5G Networking

Generations of cellular technologies

	1G	2G	3G	4G	5G
3GPP releases			4 - 7	8-9, 10-14	15, 16
era	1980s	1990s	2000s	2010s	2020s
services	analog voice	digital voice messages	WB voice packet data	voice, video Internet, apps	everything
devices					
data rate	0	100 kbps (GPRS)	10 Mbps (HSPA)	100+ Mbps (LTE/LTE-A)	10 Gbps (NR)
delay		500 ms	100 ms	10s ms	5 ms

What's wrong with 4G?

4G made possible:

- fast Internet access
- video reception and creation
- apps relying on location and identity
- always-on behavior

but suffers from numerous limitations:

- for some applications: data-rate too low
- for some applications: delay too high
- too few simultaneous connections (insufficient density)
- coverage too low / drop rate too high
- weak (if any) QoS guarantees
- price per bit too high (inefficient spectral use)
- power consumption too high (and thus battery life too low)
- poor support for new applications/markets (e.g., IoT, AR/VR, connected cars)
- no support for new mobility requirements (mobile hot spots, high speed)
- insufficient security/privacy

5G is being developed
to address 4G limitations

5G importance

5G will enable entirely new market segments

- *ubiquitous* wireless broadband (eMBB)
  - Wireless-to-the-Premises
  - broadband on high-speed trains, at events and in crowds
- *massive* IoT (20 B IoT devices connected by 2020)
- V2X (vehicle to vehicle, vehicle to infrastructure)
- Virtual Reality / Augmented Reality

A study led by Qualcomm finds that by 2035

- $3.5 trillion in 5G direct yearly revenues
- $12.3 trillion worth of goods and services will be enabled by 5G
- 22 million jobs will be attributable to 5G
- 5G will boost global GDP growth by $3 trillion (compared to 2020)
  about the GDP of India

5G is coming fast!

You may think that 5G is *futuristic*, but it is coming *fast*

In June 2016, 3GPP accelerated the standardization work-plan

- 5G phase 1 (release 15) finished by June 2018 for trials in 2019
- 5G phase 2 (release 16) finished by March 2020

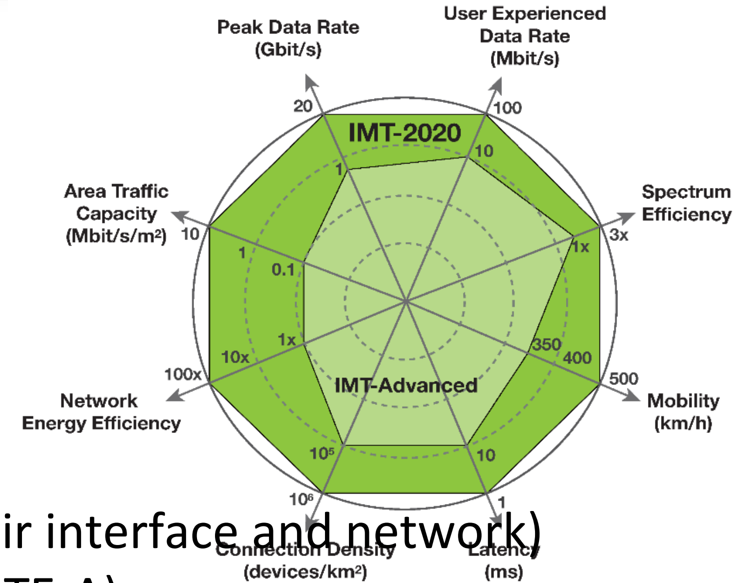
and SPs/vendors are not even waiting for 2019!

- AT&T (Ericsson, Samsung, Nokia, Intel) are in trials in 5 cities and hope to roll out nationwide by late 2018
- Verizon (Nokia, Ericsson) conducting WTTP trials (10G connectivity!)
- Telus (Huawei) demoed Wireless-to-the-Premises with Huawei
- Orange (Ericsson) announced testing of selected use cases
- Softbank and China Mobile installing thousands of massive MIMO towers
- NTT DOCOMO (Ericsson, Intel) trial environment in central Tokyo
- KT promised 5G coverage along Highway 50 ahead of 2018 Olympic games
- AirtelBSE, Vodafone India, Reliance Jio (ZTE) have conducted trials
- ***You'll see 5G in 2019 for sure*** Qualcomm CEO Steve Mollenkopf

IMT-2020 goals

The ITU is defining performance targets for 5G
that are 10 to 100 times *more* than 4G* :

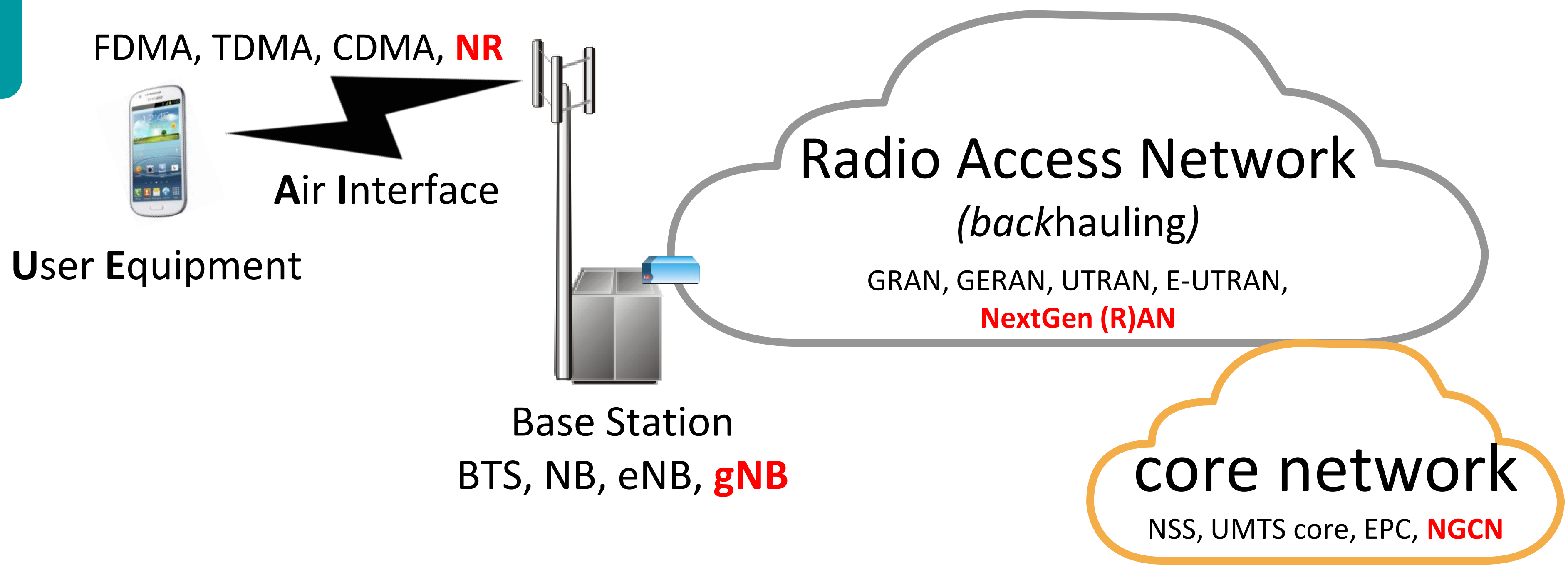
- **Peak data rate** (20 Gbps/device)
- **User experienced data rate** (100 Mbps)
- **Latency** (1 ms)
- **Mobility** (500 km/h and seamless transfer)
- **Connection density** (10^6 devices/km^2)
- **Energy efficiency** (1/100 Joule/bit for both air interface and network)
- **Spectrum efficiency** (3 times the bps/Hz of LTE-A)
- **Area traffic capacity** (10 Mbps/m^2)



