BSc in Computer Engineering
CMP4204
Wireless Technologies

Lecture 12

5G NR Cellular Mobile

Eng Diarmuid O'Briain, CEng, CISSP

Department of Electrical and Computer Engineering,
College of Engineering, Design, Art and Technology,
Makerere University

Table of Contents

Illustration Index

This page is intentionally blank

1. The future 5G NR

3GPP release 15 introduced Long Term Evolution (LTE) 5G "New Radio" by setting requirements. 5G requires that there be a major focus on data. It expects that there be:

- Significant increased data rates to support large populations.
- 1 Gb/s simultaneously to many subscribers in proximity to each other.
- Massive numbers of simultaneous connections to support the IoT.
- Enhanced spectral efficiency compared to 4G.
- Improvements in coverage compared to 4G.
- Improved signal efficiency compared to 4G.
- Reduced latency compared to 4G.

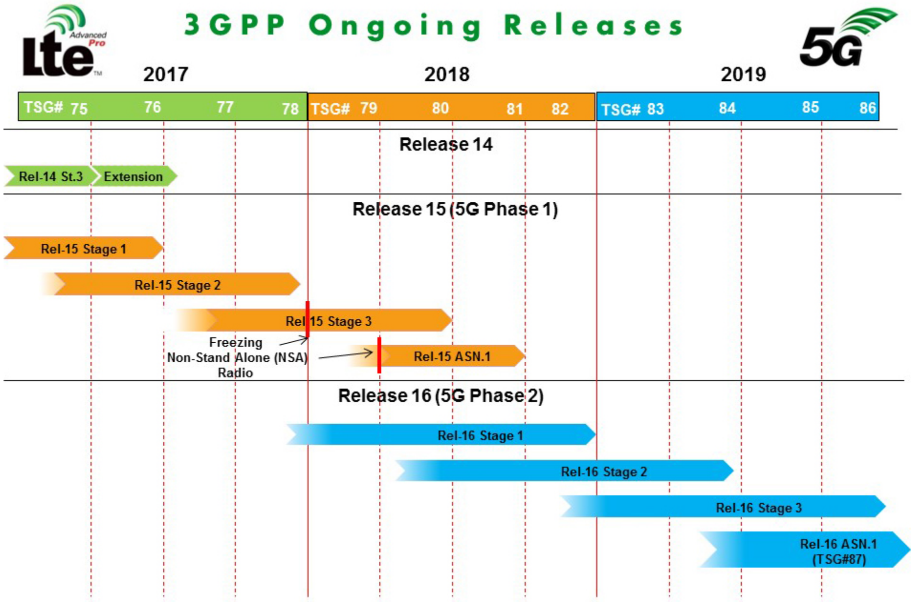

Illustration 1: 3GPP Ongoing Releases

Note that 5G is in the early stages of development with only a few initial specifications released at the end of 2017.

1.1 The 5G Use Cases

5G NR has a different set of use cases compared to 4G LTE and LTE-Advanced. These 4G technologies were focused on providing data streams carrying real time traffic like voice and video. The 5G era incorporates the additional need to supply communications to very large numbers of Internet of Things (IoT) devices. These generally require low bandwidth streams that in terms of numbers way exceed the number of traditional traffic types connections from 4G. This has forced a rethink on how 5G can cope with these very different requirements has been necessary. Additionally, as has been evidenced on this course so far the available spectrum has been hoovered first by 3G in IMT-2000 and then by 4G as part of IMT-Adv.

The design of 5G therefore split the problem into three separate use cases:

- ***enhanced Mobile BroadBand (eMBB)***
  - Essentially a fibre-like experience over a wireless radio link
  - Multi Gb/s peak rates for Downlink (DL) and Uplink (UL).
- ***massive Machine Type Communications (mMTC)***
  - Connectivity for millions of devices.
- ***Ultra Reliable Low Latency Communications (URLLC)***
  - Highly available, reliable and low latency service.

1.1.1 enhanced Mobile BroadBand (eMBB)

eMBB essentially is a 5G use case to provide a fibre-like experience over a 5G wireless radio link. It is the evolutionary path for Service Providers that have deployed Gb/s LTE network technologies today.

eMBB performance compared to 4G networks to day are expected to be in the order of

- 10 times throughput
- 10 times decrease in end-to-end latency
- 10 times connection density
- 3 times spectral efficiency
- 100 times traffic capacity
- 100 times network efficiency.

1.1.2 massive Machine Type Comms (mMTC)

The mMTC use case is about serving Device 2 Device (D2D) communication that are part of IoT scenario with projected tens of billions of connected devices and sensors. Devices like smart sensors, smart logistics and connected homes have different requirements to the Internet usage demanded by eMBB.

Typical characteristics and requirements of devices in this class are:

- Transmitting relatively low volumes
- Non-delay-sensitive data
- Low bandwidth and not latency-critical
- Low-cost devices
- Extended battery life.

1.1.3 Ultra Reliable Low Latency Communications (URLLC)

URLLC is a requirement for emerging mission-critical applications such as industrial Internet, smart grids, infrastructure protection, eHealth, public safety and Intelligent Transportation Systems (ITS). Such operations require both low latency and high reliability that demand enhanced transmission techniques to meet the reliability requirements for both data and control channels. The three key characteristics of a URLLC system are:

- *Latency*
  - Time delay between data being generated and transmitted from a device like a sensor and it being correctly received by another device like an actuator.
- *Reliability*
  - Provide a high level of certainty that a message is correctly delivered to the receiver within a latency bound.
- *Availability*
  - System endurance against outage scenarios.

Expected latencies for future systems are expected to be in the order of:

- Factory automation < 1 ms
- Motion control < 1 ms
- Remote control 5 - 100 ms
- Intelligent transport systems 5 ms
- Smart grid 3 – 5 ms
- Tactile Internet 1 ms

Latencies in 4G LTE

- Best 10%: 21 - 43 ms
- Median: 33 - 75 ms
- Worst 10%: 47 - 200 ms

1.2 5G Modulation

A typical communication on LTE and LTE-Adv networks today involve many interactive processes between the eNB and the UE before data is transmitted. This is OK for long time continuous sessions typical in functions associated with the traditional eMBB type activity. The signalling overhead is averaged over the duration of the activity and is therefore not so large. However in the mMTC use case for IoT devices there is a small volume of data transmitted over a short period of time by very large numbers of devices. The signalling overhead associated with 4G technologies is therefore too high and access efficiency becomes low.

1.2.1 Non Orthogonal Multiple Access (NOMA)

The function of Non Orthogonal Multiple Access (NOMA) methods is to deal with the issue of signals that possess significant differences in power levels. Consider a traditional data transmission and an IoT device competing for the same spectrum, this would have seen the stronger signal getting served at the expense of the weaker one. A 5G system is required to allow both signals to be handled by the system simultaneously.

It can be seen from Illustration 2 the complexity of a NOMA modulation scheme. Non-orthogonality is intentionally introduced either in time, frequency or code.

