Introduction to 5G

DD AD
Owned by David Cherney DISH •••
Last updated: Jul 25, 2023 by David Cherney DISH • Add Workflow

What puts the 5 in 5G?

5G is the 5th generation of mobile wireless technology.

- 1G: Analog Voice Technology, Introduced in 1973
- 2G: Digital voice , Introduced 1991
- 3G: Internet Access, Introduced 1998
    - enabled mobile web browsing
- 4G: Broadband Internet Access, Introduced 2009
    - enabled smartphone features like streaming video
- 5G: Huge number of devices with higher speed, Introduced 2020
    - Not just cell phones
    - Designed to support the following three kinds of use

Internet of Things (IoT)

- massive machine type communication (mMTC)
    - huge numbers of devices: a million per square kilometer
- narrow band IoT (NB-IoT)
    - connected things with low data rate, low power, and low mobility, but long battery life (as in years)
- enhanced machine type communication (eMTC)
    - medium data rate, connection maintenance while at high speeds, and still long battery life
- smart agriculture
    - e.g. the internet of fruit; a huge number of sensors monitoring fruit in the complex process of planting, growing, harvesting, shipping, stocking, and selling
    - e.g. the internet of dairy
- smart cities
    - elevator monitoring
    - infrastructure monitoring
        - e.g. monitoring growing cracks in bridges
    - intelligent traffic monitoring and traffic control
    - intelligent electrical grid
        - a massive number of things communicating the energy use of other things enables energy consumption prediction and planning
    - smart waste, recycle, and reuse management

- smart water measurement and management

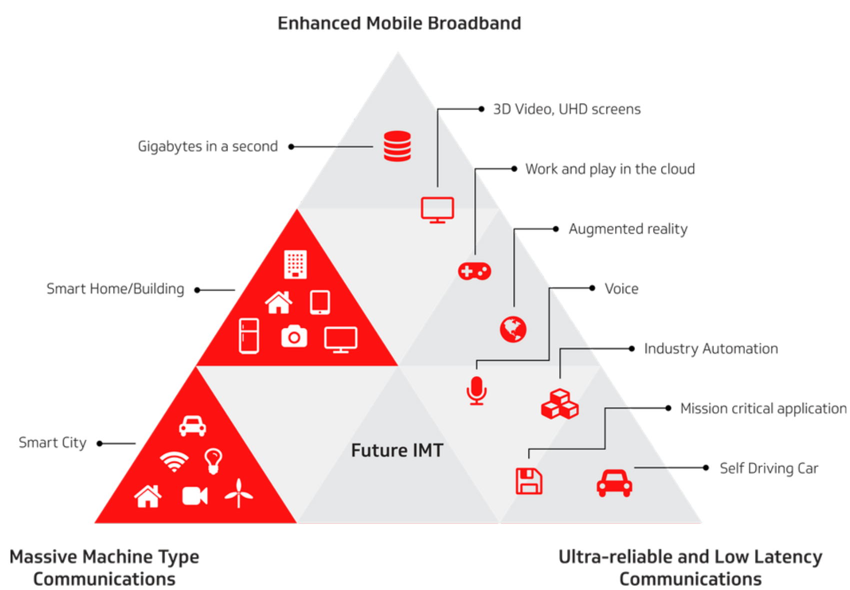
ultra reliable low latency communication (URLLC)

- tactile internet
  - augmented reality
  - virtual reality
    - social virtual reality (e.g. cloud gaming)
    - live VR video broadcasting (sit anywhere in the sport stadium you want!)
- industrial automation for industry 4.0
  - augmented workers: humans+digital technologies
    - e.g. remote repair for dangerous environments
- autonomous driving
  - cars with human passengers
  - unmanned vehicles
    - quadrupedal robots that can inspect in dangerous environments, like cell towers with active antennas
    - e.g. automated FAA air traffic control for the tens of millions of drones that will be in the US within the next few years
- remote health care
  - remote surgery
  - remote administration of pharmaceutical doses

enhanced Mobile Broad Band (eMBB)

- fixed wireless access (FWA)
  - instead of a cable to your home for internet access, connect your home computers to your in home 5G connected router.
- dense urban hotspots
  - everyone in the stadium gets a good 5G connection
- train to trackside connection
  - bullet train traveling at 200mph maintains 5G connection with trackside antennas, and provides passengers with broadband connection

Those three kinds of use, envisioned for the future of cellular mobile networks, provided the main guidance for features in the 5th generation.



There are, however, many more ideas that go into 5G. For example, 5G was designed with differentiated service in mind.

Differentiated service

- priority
  - e.g. priority service to authorized users for national security, emergency preparedness, and disaster relief
    - imagine everyone in San Francisco simultaneously trying to video chat with family just after a large earthquake, and the need for priority to be given to rescue workers.
- quality
  - guarantees for quality of service at different levels including
    - latency
    - throughput
- network slicing
  - reserving resources for different 5G use cases, or different customers
- differentiated charging for differentiated use, equipment, and user options
  - users can choose their quality of service, and modify their choice in time
  - different uses can be charged differently

Each generation had phases.

Generation	2G	3G	3G HSPA+	4G	4G LTE-A	5G	5G Advanced
Max speed	0.3Mbps	7.2Mbps	42Mbps	150Mbps	1Gbps	10Gbps	TBA
Average speed	0.1Mbps	1.5Mbps	5Mbps	10Mbps	50Mbps	50Mbps	TBA

In what follows, we will not further compare 5G to previous generations for the sake of brevity.

The 10 Pillars of 5G

What is truly different about 5G compared to previous generations? The truth is that this question is very had to answer because improvements to 4G were being made while 5G was being designed; the same ideas were put into both late versions of 4G and early versions of 5G. Thus, one needs to really get into details to see the difference. However, the following list from the 2015 book *Fundamentals of 5G Mobile Networks* provides one group's ideas of what would become the truly new things about 5G.