However, it is not possible to attain all of these *at the same time*
so 5G recognizes usage scenarios
and includes mechanisms to separate traffic types (slicing)

* ITU-R M.2083-0

5G changes all cellular segments



Multiple 4G/5G co-existence options have been proposed

- Standalone (e.g., eNB-EPC and gNB-NGCN)
- Non-Standalone (gNB-LTE-EPC, gNB-EPC, eNB-NR, eNB-NGCN, ...)

Initial 5G will be Non-Standalone gNB connecting to LTE RAN and EPC

Upgrading the air interface

How can we improve the radio link from the gNb to UE ?

- more efficient modulation (**N**ew **R**adio)
- wider spectral bands (100 Mhz, 1GHz), which requires
  - new RF bands (sub-1GHz, 24/28 GHz mmWave, 30-90GHz ?)
  - use of licensed/unlicensed unshared/shared spectrum
- higher cell density (i.e., many more antennas)
  - small cells
  - HetNets
- massive MIMO
  - massive is defined as 16 or more antennas (4-by-4 array)
  - 5G envisages up to 256 antennas
    - LTE: 4UL/8DL 802.11n: 4, 802.11ac: 8
    - in LTE's CoMP UE connects to multiple cells to reduce intercell interference
  - < 6 GHz use multipath for spatial mux and multiuser MIMO
  - > 6 GHz use *coherent beamforming* (i.e., personal cells)

NR evolved from LTE

Gen	Modulation	Duplexing	MA	MIMO
1G	analog FM	FDD	FDMA	–
2G	QPSK/MSK	FDD	TD/FD/CD-MA	–
3G	xPSK	FDD/TDD	WCDMA	–
4G - DL	OFDM QPSK/16QAM/64QAM	FDD/TDD	OFDMA	1x1/2x2/4x4
4G - UL	SC-OFDM QPSK/16QAM/64QAM		SC-FDMA	

LTE OFDM scheme (1)

LTE defined 1.4, 3, 5, 10, 15, and 20 MHz channel bandwidths

Subcarrier spacing is always 15 kHz

and thus OFDM requires the symbol rate to be 15 ksps (30, 60, or 90 kbps)

and the symbol duration to be 1/15kHz = 66.7 μsec

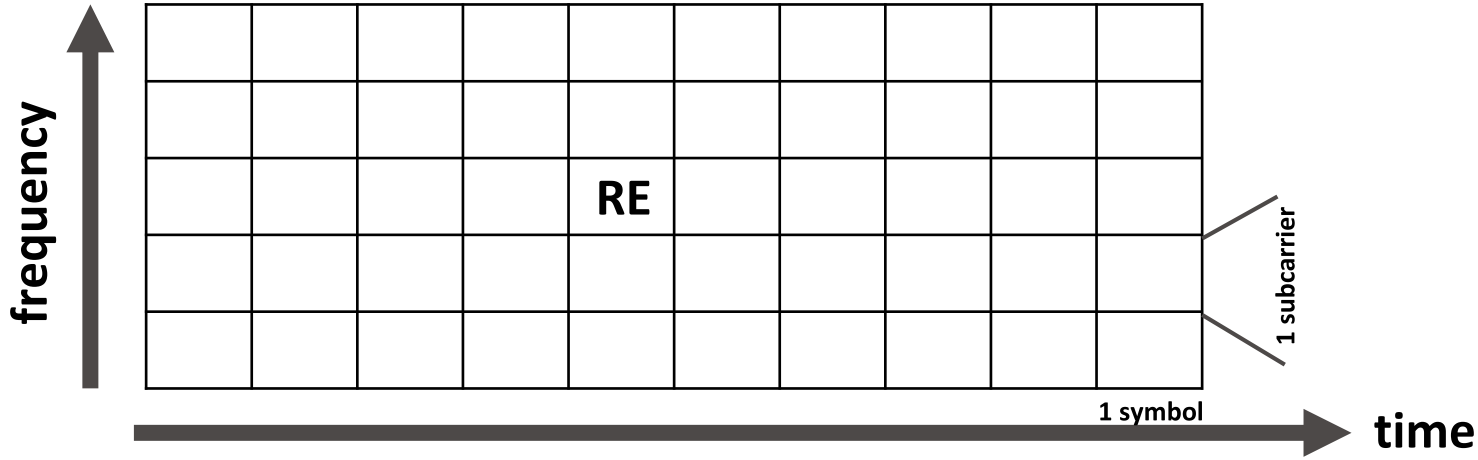
The total number of subcarriers is constrained to be a multiple of 12

(subcarriers are allocated in blocks of 12 subcarriers each)

Channel BW	subcarriers	RBs	useable BW (- guard)
1.4 MHz	72	6	1.08 MHz
3 MHz	180	15	2.7 MHz
5 MHz	300	25	4.5 MHz
10 MHz	600	50	9 MHz
15 MHz	900	75	13.5 MHz
20 MHz	1200	100	18 MHz

