Οικονομικό Πανεπιστήμιο Αθηνών
Τμήμα Πληροφορικής

Ευφυή Κινητά Δίκτυα: 5G

Εαρινό Εξάμηνο 2023-24

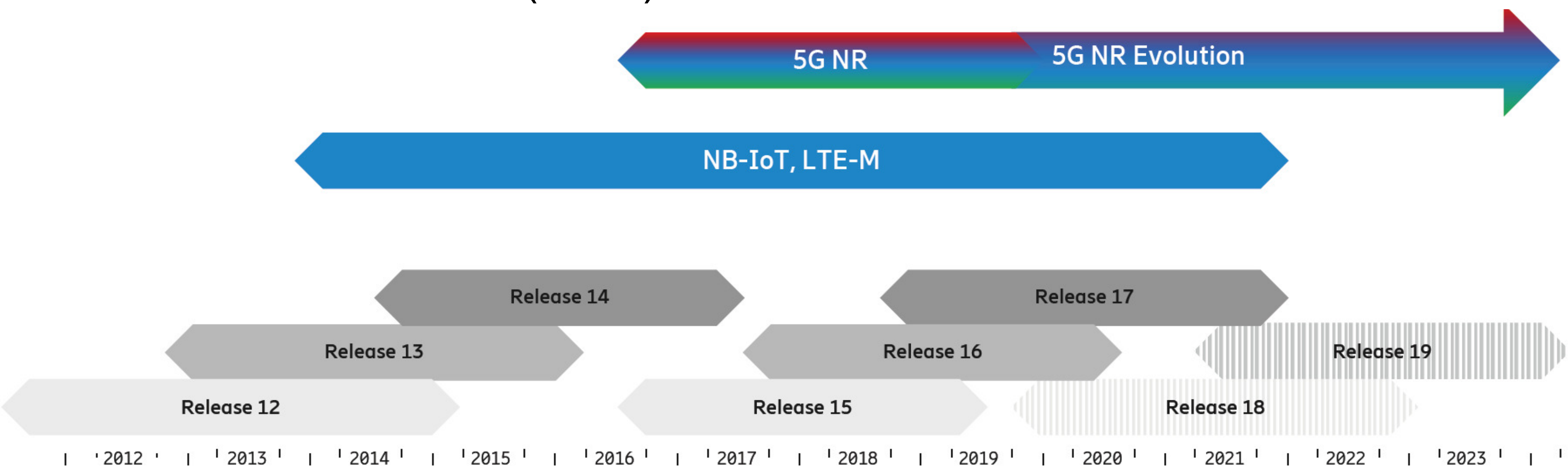
Γιάννης Θωμάς

(Βασισμένο στο “5G wireless access an overview” White-paper της Ericsson)

https://www.ericsson.com/en/reports-and-papers/white-papers/5g-wireless-access-an-overview

5G – timeline

- a.k.a. **New Radio (NR)**
- standardized by 3GPP (the 3rd Generation Partnership Project)
  - Rel 13 – Long Term Evolution for Machines: LTE-M (2016)
  - Rel 14 – NB-IoT (2017)
  - Rel 15 – 5G (2019)



https://www.ericsson.com/en/reports-and-papers/white-papers/5g-wireless-access-an-overview

5G goals and challenges

- 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.
- Lower energy per delivered bit than previous cellular networks.
- Sheer number of devices challenges signaling provisioning and connection management.

5G applications – eMBB

3 main use-case families:

- Enhanced Mobile Broadband (**eMBB**),
  - **human**-centric use cases: mobile telephony and media delivery.
  - enables large volumes of data transfer and extreme data rates.
  - Typical usage: mobile phones and mobile PCs/tablets.
- massive machine type communications (mMTC) and
- ultra-reliable low latency communications (URLLC)
  - also called critical machine type communications (cMTC)

5G applications – mMTC

3 main use case families:

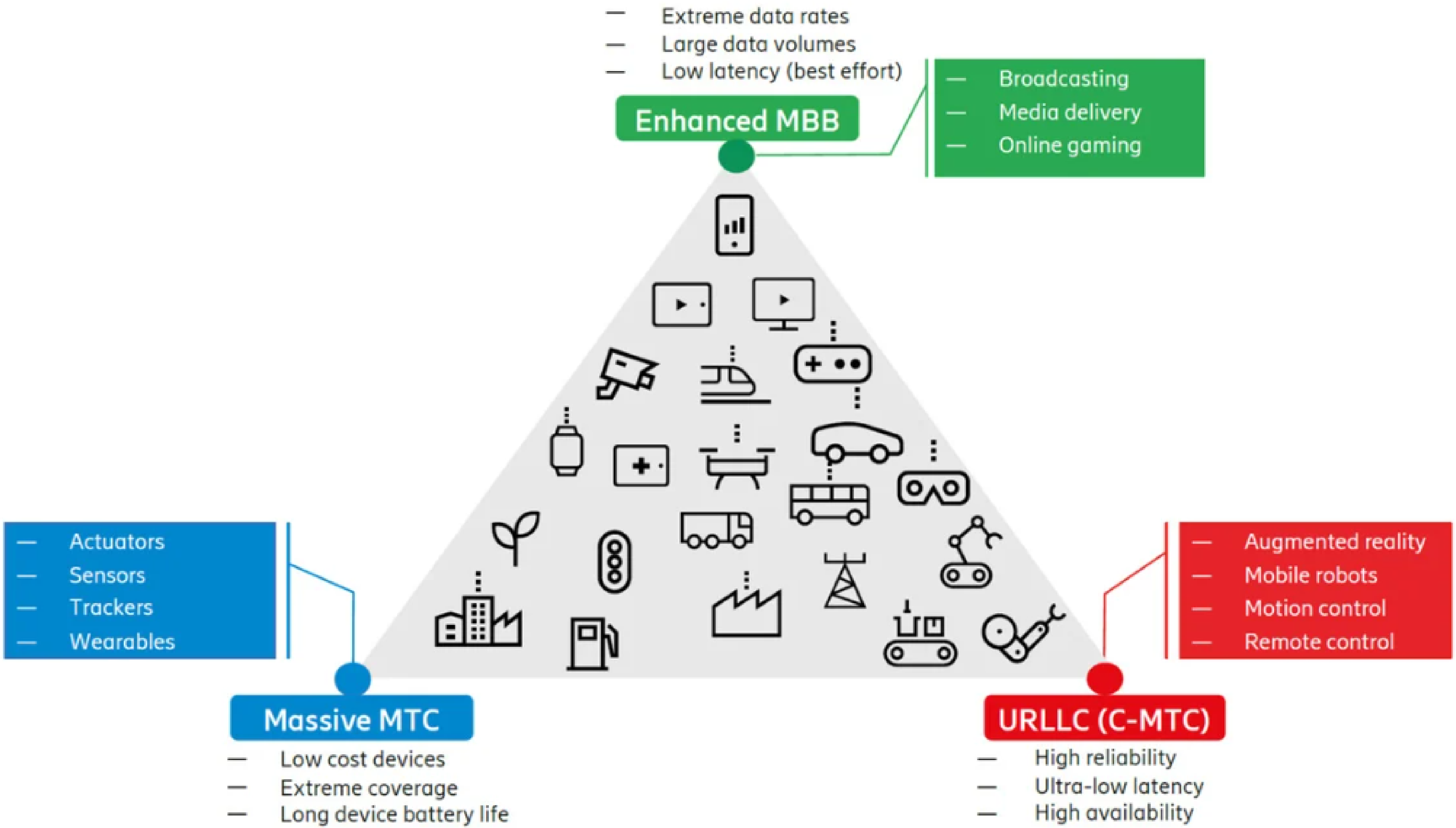
- enhanced mobile broadband (eMBB)
- massive Machine Type Communications (**mMTC**) and
  - **Machine**-centric use cases
  - Connects massive number of **low complexity narrow-bandwidth devices** that infrequently send or receive small volumes of data.
  - Devices in challenging radio conditions requiring coverage extension capabilities
  - Devices may solely rely on **battery** power supply.
  - use cases: low-cost sensors, meters, actuators, trackers, and wearables.
- ultra-reliable low latency communications (URLLC)
  - also called critical machine type communications (cMTC)

5G applications – URLLC

3 main use case families:

- enhanced mobile broadband (eMBB)
- massive machine type communications (mMTC) and
- Ultra-Reliable Low Latency Communications (**URLLC**)
  - also called critical machine type communications (cMTC)
  - Machine-centric
  - use cases: stringent requirements on reliability and latency
    - 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.