The new features of 5G over previous generations can be remembered as MOW VRS BEND (like "bend vs break", but more extreme!)

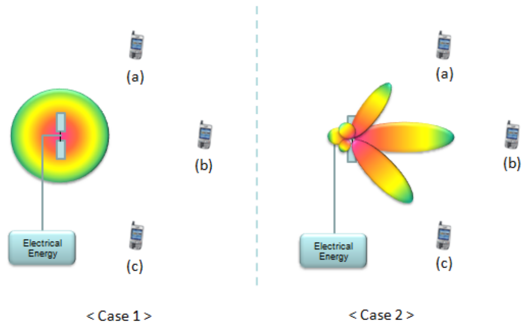
M for machine type connections: 5G networks will connect a huge number of things besides phones. For example, thousands of robotic components in dozens of factories owned by a single company in massive Machine Type Communication (mMTC).

O for self Organizing network: 5G networks will autonomously make decisions to optimize service.

W is for Wavelengths in the millimeter range: 5G networks will make use of high frequency/short wavelength radio waves not used in previous generations. Higher frequencies transmit data faster.

V for Virtualization: Software will not be sold in boxes meany to be put in racks (as in previous generations). Rather, 5G software will be required to run on consumer off the shelf computers, allowing software to be separated from hardware in 5G.

R for Radio technology: 5G networks will use a combination of new and existing radio technologies. e.g. Beam forming, in which the signal from a cell tower is sent to devices through a focused beam instead of being sent out in a circle with most of the energy of the signal wasted.



S for Spectrum Sharing: The FCC sells the right to use spectrum (meaning frequencies of electromagnetic waves). There is a finite amount of spectrum. 5G will utilize ways for different organizations to share spectrum. e.g. In some places, spectrum used for military radars might not need to be used at some times of the day, and mobile networks will be able to use that spectrum at those time of the day.

B for Backhaul: 5G networks need to carry a tremendous amount of information from cell towers to other network infrastructure, requiring a reworking of the system of fiberoptic cables used.

E for Energy Efficiency: 5G will be more efficient than previous generations thanks to new technologies like beam forming.

N for New spectrum: In anticipation of 5G networks, the FCC freed spectrum. e.g. the move from analog to digital terrestrial TV broadcasting in 2009 freed up considerable spectrum.

D for Dense network of small cells: 5G networks will, when fully built out, utilize the advantage of having a large number of antennas per square kilometer.

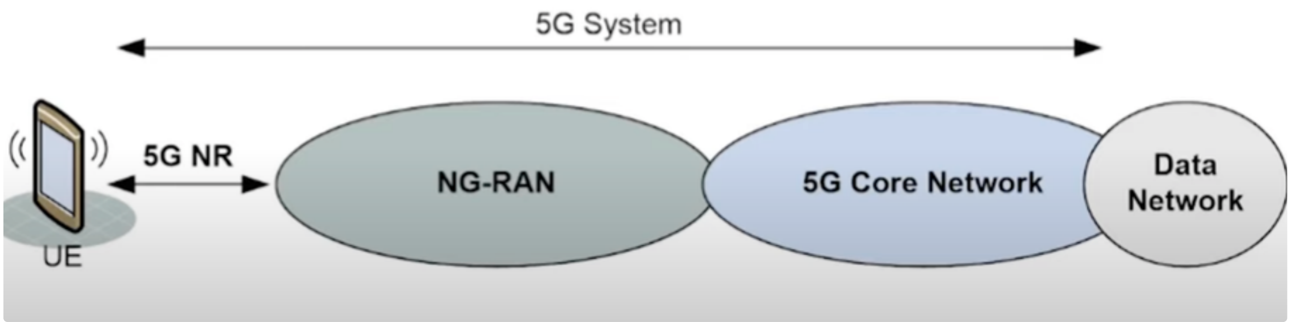
There are many other interesting facets of 5G.

That said, let's take a first look at a mobile network, focusing on 5G.

A 5G Network Includes

Note that three of the parts listed below are themselves called networks. Indeed, networks can have subnetworks.

1. User Equipment, UE (e.g. a phone)
2. Access Network, AN (RAN for Radio Access Network, but in a concept)
3. 5G Core network, 5GC
4. Data Networks, DNs
   - e.g. The public internet for web browsing is an example of a data network.
   - e.g. The data network for digital phone is not part of the public internet, but is a data network.



The following gives some details about each of those 4 parts.

1. User Equipment

Examples of user equipment (UE) in a 5G network include

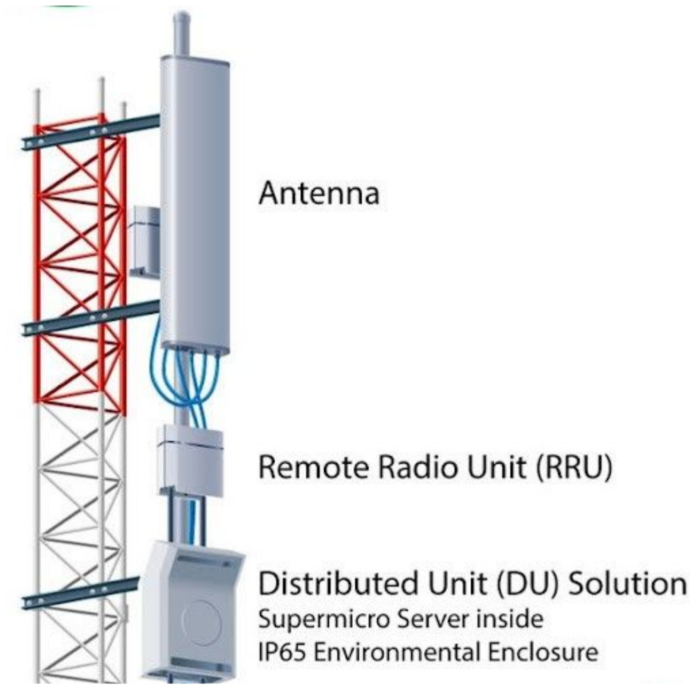
- 5G capable mobile phones
- 5G home routers
- IoT devices (much more on this below)

2. Radio Area Network (RAN)

A radio area network is the infrastructure between the user equipment and the core.