LTE OFDM scheme (2)

The time/frequency plane is naturally divided into **R**esource **E**lements
of 66.7 µsec by 15 kHz (i.e., 1 symbol on 1 subcarrier)



Each symbol is divided into 2048 times Ts, so the *sampling* rate is 30.720 MHz

Due to several µsec of multipath
we need to add a **C**yclic **P**refix to each symbol

An LTE *slot* is 0.5 ms in duration
and contains 6 or 7 symbols with their Cyclic Prefix (CP)

- normal mode: 1 long CP of 160 Ts = 5.2 µs + 6 regular CP of 144 Ts = 4.7 µs
- extended mode: 6 CP of 512 Ts = 16.7 µs

LTE OFDM scheme (3)

The smallest unit that can be allocated to a user is a

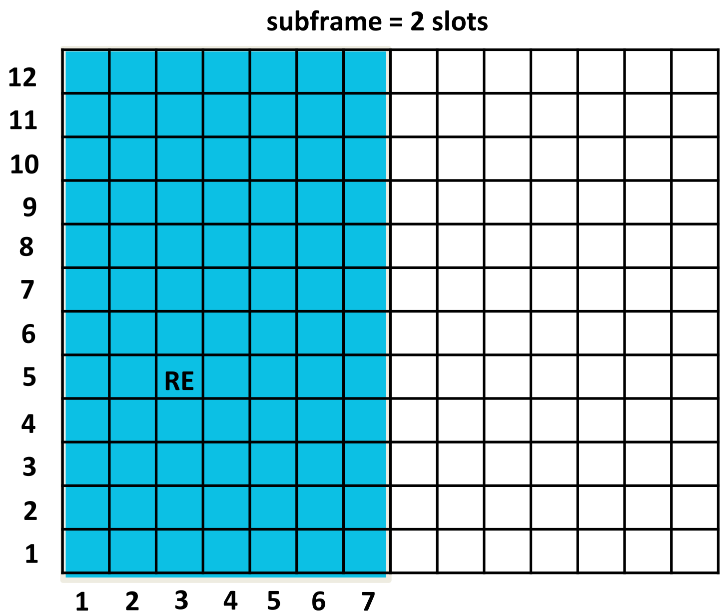
Resource **B**lock = 12 subcarriers (180 kHz) * 1 slot (0.5 ms)

1 RB = 12*7 REs = 84 REs

While allocation is per RB, scheduling is per subframe

An LTE frame is always 10 ms and subframe 1 ms

- FDD frames have ten 1 ms subframes of 2 slots each
- TDD frames have two 5 ms half-frames of 5 subframes each



5G OFDM scheme

Channel BW can be 5, 10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 200, 400 MHz
<6GHz 5 ... 100 MHz >6GHz 50, 100, 200, 400 MHz

5G has subcarrier spacing of 2^n * 15 kHz where n=0...5 (0 ... 2 for sub 6 GHz)
i.e., 15, 30, 60 (optional), 120, 240, 480 kHz

All cases use normal CP except n=2 (60 kHz) which can use an extended CP

Each subcarrier can use *higher order* modulation (up to 256 QAM)

Each slot is 14 symbols for normal CP (and 12 symbols for extended CP)

Thus a slot lasts 1 ms (15 kHz), 500 μsec (30 kHz), 250 μsec (60 kHz), ...

So, for 15 kHz 1 1 ms subframe is 1 slot, for 30 KHz it is 2 slots, ...

A slot can be DL, UL, or flexible (mixed)

Slots can be *aggregated* for higher data rates

For low latency, allocation can also be per minislot (2, 4, or 7 OFDM symbols)
to achieve low latency (LTE now has sTTI for similar reasons)

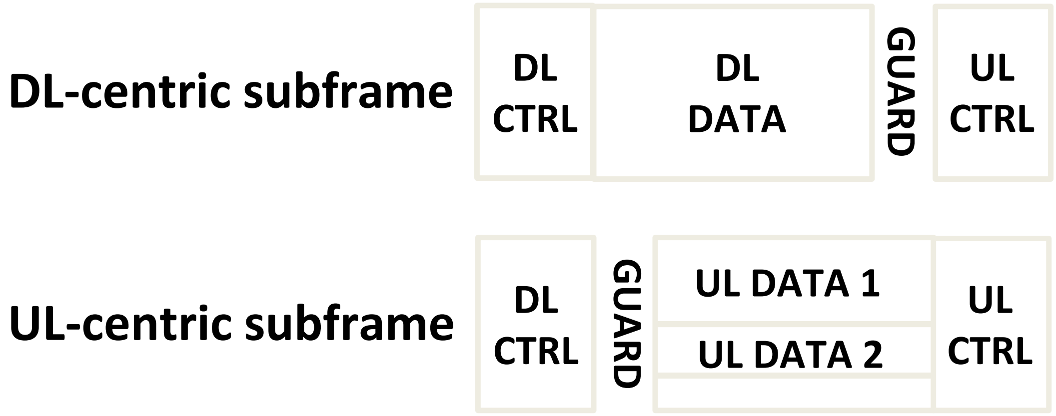
REs are grouped into **P**hysical **R**esource **B**locks of 12 subcarriers

Self contained integrated subframe

Another innovation enables transmission of both directions in one subframe

This enables for

- URLLC: quick acknowledgements
- unlicensed spectrum: blank subframes and messaging for **L**isten **B**efore **T**alk
- device-device communications: link direction and scheduling messaging
- massive MIMO: UL control and sounding



New FEC

5G needs support for

- higher reliability
- lower SNRs
- higher rates (lower complexity)

LTE used

- Turbo codes for data channels
- Reed Muller or tail-biting convolutional codes for control channels

5G uses

- **L**ow **D**ensity **P**arity **C**heck codes for data channels
- Polar codes for control channels

LTE has HARQ feedback for fast retransmission

5G has enhanced

Carrier Aggregation

CA was introduced in LTE-A Release 10 and enhanced Release 11

We saw that LTE spectrum is allocated in units of 20MHz

A mobile operator may have 2 contiguous units

but more probably has noncontiguous units – even in different RF ranges

CA enables a network operator to combine radio channels (FDD or TDD)