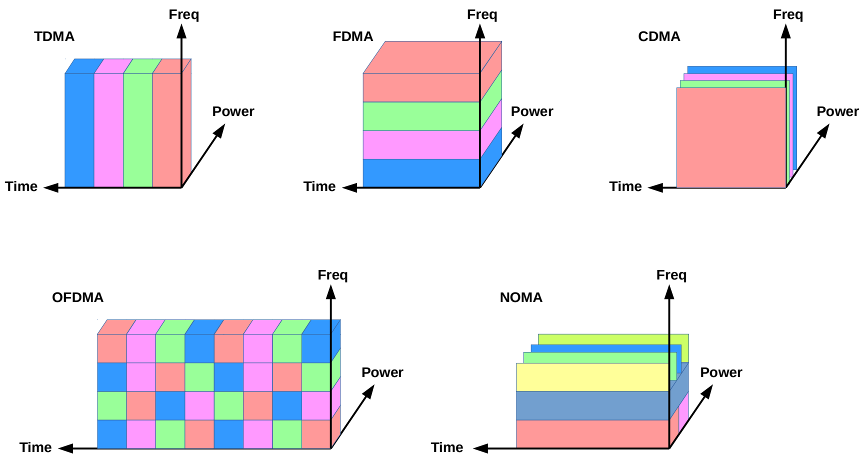

Illustration 2: Mobile modulation schemes to NOMA

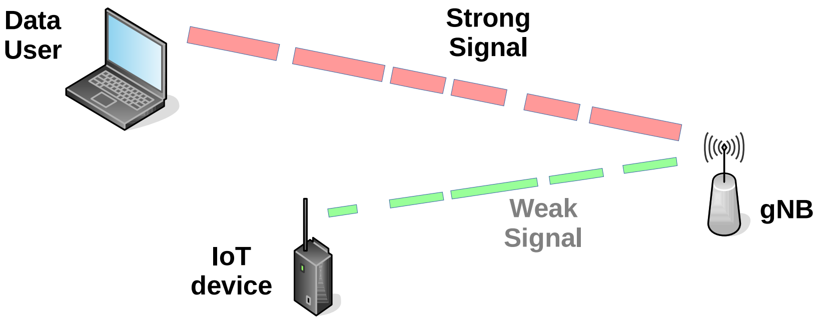

Illustration 3: Non Orthogonal Multiple Access (NOMA)

Consider for example a basestation transmitting at 40W while an IoT sensor may only operate at 25 mW. In the mobile systems currently such a collision would involve the weaker signal being ignored at the expense of the stronger signal. NOMA intentionally introduces non-orthogonality in either time, frequency or code. The signals of differing strengths can be accepted and passed using a Superposition Coding (SC) scheme despite the large power difference between the two.

Extraction of the signal requires the receiver to implement Successive Interference Cancellation (SIC). Essentially NOMA is OFDM plus SC and SIC.

1.2.2 Superposition Coding (SC)

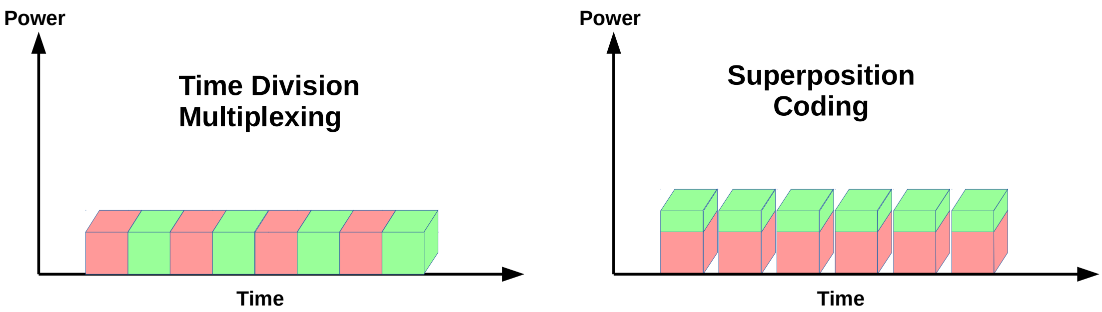

Illustration 4: Superposition Coding (SC)

Illustration 4 compares Time Division Multiplexing (TDM) with SC. In SC the strong and the weak signals are transmitted together with their individual power levels the separator.

1.2.3 Successive Interference Cancellation (SIC)

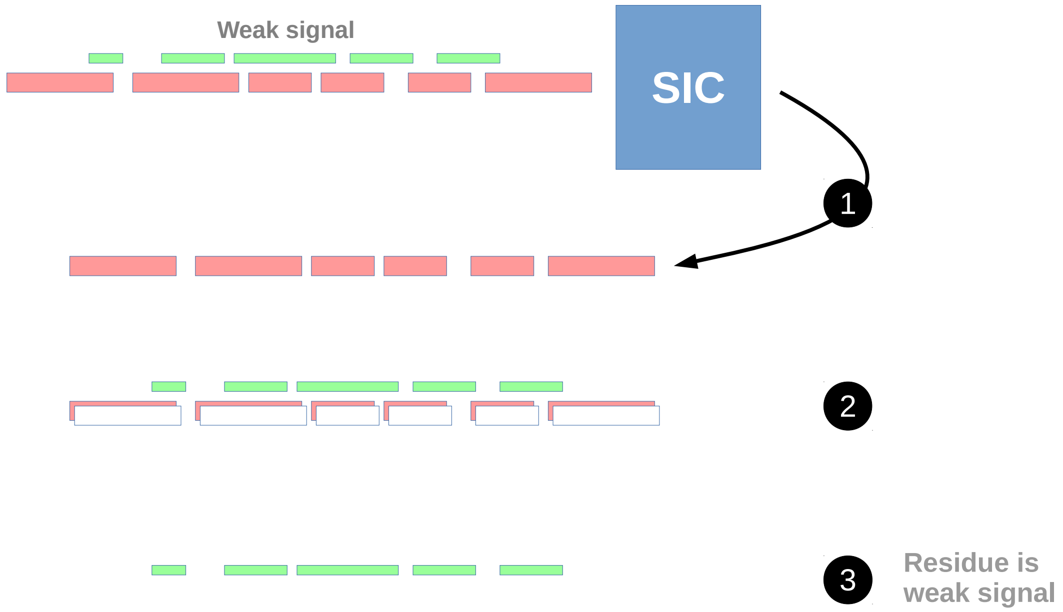

Illustration 5: Successive Interference Cancellation (SIC)

At the receiver the signal is received with the strong and weak signals sharing the same space in time. The SIC function separates the two signals by their power levels. At the physical layer the receiver decodes packets that arrive simultaneously. The SIC function decodes the strongest signal bitstream and reconstructs a copy of that signal. Taking the input, the reconstructed strong signal is cancelled leaving a residue signal. This residue signal is in fact the weak signal. In this way the receiver can extract the two separate signals transmitted together by their varying power levels.

1.3 The 5G Spectrum conundrum

Mobile generations from 1G to 4G solved spectrum issues by assigning new frequency bands as needed and increased bandwidth by using a wider spectral bandwidth per channel. For example the first four generations used:

- 1G up to 30 kHz
- 2G up to 200 kHz
- 3G up to 5 MHz
- 4G up to 20 MHz

However for 5G there is very little room for larger channel bandwidths and new frequency bands as simply moving to higher frequency bands would begin to overlap with K-band satellite transmissions. Small pockets of spectrum were acquired in the sub 1 GHz bands and spectrum blocks in 3.4 to 3.8 GHz were reallocated to IMT-2020.