A RAN is made of

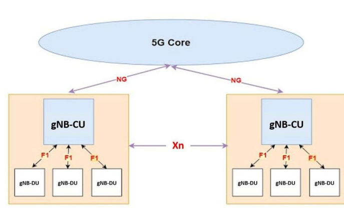
- Antennas (on towers) transmit and receive analog waves
- Radio Unit (RU), handles the conversion between analog radio waves and digital transmission



- Base stations at towers (5G base stations are called gNodeBs or gNB for “next-**g**eneration **N**ode **B**ase-transceiver-station”) which include
  - Distributed Unit (DU)
    - handles low layer data protocols
    - can be partially visualized
  - Central Unit (CU)
    - handles runs higher layer data protocols
    - controls one or more DUs
    - can be virtualized
- Fronthaul: link between RRU and DU
- Midhaul: link between DU and CU
- Backhaul: link between CU and core (fiberoptic cable)

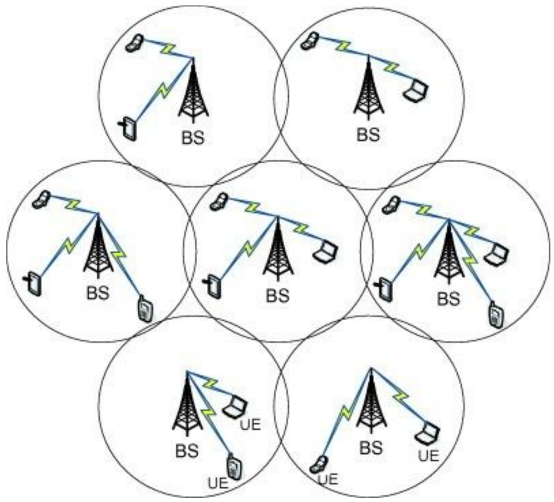
One novel feature of 5G over previous generations is that the base stations can be distributed;

- there is one CU for each base station,
- there is one DU for each cell in the cellular network,
- a single CU can be the CU for (up to) hundreds of DUs,
- a 5G base station is a CU with all of it’s DUs.

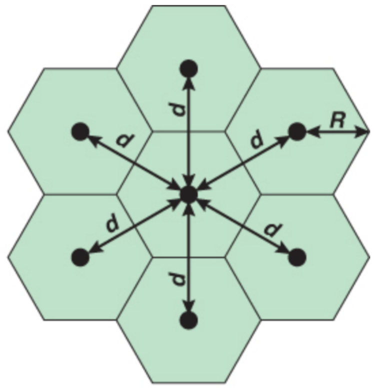


This is revolutionary in terms of what constitutes a cell:

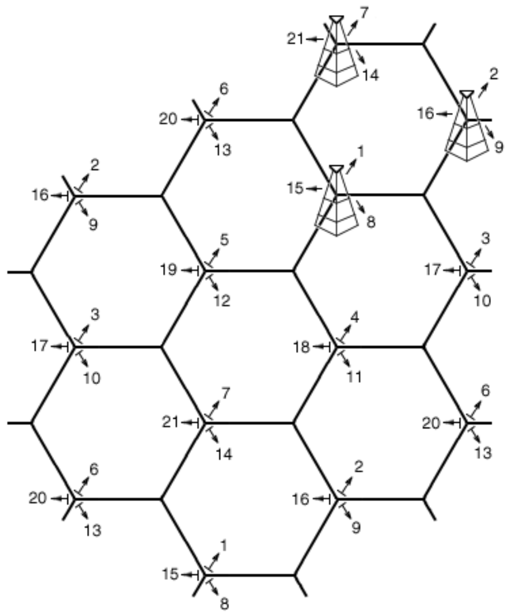
1G technology used radio towers with circular radio wave transmission, so that each tower covered a circular cell. The signal from two different antennas overlapped.



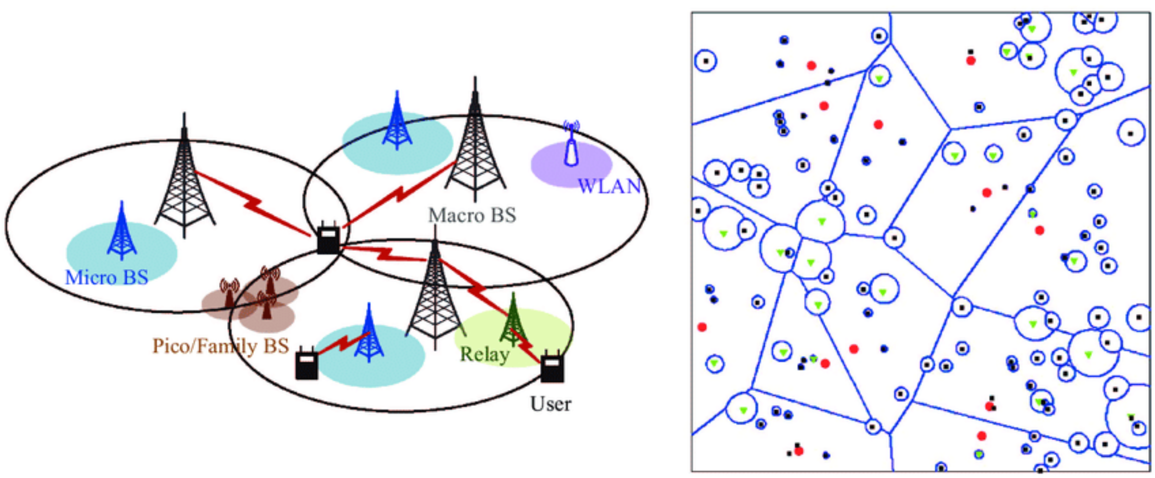
The lines demarcating where signal to UE should be handed over from one antenna to another (not shown) formed hexagonal cells. (Of course, the actual geometries are not regular hexagons, and this is just a simplifying model.)