- contiguous
- noncontiguous within the same frequency band
- across different bands
- between licensed and unlicensed bands
- across cells (e.g., to support HetNets)

Note that 4G/5G *dual connectivity* is *not* CA

Spectrum

Before 5G, cellular traffic was limited to 300 MHz – 3 GHz licensed bands

5G needs bandwidth, and can use anything it can get

This includes licensed, unlicensed, and shared spectrum

RF ranges include:

- low and mid frequencies for macro cells
- mm waves for small and ultra-small cells (LoS 300m, urban multipath 150m)

ITU-R is working on allocation, for now, bands being considered:

	low (MHz)	mid (GHz)	high (GHz)
US	600	3.5	28, 37-40, 64-71
EU	700	3.4-3.8	24.25-27.5
China		3.4-3.6	24.25-27.5
Japan		3.6-4.2, 4.4-4.9	27.25-29.5
South Korea			28, 37.5-40

Note: WiFi 802.11ad already operates in the 60 GHz band

MIMO

LTE defined 2x2 (2 transmit antennas + 2 receive antennas) and 4x4 MIMO

MIMO was used for

- single user – to exploit spatial diversity
- multiple user (MU-MIMO) – to conserve bandwidth

5G allows more advanced MIMO

- spectral reuse and cell edge improvement
- massive MIMO – up to 256 antennas
- >6GHz
  - smaller antennas so higher order MIMO
  - beamforming and beam-tracking
- 2D antenna arrays for 3D beamforming

Device centric mobility

LTE uses *network-centric mobility*

- Nbs periodically transmit reference signals
- UE receives reference signals from all cells in its neighborhood
- UE sends measurements to network
- network triggers handover

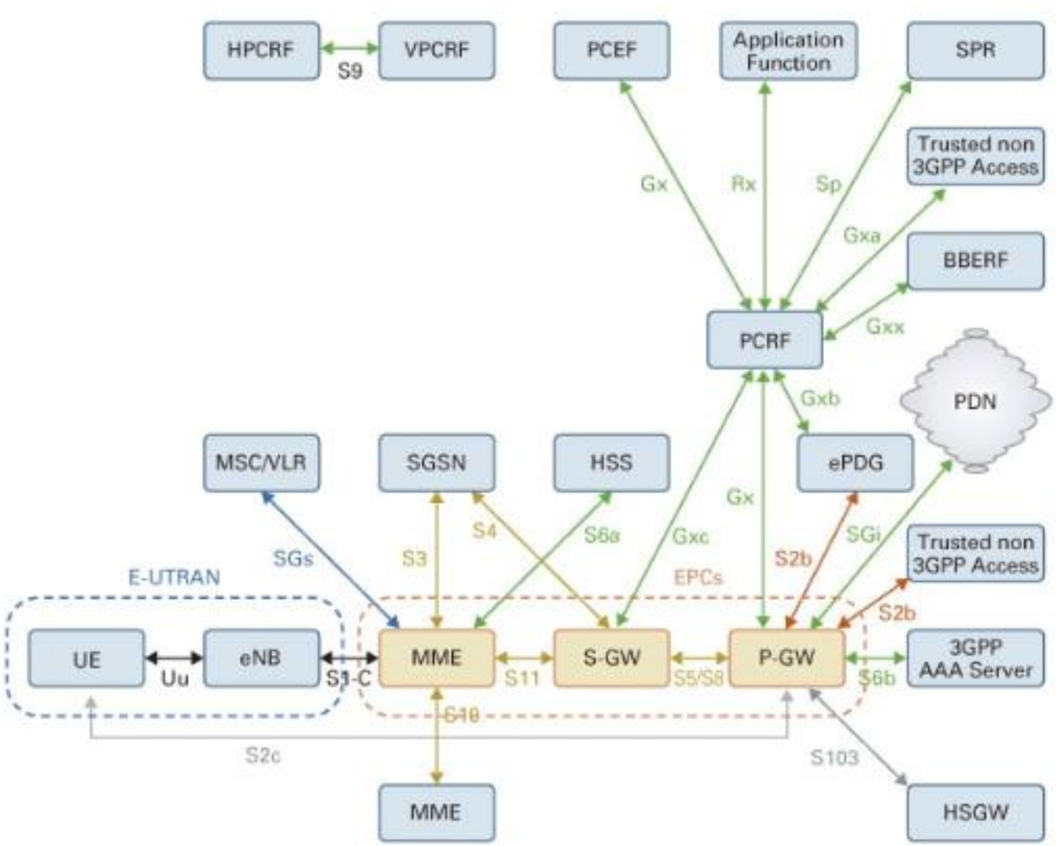
5G uses *device-centric mobility*

- UE periodically transmits reference signals
- gNbs measure strength and trigger handover

Since gNbs don't need to transmit reference signals
    they transmit only initial discovery beacon
    full information transmitted on-demand

LTE Architecture and Interfaces

Policy & **C**harging **R**ules **F**unction **P&C E**nforcement **F**unction **S**ubscriber **P**rofile **R**epository
Serving **G**PRS **S**upport **N**ode **H**ome **S**ubscriber **S**erver **P**acket **D**ata **G**ateway / **N**etwork
Evolved **U**MTS **T**errestrial **R**adio **A**ccess **N**etwork **B**earer **B**inding & **E**vent **R**eporting **F**unction
User **E**quipment **M**obility **M**anagement **E**ntity **S**erving **– G**ate**W**ay **P**DN **- G**ate**W**ay



5G Architecture Principles

Modular Function design based on **N**etwork **F**unctions – not boxes

- NF can be hardware or (virtual) software – even in cloud
- reference points between NFs
- function separation for scaling (e.g., AMF/SMF, AUSF/PCF)

Service based architecture

- define procedures as services (enable reuse)
- NFs in **C**ontrol **P**lane enable authorized functions to directly access services

Minimize dependencies between Access Network and Core Network

Multiple access technologies (multi-RAT)

