NR – The new 5G radio-access technology

Erik Dahlman and Stefan Parkvall
Ericsson AB
Stockholm, Sweden

***Abstract*—This paper provides a detailed overview of the key technology features of the new 5G/NR radio-access technology, the first release of which has recently been published by 3GPP**

Keywords—5G radio access, NR

I. Introduction

NR ("New Radio") is the new 5G radio-access technology being developed by 3GPP. The technical work on NR was initiated in the spring of 2016 with the first release, being part of the 3GPP release 15, of the NR specifications finalized by the end of 2017. This first release is limited to non-standalone NR operation, implying that NR devices rely on LTE for initial access and mobility. The final release-15 specifications, to be available in June 2018, will also support stand-alone NR operation. The difference between stand-alone and non-stand-alone operation is primarily affecting higher layers and the interface to the core network; the basic radio technology is the same in both cases.

This paper will give a detailed overview of the NR radio-access technology with focus on the key features that distinguish it from 4G LTE.

II. NR key features

A. Higher frequencies and spectrum flexibility

One key feature of NR is a substantial expansion of the range of spectrum in which the radio-access technology can be deployed with support for operation from below 1 GHz up to roughly 50 GHz, that is, well into the mm-wave bands, already in the first release.

Operation at mm-wave frequencies offers the possibility for large amounts of spectrum and very wide transmission bandwidths, enabling very high traffic capacity and extreme data rates. However, higher frequencies are also associated with higher radio-channel attenuation. Although this can partly be compensated for by means of advanced multi-antenna transmission/reception, a substantial coverage disadvantage remains, especially in non-line-of-sight and outdoor-to-indoor propagation conditions. Thus, operation in lower frequency bands will remain important also in the 5G era. Furthermore, joint operation in lower *and* higher spectrum can provide substantial benefits. A higher-frequency layer, with access to a large amount of spectrum can provide service to a large fraction of the mobile devices despite the more limited coverage. This will reduce the load on the more bandwidth-constrained lower-frequency spectrum, allowing this to focus on devices in bad coverage situations. [1]

B. Ultra-lean design

An issue with current mobile-communication technologies is the amount of transmissions carried by network nodes regardless of the amount of user traffic. Such "always-on" transmissions include, for example, signals for base-station detection, broadcast of system information, and always-on reference signals for channel estimation. Under 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. However, in very dense networks deployed for high peak data rates, 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 always-on transmissions have two negative impacts:

- they impose an upper limit on the achievable network energy performance, and
- they cause interference to other cells, thereby reducing the achievable data rates.

The *ultra-lean-design* principle aims at minimizing the always-on transmissions, thereby enabling higher network energy performance and higher achievable data rates.

C. Transmission scheme, bandwidth parts, and frame structure

Similar to LTE, NR is based on OFDM. However, unlike LTE, where DFT-precoded OFDM is the sole uplink transmission scheme, NR uses conventional (non-precoded) OFDM as the baseline uplink transmission scheme due to less complex receiver structure, especially in combination with uplink spatial multiplexing[1] and a general desire to have the same transmission scheme in uplink and downlink. Nevertheless, uplink DFT-precoding can be used as a complement to enable higher device power-amplifier efficiency in case of single-layer transmission.

---

[1] Note that the first releases of LTE did not support uplink spatial multiplexing

978-1-5386-6355-4/18/$31.00 ©2018 IEEE

To support a wide range of deployment scenarios, from very large cells operating below 1 GHz up to mm-wave deployments with very wide spectrum allocations, NR supports a flexible numerology with subcarrier spacings ranging from 15 kHz up to 120 kHz with a proportional change in cyclic-prefix duration. A smaller subcarrier spacing has the benefit of providing a relatively longer cyclic prefix with a reasonable overhead, while higher subcarrier spacings are needed to handle, for example, the increased phase noise at higher carrier frequencies. A carrier may consist of up to 3300 subcarriers, resulting in a maximum carrier bandwidths of 50/100/200/400 MHz for subcarrier spacings of 15/30/60/120 kHz, respectively. To realize even larger bandwidths, carrier aggregation can be used.

Although the NR physical-layer specification is band-agnostic, not all supported numerologies are relevant for all frequency bands. For each frequency band, radio requirements are therefore defined for a subset of the supported numerologies as illustrated in Figure II-1. Currently, there is no spectrum identified for NR between 6 GHz and 24.25 GHz but requirements can easily be added at a later stage if such spectrum becomes available.