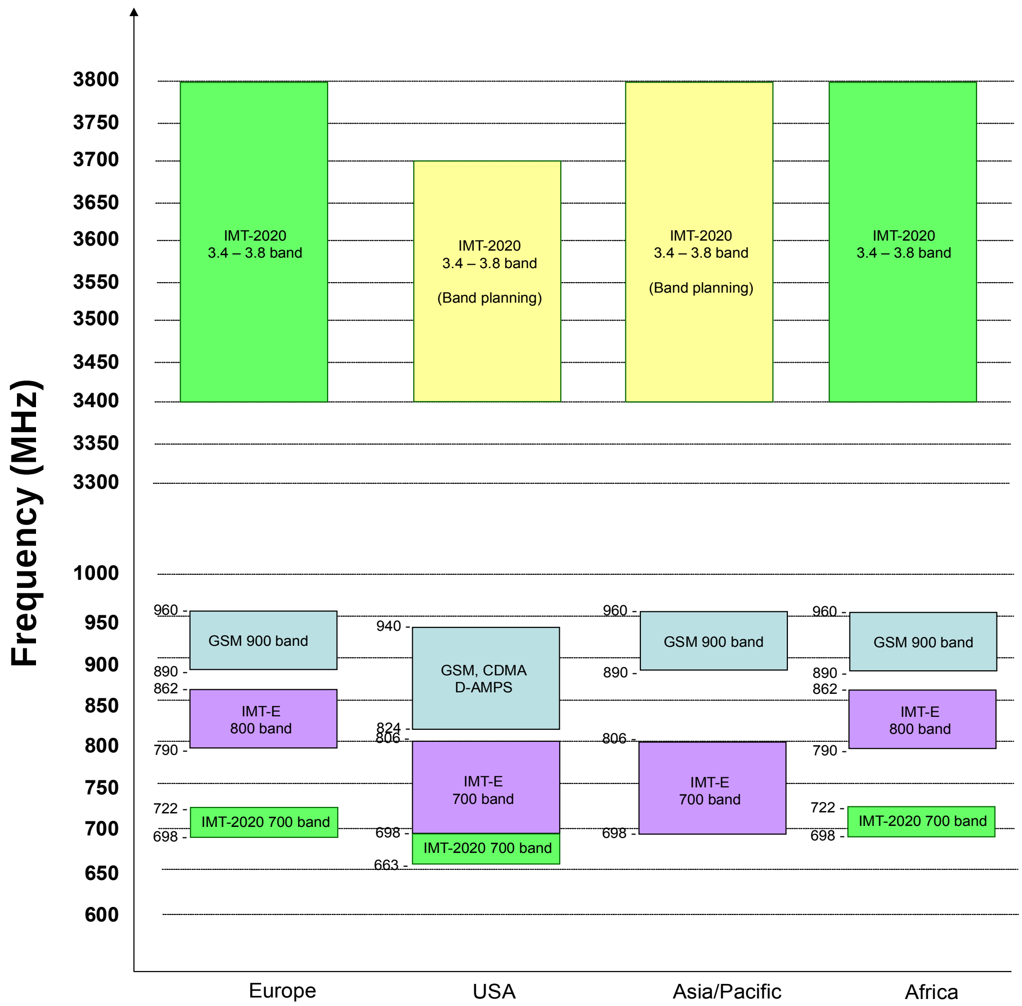

Illustration 6: 5G sub 6 GHz Spectrum

1.3.1 The mmWave bands

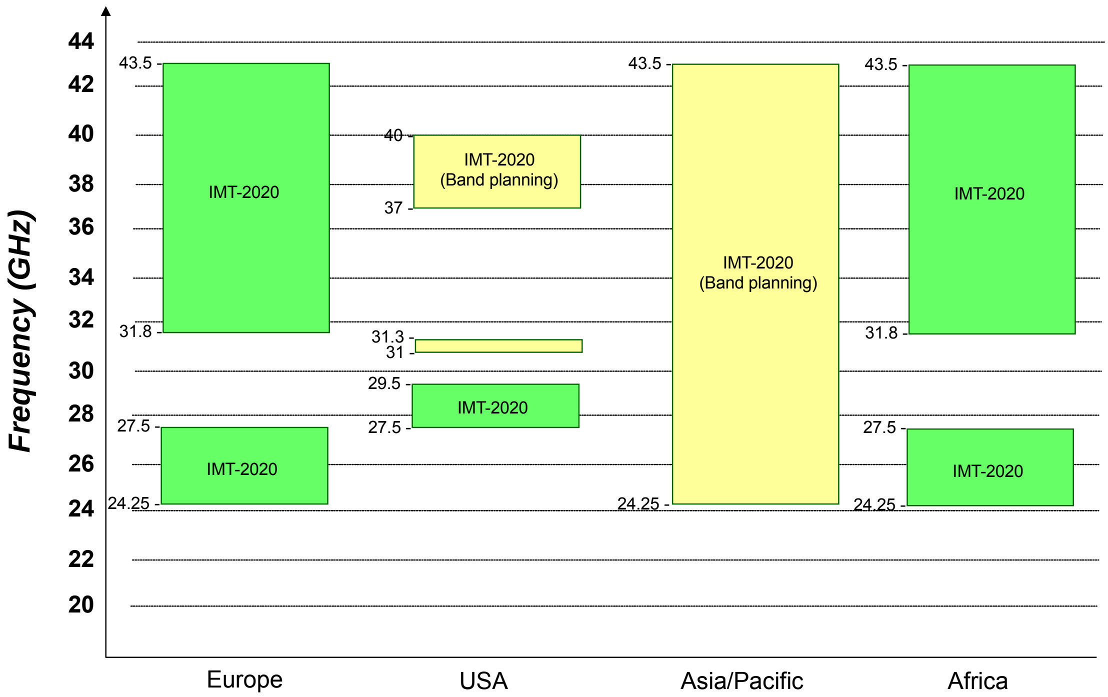

Illustration 7: mmWave bands

The mmWave bands are new spectrum assigned to IMT-2020 for 5G. The frequencies are very high indeed and are only useful for short range but high throughput applications. Additionally propagation at these frequencies are susceptible to interference.