- interwork with previous 3GPP generations
- interwork with non-3GPP networks

Support diverging architectures and new services

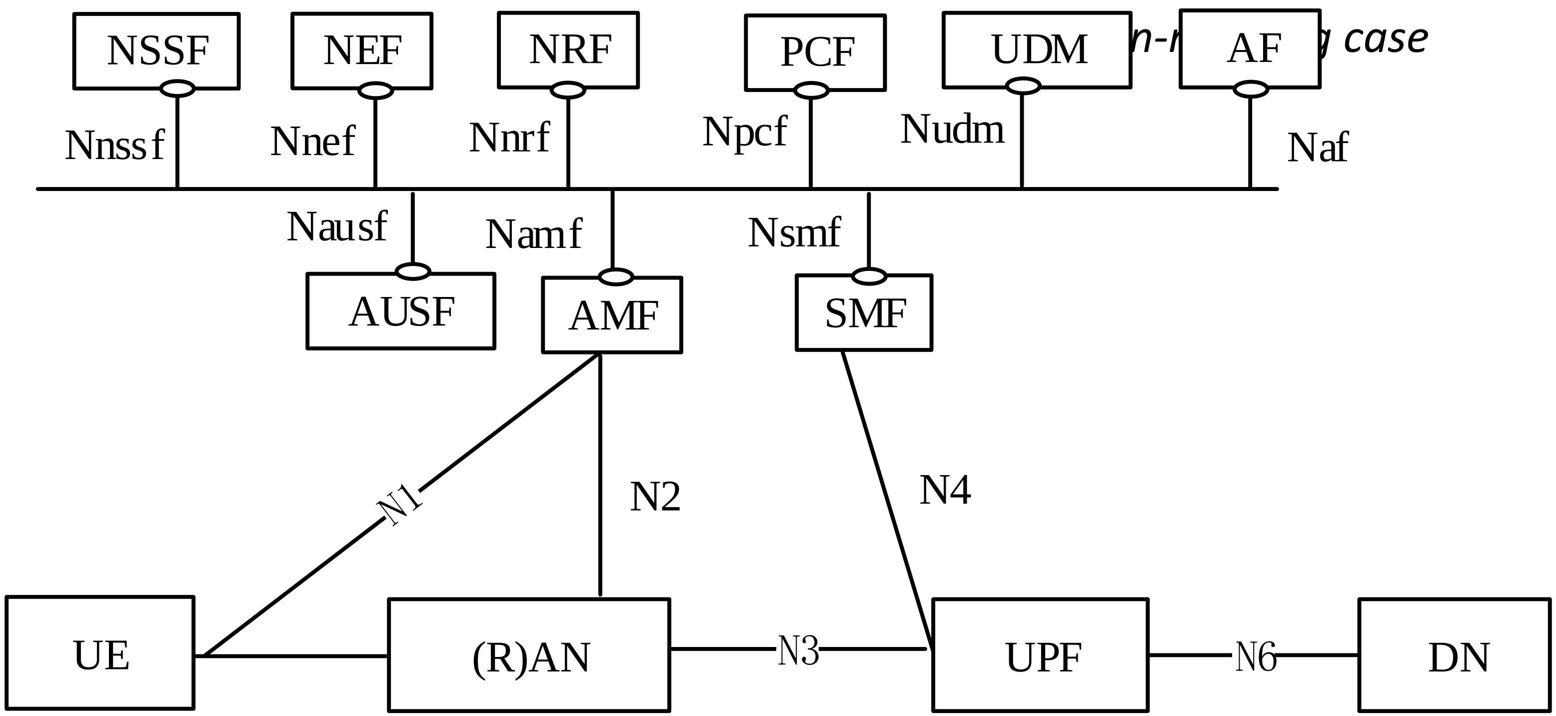
Automation and programmability designed into architecture

5G Architecture and Interfaces

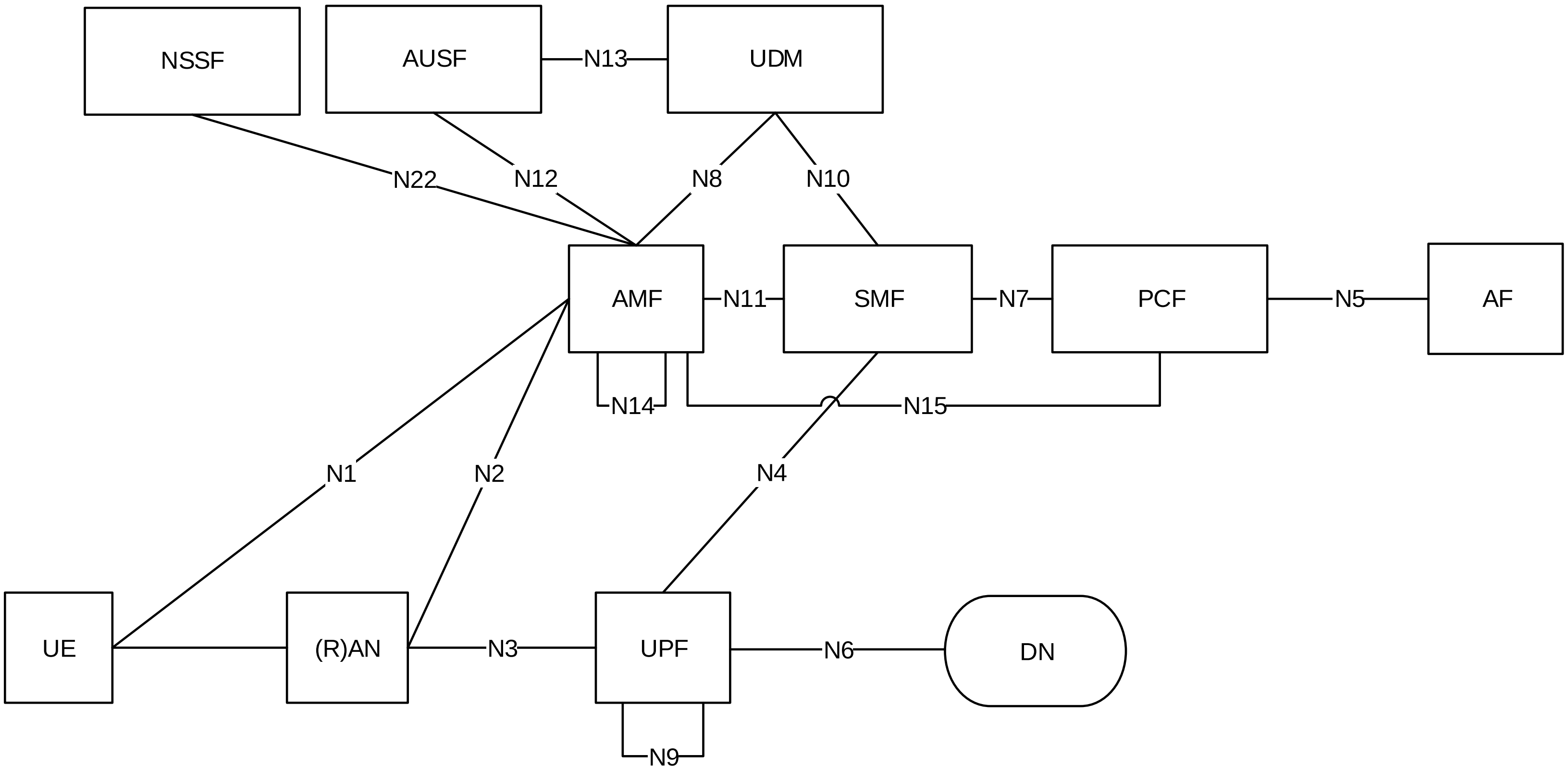
Network **S**lice **S**election **F**unction **N**etwork **E**xposure **F**unction **NF R**epository **F**unction
Policy **C**ontrol **F**unction **U**nified **D**ata **M**anagement