5G applications – all together



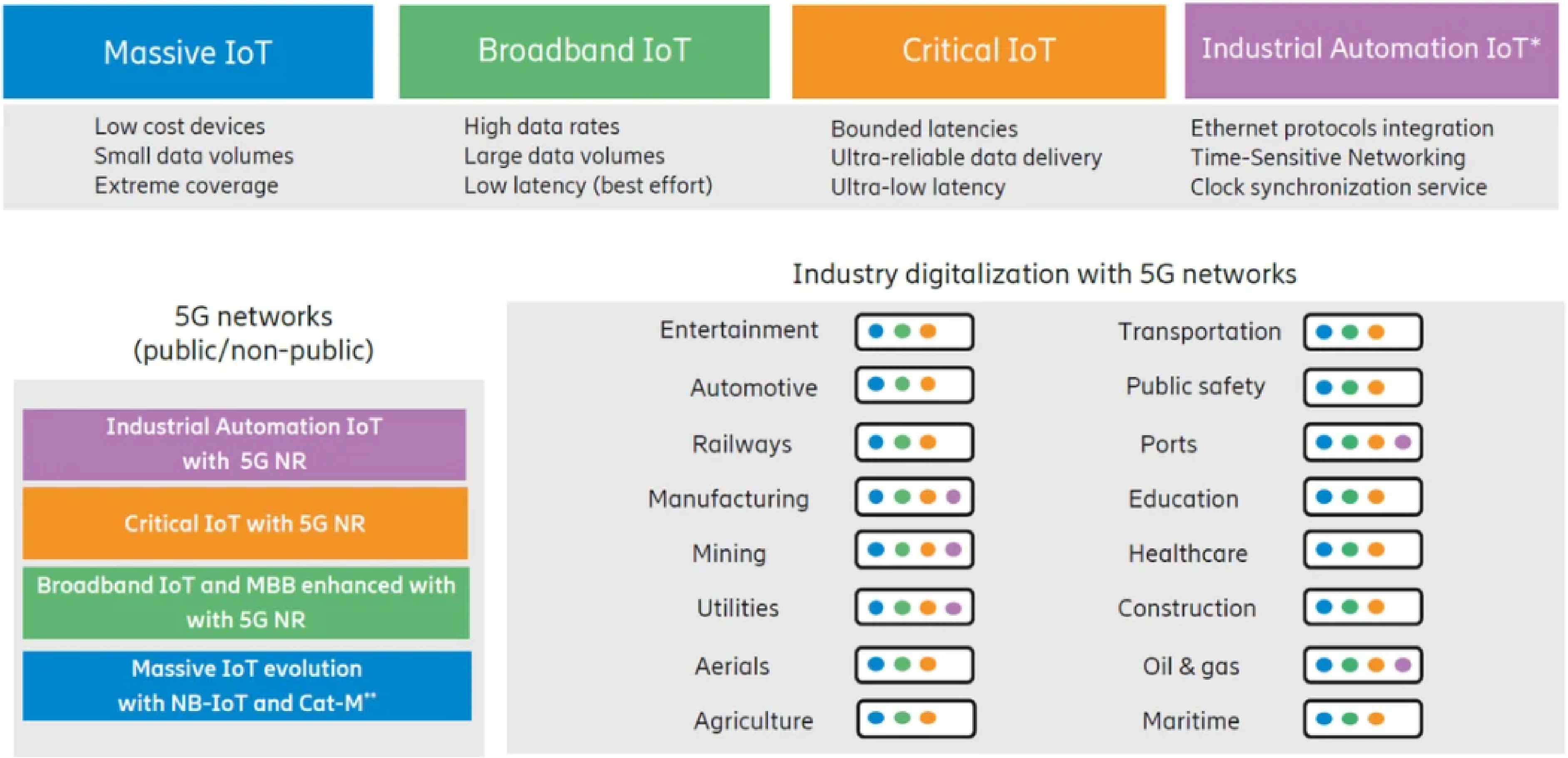
5G and IoT

There are over one billion cellular Internet of Things (IoT) connections in 2020 and Ericsson forecasts around 5 billion connections by 2025. https://www.ericsson.com/en/reports-and-papers/mobility-report

4 IoT connectivity segments:

- Massive IoT,
- Broadband IoT,
- Critical IoT, and
- Industrial Automation IoT

5G and IoT (cont.)



5G eMBB Reqs

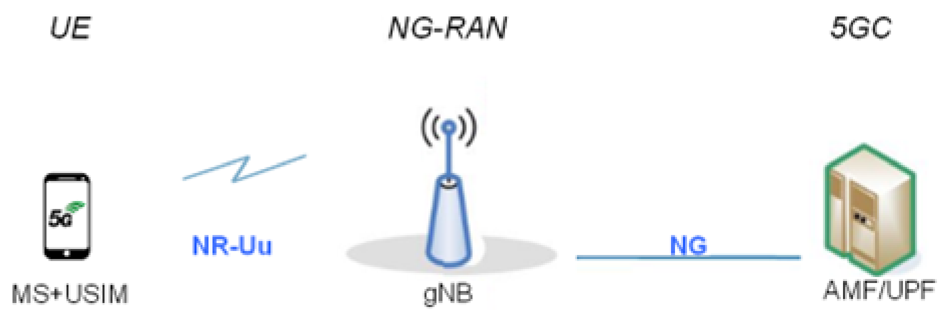
- **Peak data rate:** 20 / 10 Gbps (*down/up link)
  - Carrier aggregation (16 x 17.5 Gbps in 400 MHz-wide carrier), higher modulation, Massive SU MIMO
- **Peak spectral efficiency:** 30 / 15 bps/Hz
  - Higher modulation (256QAM), Massive MIMO
- **User-experienced data rate** (fifth percentile user data rate): 100 / 50 Mbps
  - Multi-antenna
- **Cell average spectral efficiency:** Scenario dependent
  - Multi-antenna
- **Fifth percentile user spectral efficiency:** Scenario dependent
  - Multi-antenna
- **Area traffic capacity:** 10 Mbps/m2 (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
- **Control plane latency:** 10 ms
  - Mini-slot, scalable numerology, fast HARQ, pre-emption, RRC inactive

5G URLLC and mMTC Reqs

- **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:** 1,000,000 devices per sq km2
- **Reliability:** 99.999% success rate
  - Multi-antenna, robust control and data design (low MCS/CQI), multiconnectivity (including duplication), retransmissions (HARQ)
- **Mobility interruption time:** 0 ms
  - multi-connectivity (dual connectivity)
- **Battery life:** 10 years RRC inactive,
  - power saving mode, enhanced Discontinuous Reception (DRX)
  - RRC: in CM-CONNECTED mode

Reference Architecture

- same elements as the previous generations:
  - User Equipment (UE),
  - Radio Access Network (NG-RAN) and
    - GNB (instead of eNode)
  - Core Network (5GC)
    - User Plane Function (**UPF**), handling the user data
    - Mobility management Function (**AMF**) that accesses the UE and the (R)AN
  - NG: interfaces between the access and the core networks.

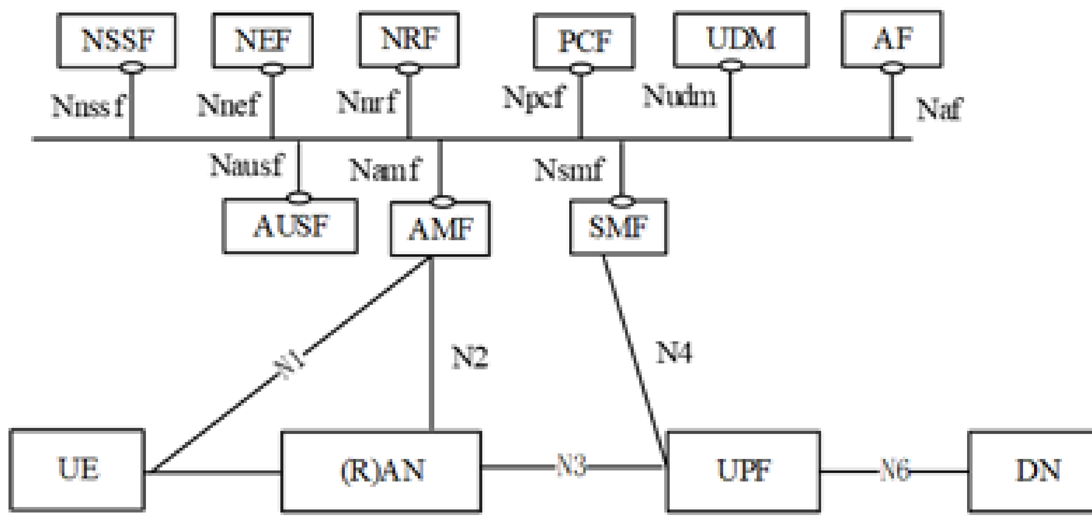


https://www.3gpp.org/technologies/5g-system-overview

Reference Architecture - 5GC

- Network entities are now **Network functions (Nfs)**
  - External Data Network (**DN**), mostly in the User Plane
  - Application Function (**AF**): controls the application(s)
  - Session Management Function (**SMF**) handles the calls and sessions, and contacts the UPF accordingly
  - Unified Data Management (**UDM**), functionally similar to 3G and 4G's HSS
  - Policy Control Function (**PCF**) controls that the user data traffic does not exceed the negotiated bearer(s) capacities

