defined for low-, mid-, and high-band spectrum currently up to 43.5 GHz.

It is important to understand that the combination of spectrum in the three different frequency ranges together will provide the full 5G experience and capabilities. Figure 4 provides an example of how to employ different frequency bands efficiently in an urban environment. The first picture shows downlink performance with only 26 GHz spectrum,

providing very high bitrates in certain areas but lacking in coverage, whereas the second picture demonstrates the benefits of combining 0.8 GHz, 2.6 GHz, 3.5 GHz, and 26 GHz to obtain very high bitrates as well as very good coverage.
In addition to eMBB use — primarily based on traditional nationwide licensing to service providers — 5G will also be of critical importance for different vertical industries, as

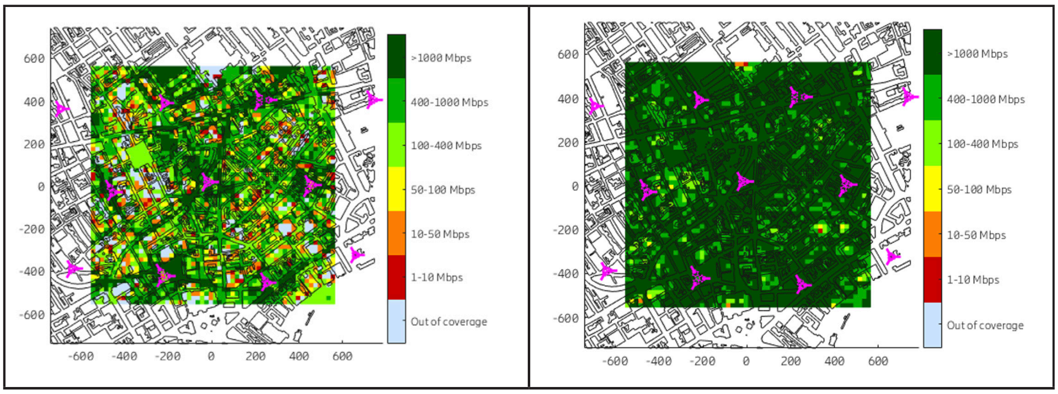


Figure 4 Efficient use of spectrum by combining different frequency bands.

described in the preceding section. Massive MTC will be handled primarily in the low- and mid-band spectrum range, where there is already sufficient spectrum available in 3GPP bands to provide the capacity needed. Critical MTC will, on the other hand, require larger amounts of spectrum available only in mid- and high-band spectrum (in the range of tens or even hundreds of MHz, depending on the application).

5G technology components

Beyond extending operation to higher frequencies, there are several other key technology components relevant for the evolution to 5G wireless access. These components include [5]:

— Exploitation of much higher frequency bands as a means to obtain additional spectrum needed to support very wide transmission bandwidths and the associated high data rates
— Ultra-lean design to enhance network energy performance and reduce interference,
— Interworking, and LTE co-existence
— Forward compatibility to prepare for future, as yet unknown use cases and technologies
— Low latency and high reliability to improve performance and enable new use cases
— Beam-centric design, enabling extensive usage of beamforming and a massive number of antenna elements not only for data transmission (which to some extent is possible in LTE) but also for control-plane procedures such as initial access

In addition, there are numbers of technology areas and use cases that are in focus to the 5G wireless access in particular access and backhaul integration, unlicensed spectrum, vehicle-to-anything(V2X), industrial IoT and positioning.

Higher frequency operation and scalable numerology

One key feature of NR is a substantial expansion in terms of the range of spectrum in which the radio-access technology can be deployed, from below 1 GHz up to 52.6 GHz, divided into two frequency ranges:
— Frequency Range 1 (FR1): 450 MHz – 7.125 GHz, commonly referred to as "sub-6 GHz"
— Frequency Range 2 (FR2): 24.25 GHz – 52.6 GHz, commonly referred to as "millimeter wave"