Authentication **S**erver **F**unction **A**ccess & Mobility **M**ngment **F**unction **S**ession **M**ngment **F**unction

User **E**quipment **(R**adio**) A**ccess **N**etwork **U**ser **P**lane **F**unction **D**ata **N**etwork



Another way to look at it



What the NFs do

AMF provides authentication, authorization, mobility management of UE

- UE (even with multiple access technologies) connected to single AMF

SMF provides session management, IP address allocation, selects/controls UPF

- UE with multiple sessions may connect to multiple SMFs

UPF provide data transfer functionalities, e.g., routing, NAT, DPI, firewall

AF provides packet flow information to PCF to support QoS

PCF determines (mobility, session) policies so AMF and SMF work properly

AUSF stores UE authentication data

UDM stores UE subscription data (replaces HLR/HSS)

DN provides Internet access.

NEF expose capabilities to partners

NSSF Network Slice Selection Function

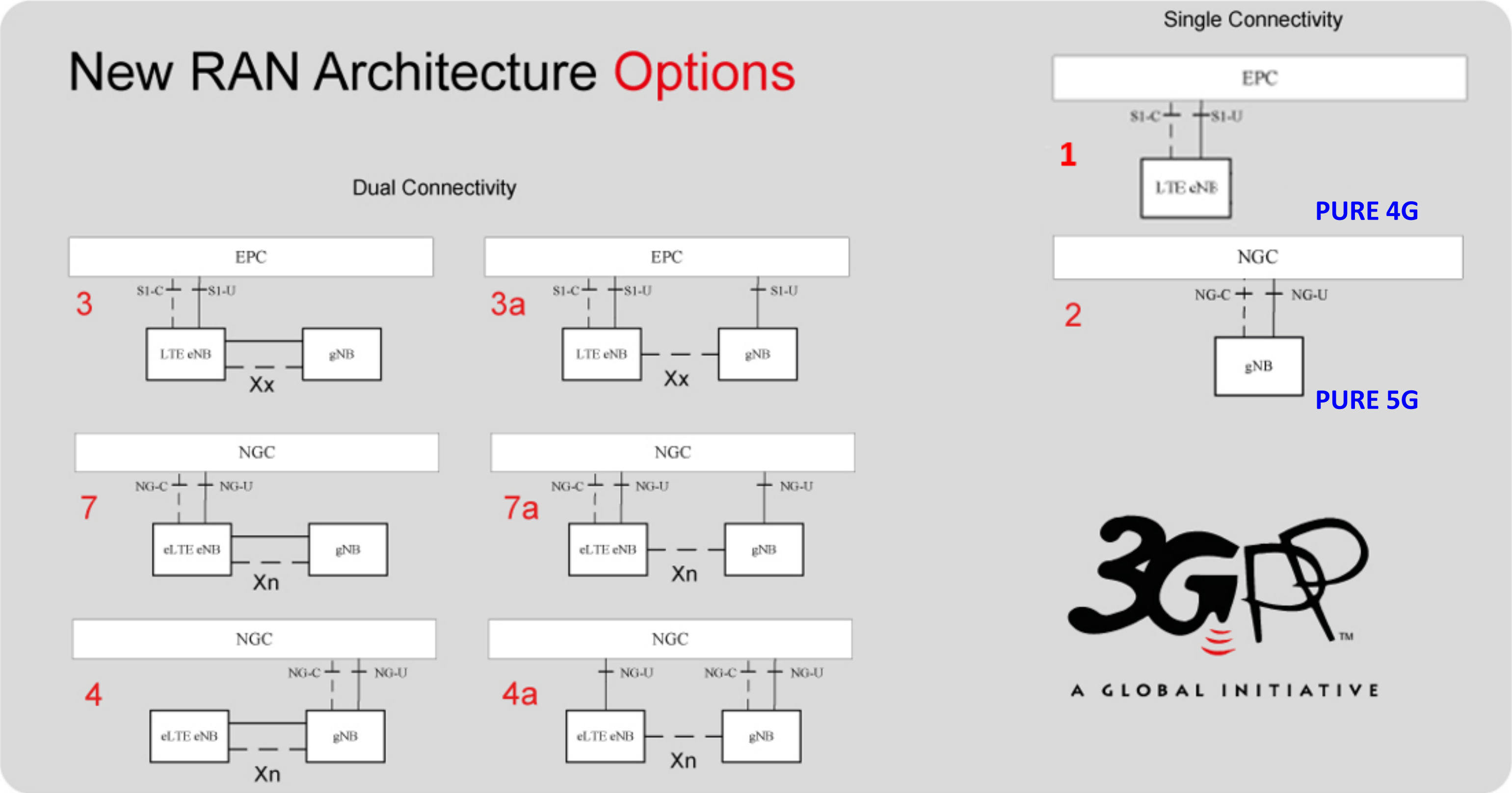
NRF Network repository function for subscribe to/for service