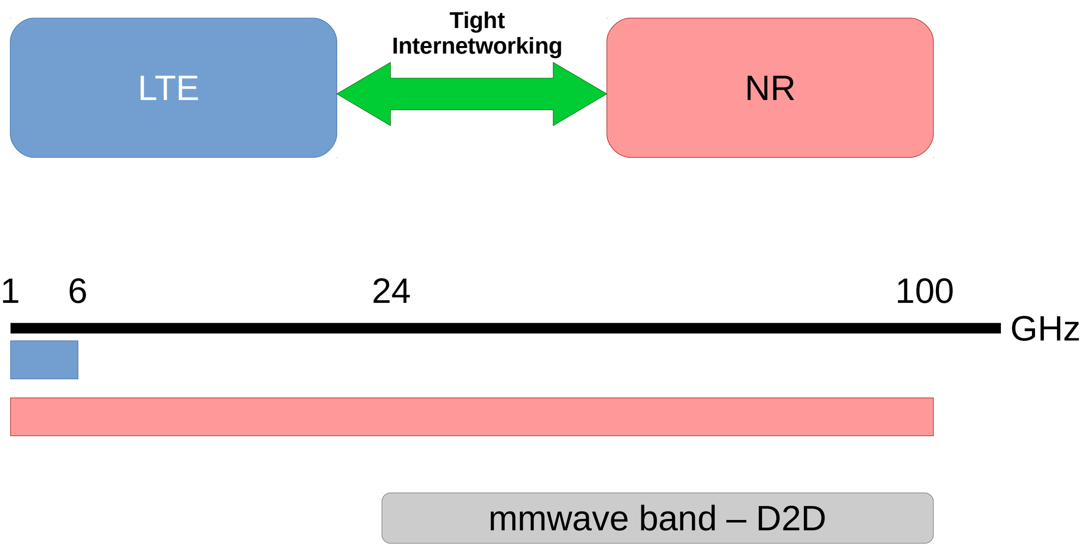

Illustration 8: LTE -vs- NR spectrum

Illustration 8 Gives a representation of the spectrum variance across IMT-2020. It will be seen that all spectrum blocks are not suitable for all applications. For example the sub 1 GHz blocks are highly suitable for IoT devices in a wide area, the 3 GHz band is more suitable for dense urban scenarios and the mmWave bands are best use for short D2D and Basestation to Device (B2D) applications at hotspots, shopping centres etc..

1.4 Massive MIMO (mMIMO)

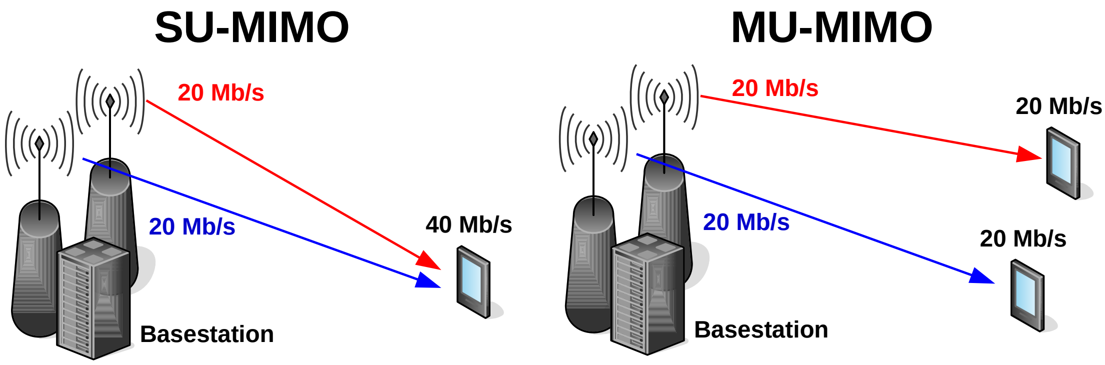

Illustration 9: SU and MU MIMO

Before exploring massive MIMO (mMIMO) consider Single User and Multi User MIMO. With SU-MIMO each device waits its turn to send and receive data to the transmitter. This increases the throughput of one user at a time. With MU-MIMO the transmitter communicates with devices simultaneously and this increases the overall system capacity as opposed to increasing the data transfer rate of one user.

mMIMO is a form of MU-MIMO. The difference is the number of antennas at the basestation is much larger than the number of mobile stations per signalling resource. For example a 4G system might have 64 antennas at the transmitter but due the the frequencies in use is physically limited to two antennas on device. Spectral efficiency therefore for LTE and LTE-Adv are in the region of:

- 16.32 b/s/Hz for LTE with 4x4 MIMO
- 30 b/s/Hz for LTE-Advanced with 8x8 MIMO.

1.5 Beamforming

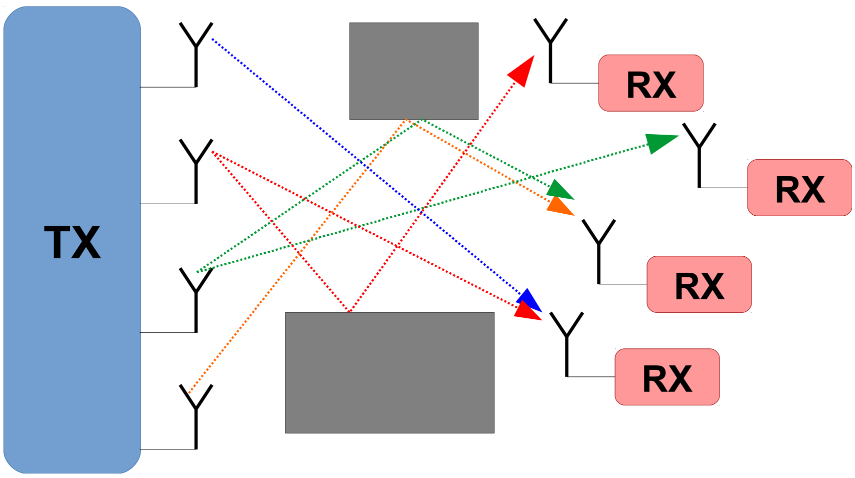

Illustration 10: Beamforming

Beamforming is a signal processing technology that is used to direct the reception or transmission (the signal energy) on a transducer array in a chosen angular direction.

It identifies the most efficient data-delivery route to a particular user, and it reduces interference for nearby users in the process. The primary problem associated with massive MIMO is how interference can be reduced while transmitting more information from many more antennas simultaneously. Each base station uses signal-processing algorithms to plot the best transmission path through the air to each user. Once plotted individual data packets can be sent in many different directions, these may be reflected off buildings and other objects in a precisely coordinated pattern as determined by the signal-processing algorithm. By managing the packets' movements and arrival time, beamforming allows many users and antennas on a massive MIMO array to exchange much more information at once.

With 5G it is quite feasible to have as many as 256 antennas on the transmitter and as the frequencies in use are much higher the antenna size is much smaller than LTE. It is expected that devices will accommodate as many as 64 antennas. This supports very high spectral efficiencies in the region of 145 b/s/Hz. This demonstrates that 5G can thus have a gain of 6 dB compared to 2x2 MIMO.