Circular radiation is wasteful of energy. So, later, three antennas per tower were used, and towers were placed on the vertices of hexagonal tilings.



With the separation of base stations into a CU and many DUs, and with the introduction of many different antenna technologies with different ranges and frequencies, much more complicated cellular geometries are possible.

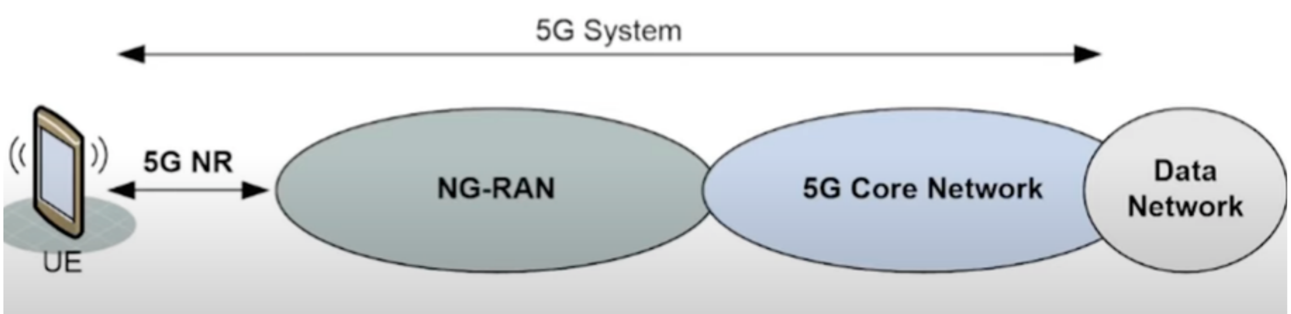


(a) (b)

5G is designed to include lots of different kinds of antennas and radio waves. This heterogeneous collection of Radio Area Network equipment is called a HetNet. There are a lot of possibilities, and a lot of complexity, in building out a HetNet. And 5G is designed to allow handling of that complexity.

3. 5G Core

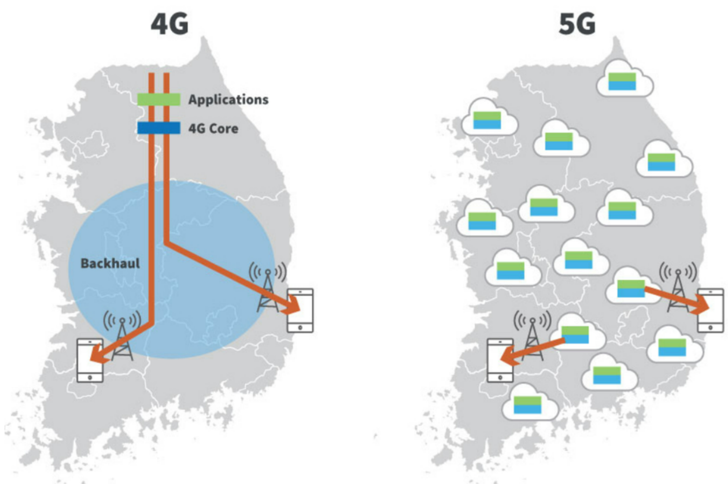
The core is, conceptually but not spatially, between the RAN and the Data Networks.



The core is the part of the 5G network that the UE stays connected to while the UE's radio connection changes from one cell to another, to another. In older generations the core was a place.

3.1) A 5G Core is Not A Place

It is, rather, a collection of instances of functions. Those functions can run on off the shelf hardware, like the hardware in cloud data centers. Those cloud data centers are spatially spread out, and a network operator can make decisions about where it is best to run each kind of function. In this way, 5G decentralizes the core of a mobile network. Core functions can be co-located with application servers in an edge datacenter, making communication paths from UE to application servers short and thus improving end-to-end speed and latency.



3.2) The core is comprised of Network Functions (NF) Instances

Instances?

One NF, for example the user plane function (UPF), might have 5 instances. Each UPF instance might run at a different data centers. It is also possible for multiple instances of a NF to run at the same data center.