To support this wide range of carrier frequencies (sub-1 GHz for very large cells up to mm-wave deployments with smaller cells but very wide spectrum allocations) NR supports a flexible orthogonal frequency-division multiplexing (OFDM) numerology, with subcarrier spacings ranging from 15 kHz up to 120 kHz, for data transmission with a proportional change in cyclic prefix duration. A small subcarrier spacing has the benefit of providing a relatively long cyclic prefix in absolute time at a reasonable overhead, while higher subcarrier spacings are needed to handle, for example, the increased phase noise at higher carrier frequencies. Up to 3,300 subcarriers are used, although the maximum total bandwidth is limited to 400 MHz, resulting in the maximum carrier bandwidths of 50/100/200/400 MHz for subcarrier spacings of 15/30/60/120 kHz, respectively. As mentioned earlier, larger bandwidths are supported through carrier aggregation.

Ultra-lean design

Past mobile networks constantly broadcast their presence in the network, thus supporting network discovery. These signals are sometimes referred to as "always-on" signals. Under the typical traffic conditions for which LTE was designed, such transmissions constitute only a minor part of the overall network transmissions and, thus, have relatively small impact on the network performance. In very dense networks deployed for high-peak data rates, however, the average traffic load per network node can be expected to be relatively low, making the always-on transmissions a more substantial part of the overall network transmissions.

The ultra-lean-design principle in NR minimizes the always-on transmissions, thereby enabling higher network energy performance and higher achievable data rates. Several procedures, such as cell search and random access, have been redesigned compared to the LTE in order to support the ultra-lean paradigm. Another example is the demodulation reference signal structure, where NR relies heavily on reference signals being present only when data is transmitted but not otherwise.

Interworking and LTE coexistence

Due to the difficulty inherent in providing full wide-area coverage at higher frequencies, interworking with systems operating at lower frequencies is important; however, lower frequency bands are often already occupied by LTE. Therefore, the possibility for a service provider to deploy NR in the same spectrum as an already existing LTE deployment (initiating LTE/NR spectrum coexistence) is of the utmost importance.

The possibility for an LTE-compatible NR numerology based on 15 kHz subcarrier spacing, which enables identical time/frequency resource grids for NR and LTE, is a fundamental tool for such coexistence. The flexible NR scheduling, with a scheduling granularity as small as one symbol, can then be used to avoid scheduled NR transmissions to collide with key LTE signals. Reserved resources, introduced for forward compatibility and discussed below, can also be used to further enhance LTE/NR coexistence. It is possible to configure reserved resources matching the cell-specific reference signals in LTE, thereby enabling an enhanced LTE/NR overlay in the downlink.

Forward compatibility

An important aim in the development of the NR specification has been to ensure a high degree of forward compatibility in the radio-interface design. In this context, forward compatibility implies a radio interface design that allows for substantial future evolution (in terms of introducing new technology and enabling new services with as yet unknown requirements and characteristics) while still supporting legacy devices on the same carrier.

Forward compatibility is inherently difficult to guarantee, however, based on experience from the evolution of previous generations, NR has been designed to:

— Maximize the amount of time and frequency resources that can be flexibly utilized or that can be left blanked without causing backward compatibility issues in the future
— Minimize the transmission of always-on signals
— Confine signals and channels for physical layer functionalities within a configurable/allocable time/frequency resource

Note that these design principles partly coincide with the aim of ultra-lean design as described above. There is also the possibility with NR to configure reserved resources (that is, time-frequency resources that, when configured, are not used for transmission and, thus, are available for future radio-interface extensions).

Low-latency, high-reliability support, and flexible frame structures

Low-latency transmission modes are an important characteristic of NR and has informed many NR design choices. Control signals have been placed immediately before user data to allow the device to start processing received data immediately. It is also possible to transmit using a fraction of a slot, using "mini-slot" transmission, to reduce latency. The support for higher subcarrier spacing, and hence a correspondingly shorter slot duration, is also helpful from a latency perspective in some scenarios, assuming the resulting cyclic prefix is long enough. In high-load scenarios, latency can also be reduced by scheduling high-priority users on resources originally intended for low-priority users, a process known as "preemption." The requirements on the device (and network) processing times (in terms of scheduling latencies and hybrid ARQ (HARQ) retransmission latencies) are tightened significantly in NR as compared to LTE. The higher layer protocols MAC and RLC have also been designed with low latency in mind, with header structures chosen to enable processing without knowing the amount of data to transmit.