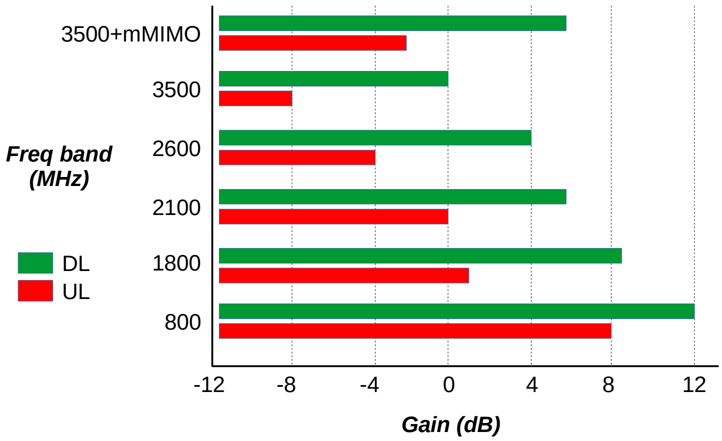

Illustration 11: Outdoor coverage gain for various bands

1.6 5G NR Tiers of coverage

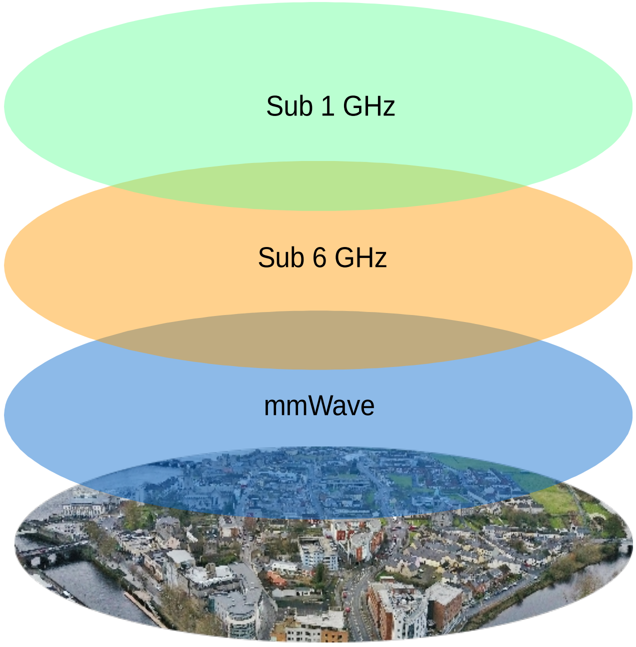

Illustration 12: 5G NR tiers of coverage

The reality of 5G is the need for tiers of coverage. The sub 1 GHz characteristics give excellent wide area coverage with indoor penetration. This is very suitable for massive IoT and URLLC. It also allows Service Providers the opportunity to reuse existing sites for 800/900 MHz.

The next tier, the characteristics of the sub 6 GHz tier presents solutions for dense urban coverage that supports eMBB and the opportunity for Service Providers to reuse existing sites for 2 GHz.

The final tier, the mmWave characteristics present options for hotspots in places like airports and stadiums to support full eMBB with data rates exceeding 10 Gbps.

1.6.1 Sub 1 GHz bands

The sub 1 GHz band characteristics provide for deep indoor penetration, reliable UL and large coverage area. This is ideal for large numbers of IoT devices spread across an urban or rural area. In the US the available band is 600 MHz whereas 700 MHz is the spectrum available in Europe. The 900 MHz band is already occupied with 2G and 3G today and therefore will not become available for some time as Service Providers are likely to keep 2G and 3G running until 2020. These bands present limited options to use features like mMIMO. Handheld devices are typically not large enough to accommodate more than two sub-1 GHz antennas. Due to the low frequencies and the general congestion in the spectrum area there is typically only two 10 MHz Frequency Division Duplex (FDD) challenges available.

1.6.2 Sub 6 GHz bands

	LTE 1800 with 2X2 MIMO		5G NR 3500 with mMIMO
Frequency band	1.8 GHz		3.5 GHz
Channel size	20 MHz	10 – 20 X	100 MHz
Spectral efficiency	2 bps/Hz		4-8 bps/Hz
Cell throughput	40 Mb/s		400 – 800 Mbps

Illustration 13: 5G vs 4G capacity per cell

The available spectrum in the 2GHz area was taken by IMT-Adv for LTE and LTE-Adv. As such the only bandwidth available for assignment to IMT-2020 is the 3.4 – 3.8 GHz. This is suitable for services in dense urban coverage areas. High rates are possible in this band with capacity supporting eMBB and Service Providers have the option to reuse existing 2 GHz sites.

1.7 mmWave bands

mmWave bands take their name from the fact that the wavelength (λ) ranges from 1 mm - 10 mm. There is an abundance of spectrum available that is capable of delivering extreme data speeds and capacity.

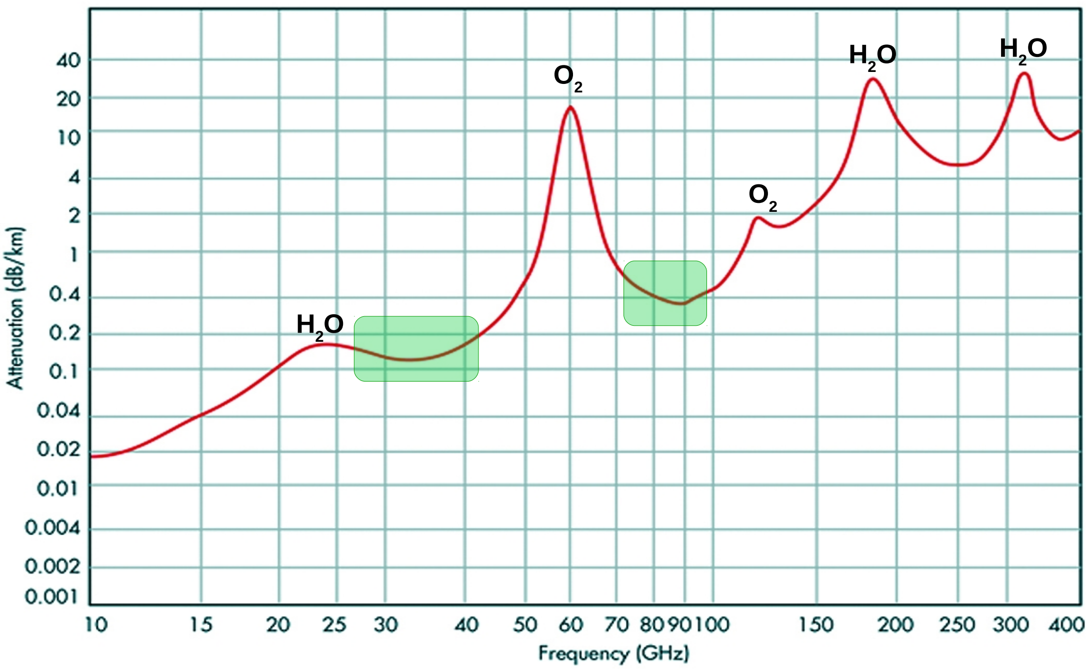

Illustration 14: mmWave attenuation -vs- frequency

Electromagnetic waves are absorbed in the atmosphere according to wavelength. The majority of signal absorption is caused by oxygen (O_2) and water vapour (H_2O). Referring to Illustration 14 it can be seen that the first peak occurs at 22 GHz due to water, and the second at 63 GHz is due to oxygen. The actual amount of water vapour and oxygen in the atmosphere normally declines with an increase in altitude because of the decrease in pressure, so these graphs apply from sea level to around 1 km altitude.

Total attenuation through the atmosphere at any frequency through unobstructed atmosphere is the sum of free space path loss, attenuation caused by oxygen absorption and attenuation caused by water vapour absorption. Rain attenuation, when present adds an additional element.

So,

$$Atten_{Total} = Atten_{FreeSpacePathLoss} + Atten_{Oxygen} + Atten_{WaterVapor} + Atten_{Rain}$$

Illustration 14 demonstrates also that there is less atmospheric attenuation at areas marked in green. The frequencies have been allocated to 5G NR and their corresponding bandwidths are given in the table below. Up to 40 GHz carriers are aggregated to achieve higher bandwidth of 1 GHz. Whereas above 40 GHz bandwidths from 500 MHz to 2 GHz can be achieved without carrier aggregation.

Band	Bandwidth
28 GHz	500 MHz
38 GHz	1 GHz
72 GHz	2 GHz

A characteristic of propagation at these frequencies is that transmission signals suffer from much higher path loss.