Migration options for core

The 5G NGC core will only be defined in Release 16

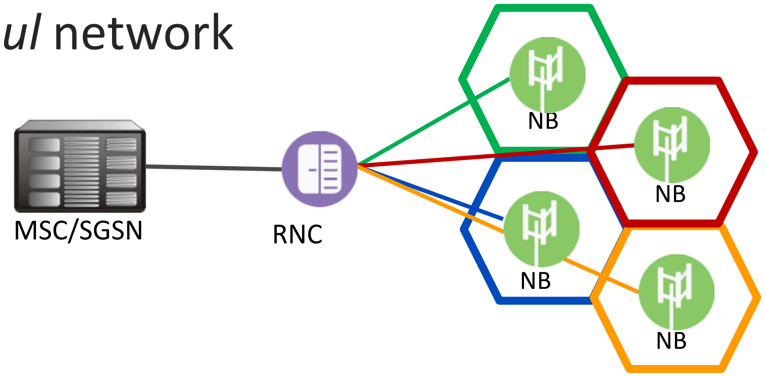
so in the meantime 5G is based on *non-standalone* (NSA) mode where the gNb connects to a 4G EPC

Once the 5G core is introduced, multiple options have been defined



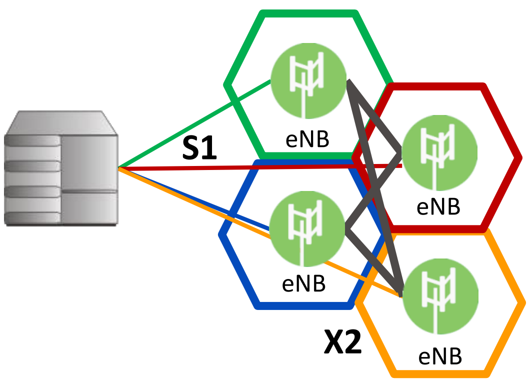
Splitting up the RAN (1)

In 3G the RAN was a pure *backhaul* network

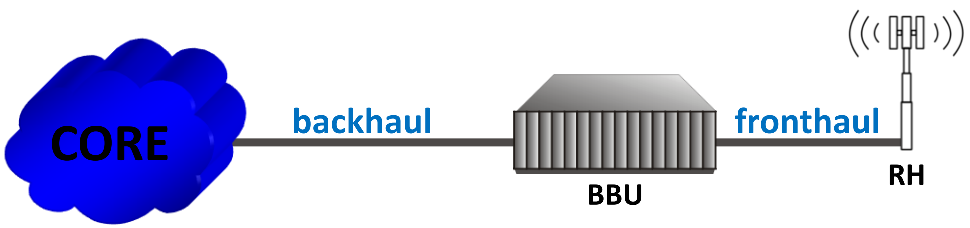


In 4G this changed in 2 ways

- the X2 interface interconnected eNBs (at least logically)



- fronthaul (CPRI)



In 5G this must change more drastically

Splitting up the RAN (2)

Fronthaul originated as a digital interface
to connect the antenna to the BBU at the foot of the tower

But then it was realized that is advantageous to fronthaul further

- share processing resources
- reduce effective delay for CoMP

This led to new cost/energy-saving network architectures:

- cRAN / cloud-RAN (BBU pooling, BBU hosteling)
- vRAN / virtualized RAN (software BBU)

Unfortunately, there is a problem with 5G fronthaul ...

The *sampling theorem* tells us that we need to sample at least twice the BW
so a single 100 MHz signal requires 200 Msps or 3.2 Gbps (without overhead)

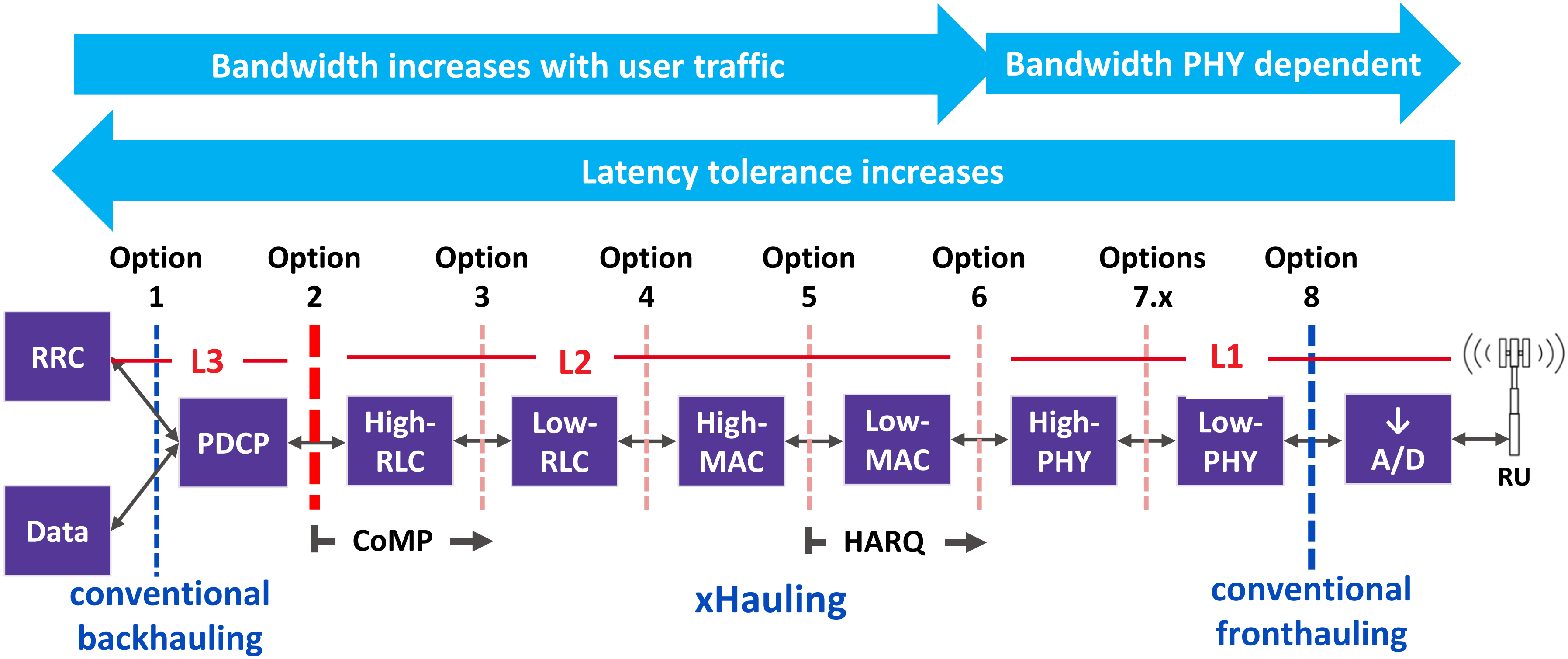
Assuming 3 sectors each with 16 MIMO antennas : > 150 Gbps