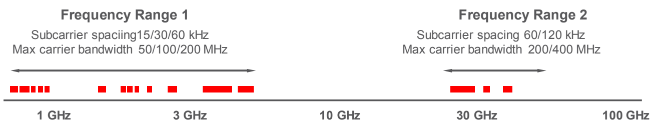

Figure II-1 NR spectrum

In LTE, all devices support the maximum LTE carrier bandwidth of 20 MHz. However, given the very wide maximum bandwidth, it is not reasonable to require all NR devices to support the maximum NR carrier bandwidth. Furthermore, NR allows for device-side *receiver-bandwidth adaptation* as a mean to reduce the device energy consumption. Bandwidth adaptation refers to the use of a relatively modest bandwidth for monitoring control channels and receiving medium data rates, and dynamically open up a wideband receiver only when needed to support very high data rates.

To handle this, NR defines *bandwidth parts* that indicate the bandwidth over which a device is currently assumed to receive transmissions of a certain numerology.

The NR time-domain structure is illustrated in Figure II-2, with a 10 ms radio frame divided into ten 1 ms subframes. A subframe is, in turn, divided into slots consisting of 14 OFDM symbols each, that is, the duration of a slot in milliseconds depends on the numerology. Thus, for the 15 kHz subcarrier spacing, an NR slot has the same structure as an LTE subframe which is beneficial from a coexistence perspective. Since a slot is defined as a fixed number of OFDM symbols, a higher subcarrier spacing leads to a shorter slot duration. In principle this could be used to support lower-latency transmission, but as the cyclic prefix also shrinks when increasing the subcarrier spacing, it is not a feasible approach in all deployment scenarios. Therefore, NR supports a more efficient approach to low latency by allowing for transmission over a fraction of a slot, sometimes referred to as "mini-slot" transmission. Such transmissions can also preempt an already ongoing slot-based transmission to another device, allowing for immediate transmission of data requiring very low latency.

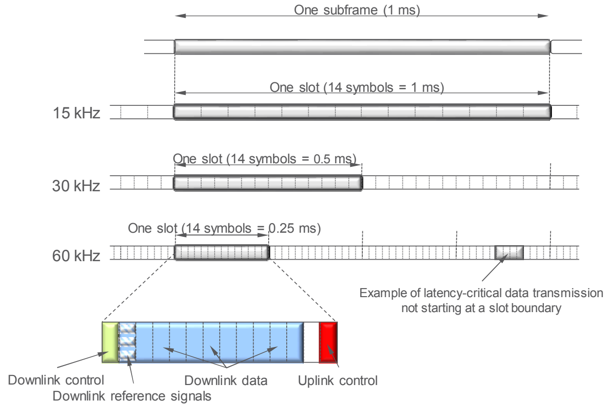

Figure II-2 NR time-domain structure

Having the flexibility of starting a data transmission not only at the slot boundaries is also useful when operating in unlicensed spectrum. In unlicensed spectrum the transmitter is typically required to ensure that the radio channel is not occupied by other transmissions prior to starting a transmission, a procedure commonly known as "listen-before-talk". Once the channel is found to be available it is beneficial to start the transmission immediately, rather than wait until the start of the slot, in order to avoid that some other transmitter initiates a transmission on the channel.

Operation in the mm-wave domain is another example of the usefulness of "mini-slot" transmissions as the available bandwidth in such deployments often is very large and even a few OFDM symbols can be sufficient to carry the available payload. This is especially beneficial in conjunction with *analog beamforming* where transmissions to multiple devices in different beams must be separated in time.

D. Duplex schemes

The duplex scheme is typically given by the spectrum allocation at hand. For lower frequency bands, allocations are often paired, implying frequency-division duplex (FDD). At higher frequency bands, unpaired spectrum allocations are increasingly common, calling for time-division duplex (TDD). Given the significantly higher carrier frequencies supported by, support for unpaired spectrum is thus even more pronounced in NR compared to LTE.

In contrast to LTE, NR can operate in both paired and unpaired spectrum using a common frame structure. The basic frame structure is designed to support both half-duplex and full-duplex operation. In half duplex, the device cannot transmit and receive at the same time. Examples hereof are TDD and half-duplex FDD. In full-duplex operation, on the other hand, simultaneous transmission and reception is possible with FDD as a typical example.