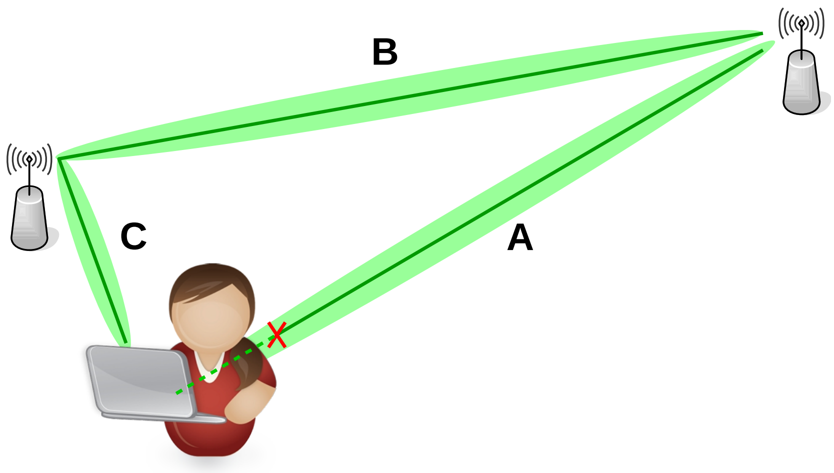

Illustration 15: mmWave communication paths

Referring to Illustration 15 it can be seen that simply the body position of the user can block the mmWave signal A, however the signal can be received by relaying via paths B and C.

Signals can be blocked by:

- Buildings
- Hand blocking
- People blocking
- Self-body blocking

So the characteristics of mmWave communications are short transmission paths, high propagation losses and tiny antennas which support multi-element, dynamic beamforming small enough to fit into handsets. By using an array of antennas that can relay to each other it is possible to provide eMBB services as hotspots in shopping centres, stadia and public areas.

1.8 mmWave 5G cellular network architecture

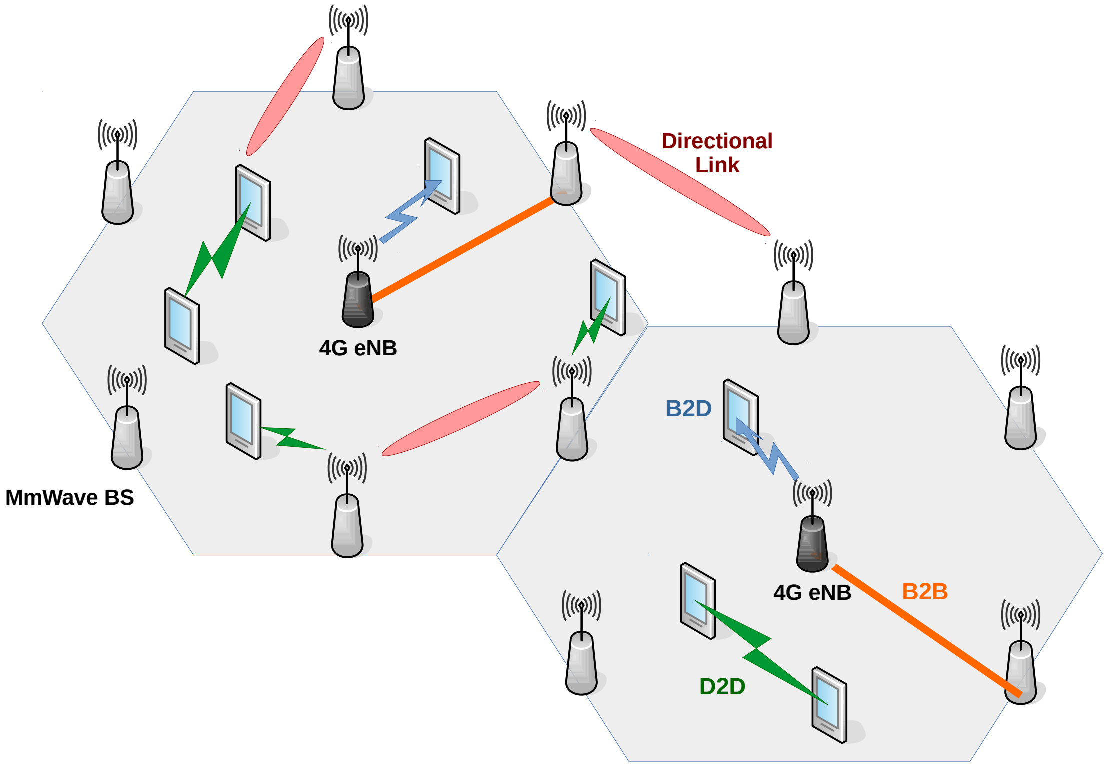

Illustration 16: mmWave 5G cellular network architecture

Illustration 16 demonstrates the mix of communications paths that the 5G NR network must support. D2D communication provides the connection between two wireless devices either directly or by hopping. Local D2D communications are formed by establishing paths between two wireless devices associated with the same basestation. Global D2D communications connect two wireless devices associated with different basestations by hopping via backbone networks. Basestation-to-Basestation (B2B) on 4G LTE is provided by fibre links however with mmWave and beamforming with a highly directional antenna on 5G NR Device-to-Basestation (D2B).

1.9 5G NR Architecture

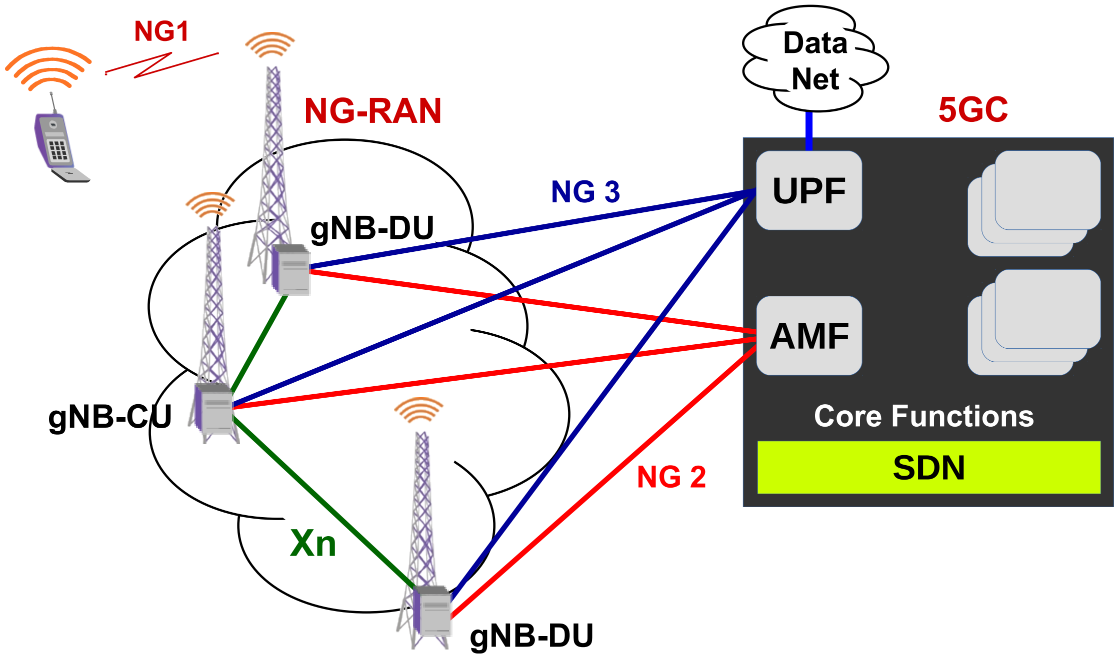

Illustration 17: 5G NR Reference Architecture

The 5G architecture will support data connectivity and services enabling deployments to use modern networking techniques such as Software Defined Networking (SDN) and Network Functions Virtualisation (NFV). Some key principles and concept are to:

- Separate the User Plane (UP) functions from the Control Plane (CP) functions, allowing independent scalability, evolution and flexible deployments such as deployments at either a centralised or distributed location.
- Modularise the function design to enable flexible and efficient network slicing.
- Wherever applicable, define procedures between network functions as services to facilitate re-use is possible.
- Enable each network function to interact with other network functions directly if required.
- Minimise dependencies between the Next Generation Radio Access Network (NG-RAN) and the 5G Core Network (5GC). The architecture is defined with a converged 5GC with a common NG-RAN to 5GC interface which integrate different Access Types e.g. 3GPP access and non-3GPP access.
- Support a unified authentication framework.
- Support *stateless* Network Functions, where the *compute* resource is decoupled from the "storage" resource.

- Support concurrent access to local and centralised services. To support low latency services and access to local data networks, User Plane Functions (UPF) can be deployed close to the NG-RAN.
- Support roaming with both home routed traffic as well as local breakout traffic in the visited Public Land Mobile Network (PLMN).

1.10 New Generation Access Network (NG-RAN)

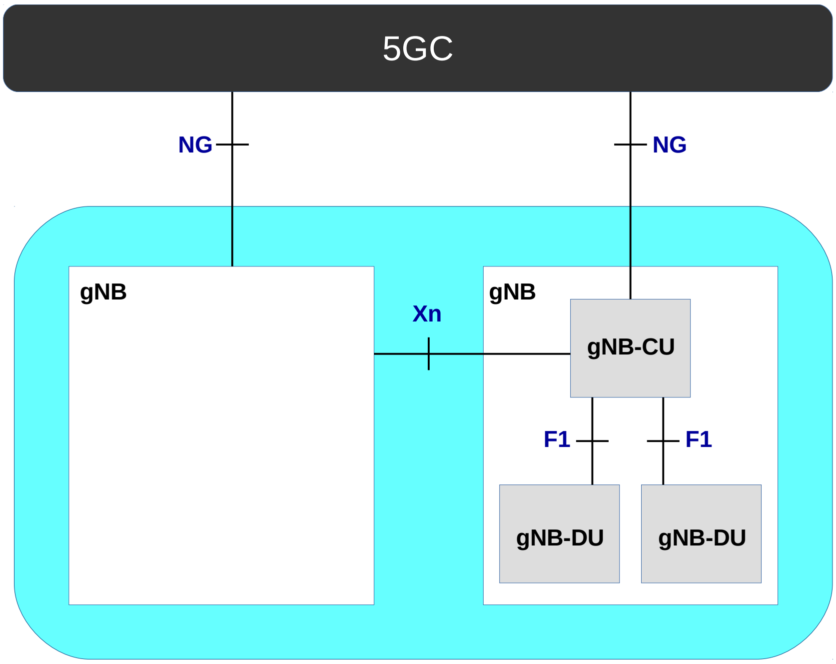

Illustration 18: NG-RAN

The NG-RAN consists of a set of Next Generation NodeBs (gNB) connected to the 5GC through the NG interface.

A gNB can support both FDD mode, Time Division Duplex (TDD) mode or a dual mode operation supporting both. gNBs can be interconnected through the Xn interface.

There are two models, stand alone gNBs connected to the 5GC via the NG interfaces or a model where a gNB may consist of a gNB Central Unit (gNB-CU) and one or more gNB Distributed Units (gNB-DU). A gNB-CU and a gNB-DU are connected via the F1 interface. One gNB-DU is connected to only one gNB-CU. Note that for resiliency, a gNB-DU may be physically connected to multiple gNB-CU by appropriate implementation, however only one is active.

The NG and Xn-C (Control) interfaces for a gNB consisting of a gNB-CU and gNB-DUs, terminate in the gNB-CU. The gNB-CU and connected gNB-DUs are only visible to other gNBs and the 5GC as a single gNB.

1.10.1 Function of the gNB

The gNB hosts the following functions:

- Functions for Radio Resource Management: Radio Bearer Control, Radio Admission Control, Connection Mobility Control, Dynamic allocation of resources to UEs in both uplink and downlink
- IP header compression and encryption of user data stream
- Selection of a Core Access & Mobility Management Function (AMF) at UE attachment when no routing to an AMF can be determined from the information provided by the UE
- Routing of UP data towards UPF
- Scheduling and transmission of paging messages originated from the AMF
- Scheduling and transmission of system broadcast information
- Measurement and measurement reporting configuration for mobility and scheduling.

1.11 Cloud Native 5G Core Network (5GC)

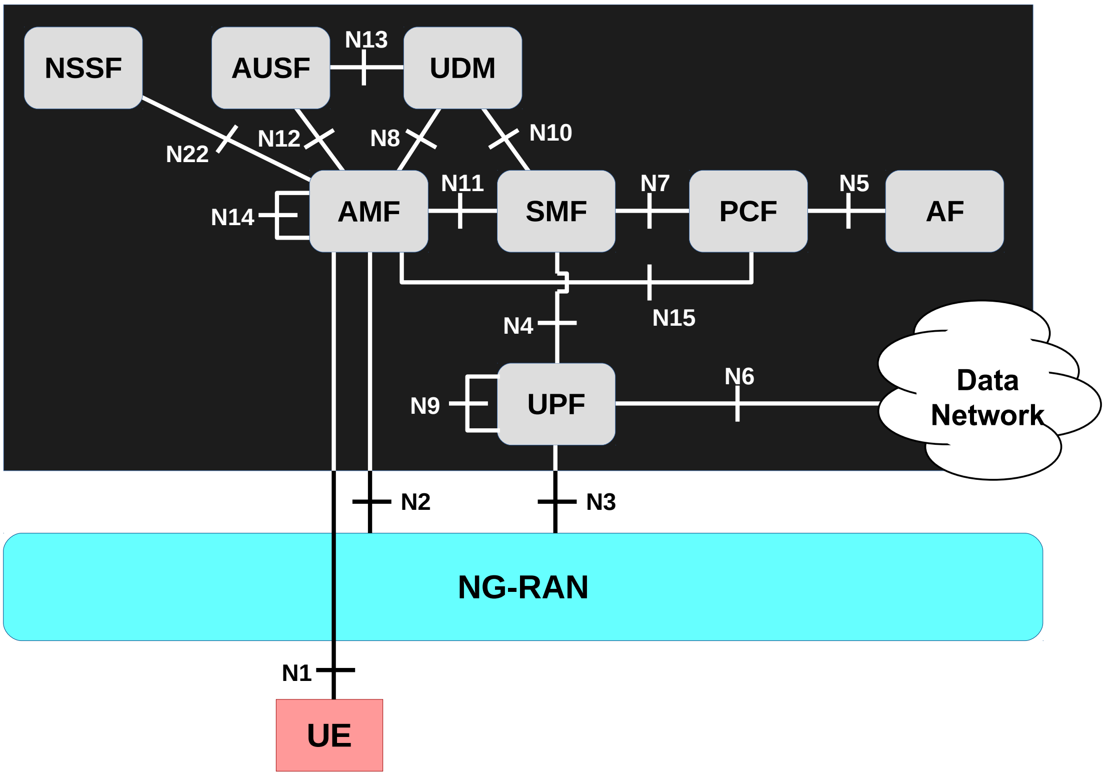

Illustration 19: 5GC Basic Architecture

Illustration 19 of the 5GC basic architecture is a representation of both a service based and a reference point based view of the interaction between network functions is represented in two ways.