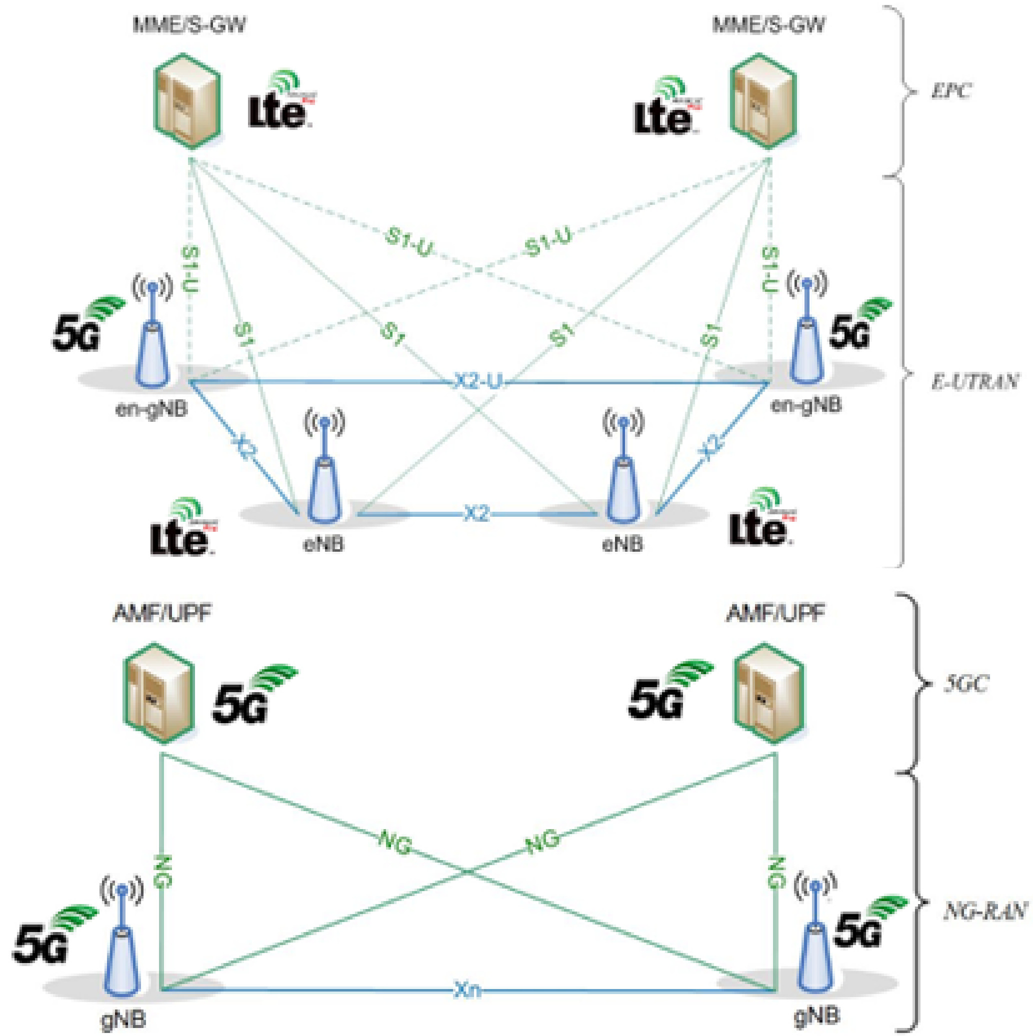
- Network Repository Function (**NRF**) "controls" the other NFs (register, deregister and update).
- security-related NFs: Network Exposure Function (**NEF**), Authentication Server Function (**AUSF**), Security Anchor Functionality (**SEAF**)
- The Network Slice Selection Function (**NSSF**)



https://www.3gpp.org/technologies/5g-system-overview

Non Stand Alone (NSA) Arch

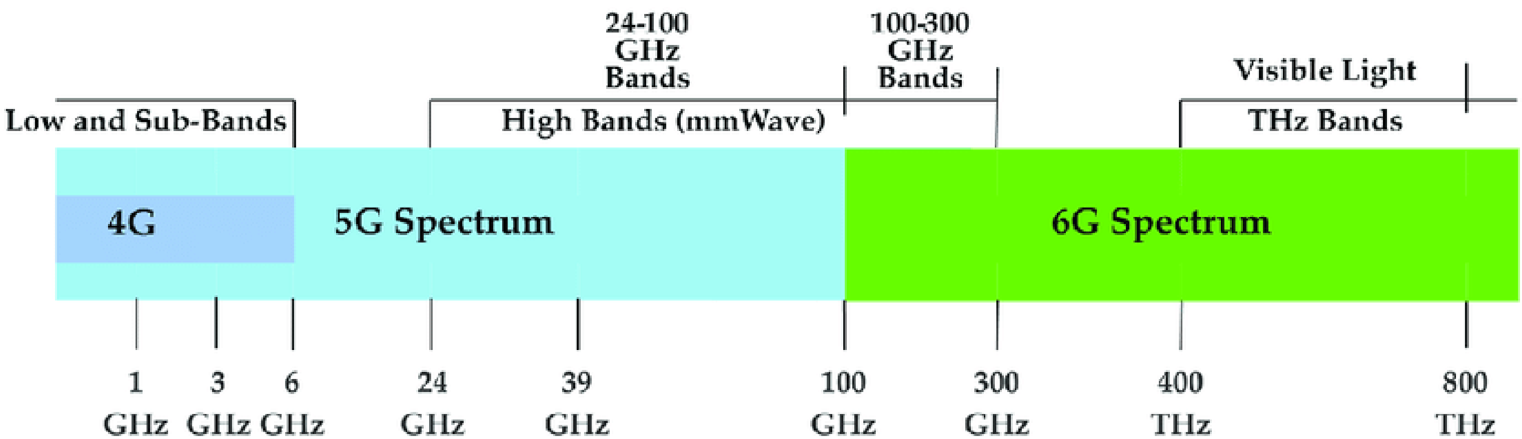
- step towards 5G
- making the NR technology available without network replacement
- In NSA, only the 4G services are supported, but they enjoy the capacities offered by the 5G New Radio (lower latency, etc).
  - No 5GC!



Spectrum

much wider range of frequencies than before, from below 1 GHz to 100 GHz.

- **Low-band** (below 2.5 GHz): excellent coverage,
- **mid- band** (2.5–10 GHz): good coverage and very high bitrates,
- **High-band** (10–100 GHz): highest bitrates and lowest latencies envisioned for 5G.



https://www.researchgate.net/publication/358001785/figure/fig2/AS:1122614135267330@1644663475488/4G-5G-and-6G-spectrum-bands.png

Multi-antenna

- large number of steerable antennas
- higher frequency bands: beamforming to extend coverage, reduce interference
- lower frequency bands: full-dimensional MIMO (“massive MIMO”) and interference avoidance by spatial separation.
- NR channels and signals have all been designed to support beamforming.
  - required to fully support operation at the higher frequency bands.