As already mentioned, TDD increases in importance when moving to higher frequency bands where unpaired spectrum allocations are more common. These frequency bands are less useful for wide-area coverage with very large cells but are highly relevant for local-area coverage with smaller cell sizes. Furthermore, some of the problematic interference scenarios in wide-area TDD networks are less pronounced in local area deployments with lower transmission power and below-rooftop antenna installations. At the same time, in such denser deployments with smaller cell sizes, the per-cell traffic variations are more rapid compared to large-cell deployments with a large number of active devices per cell. To address such scenarios NR supports *dynamic TDD* where the allocation of time-domain resources can be dynamically assigned to downlink or uplink depending on the instantaneous traffic needs. Dynamic TDD enables following rapid traffic variations which are particularly pronounced in dense deployments with a relatively small number of users per cell.

The basic approach to dynamic TDD is for the device to monitor for downlink control signaling and follow the scheduling decisions. If the device is instructed to transmit, it transmits in the uplink, otherwise it will attempt to receive any downlink transmissions. The uplink-downlink allocation is then completely under control of the scheduler and any traffic variations can be dynamically tracked. There are deployment scenarios where dynamic TDD may not be useful, but it is much simpler to restrict the dynamics of a dynamic scheme in those scenarios when needed rather than trying to add dynamics to a fundamentally semi-static design as LTE. For example, in a wide-area macro network with above-rooftop antennas, the inter-cell interference situation requires coordination of the uplink-downlink allocation between the cells. In such situations, a semi-static allocation is appropriate with operation along the lines of LTE.

E. Low latency support

Support for low latency is an important part of NR that impacts many of the design details. One example is the use of “front-loaded” reference signals and control signaling as illustrated in Figure II-2 above. By locating the reference signals and downlink control signaling carrying scheduling information at the beginning of the transmission and not using time-domain interleaving across OFDM symbols, a device can start processing the received data immediately without prior buffering, thereby minimizing the decoding delay.

The requirements on the device (and network) processing times are significantly tighter in NR compared to LTE. As an example, a device is assumed to be able to respond with a HARQ acknowledgement approximately one slot (or even less for some device categories) after receiving downlink data. The time from a scheduling grant to uplink data transfer is in the same range.

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. This is especially important in the uplink direction as the device may only have a few OFDM symbols after receiving the uplink grant until the transmission should take place.

F. Scheduling and data transmission

One key characteristic of mobile radio communication is the large and typically rapid variations in the instantaneous channel conditions stemming from frequency-selective fading, distance-dependent path loss, and random interference variations due to transmissions in other cells and by other devices. Instead trying to combat these variations, they can be exploited through *channel-dependent scheduling* where the time-frequency resources are dynamically shared between users. On a high level, the NR scheduling framework is similar to the one in LTE. The scheduler, residing in the base station, takes scheduling decisions based on channel-quality reports obtained from the devices. It also takes different traffic priorities and quality-of-service requirements into account when forming the scheduling decisions.

Each device monitors a number of *physical downlink control channels* (PDCCHs), typically once per slot although it is possible to configure more frequent monitoring to support traffic requiring very low latency. Upon detection of a valid PDCCH, the device follows the scheduling decision and receives (or transmits) one unit of data (a transport block).

Given the very high data rates supported by NR, channel-coding for data transmission is based on low-density parity-check (LDPC) codes [2]. LDPC codes are attractive from an implementation perspective, especially at higher code rates where they can offer a lower complexity than the Turbo codes used in LTE.

Hybrid automatic repeat-request (ARQ) retransmission using incremental redundancy is used where the device reports the outcome of the decoding operation to the base station. In case of erroneously received data, the network can retransmit the data and the device combines the soft information from multiple transmission attempts. However, retransmitting the entire transport block may sometimes be inefficient. NR therefore supports retransmissions on a finer granularity known as *code-block group* (CBG). This can also be useful when handling *preemption*. An urgent transmission to a second device may use only one or a few OFDM symbols and therefore cause high interference to the first device in some OFDM symbols only. In this case it may be sufficient to retransmit the interfered CBGs only and not the whole data block. Handling of preempted transmission can be further assisted by the possibility to indicate to the first device the impacted time-frequency resources such that it can take this information into account in the reception process.

Although dynamic scheduling is the basic operation of NR, operation without a dynamic grant can be configured. In this case, the device is configured in advance with resources that can be used for uplink data transmission (or downlink data reception). Once a device has data available it can immediately commence uplink transmission without going through the scheduling request-grant cycle, thereby enabling lower latency.