- Service based representation, where network functions like the AMF within the CP enables other authorised network functions to access their services. This representation also includes point-to-point reference points where necessary.
- Reference point representation, shows the interaction that exists between the network functions. For example the reference point N11 between the AMF and the Session Management Function (SMF).

1.11.1 Network Functions and entities

The 5G System architecture consists of the following network functions (NF).

- Authentication Server Function (AUSF)
- Access and Mobility Management Function (AMF)
- Unstructured Data Storage Function (UDSF)
- Network Exposure Function (NEF)
- Network Function Repository Function (NRF)
- Network Slice Selection Function (NSSF)
- Policy Control Function (PCF)
- Session Management Function (SMF)
- Unified Data Management (UDM)
- Unified Data Repository (UDR)
- User Plane Function (UPF)
- Application Function (AF)
- User Equipment (UE)
- Next Generation - Radio Access Network (NG-RAN)
- 5G-Equipment Identity Register (5G-EIR)
- Security Edge Protection Proxy (SEPP)

1.11.2 Reference Points

The 5G System Architecture contains the following reference points:

- N1: Reference point between the UE and the AMF.
- N2: Reference point between the NG-RAN and the AMF.
- N3: Reference point between the NG-RAN and the UPF.
- N4: Reference point between the SMF and the UPF.
- N6: Reference point between the UPF and a Data Network.
- N9: Reference point between two UPFs.

The following reference points show the interactions that exist between the Network Function services.

- N5: Reference point between the PCF and an AF.
- N7: Reference point between the SMF and the PCF.
- N24: Reference point between the visited network PCF and the home network PCF.
- N8: Reference point between the UDM and the AMF.
- N10: Reference point between the UDM and the SMF.
- N11: Reference point between the AMF and the SMF.
- N12: Reference point between AMF and AUSF.
- N13: Reference point between the UDM and Athe AUSF.
- N14: Reference point between two AMFs.
- N15: Reference point between the PCF and the AMF.
- N16: Reference point between two SMFs.
- N17: Reference point between AMF and 5G-EIR.
- N18: Reference point between any NF and UDSF.
- N22: Reference point between AMF and NSSF.
- N27: Reference point between visited network NRF and the home network NRF.
- N31: Reference point between visited network NSSF and the home network NSSF.

1.11.3 5GC main functions

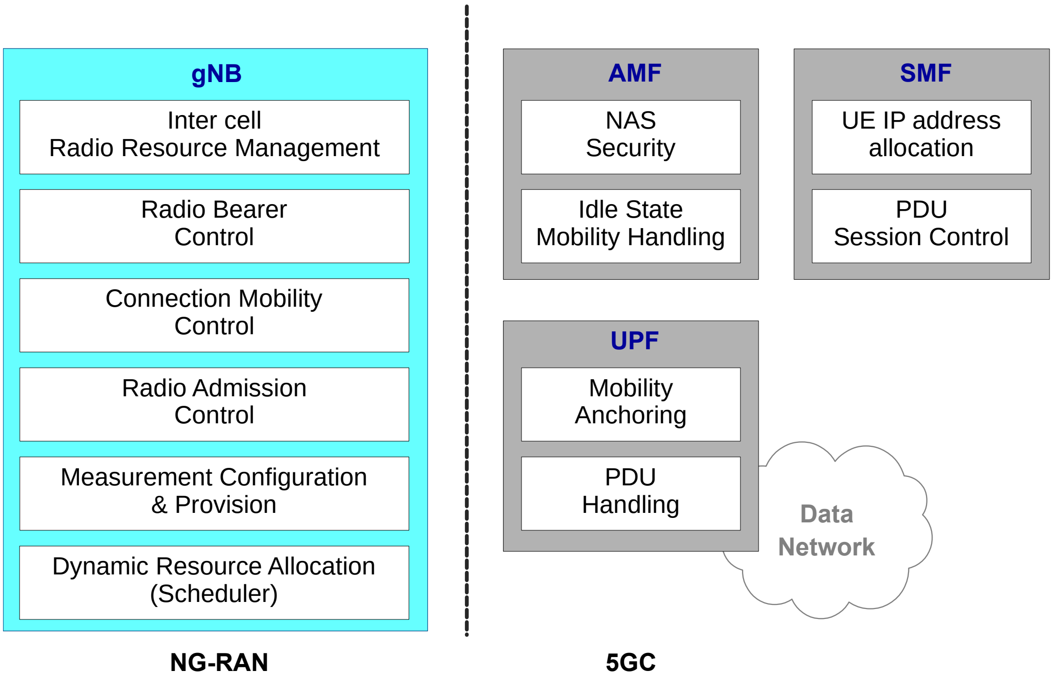

Illustration 20: Functional Split between NG-RAN and 5GC

The AMF is responsible for communications between the UE and the 5GC. This functional layer is called the Non Access Stratum (NAS). The AMF handles UE reachability, access authentication, authorisation, roaming rights as well as signalling termination and security to the UE.

The UPF has the responsibility for the user plane. It is a router the carries out functions like mobility anchoring, Packet Data Unit (PDU) routing & forwarding, policy rule enforcement, traffic usage reporting, Quality of Service (QoS), packet marking in the transport layer, packet buffering as well as the interconnection to the Data Network.

The SMF manages sessions. It allocates and manages Internet Protocol (IP) addresses. It selects and controls the user plane functions. It steers traffic at the UPF to ensure traffic is routed to its proper destination. It handles the control part of policy enforcement and QoS;

1.11.4 Network Slicing

Network slicing allows multiple virtual networks to be created on top of a common shared physical infrastructure. These virtual networks are then customised to meet the specific needs of applications, services, devices, customers or operators.

In 5G, a single physical network can be sliced into multiple virtual networks that can support different NG-RAN, or different service types running across a single NG-RAN. It is envisaged that network slicing will primarily be used to partition the 5GC, but it may also be implemented in the NG-RAN.

1.12 Comparison with 4G LTE

Compared to LTE the big differences are:

- The 5GC control plane has been split into AMF and SMF nodes. A given device is assigned a single AMF to handle mobility and Authentication, authorisation, and Accounting (AAA) roles but can then have multiple SMF each dedicated to a given network slice.
- The 5GC user plane is handled by single node UPF with support for multiple UPF serving the same device and hence avoiding the need for a common SGW used in LTE. UPF nodes may be daisy chained to offer local breakout and may have parallel nodes serving the same APN to assist seamless mobility.

2. Self-test Quiz

1. Describe the three use cases for 5G.
2. What is mmWave and what 5G use case is it most appropriate to ?
3. Describe the 5GC functions:
    - AMF
    - UDP
    - SMF.
4. How does 5G avoid the need for a common SGW used in LTE ? And why does it need too ?