NR also defined a third UE state for inactive UEs in addition to the idle and connected states already present in LTE. In the inactive radio resource control (RRC) state, the RRC connection is kept while the UE is sleeping to conserve battery. This allows the UE to skip the request for an RRC connection setup and also wake up to transmit and receive data much faster as compared to moving from the idle to the active state.

In addition to exploiting multi-antennas to improve the transmission reliability of the communication, carrier aggregation (CA) and dual connectivity (DC) are used. This can be achieved by packet duplication on the packet data convergence protocol (PDCP) layer. (It should be noted that CA and DC are sometimes referred to as "multi-connectivity".)

Beam-centric design and multi-antenna transmission

The support for a large number of steerable antenna elements for both transmission and reception is a key feature of NR. At higher frequency bands, the large number of antenna elements are primarily used for beamforming to extend coverage, while at lower frequency bands, they enable full-dimensional MIMO (sometimes referred to as "massive MIMO") and interference avoidance by spatial separation.

NR channels and signals, including those used for control and synchronization, have all been designed to support beamforming. This is different compared to previous generations and is required to fully support operation at the higher frequency bands.

Duplex schemes and dynamic TDD

The duplex scheme is pre-determined by the spectrum allocation at hand. For lower frequency bands, allocations are typically paired, implying frequency division duplex (FDD) operation. At higher frequency bands, unpaired spectrum allocations are increasingly common, resulting in time division duplex (TDD) operation. NR supports both duplex schemes in a unified manner. Unlike LTE, where the TDD uplink-downlink allocation does not change over time, NR supports dynamic TDD as a key feature. In dynamic TDD, (parts of) a slot can be dynamically allocated to either uplink or downlink as part of the scheduler decision. This enables tracking rapid traffic variations, which are particularly pronounced in dense deployments with a relatively small number of users per base station. It is also possible to configure a static uplink-downlink allocation in a similar manner as in LTE, which is useful in scenarios such as wide-area deployments where dynamic TDD is less feasible due to the interference characteristics.

Access and backhaul integration

Integrated access backhauling (IAB) extends NR to support short-haul transport (that is, to enable the use of NR for a wireless link from central locations to distributed cell sites and between cell sites— for example, as an alternative to fiber backhauls). This can simplify deployments of, for example, small cells in dense urban networks or temporary sites deployed for a specific event.

IAB can be used in any frequency band in which NR operates. It is expected however that the mm-wave frequencies are the most appropriate choice due to the sheer amount of spectrum available. As higher-frequency spectrum typically is unpaired spectrum, this also means that IAB can be expected to primarily operate in TDD mode on the backhaul link. IAB may either operate in the same frequency band as the access link (known as in-band operation) or in a separate frequency band.

Unlicensed spectrum

Operation of 3GPP technologies in unlicensed spectrum was first introduced in LTE Release 13 with licensed assisted access (LAA). In LAA, the device is attached to the network using a licensed carrier and —optionally — can use one or more unlicensed carriers to boost data rates. NR will support a similar setup, but, unlike LTE, will also support standalone operation without support from a carrier in licensed spectrum. This greatly adds to the deployment flexibility of NR in unlicensed spectrum compared to LTE, providing a single global framework in Release 16, where operations are not only possible in the existing 5 GHz unlicensed bands (5150 – 5925 MHz) but also in new bands when they become available. The most important new band is 6 GHz, where there are potentially hundreds of MHz available. For example, the following frequency ranges are being discussed for unlicensed operation within various regulatory bodies: 5925–7125 MHz in the US and the lower part of the 6 GHz band (5925–6425 MHz) in Europe. While both NR and Wi-Fi will be new to the 6 GHz band, it is assumed that regulations will provide a framework for the protection of incumbent services (for example, fixed services) while unlicensed technologies operate in this band.