G. Control channels

Operation of NR requires a set of control channels to carry the scheduling decisions in the downlink and to provide feedback information in the uplink.

Downlink control channels are known as PDCCHs (*Physical Downlink Control Channel*). One major difference compared to LTE is the more flexible time-frequency structure of downlink control channels where PDCCHs are transmitted in one or more *control resource sets* (CORESETs) which can be configured to span only part of the carrier bandwidth.

Another major difference compared to LTE is the support for beamforming of the control channels, which has required a different reference-signal design with each control channel having its own dedicated reference signal.

Uplink control information such as hybrid-ARQ acknowledgements, channel-state feedback for multi-antenna operation, and scheduling request for uplink data awaiting transmission, are transmitted using the *physical uplink control channel* (PUCCH). There are several different PUCCH formats, depending on the amount of information and the duration of the PUCCH transmission. The *short PUCCH* is transmitted in the last one or two symbols of a slot and can support very fast feedback of hybrid-ARQ acknowledgements in order to realize so-called self-contained slots where the delay from the end of the data transmission to the reception of the acknowledgement from the device is in the order of an OFDM symbol, corresponding to a few ten microseconds depending on the numerology used. This can be compared to almost 3 ms in LTE and is yet another example on how the focus on low latency in NR has impacted the design. For situations when the duration of the short PUCCH is too short to provide sufficient coverage, there are also possibilities for longer PUCCH durations.

H. Beam-centric design and multi-antenna transmission

Support for a massive number of steerable antenna elements for both transmission and reception is a key feature of NR. At higher frequency bands, the large number of antennas 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. Channel-state information (CSI) for operation of massive multi-antenna schemes can be obtained by feedback of CSI reports based on transmission of CSI reference signals in the downlink as well as using uplink measurements exploiting channel reciprocity.

To provide implementation flexibility, NR is deliberately supporting functionality to support analog beam-forming as well as digital precoding/beam-forming. At high frequencies, analog beamforming, where the beam is shaped after digital-to-analog conversion, may be necessary from an implementation perspective, at least initially. Analog beamforming results in the constraint that a receive or transmit beam can only be formed in one direction at a given time instant and requires beam-sweeping where the same signal is repeated in multiple OFDM symbols but in different transmit beams. By having beam-sweeping possibility, it is ensured that any signal can be transmitted with a high gain beamformed transmission to reach the entire intended coverage area.

Signaling to support beam-management procedures is specified, such as indication to the device to assist selection of a receive beam (in case of analog receive beamforming) to be used for data and control reception. For a large number of antennas, beams are narrow and beam tracking can fail, therefore beam-failure/recovery procedures have been defined by means of which a device detecting beam failure can trigger a beam-recovery procedure to rapidly re-establish connectivity. Additionally, uplink-centric and reciprocity-based beam management is possible by utilizing uplink signals.

With the use of a massive number of antenna elements also for lower frequency bands, the possibility to separate users spatially increases both in uplink and downlink but requires that the transmitter has channel knowledge. For NR, extended support for such multi-user spatial multiplexing is introduced, either by using a high-resolution channel-state-information feedback using a linear combination of DFT vectors, or uplink sounding reference signals targeting the utilization of channel reciprocity.

Twelve orthogonal demodulation reference signals are specified for multi-user MIMO transmission purpose while an NR device can maximally receive eight MIMO layers in the downoink and up to four layers in the uplink. Moreover, additional configuration of a phase tracking reference signal is introduced in NR since the increased phase noise power at high carrier frequency bands otherwise will degrade demodulation performance for higher-order modulation constellations such as. 64 QAM.

I. Initial access

Initial access is the procedures by which a device finds a cell to camp on, receive the necessary system information, and to request a connection through random access. The basic structure of NR initial access is similar to the corresponding functionality of LTE [3] with a *Primary Synchronization Signal* (PSS) and *Secondary Synchronization Signal* (SSS) used to find, synchronize to, and identify a network, and a *Physical Broadcast Channel* (PBCH) that carries a minimum amount of system information. In the context of NR, the PSS, SSS, and PBCH are jointly referred to as a *Synchronization Signal* (SS) *block*.

Similar to LTE there is also a four-stage random-access procedure, commencing with the uplink transmission of a *random-access preamble*

However, there are some important differences between LTE and NR in terms of initial access. These differences mainly come from the ultra-lean principle and the beam-centric design, both which has required significant changes to the initial access procedures compared to LTE.