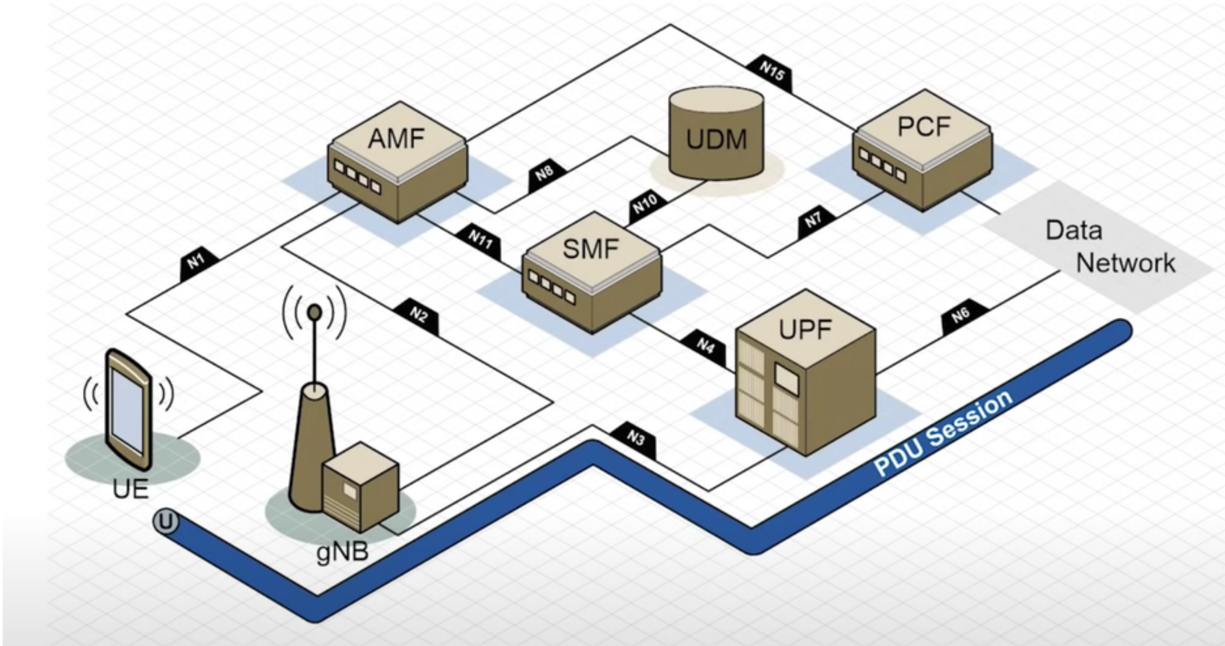


Diagram of a 5G core showing some of the mandatory core network functions.

It will be worth learning the names and initialisms for the functions that have to be in a 5G core.

Mandatory Network Functions

- **Access and Mobility management Function** (AMF) is like the UE's best friend; it always knows where the UE is, and is the first one (in the core) to hear whenever something happens to the UE, or when the UE needs something.
  - Access
    - the AMF is the proxy for all communication between the UE and the core NFs. (Except the UPF. More later.)
    - Facilitates initial contact with the core through an always listening feature
    - Facilitates authentication with the core. This is called *primary* authentication.
    - Provides a temporary ID, GUTI (Globally Unique Temporary ID) to UE while authenticating.
  - Mobility management
    - The Keeps track of the location of UE based the ID of the cell the UE is in.
    - Manages handing off connection from one cellular antenna to another and thus from one DU to another, including when CU needs to change.
    - If the UE is in idle mode (is off), the AMF approximates location.
- **Authentication Server Function** (AUSF) is like the badge scanner next to the reception desk.
  - After the AMF initiates connection, the AUSF is called to pick an authentication method to authenticate UE.
    - Different UE will have different capabilities so the AUSF looks up the UE's methods of identification in the Unified Data Management function (UDM).
    - The UDM has the authentication key for each UE. The UE has the same key in its memory.
  - The AUSF performs hashing calculations on the core side using the UE's keys and a random number it sends to the UE. The UE uses its keys and the random number from the AUSF. The two check to see if they have generated the same hashed value in a mutual authentication.
- **Session Management Function** (SMF) is like lawyer; it looks up rules and tells other NFs what those rules are.
  - Interacts with Policy Control Function (PCF) to
    - retrieve policies about to the sessions a particular UE is allowed to have,
    - retrieve charging instructions for the various sessions a UE has.
  - Responsible for creation, deletion, and modifications of data sessions for the UE by
    - telling the UE the available entry points to the core (UPFs) so the UE can decide which to use.

            - telling the chosen UPF the rules for moving data for the various sessions,
                - including how to report data usage for charging.
        - Allocates IP address for UE, If needed, by looking up addresses on the relevant data network.
        - Performs authentication between UE and an application on a data network, if needed, by looking up secondary authentication data in the UDM. This is called *secondary* authentication.
- **User Plane Function** (UPF) is like a traffic officer; it directs the flow of data packet by packets, enforcing the laws.
    - User data flows from UE to gNB to UPF to DN, and back; the UPF is the only connection for user data flow in the core.
    - The UPF is the anchor point throughout mobility; no matter what cell/DU or CU the mobile UE connects to, the UPF instance that serves a data flow remains constant.
    - The UPF enforces Quality of Service (QoS) rules, for example that the data packets a UE's session for an emergency call should be forwarded before packets for the same UE's session for a video game update.
    - The UPF follows rules about how to measure traffic flow for each session, and generates reports for other functions to use in calculating charging.
- **Unified Data Repository** (UDR) is the book of all UE information.
    - For each subscriber, the UDR has
        - authentication information and state (what stage of authentication the UE is in)
        - authorization information
        - a list of services have they subscribed to, and
        - the data network details relevant to each.
    - The UDR participates in registration by holding a copy of private keys necessary during authentication. The UE holds another, identical pair. The UE and AMF make sure the keys are hashed to the same value.
- **Unified Data Management** (UDM)
    - The UDM is the front end to the UDR; it is where user subscription data can be created, modified, or deleted.
    - The UDM has a module called Subscriber Identity De-Concealing Function (SIDF) for decrypting the subscriber concealed identity, allowing identification for accurate look up of a subscriber without broadcasting the identity of the subscriber over radio waves.
- **Policy Control Function** (PCF) is the judge: it knows the law and interprets it when needed.
    - Policies are retained here and SMF generally implements those policies.
    - The policies are about
        - quality of service (QoS) guaranteed by the network operator to the UE,
        - charging (as in money, not battery power), and
        - access control (for example giving access to high priority channels to fire departments for disaster relief).
    - The PCF has the ability to make real time (dynamic) decisions based current network conditions.
        - e.g. The PCF can ensure SMF denies a session to a UE to prevent exceeding the maximum number of allowed UE sessions of a particular kind.