V2X direct

Intelligent transportation systems (ITS) require communications not only with fixed infrastructure but also between vehicles —one example of the new verticals in focus for 5G NR Release 16. The term "vehicle-to-anything" (V2X) is commonly used to cover all kinds of communication to and from a vehicle. Communication with fixed infrastructure is already catered to in the NR Release 15 via the access link, while the sidelink required for direct vehicle-to-vehicle communication is added in Release 16. Besides sidelink, many of the enhancements introduced for the cellular uplink/downlink interface of NR in Release 16 (for example, enhancements in the URLLC area) are also relevant for supporting advanced ITS services.

Industrial IoT and ultra-reliable low-latency communication (URLLC)

Industrial Internet of Things (IoT) is the other major new vertical in focus for NR Release 16. While Release 15 can provide very low air-interface latency and high reliability, further enhancements to latency and reliability are introduced in Release 16. This is to enable widening of the set of industrial IoT use cases and to include increased demand for new use cases, such as factory automation, electrical power distribution, and transport industry (including the remote driving use case). Time-sensitive networking (TSN), where latency variations are as important as low/average latency, has been one of the targeted areas of enhancements ensuring clock synchronization across nodes. Another example is a mechanism to prioritize traffic flows within and between UEs. In general, many of the additions can be viewed as a collection of technical refinements that together significantly enhance NR in the area of URLLC.

Industrial IoT

Industrial Internet of Things (IoT) is the other major new vertical in focus for NR Release 16. While Release 15 can provide very low air-interface latency and high reliability, further enhancements to latency and reliability are introduced in Release 16. This is to enable widening of the set of industrial IoT use cases and to include increased demand for new use cases, such as factory automation, electrical power distribution, and transport industry (including the remote driving use case). Time-sensitive networking (TSN), where latency variations are as important as low/average latency, has been one of the targeted areas of enhancements ensuring clock synchronization across nodes. Another example is a mechanism to prioritize traffic flows within and between UEs. In general, many of the additions can be viewed as a collection of technical refinements that together significantly enhance NR in the area of URLLC.

Positioning

Global Navigation Satellite Systems (GNSS), assisted by cellular networks, have for many years been used for UE positioning. This provides accurate positioning but is typically limited to outdoor areas with satellite visibility. Currently, there is a range of applications requiring accurate positioning not only outdoors but also indoors. For example, the US Federal Communications Commission requires 50 m horizontal accuracy and floor-level vertical accuracy for positioning of emergency calls. Industrial applications, such as manufacturing and logistics, may sometimes require very accurate positioning in the order of a fraction of a meter. Additional positioning methods beyond GNSS are, therefore, of great importance (especially for indoor applications) as well as the enhancements in this area (enabled by Release 16) allowing down to 3 m (for indoors) and 10 m (for outdoors).

Architecture-wise, NR positioning is based on the use of a location server, similar to LTE. The location server collects and distributes information related to positioning (UE capabilities, assistance data, measurements, position estimates, and so on) to the other entities involved in the positioning procedures (base stations and UEs, for example). A range of positioning methods, such as Enhanced Cell Identity (E-CID) and Observed Time Difference of Arrival (OTDOA) — both downlink-based and uplink-based — are used separately or in combination to meet the accuracy requirements for different scenarios.

Future plans

Although the recently completed Release 16 is a significant enhancement to the NR standard, the evolution of the standard will continue for many years [14]. On a high level, these enhancements can be grouped into two categories: general enhancements and improvements of already existing features and new features addressing new verticals and deployment scenarios.

The next step in the evolution is Release 17, work upon which has recently started and scheduled for completion by the end of 2021. The enhancements for the release have been identified (see Table 33). Some of the them (for example, MIMO enhancements and IAB enhancements) are improvements to features present in Release 16, while others (for example, NR beyond 52.6 GHz and non-terrestrial networks) address new scenarios.

Table 3: Enhancements planned for inclusion in Release 17.