To enable higher NR network energy performance in line with the ultra-lean principle, the SS block is, by default, transmitted once every 20 ms, compared to every 5 ms for the corresponding signals of LTE. Due to the longer period between consecutive SS blocks, compared to the corresponding signals/channels in LTE, a device searching for NR carriers must dwell on each possible frequency for a longer time. To reduce the overall search time while keeping the device complexity comparable to LTE, NR supports a *sparse frequency raster* for SS block. This implies that the possible frequency-domain positions of the SS block could be significantly sparser, compared to the possible positions of an NR carrier (the *carrier raster*). The sparse SS-block raster enables significantly reduced time for initial cell search, at the same time as the network energy performance can be significantly improved due to the longer SS-block period.

Network-side beam-sweeping is supported for both downlink SS-block transmission and uplink random-access reception as a means to improve coverage, especially in case of operation at higher frequencies. It is important to realize that beam sweeping is a *possibility* enabled by the NR design. It does not imply that it must be used. Especially at lower carrier frequencies, beam sweeping may not be needed.

J. *Interworking and LTE coexistence*

As it is difficult to provide full coverage at higher frequencies, interworking with systems operating at lower frequencies is important. In particular, a coverage imbalance between uplink and downlink is a common scenario. The higher transmit power for the base station compared to the mobile device results in the downlink achievable data rates often are bandwidth limited, making it more relevant to operate the downlink in higher spectrum where wider bandwidth may be available. In contrast, the uplink is more often power limited, reducing the need for wider bandwidth. Instead, higher data rates may be achieved on lower-frequency spectrum, despite less available bandwidth, due to less radio-channel attenuation.

Through interworking, a high-frequency NR system can complement a low-frequency system. The lower frequency system can be either NR or LTE, and NR supports interworking with either of these. The interworking can be realized at different levels, including intra-NR carrier aggregation, dual connectivity[2] with a common packet data convergence protocol (PDCP) layer, and handover.

However, the lower frequency bands are often already occupied by current technologies, primarily LTE. Furthermore, additional low-frequency spectrum is planned to be deployed with LTE in a relatively near future. LTE/NR *spectrum co-existence*, that is, the possibility for an operator to deploy NR in the same spectrum as an already existing LTE deployment has therefore been identified as a way to enable early NR deployment in lower frequency spectrum without reducing the amount of spectrum available to LTE.

Two co-existence scenarios were identified in 3GPP and guided the NR design.

In the first scenario (left part of Figure II-3) there is LTE/NR co-existence in both downlink and uplink. Note that this is relevant for both paired and unpaired spectrum although paired spectrum is used in the illustration.

In the second scenario (right part of Figure II-3) there is co-existence only in the uplink transmission direction, typically within the uplink part of lower-frequency paired spectrum, with NR downlink transmission taking place in spectrum dedicated to NR, typically at higher frequencies. This scenario attempts to address the uplink-downlink imbalance discussed above. NR supports a *supplementary uplink* (SUL) to specifically handle this scenario.

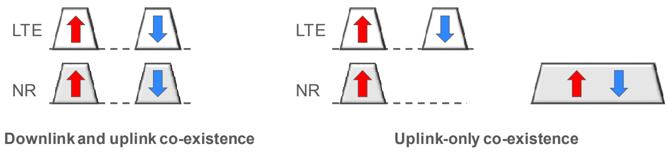

Figure II-3

The possibility for an LTE-compatible NR numerology based on 15 kHz sub-carrier spacing, enabling identical time/frequency resource grids for NR and LTE, is one of the fundamental tools 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 such as cell-specific reference signals, CSI-RS, and the signals/channels used for LTE initial access. *Reserved resources,* introduced for forward compatibility, can also be used to further enhance NR-LTE co-existence. It is possible to configure reserved resources matching the cell-specific reference signals in LTE, thereby enabling an NR-LTE overlay in the downlink.

REFERENCES

[1] S. Parkvall, E. Dahlman, A. Furuskär, M. Frenne, " NR: The New 5G Radio Access Technology", IEEE Communications Standards Magazine, December 2017, pp. 24 – 30

[2] T. Richardson, R. Urbanke, "Modern coding theory", Cambridge University Press, 2008

[3] E. Dahlman, S. Parkvall, J. Sköld, "4G LTE-Advanced Pro and the Road to 5G", Elsevier, 2016

[2] In release 15, dual connectivity is only supported between NR and LTE. Dual connectivity within NR and NR will be added in a later release.