Beamforming

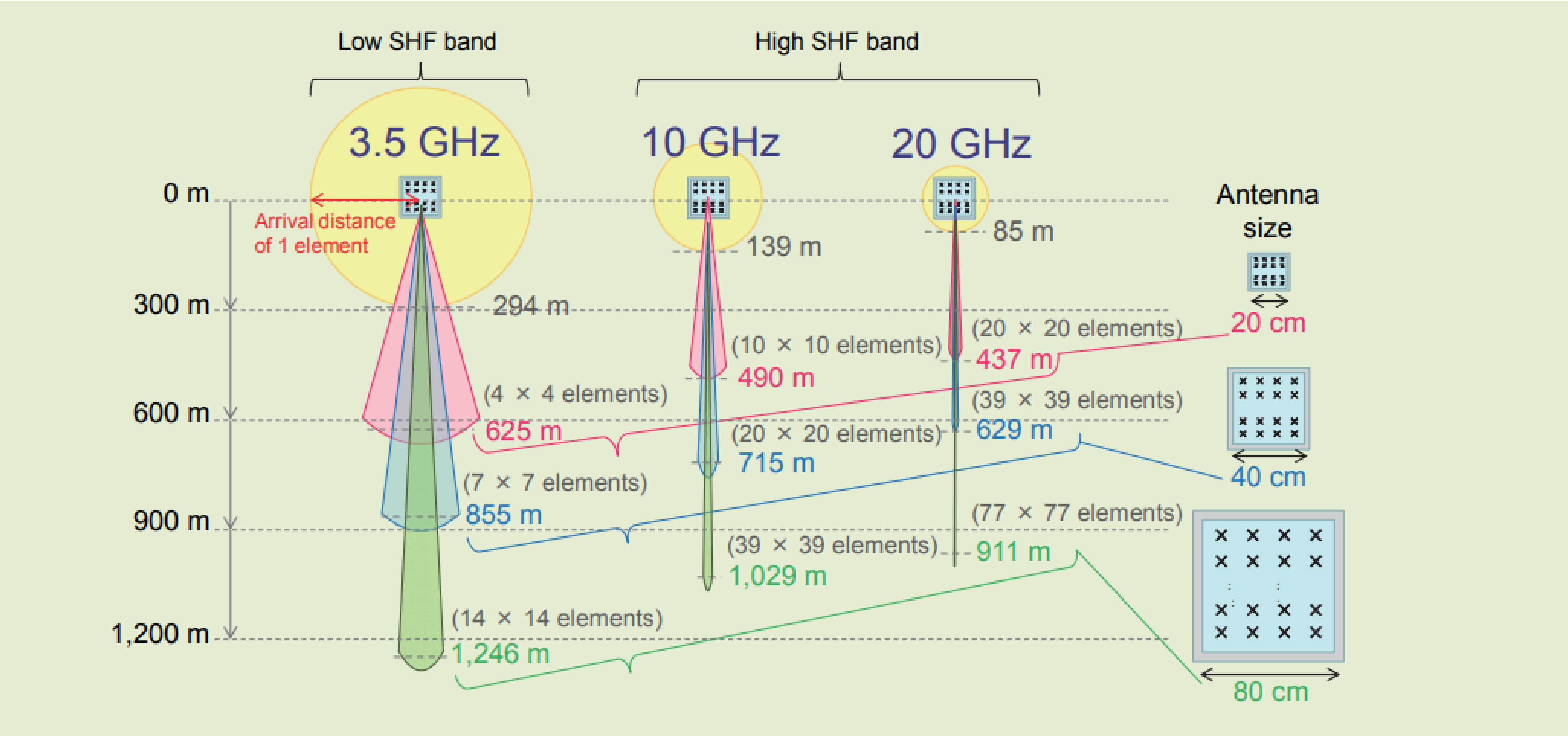
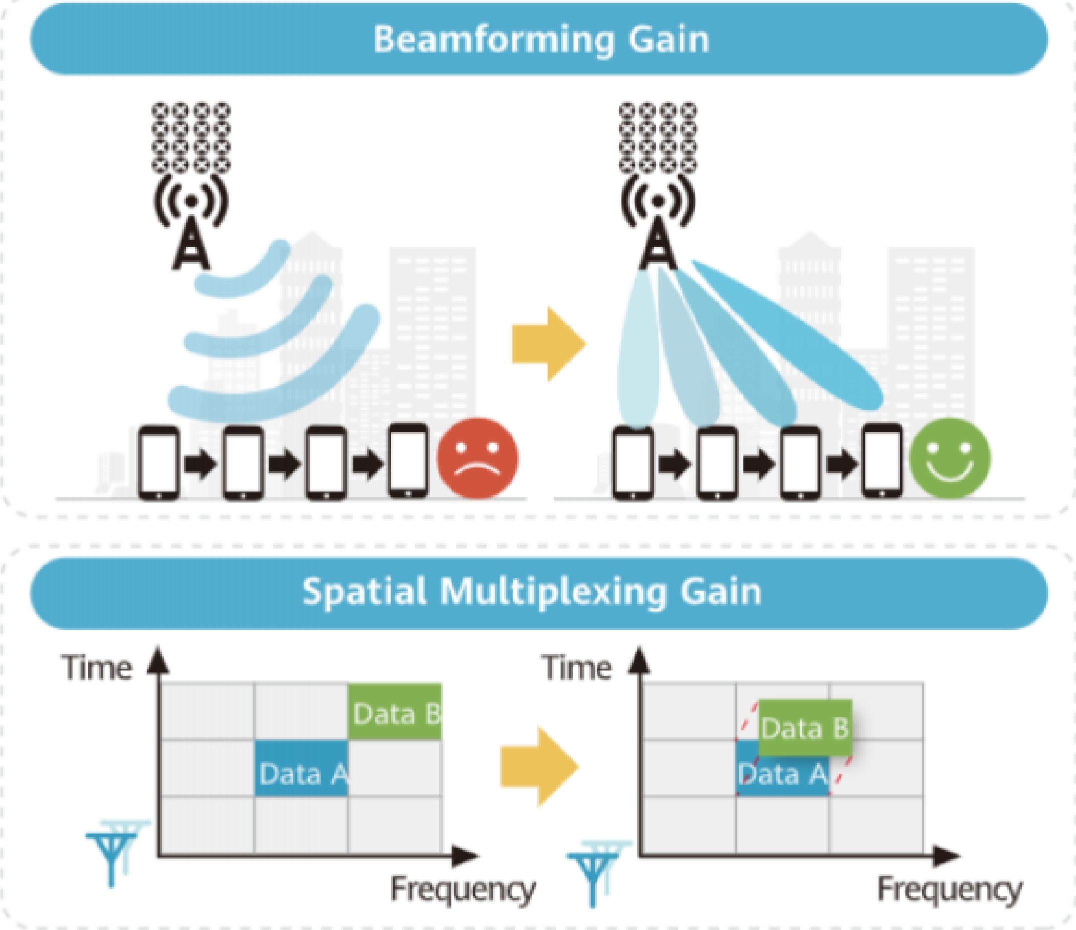


Figure 3 Beamforming effect

Suyama, Satoshi, et al. "5G multi-antenna technology and experimental trials."Â 2016 IEEE International Conference on Communication Systems (ICCS). IEEE, 2016.

Massive MIMO

- “massive” refers to the number of antennas (not size)
- propagation loss compensation
  - Dynamic antenna directivity
  - Expanded communication range
- User/spatial multiplexing and simultaneous connection of many users
  - increased system capacity bands
- Precoding to avoid cross-user/stream interference
- Complex signal processing to support the large number of antennas.



https://forum.huawei.com/enterprise/en/what-is-massive-mimo-su-mimo-and-mu-mimo/thread/667284186240794625-667213872962088960

Massive Single-User (SU) MIMO

- SU: all resources was allocated to one user and will improve single-user throughput.
  - In Multi-User (MU) MIMO, multiple user use a carrier (OFDMA)
- BW aggregation
- Similar to SU MIMO used in LTE, BUT:
  - In theory: more (up to 64) narrower layers
  - In practice: up to 4 due to UE limitations

Scallable Numerology

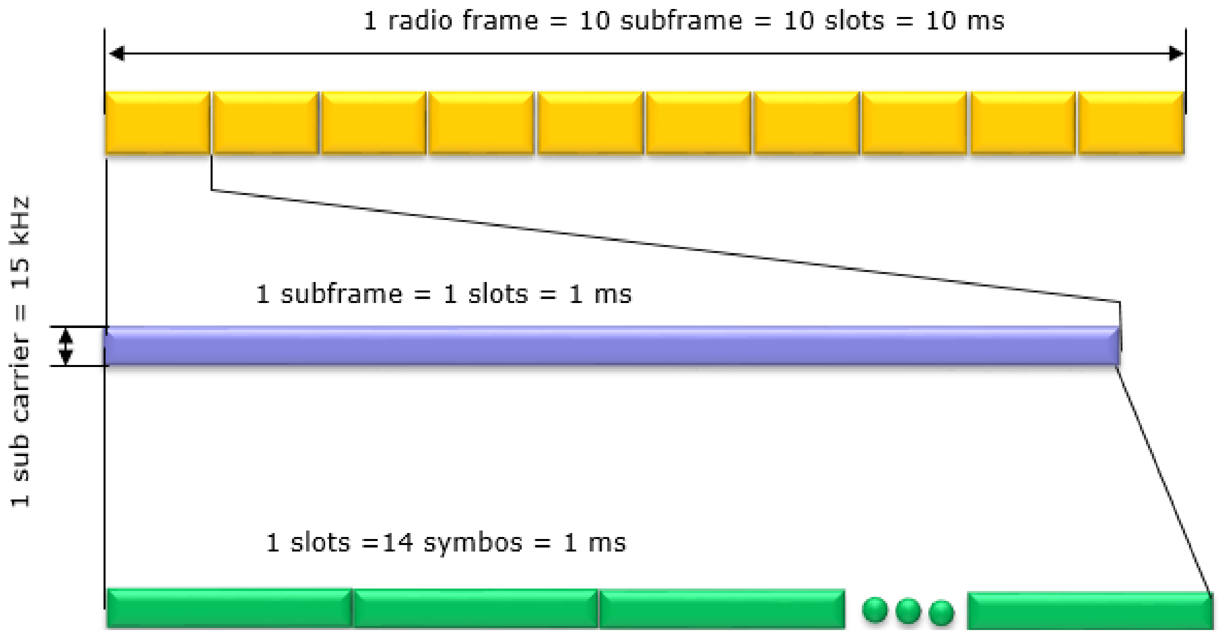
- Numerology: Set of parameters that define the physical layer structure,
  - the subcarrier spacing,
  - symbol duration, and
  - cyclic prefix length in an OFDM system.
- Challenge to support so wide spectrum
- OFDM, 15-120KHz spacing
  - LTE uses a single spacing (15KHz)
- 3.300 subcarriers: 50/100/200/400 MHz
  - Correlated spacings: 15/30/60/120 kHz