Area	Brief description
Supporting NR above 52.6 GHz	Extending NR to the frequency range 52.6 – 71 GHz to allow even more spectrum to be exploited, including the 60 GHz unlicensed band.
Multicast/ broadcast services	Introduction of broadcast/multicast capabilities to NR, primarily targeting V2X, public safety, IP multicast, software delivery, and IoT applications.
Support for non-terrestrial networks	Introducing support for satellites and high-altitude platforms as an additional mean to provide coverage in rural areas.
Dual-Connectivity enhancements	Continuation of Release 16 dual connectivity enhancements, for example improvements to activation/deactivation of secondary cells.
Integrated Access and Backhaul (IAB) enhancements	Enhancements to Release 16 IAB to support (limited) network topology changes and improved duplexing of access and backhaul links.
Coverage enhancements	Study possibilities to provide enhanced (wide-area) coverage.
RAN data collection enhancements	Improved mechanisms for collecting measurement information from Ues and network in order to simplify deployment and enhance the support self-optimized networks.
MIMO enhancements	Improvements to Release 16 MIMO support based on experience from commercial networks.
UE Power-saving enhancements	Enhanced mechanisms for UE power saving, for example improved DRX handling and improvements to blind decoding.
Dynamic spectrum sharing enhancements	Dynamic spectrum sharing, facilitating NR and LTE to share the same carrier, is supported already in Release 15 but will be enhanced with e.g. cross-carrier scheduling and other scheduling enhancements.
Dual-Connectivity enhancements	Continuation of Release 16 dual connectivity enhancements, for example improvements to activation/deactivation of secondary cells.
Sidelink enhancements	Enhancements to the Release 16 sidelink to improve resource efficiency and reliability, as well as studying sidelink relaying.
Industrial IoT enhancements	Improved support for factory automation and URLLC, including synchronization enhancements for time-sensitive networking.

Positioning enhancements	Enhancements to Release 16 positioning mechanism, in particular sub-meter accuracy for industrial use cases.
Anything Reality (XR)	Evaluation of the needs in terms of simultaneously providing very high data rates and low latency in a resource-efficient manner to support various forms of augmented reality (AR) and virtual reality (VR), collectively referred to as XR.
RAN slicing (also relevant for mMTC)	Mechanisms to enable UE fast access to the cell supporting the intended slice, and to support service continuity for intra-radio-access technology handover service interruption.
Small data transmission	Enhancements targeting transmission of small data packets in in-active state to reduce the overhead from connection establishment. Use cases include keep-alive messages, wearables, and various sensors.
Reduced Capability (NR light)	Reduced-capability NR, targeting mid-tier applications such as MTC for industrial sensors, video surveillance, and wearables. Data rates targeted are between NB-IoT/LTE-M data rates and “full” NR data rates. Extended battery lifetime is important at least for some applications.

References

[1] Ericsson, Ericsson Mobility Report, November 2019, available at: https://www.ericsson.com/en/mobility-report

[2] 5G Alliance for Automotive, accessible at: https://www.5gaa.org

[3] 5G Alliance for Connected Industries & Automation, available at: https://www.5g-acia.org

[4] A. Osseiran, J. Monserrat and P. Marsch, 5G Mobile and Wireless Communications Technology, Cambridge University Press, May 2016

[5] E. Dahlman, S. Parkvall and J. Sköld, 5G NR: The Next Generation Wireless Access Technology, Academic Press, August 2018

[6] ITU-R, Framework and overall objectives of the future development of IMT for 2020 and beyond Recommendation ITU-R M.2083-0, 2015, available at: http://mioa.gov.mk/sites/default/files/pbl_files/documents/strategies/recommendation_itu-r_m.2083-0.pdf

[7] Ericsson, Cellular IoT in the 5G era, White Paper, January 2020, available at https://www.ericsson.com/en/reports-and-papers/white-papers/cellular-iot-in-the-5g-era

[8] J. Sachs, K. Wallstedt, F. Alriksson and G. Eneroth, Boosting smart manufacturing with 5G wireless connectivity, Ericsson Technology Review, February 2019, available at: https://www.ericsson.com/en/ericsson-technology-review/archive/2019/boosting-smart-manufacturing-with-5g-wireless-connectivity

[9] J. Farkas, B. Varga, G. Miklós and J. Sachs, 5G-TSN integration meets networking requirements for industrial automation, Ericsson Technology Review, August 2019, available at: https://www.ericsson.com/en/ericsson-technology-review/archive/2019/5g-tsn-integration-for-industrial-automation