Assuming 3 sectors each with 256 MIMO antennas : > 2.5 Tbps

Assuming 1 GHz bandwidth, 3 sectors, 256 MIMO antennas : > 25 Tbps

So, we need a new kind of *front*haul

5G RAN Functional Splits



LTE uses splits 1 and 8

3GPP is working on split 2

xRAN (ORAN) is working on split 7.2

Decomposed base station transceiver

- PHY – Physical layer processing
- MAC – Medium Access Control
- RLC – Radio Link Control
- PDCP – Packet Data Convergence Protocol

Network based control functionality

- RRC – Radio Resource Control

Potential RAN transport technologies

What transport functionality is required for the Radio Access Network* ?

- XG-PON/XGS-PON/NG-PON2
  - but 10 Gbps will only be satisfactory for initial 5G deployments
- 25 GbE (802.3by), 1-lane 50 GbE (802.3cd), NG 100/200/400 GbE (802.3bs)
- FlexE (bonding, sub-rating, channelization, TDM *calendar*)
- eCPRI and IEEE 1914 Radio over Ethernet
- Time Sensitive Networking
  - IEEE 802.1CM for Ethernet fronthaul
  - IEEE 802.1Qbv scheduled traffic enhancements
  - IEEE 802.1Qbu frame pre-emption
  - IETF DetNet (Deterministic Networking) for IP

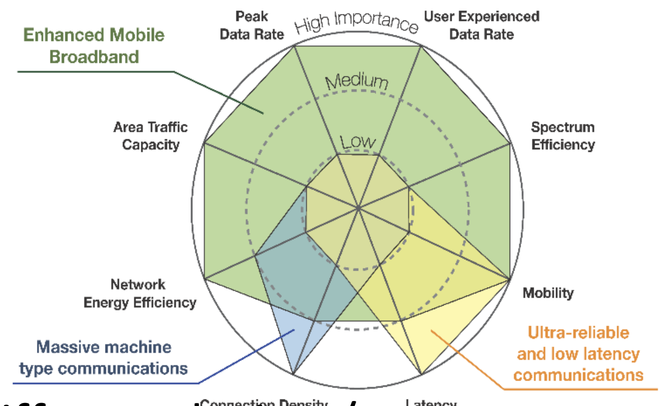
* we'll mention upgrading the core later on in the context of slicing/SDN/NFV

Network slicing

We already said that 5G can't reach all of its goals simultaneously
but it doesn't have to

For example:

- enhanced mobile broadband
  - needs high data rates
  - doesn't need very low latency
- massive IoT
  - needs high connection density
  - doesn't need high data rates



So, 5G using *network slicing* to satisfy needs of different devices/apps

Network slicing means:

- *on-demand* assignment of networking/computational resources
  - resources: bandwidth, forwarding tables, processing capacity, etc.
- resources can be physical or virtual, but
- each slices acts as a strongly isolated network or cloud
  - isolation of performance, security, and management aspects

Why slicing in 5G?

Different network slices cater to different service needs

- latency (1 ms – 10 ms – non-delay sensitive)
- data rate (kbps – Mbps – Gbps)
- QoS parameters (or Best Effort)
- battery usage
- mobility (stationary – mobile – high speed)
- range (150m – 1 km – long range)
- service pricing

23.501: S-NSSAI identifies a slice, and is composed of:

- Slice/Service type (SST) expected slice behaviour (features and services)
- Slice Differentiator (SD) optional information to differentiate between slices of the same Slice/Service type

Slice/Service type	SST value	Characteristics.
eMBB	1	slice suitable for the handling of 5G enhanced Mobile Broadband.
URLLC	2	slice suitable for the handling of ultra- reliable low latency communications.
MIoT	3	slice suitable for the handling of massive IoT.

Implementing slicing

In 5G, slicing is end-to-end, i.e., both RAN and core network must support slicing

Slicing requires programmability, flexibility, and modularity
 in order to create multiple virtual networks, each tailored for a given use case
on top of a common network

Different slices can be labeled using VLAN ID, VXLAN, IPv6 flow label, DSCP
 but a slice is different from a separate physical network or a VPN
because of dynamic set-up / release requirement

SDN techniques (AKA network softwarization) will be used to achieve slicing

- APIs provided to specify requirements
- rich service includes NFV/MEC elements
  - low latency apps benefit from server allocation close to UE
  - parts of mobile L1/L2/L3 stack itself can be virtualized and optimally located
- global orchestrator
  - has universal knowledge of network allocations and loads
  - enables rapid yet highly optimized slice creation

Timing for 5G (1)

RAN timing requirements are becoming stricter from generation to generation

Frequency and Time requirements are defined
to assure efficient and proper functioning of the air interface

Base stations require more and more accurate timing in order to

- maximize data-rates
  - minimize guard frequencies/times in order to maximize spectral efficiency
  - utilize BW boosting technologies like Carrier Aggregation (CA) and MIMO
- optimize user experience
  - smooth handover
  - reduced delay
  - Location Based Services (LBS)

Base stations (including NodeB, eNodeB, gNodeB) obtain timing from the RAN
(unless they have a local source of timing, e.g., GNSS)

So the RAN must deliver ever more accurate timing!

Timing for 5G (2)

5G requirements will be at least as strict as 4G
and some experts are speaking of them becoming significantly stricter

CPRI interfaces require very strict transport delay accuracy of 16 nanoseconds

Even if this requirement will be relaxed for some 5G splits
it is still likely to be a few dozens of nanoseconds

So 5G xhaul will need to support extremely accurate time distribution

The centralized GrandMaster model will require on-path support

The Distributed GM (MiCLK) approach is advantageous if GNSS can be secured

- < 50 nsec accuracy guaranteed for gNBs served by same MiCLK
- ±125 nsec accuracy guaranteed for gNBs served by different MiCLK
- Note that GNSS at every site
  only guarantees ± 100 nsec accuracy between neighboring eNBs!