Radio Frame (μ = 0)

- length of a Radio Frame is always 10 ms
- length of a subframe is always 1 ms.

< 38.211 - Table 4.3.2-1 >

μ	N_{symb}^{slot}	$N_{slot}^{frame,\mu}$	$N_{slot}^{subframe,\mu}$
0	14	10	1
1	14	20	2
2	14	40	4
3	14	80	8
4	14	160	16



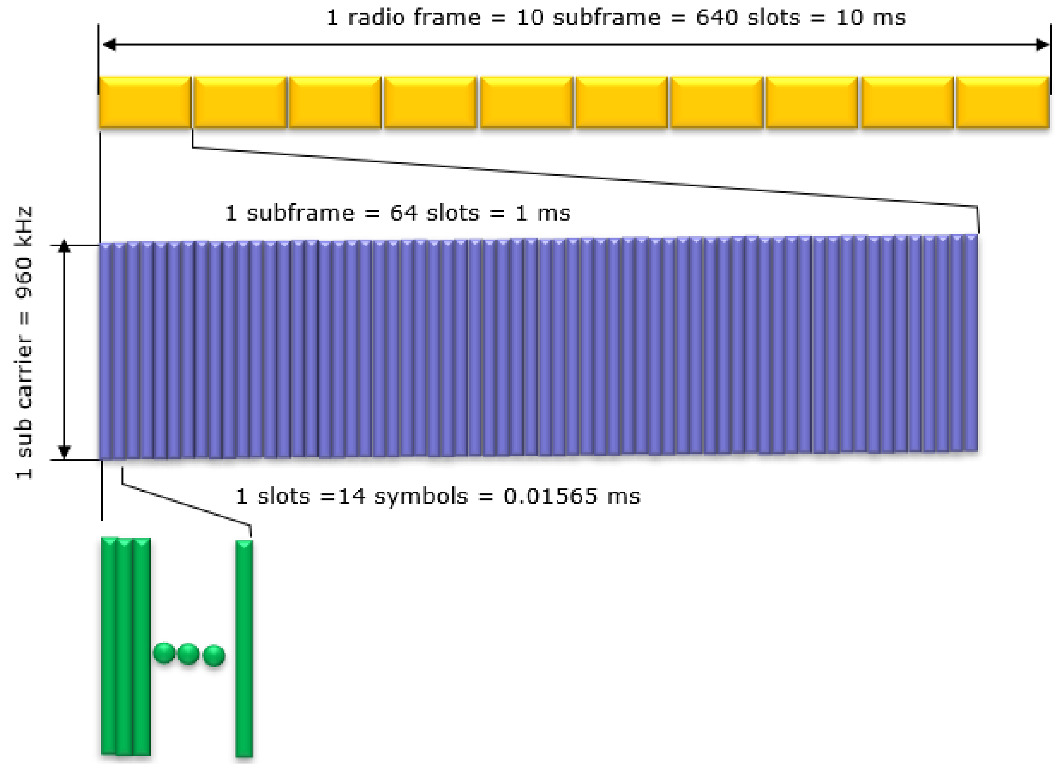
https://www.sharetechnote.com/html/5G/5G_FrameStructure.html

Radio Frame (μ = 6)

- length of a Radio Frame is always 10 ms
- length of a subframe is always 1 ms.

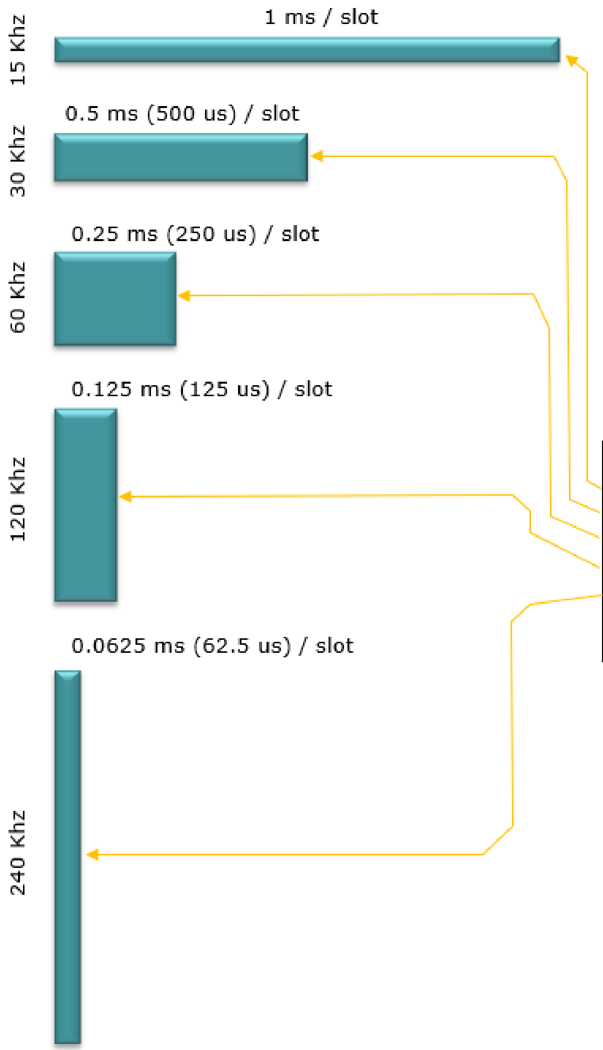
< 38.211 – v17.2.0 Table 4.3.2-1 >

μ	N_{symb}^{slot}	$N_{slot}^{frame,\mu}$	$N_{slot}^{subframe,\mu}$
0	14	10	1
1	14	20	2
2	14	40	4
3	14	80	8
4	14	160	16
5	14	320	32
6	14	640	64



https://www.sharetechnote.com/html/5G/5G_FrameStructure.html

Slot length



< 38.211 – v17.2.0 Table 4.3.2-1 >

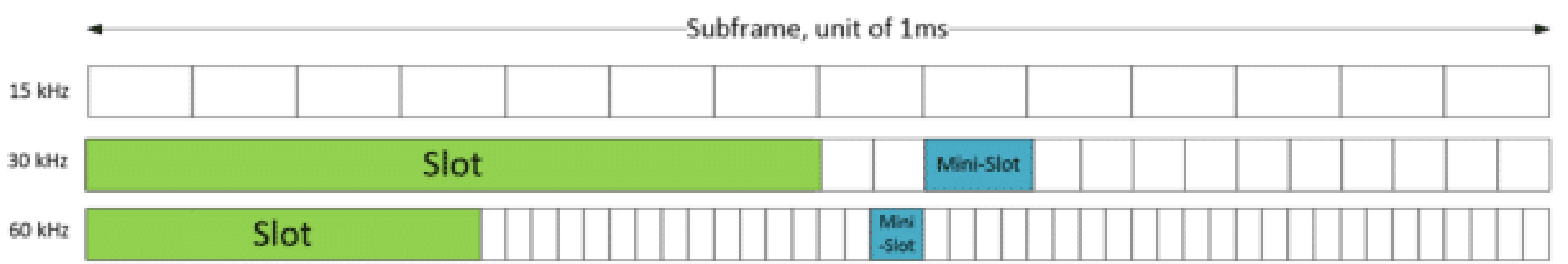
μ	N_{symb}^{slot}	$N_{slot}^{frame,\mu}$	$N_{slot}^{subframe,\mu}$
0	14	10	1
1	14	20	2
2	14	40	4
3	14	80	8
4	14	160	16
5	14	320	32
6	14	640	64

https://www.sharetechnote.com/html/5G/5G_FrameStructure.html

Mini-slot

https://www.etsi.org/deliver/etsi_tr/138900_138999/138912/15.00.00_60/tr_138912v150000p.pdf