[10] T. Lohmar, A. Zaidi, H. Olofsson and C. Boberg, Driving transformation in the automotive and road transport ecosystem with 5G, Ericsson Technology Review, Sep. 2019, available at:

https://www.ericsson.com/4a61e3/assets/local/reports-papers/ericsson-technology-review/docs/2019/etr-transforming-transportation-with-5g.pdf

[11] 5G-ACIA, 5G for Automation in Industry: Primary use cases, functions and service requirements, White Paper, July 2019, available at: https://www.5g-acia.org/publications/5g-for-automation-in-industry-white-paper/

[12] Ericsson, IMT-2020 self-evaluation: On eMBB spectral efficiency — - NR, 3GPP Document R1-1809268, 2018

[13] Ericsson, IMT-2020 self-evaluation: Mobility evaluations for NR, 3GPP document R1-1809265, 2018

[14] J. Peisa, Patrik P., Stefan P., Erik D., Asbjörn G., Christian H., Dirk G. , 5G evolution: 3GPP Releases 16&17 overview, Ericsson Technology Review, March 2020. Available at : https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/5g-nr-evolution

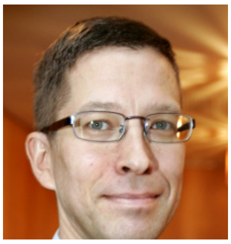
Glossary

3GPP (3rd Generation Partnership Project)
5GAA (5G Automotive Association)
5G-ACIA (5G Alliance for Connected Industries and Automation)
5GC (5G Core)
CP (Cyclic prefix)
DRX (discontinuous reception)
eMBB (enhanced Mobile BroadBand)
E-CID (Enhanced Cell Identity)
EPC (Evolved Packet Core)
FDD (Frequency Division Duplex)
gNB (5G gNodeB)
GNSS (Global Navigation Satellite Systems)
IAB (integrated Access and Backhaul)
ICT (Information and Communication Technology)
IMT-2020 (International Mobile Telecommunications 2020)
ITS (Intelligent Transportation Systems)
ITU (International Telecommunication Union)
IoT (Internet of Things)
LTE (Long-Term Evolution)
LTE-M (LTE for Machine type communication)
MAC (MediumAaccess Control)
mmW (millimeter Wave)
NB-IoT (NarrowBand IoT)
MTC (Machine-Type Communication)
NPRM (Notice of Proposed RuleMaking)
NR (New Radio)
NSA (Non-StandAlone)
OFDM (Orthogonal frequency-division multiplexing)
OT (Operation Technology)
OTDOA (Observed Time Difference of Arrival)
PSM (Power Saving Mode)
RLC (radio link control)
SA (StandAlone)
TDD (Time Division Duplex)

TSN (Time-Sensitive Network)
URLLC (Ultra-Reliable Low-Latency Communication)
V2X (vehicle-to-anything)
WRC (World Radiocommunication Conference)

Author biographies

Afif Osseiran is director of Industry Engagements & Research at Ericsson's headquarter in Stockholm. He is also vice chairman of the 5G-ACIA alliance and a senior member of IEEE. His main responsibility is to bridge insights, tactics and strategies between technology, research/standardization and industrial sectors such as manufacturing. He holds a doctorate degree from the Royal Institute of Technology (KTH), Sweden, a master's degree from École Polytechnique de Montreal, and another one from INSA Rennes, France. Since 1999 he has held several positions at Ericsson in various units. From 2014 to 2017, he was director of radio communications at the Ericsson CTO office with the main responsibility to develop the strategy for 5G and Industrial IoT. From 2012 to 2014, he managed METIS, the EU 5G flagship project. He co-authored the first comprehensive book on 5G with Cambridge University Press and two books on IMT-Advanced with Wiley.