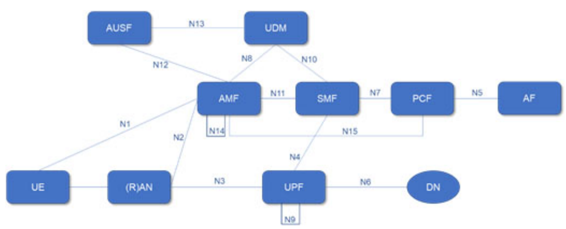
DISH Wireless's 5G core is thus a set of NF instances. Those NFs are software. Note that those NFs are updated frequently via in service software updates. Those NFs are software written/made by many companies: Nokia, Oracle, Matrixx, Mavenir, SS8, Palo Alto, and Altiostar. Similarly, T-Mobile's 5G core is made by Nokia and others. AT&T has at least 15 suppliers for its core functions as of January 2022. These core functions are software made to meet international standards set by two organizations, the 3rd Generation Partnership Project (3GPP) and the International Telecommunications Union (ITU). The latter is a United Nations agency with 193 members. Those members are nations. These standards ensure that UE work in all countries that have agreed (in a binding treaty) to the Radio Regulations set up by the agency. Network operators in those nations are required to conform to the regulations.

In the next section we present two ways to visualize the of the 5G core network links.

3.3) 5G Core Networking

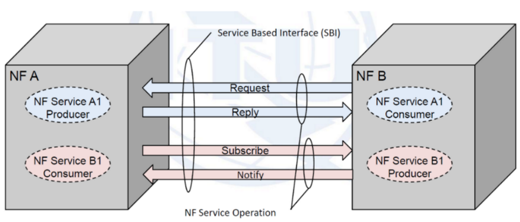
There are two ways to look at the communication between 5G core functions:

- *5G Reference Point Architecture* is suggestive of connections between pairs of network functions. Connections are characterized by addresses on the ends. This diagram does not mean to convey that idea; it uses an idea of “reference points”.
  - These reference points are conceptual with all instances of a network function shrunk down to a single point. A core network might have SMF1, SMF2, UPF1 and UPF2 and a connection between SMF1 and UPF1 and a different connection between SMF2 and UPF2. The concept of the SMFs being connected to the UPFs is captured in the N4 reference point, while the individual connections should be given distinct names. Reference points denoted as N_{number} (e.g. N11) between any two network functions (e.g. N11 between AMF and SMF).
  - Reference architecture is suggestive of long term session setup between the NFs, but the 5G core was designed so that most of the communication between NFs is done via short term application-program interface (API) call-response communication.
  - Further, any of the network functions can make API calls to each other. The reference point architecture indicates which of those calls will be accepted and which rejected. This makes for a rather complicated network diagram. However the idea that all NFs can try to make calls to any other NF motivates a second way to look at the architecture.

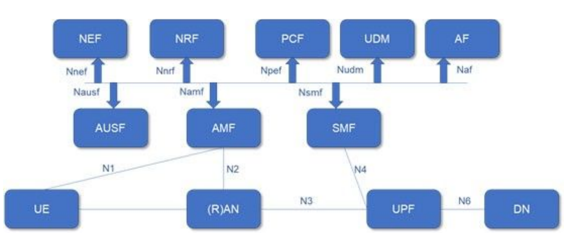


Reference Architecture

- *Service Based Architecture (SBA)* emphasizes API based communication between any two Network Functions. The main idea here is that NFs enable other NFs to access their services through *Service Based Interfaces* (SBI), which are APIs. SBIs are denoted as $N_{NetworkFunctionName}$ (e.g. N_{af}).



Service-based interface (request-reply and subscribe-notify) (Credit: ITU)



Service Based Architecture

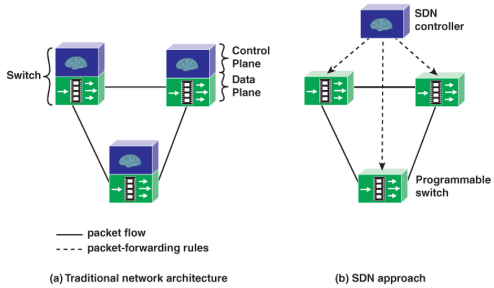
Note that in both service based architecture and reference architecture, the N2 and N3 interfaces are the two paths from the RAN to the 5g Core. The connections on these reference points are not APIs. Thus, SBA describes communication within the core only, and reference points describe connections between UE, core, and DN.

3.4) The 5G core is designed for

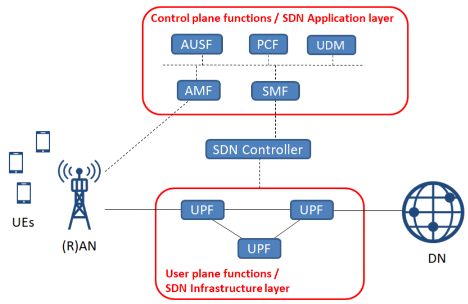
three enhancements.

Enhancement 1: Control Plane and User Plane Split (CUPS)

1. One of the technological trends that motivated the design of the 5G core is software defined networking (SDN).



This kind of networking takes advantage of the fact that the different network functions are not running necessarily on different computers connected with wires. Rather, several of them might be running at the same data center. So, the connections between the NFs is software defined, not physically defined. This allows the networking software to be decoupled from the things in the network; the network controller can then run independently, and set up networking though APIs.