- Low latency communication
- 2, 4, 7 or 14 OFDM symbols
- Target slot lengths are at least 1ms, 0.5ms
- finer TDM granularity of scheduling for the same/different UEs **within a slot**
- NR-LTE coexistence



https://www.5gworldpro.com/blog/2020/05/04/216-what-is-mini-slot-in-5g/

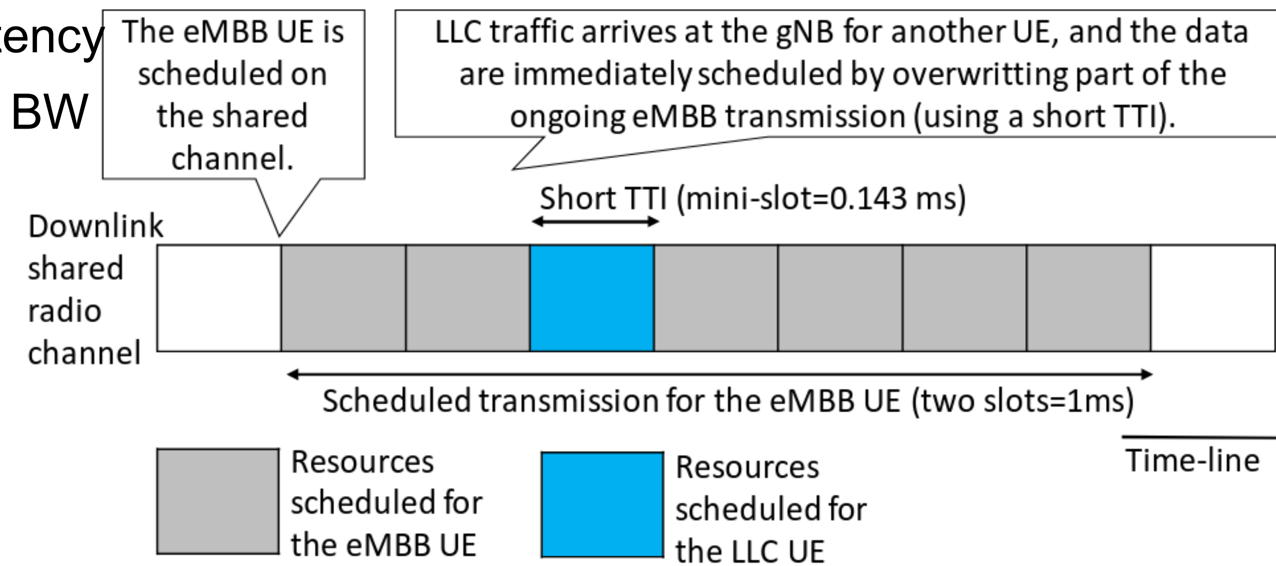
Pre-emption

Problem:

- The LLC traffic takes priority over the best effort eMBB data flows
- LLC traffic is quite unpredictable
  - Reservation is inefficient → less BW for eMBB
  - If not reserved, eMBB monopolizes channel causing latency to LLC

Solution: Override eMBB minislots when LLC appears →

- lower LLC latency
- Lower eMBB BW

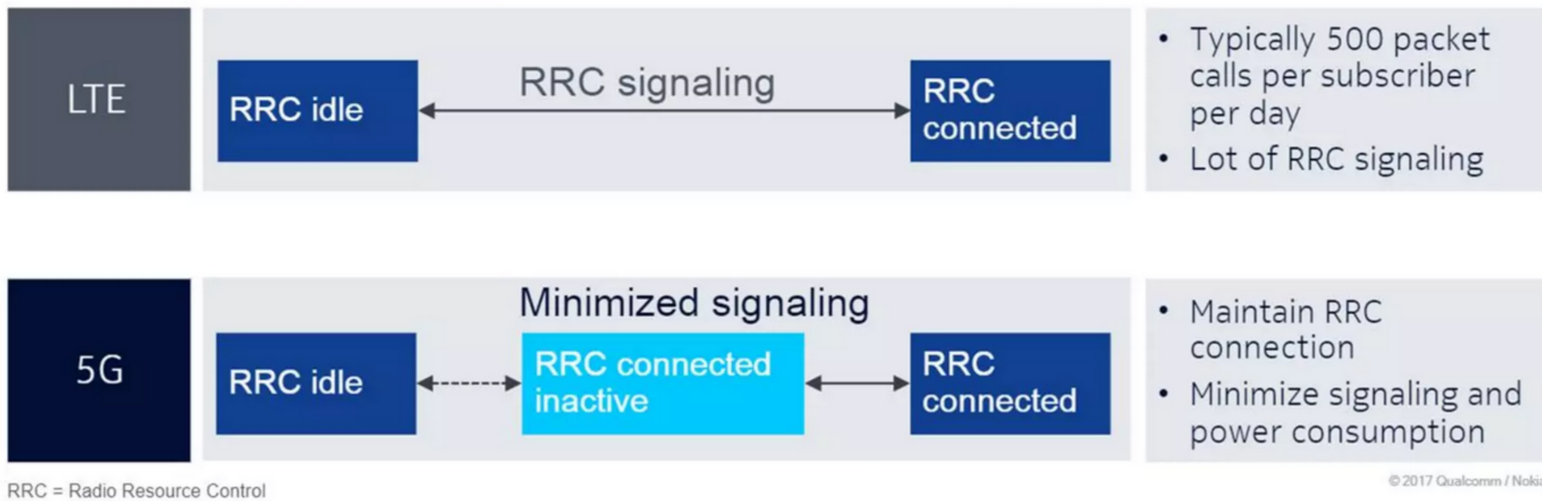


Pedersen, Klaus I., et al. "Punctured scheduling for critical low latency data on a shared channel with mobile broadband." 2017 IEEE 86th Vehicular Technology Conference (VTC-Fall). IEEE, 2017.

Fast Hybrid Automatic Repeat reQuest (HARQ)

- Enhanced reliability through error detection and retransmission
  - Forward error correction and ARQ
- Compared to 4G:
  - multiple HARQ-ACK feedback channels
  - Shorter Transmission Time Interval (TTI)
  - Soft Combining
    - Combine multiple received signal instances, such as erroneously received packets
  - advanced beamforming techniques
  - Channel coding with better error correction capabilities

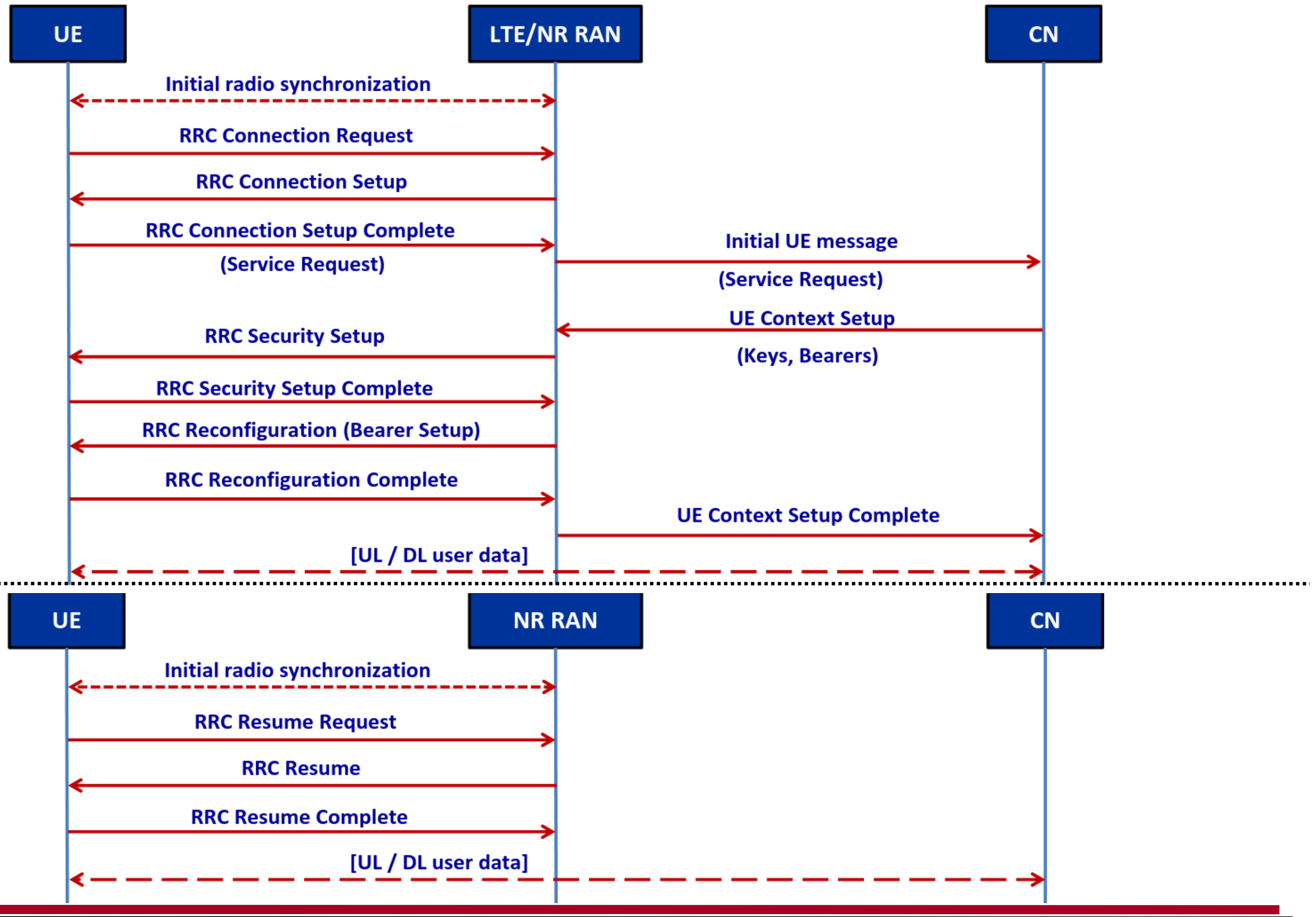
Energy: RRC inactive, Discontinuous reception (DRX), WakeUp Signal (WUS)