Stefan Parkvall is a senior expert Researcher at Ericsson Research working with 5G and future radio access. He is one of the key persons in the development of HSPA, LTE and NR radio access and has been deeply involved in 3GPP standardization for many years. Dr Parkvall is a fellow of the IEEE and served as an IEEE Distinguished lecturer 2011-2012. He is co-author of the popular books "3G Evolution – HSPA and LTE for Mobile Broadband", "HSPA evolution – the Fundamentals for Mobile Broadband", "4G – LTE/LTE-Advanced for Mobile Broadband", "4G, LTE Advanced Pro and the Road to 5G", and "5G NR – The Next Generation Wireless Access". Dr. Parkvall has more than 1000 patents in the area of mobile communication. In 2005, he received the Ericsson "Inventor of the Year" award, in 2009 the Swedish government's Major Technical Award for contributions to the success of HSPA, and in 2014 he and Ericsson colleagues were among the finalists for the European Inventor Award for their contributions to LTE. Dr Parkvall holds a Ph.D. in electrical engineering from the Royal Institute of Technology (KTH) in Stockholm, Sweden. Previous positions include assistant professor in communication theory at KTH, and visiting researcher at University of California, San Diego, USA.

Patrik Persson is a Principal Researcher and joined Ericsson Research in 2007. Currently he holds a position as the program manager for the Ericsson Research program on 5G evolution and 6G being responsible for the research activities including 3GPP RAN standardization, proprietary evolution of 5G and the 6G research activities including both RAN and CN aspects. From 2014 to 2019. Patrik was responsible for the Ericsson back-office work in the 3GPP RAN standardization of 4G and 5G. Previously, he has been working with, and leading the work, in the areas of antennas and propagation as well as proprietary development of LTE within Ericsson Research. Patrik holds a Ph.D. (2002) and docent degree (2011) in electrical engineering from the Royal Institute of Technology (KTH) in Stockholm, Sweden. Patrik has authored several publications and books in the area of antennas and electromagnetics. He is also the recipient of the 2002 R. W. P. King paper award given by IEEE Transactions of Antennas and Propagation.

Ali Zaidi is a Strategic Product Manager for Cellular IoT at Business Area Networks, Ericsson. He received an MSc and a PhD in Telecommunications from KTH Royal Institute of Technology, Stockholm, Sweden, in 2008 and 2013, respectively. Since 2014, he has been working with technology and business development of 4G and 5G radio access at Ericsson. Ali is currently responsible for LTE-M, URLLC, Industrial IoT, V2X and local industrial networks. He is also Head of IoT Competence at Ericsson. Ali has co-authored more than 50 peer-reviewed research publications and two books, filed over 20 patents and made several 3GPP and 5G-PPP contributions, spanning communications, control and automation technologies.

Sverker Magnusson is Head of Spectrum and Technology Regulations at Ericsson. He holds a M.Sc. in Engineering Physics from the Royal Institute of Technology in Stockholm and a Ph.D. in Operations Research from Cornell University. He has been employed by Ericsson since 1995, holding various positions at Ericsson Research and now at the CTO office. Over the last decade he has focused on spectrum matters with involvement in numerous groups of CEPT and ITU-R

Kumar Balachandran is an Expert in Wireless Communications Networks and has been with Ericsson Research since 1995. He has a BE (Hons) in Electronics and Communications Engineering from the Regional Engineering College, Tiruchirappalli, acquired in 1986 and holds an M.S. and PhD in Computer and Systems Engineering from Rensselaer Polytechnic Institute in Troy, NY, awarded in 1988 and 1992 respectively. Kumar has worked on a variety of areas in mobile communications spanning the Physical Layer, signal processing, radio resource management, spectrum sharing, protocol design, and systems engineering spanning all five generations of mobile cellular technologies. His recent contributions have been in the area of shared spectrum and he has been a prominent contributor to the specification of the Citizens Broadband Radio Service (CBRS) in the WInnForum and the CBRS Alliance. He is currently working on research problems pertaining to radio resilience and system reliability. He is active in working with Ericsson's technology strategy and takes an interest in competitive analysis of mobile technology. He has served as a technical expert on the FCC's Technological Advisory Council on spectrum topics, receiver performance, and 5G&IoT in previous years. He has contributed to Ericsson's outreach to the FCC and the NTIA on several occasions with technical arguments favoring the release of spectrum for use by the mobile industry. He has served as panelist and invited speaker at several prominent conferences, is well published, has contributed to several books, and has been named on 100 issued US patents as inventor.

Acknowledgments
The authors would like to thank Marie Hogan and Håkan Ohlsén for their help.