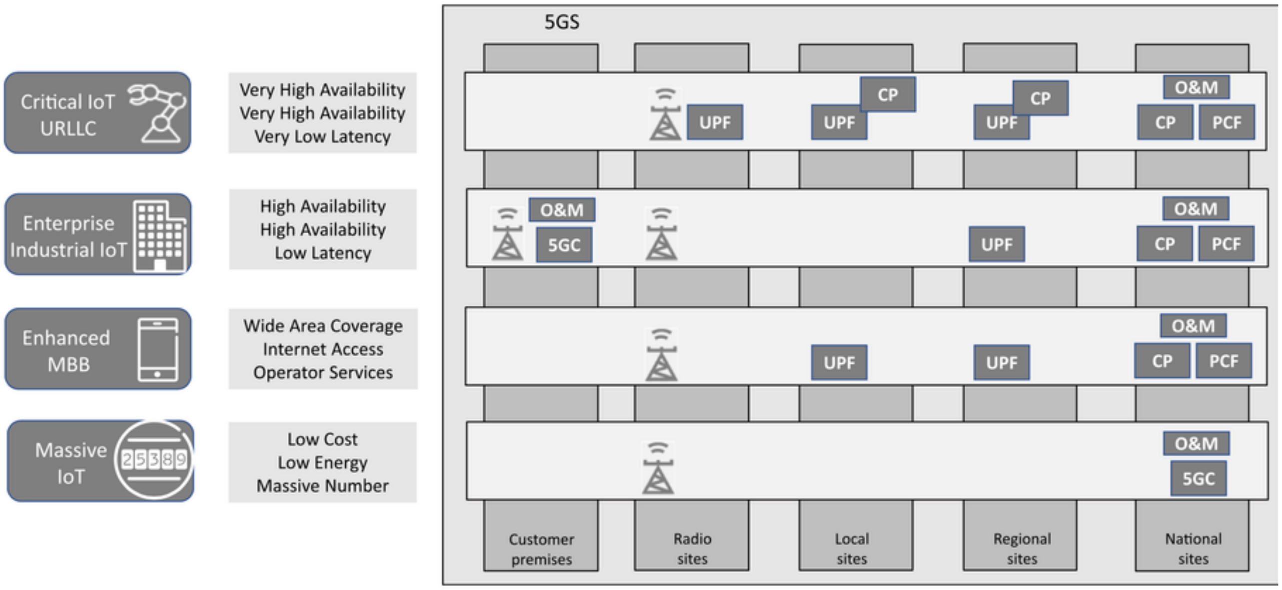
The diagram above is misleading because it wants a separation between the north bound interface (NBI) and south bound interface (SBI); the SDN controller in the 5G core is the SMF.

- in SDN, the north bound interface (NBI) connects the SDN controller to network functions that involve data exchange between NFs (the control plane). In the 5G core, these functions are all the functions except the UPF.
- in SDN, the south bound interface (SBI) connects the SDN controller to the network functions that involve data transfer to and from the user (the user plane). In the 5G core, these functions are instances of the UPF.

- **Control plane** is the set of NFs that deal with
  - UE connection management,
  - defining of quality of service policies, and
  - authentications.
- **User plane** is the set of network functions that deal with
  - data traffic from the UE to the data networks, and
  - data traffic from the data networks to the UE.
- The user plane consists of many instances of the UPF that are at geographically different data centers.
- When the user plane and control plane are separated they can be scaled separately, increasing efficiency.
  - e.g. An increase in user plane traffic can be supported by adding more UPF instances at strategic data centers, and no additional control plane resources need to be created with it.

Enhancement 2. Native support for Network Slicing

A network slice is (according to the ITU) “a logical network that provides specific network capabilities and network characteristics.” That is, two slices might use the same underlying hardware infrastructure, but software keeps the virtual infrastructure logically separated so that slices act as independent networks. This facilitates differentiated use of the underlying hardware, as in the diagram below.



- A 5G network operator may offer one slice that is optimized for high bandwidth applications, another slice that's more optimized for low latency, and a third that's optimized for a massive number of IoT devices.
- Some slices might require UPF instances very close to radio sites for low latency.
    - Other slices might not require this low latency and can place UPFs at more distant data centers.
    - The control plane (CF) functions usually can run at more distant data centers. This includes the policy control function (PCF).
- Some slices might want the entire core at a national data center.
- Each slice can have its own operations and management (O&M).
- Not every slice must have exactly the same capabilities; the available computing power is used more efficiently via slicing.
    - e.g. if you are only servicing IoT devices, you would not need the voice function that is necessary for mobile phones.
- Two slices can share some network function instance.
    - e.g. The same AMF instance can be used for a slice serving IoT and mobile phones.

The following shows the official enumeration scheme for slice types; an operator (like Dish Wireless) can have up to 255 slice types, and a huge number (16^6) slices within each slice type. These slices within slice type are distinguished with a 6 hexadecimal digit slice differentiator. A single UE can be on up to 8 slices.

SST (Slice/Service Type)	SD (Slice Differentiator)

SST Value	Slice/Service Type
1	eMBB
2	URLLC
3	Massive IoT (mIoT)
4 – 127	Standard (TBD)
128 - 255	Operator-specific

Enhancement 3: Service Based Architecture

As presented above, in SBA the control plane network functions access each others' services via APIs.

Note: The 5G core is enormously complex.

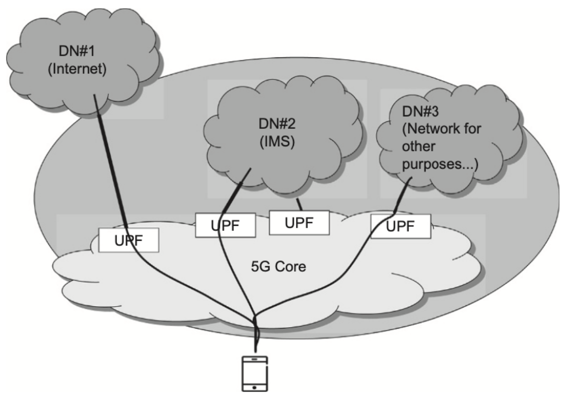
"The interworking of these various NFs to implement the various procedures performed by the core network is extraordinarily complex. TS 23.502 (Technical Specification Group Services and System Aspects; Procedures for the 5G System [5GS]; Stage 2 [Release 16], December 2020) lists dozens of these procedures. The current version of the document is 603 pages long.”