- 3 states:
  - **Idle**: no active connection with gNB – sleeping/unreachable
    - DRX: checks periodically the physical downlink control channel
    - Synchronization needed, DRX period managed by 5GC
  - **Connected**: active control and data connection with gNB
  - **Inactive**: active control bearer, UE context not removed from RAN
- “Inactive” offers faster transition to “connected” – for ultra-low latency apps
- WUS: more frequent than DRX, wake up only when packet is waiting
- BDW adaptation and dynamic antenna activation

https://blog.huawei.com/2022/05/19/how-5g-can-improve-the-battery-life-of-user-equipment/

Reconnection message diagram

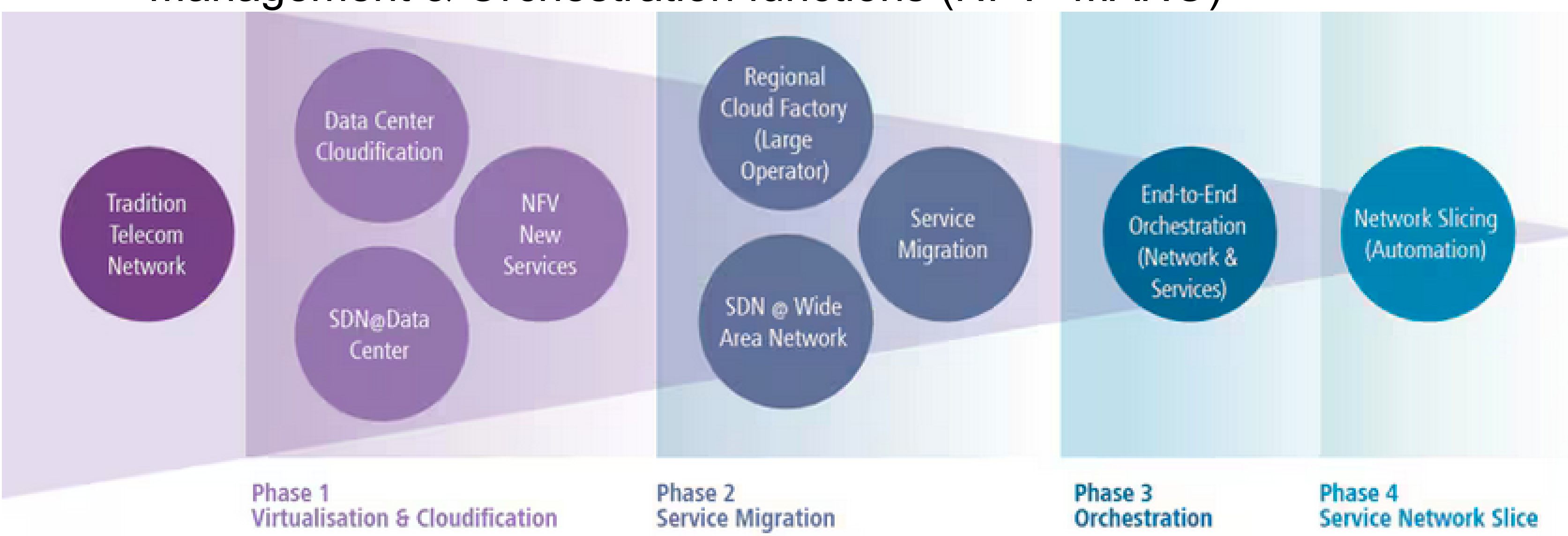


LTE/NR spectrum coexistence

- NR in the same spectrum with LTE deployment
- LTE-compatible NR numerology based on 15 kHz subcarrier spacing,
  - identical time/frequency resource grids for NR and LTE
- Enhanced LTE/NR overlay in the downlink
  - Through reserved resources
    - Forward compatibility

Important Technologies: NFV

- Network Function Virtualization
  - Separation of HW and SW
    - See "5GC architecture"
- Densification makes RAN management challenging.
  - Cry for flexibility and automation
- NFV provides full automation of network management, NFV Management & Orchestration functions (NFV--MANO)

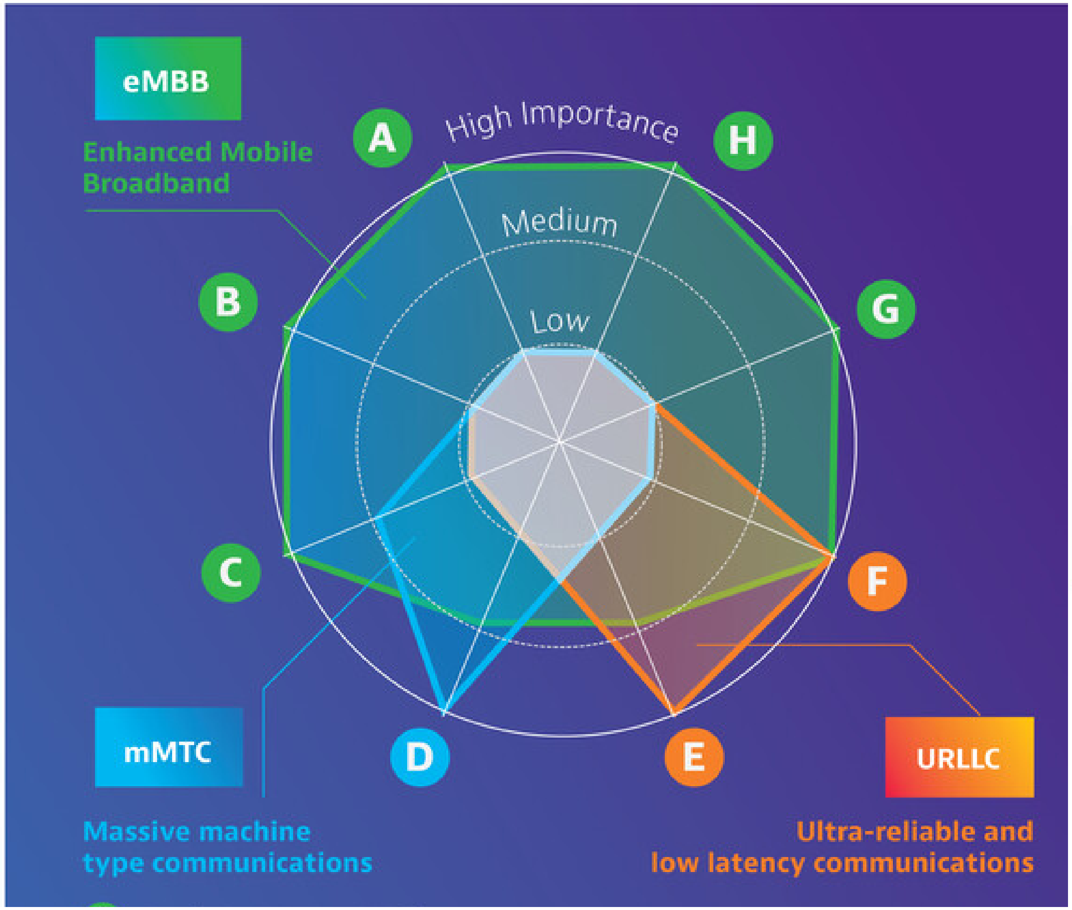


Important Technologies: Slicing

- Replacement of QoS identifiers
- Transforming the physical network into multiple **logical** networks on top of a shared infrastructure
- A network slice spans from the UE to the Data Network (DN),
- A slice may belong to a single customer/group or may be shared among multiple customers
- Slices may be based on
  - latency, bandwidth, reliability, security
- The O-RAN (Open RAN) ALLIANCE specifies solutions for an open RAN to support end-to-end network slicing for 3GPP-specified 5G systems

https://www.5gamericas.org/wp-content/uploads/2022/07/Commercializing-5G-Network-Slicing-Jul-2022.pdf

Slicing and use-case families

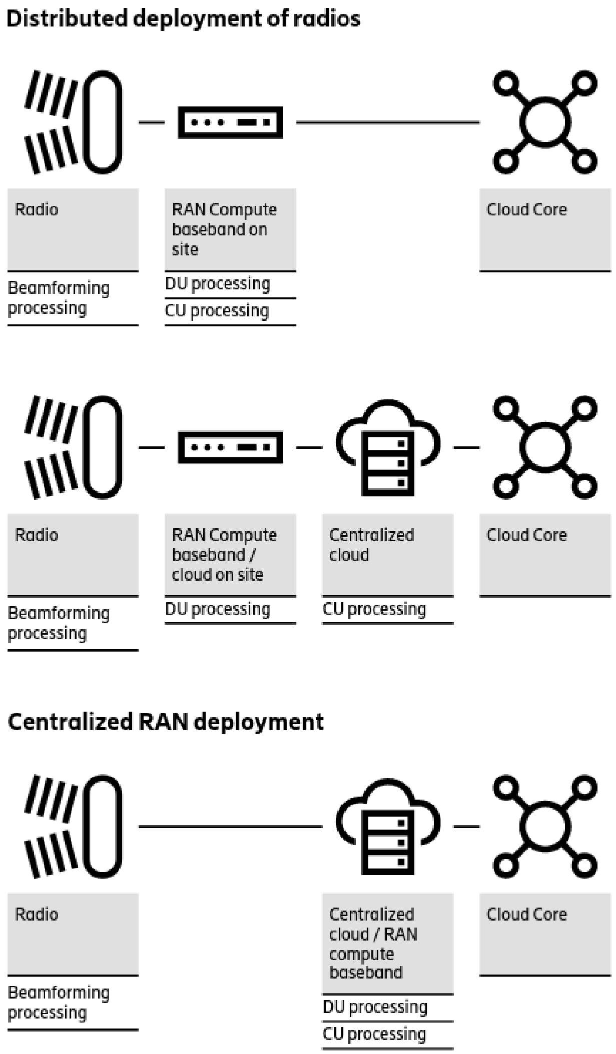


- A Peak Data Rate - 20 Gbps
- B Area Traffic Capacity - 10 Mbps/m²
- C Network Energy Efficiency - 100 Times
- D Connection Density - 10^6/km² (1 per m²)
- E Latency - 1ms(Radio segment)
- F Mobility - 500 km/h
- G Spectrum Efficiency - 3 Times
- H User Experienced Data Rate - 100Mbps

Source: Rec. ITU-R M.2083-0 Figure 3 and 4

Important Technologies: RAN cloudification

- **C-RAN: separation** of the RAN baseband software and the RAN baseband hardware
- software can run on off-the-shelf hw
  - Easier scale-up, protocol update, openess?
  - @ local RAN or provider's DC or public cloud
- Connect C-RAN and antenna site with dark fiber.
- C-RAN drivers:
  - Cloud economics – flexibility in deployment, benefits of scale,
  - common and unified and automated operation (AI)
  - Development of new enterprise business



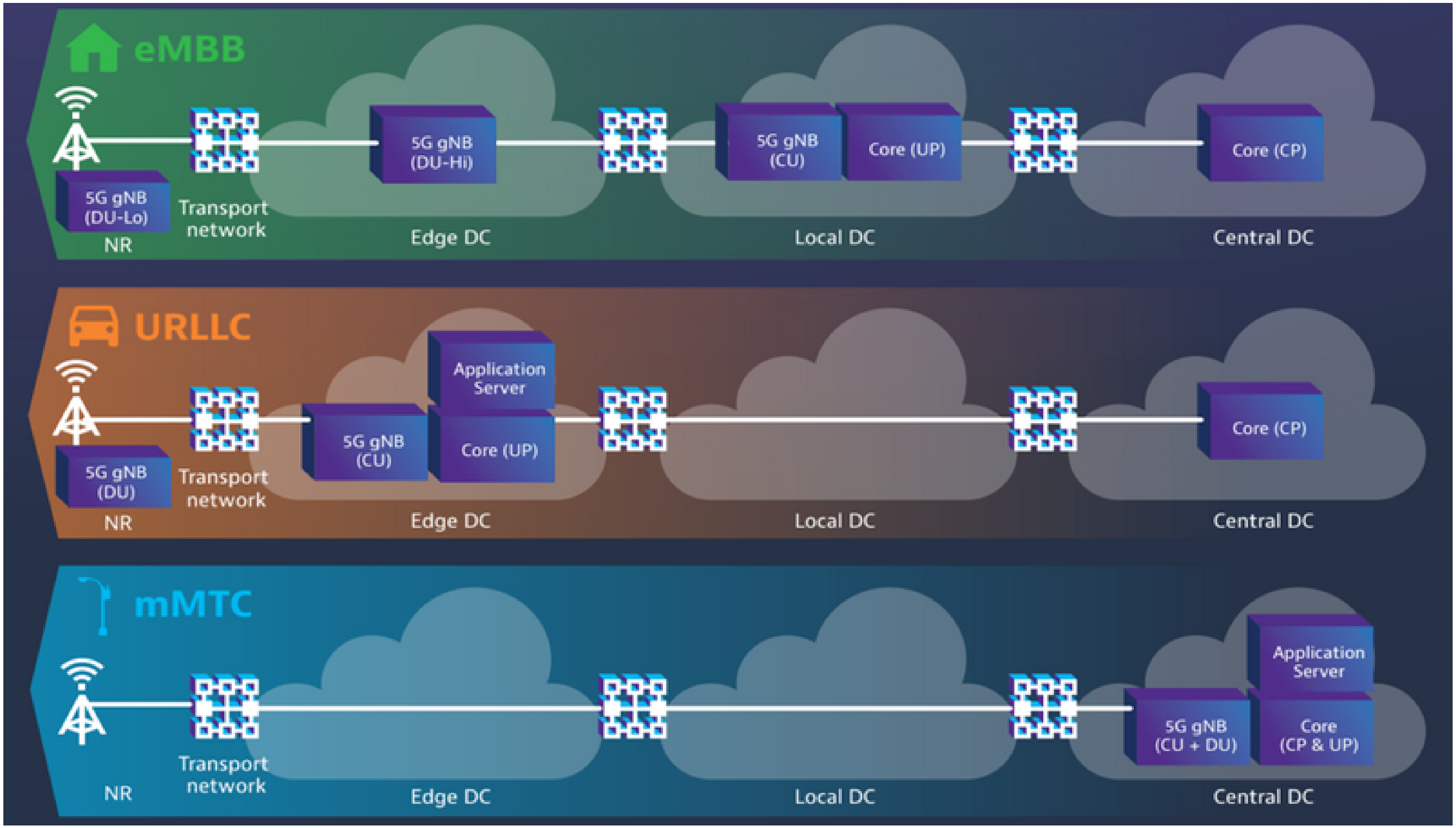
https://www.ericsson.com/4ab4fb/assets/local/reports-papers/further-insights/doc/whats-next-for-ran.pdf

Important Technologies: Edge Computing

- a.k.a. Mobile Edge Computing (MEC) or Multi-Access Edge Computing,
- bring processing and storage closer to where data is generated or consumed.
- ‘edge’ is within or at the boundary of an access network or at the UE premises (train, airplane, enterprise etc.)
- virtualize the cloud beyond the four walls of the data center
- Gains:
  - Better data control and lower costs: less transport to central hubs and reducing vulnerabilities
  - Faster insights and actions
  - Continuous operations even when disconnected
- Empowered by the shared infrastructure from application perspective
  - network slicing!

https://www.ibm.com/cloud/smartpapers/5g-edge-computing/

All together



https://www.viavisolutions.com/en-us/5g-network-slicing