- Excerpt From *5G Wireless: A Comprehensive Introduction,* by Stallings

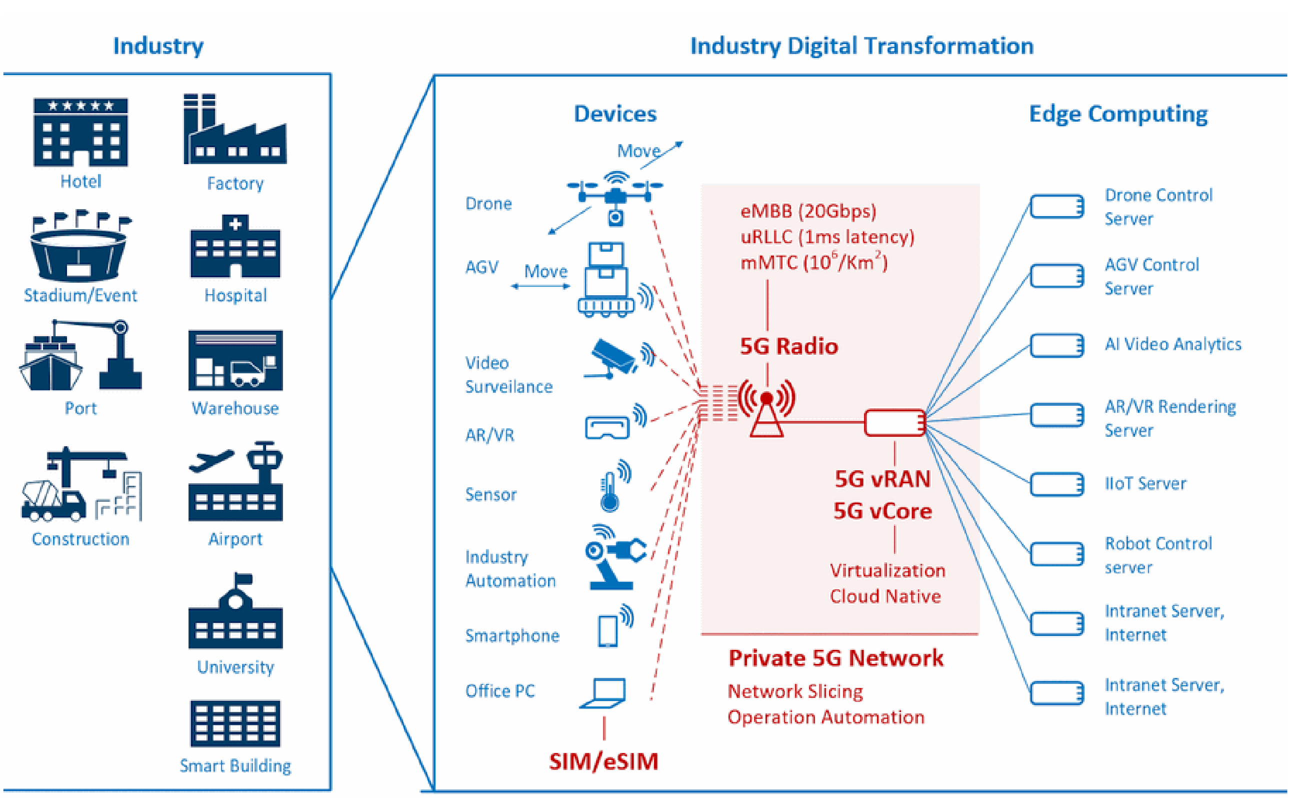
4. Data Networks

Examples of data networks include

- the internet
- the IP Multimedia System (IMS), the network for digital voice and/or video transmission; both UE in the call connect to the the same data network.



- a local data area network (LADN) accessible by UE only in a specified area like a campus or stadium,
- a data network of edge data centers with servers for applications,
- a data network for self driving cars (someday), and
- a data network for orchestration of millions of devices in dozens of factories owned by a single company (someday).



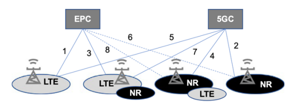
Recap

You have now seen the 4 parts of a 5G network.

1. User Equipment (e.g. a phone)
2. Radio Area Network (NG for Next Generation)
3. 5G Core (the main focus below)
4. Data Networks

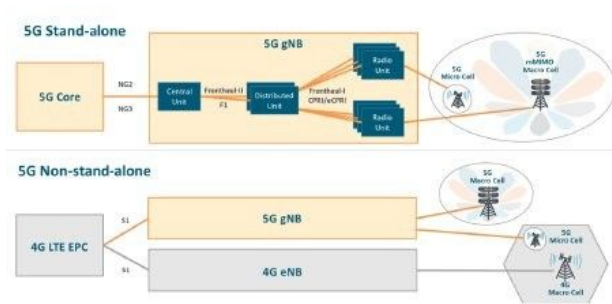
Non-Stand Alone

When 5G was first introduced, its initial feature set included ways to combine 5G radio technology with 4G core. The table below shows the possible combinations of 5G radio with 4G and 5G cores. The 4G core is called the evolved packet core (EPC) while 4G radio is called long term evolution (LTE). By contrast 5G radio is called new radio (NR) and the 5G core is denoted 5GC.



Access network:	LTE only	NR only	LTE with NR for data only	NR with LTE for data only
EPC core network	Option 1 (=4G)	Option 6 (disregarded)	Option 3	Option 8 (disregarded)
5GC core network	Option 5	Option 2	Option 7	Option 4

A network that supports any option besides option 2 is called a non-standalone network. A network that supports only option 2 is called a stand alone 5G network.



DISH: Cloud Native Standalone 5G

The number of problems to tackle in integrating new systems into old systems when both of those systems are extremely complex… is huge. DISH Wireless is building the first standalone 5G network in the United States of America.

Since DISH is not building on top of existing infrastructure, but rather building a brand new network, DISH’s 5G network is called a greenfield network. The greenfield advantage is the advantage of not needing to tackle the problems associated with interweaving old and new technologies.

Further, DISH is leading the way in the USA by having that standalone 5G network be cloud native, running in AWS.