How 5G Addresses the Needs of High Performance Communication Networks

Anthony Kabala

Abstract

Existing mobile network technology, 4G, needs to be updated to address and meet the needs of the current day high performance communication networks and future expected needs. 5G has been developed and is currently being deployed to meet these needs. This paper will survey the 5G network and assess how it meets the needs of increased high performance communication networks with relation to current technology and the changes that 5G technology provides. There is a need for more devices to be supported on mobile networks as well as an increased demand for more bandwidth with less latency and higher reliability depending on the use. 5G technology also provides new use case possibilities not previously capable of being handled by the previous cellular wireless technology of 4G. The demand of ultra-reliable low latency communication is now possible to be met using 5G technology and will be surveyed and assessed with a comparison to existing technology shortcomings. Several key technologies have been enhanced to meet the current needs including mMIMO (Massive Multiple Input Multiple Output) with beamforming capabilities and mmWave frequency band usage to increase the bandwidth throughput to devices. There are limitations to several of the technologies that are partially addressed through a hybrid of technologies including network virtualization, and changes to the RAN with mmWave technology for shortwave small cell communications paired with macro cell communication technology including mMIMO to provide an overall increase in meeting the high-performance communication networking needs currently and those expected into the future.

Problem and its importance

High performance communication networks are in essence trying to utilize the existing communication networks to their full capacity and expand them to meet new network communication needs that are limited by the previous technologies. The main things needed for this high level of performance is high rates of operation, multi-service integration, and the quality of service for these multiple types of services. The current cellular communication technology that is the most widely used today and provides the majority of cellular service over the globe is 4G [12]. The usage of cellular services has also expanded to be one of the leading ways in which people globally connect and interact on a professional and personal level [12]. 4G which stands for 4th generation is just that, it is designed to evolve with time as the cellular services available to be used expand and are updated with new technology. However, current trends in use cases for cellular service has expanded the high-performance expectations for 4G beyond what its design standards are capable of delivering.

There are 3 main use case categories that are pushing the evolution of 4G to the next generation of cellular technology of 5G and that are also developing as a result of the technological advances that 5G offers to different performance measurements. These are xMBB (Extreme Mobile Broadband), mMTC (Massive Machine-Type Communication), and uMTC (Ultra reliable Machine-Type Communication) [5,7]. All of which demand that the qualities of high-performance communication networks be met by cellular networks. The demands of streaming and telecommuting increasing the needs for increased data rates are categorized in xMBB [5]. Another trend in networking is mMTC that has different needs from the cellular network than was focused on in 4G. It needs the support of many devices, up to a million, with many of them communicating at the same time, with wide area coverage, but not a lot of bandwidth like xMBB, that are low cost, and have low energy usage [5,7]. This would require the use of some similar services and different services than before. The last main use category is for uMTC that need high reliability and ultra-low latency. This is because it needs to connect devices for real-time control for services such as autonomous vehicle control systems to the cellular network that must have these needs meet for the safety of passengers in vehicles [3]. With uMTC the possibility of remote medical treatment can also be pursued with the new capabilities presented by 5G [7]. We will see later how 5G can meet these needs through an analysis of the technology of 5G and its application.

Survey of protocols/network

<u>5G Cellular Standard (3GPP)</u>

5G is the fifth generation of cellular networks, 4G is the predecessor of 5G. The 3GPP (3rd Generation Partnership Project) is the standards group that unifies several different groups of standardizing organizations to develop cellular technologies that includes 5G. The latest release of standards for 5G are release 17 state 2 by 3GPP which will be released Dec. 12, 2021 [10].

Network Evolution

The 5G network architecture is based on 4G and has many similarities and functions that are taken from 4G as well as several functions and technologies that are new to cellular networks. A significant driving force behind the evolution of 4G to 5G is the intent to implement the use cases of xMBB, mMTC, and uMTC. The 4G network has a monolithic design that is not adjustable to provide significantly different services [5]. 4Gs main design is to provide high speed mobile broadband across a large area to many users. The main design of the 4G network is composed of the RAN (Radio Access Network), which is the physical network, in which devices connect to the cellular network. The other main component of the 4G network is the 4G mobile core that processes the network protocol that separates the user plane and the control plane that will be described later when describing the 5G core components.

The RAN of 4G is still utilized in 5G but with more advancements in capabilities that can accommodate more users, different bandwidths, and higher bitrates. The core however has been redesigned from the ground up. The 5G core still uses many similar functions as the 4G core but also adds several functions to help expand the capabilities and use cases that 5G can accommodate. The most significant change from the 4G core to the 5G core, and the most significant change altogether, is the way the core is distributed. The original 4G core is centrally located geographically. This causes the backhaul portion of the 4G RAN to result in levels of latency that restrict the types of services that could be offered over 4G cellular networks. Specifically, ultra-low latency applications such as the smart grid, autonomous vehicle control, and industrial automation [3,9,13]. The 5G core is a revolutionary approach to solve the issue of latency as well as increases in reliability and security. The 5G core network is virtualized on distributed cloud servers [5,7,8]. The cloud servers can be located physically close to the point in the RAN that the device, which is referred to as UE (User Equipment) under the 3GPP specification, is connected to the 5G cellular RAN through the local gNB (Next Generation Node B) as can be seen in Figure 1 from [7] below [7,10]. The 5G core network is further enhanced by virtualization to allow separate virtual network connections that can have different parameters than other virtual network connections that operate on the same physical hardware. This is known as slicing and allows for separate virtual instances that perform at different parameters that include bandwidth, latency, reliability, and many others as well as the potential for new virtual instance schemes to be developed as well. These core changes are the driving force behind 5G that allows it to provide functionality that is limited in the 4G network architecture.

As 4G transitions to 5G there is a need to still support the 4G network of UE and 4G network itself. This is because the transition to 5G requires updating the 4G RAN as well as developing support for the 5G core network. This means that the 4G and 5G networks need to be updated in a way that allows for both systems to exist and work together seamlessly at the same time as the 5G network is deployed.

There are several ways for this transition to happen that different providers can invest into. The quickest first step towards deploying the 5G network and integrating it into the existing 4G network is to continue to use the 4G RAN and the 4G mobile core for the control plane and the user plane, but add the use of the 5G RAN for the user plane [14]. This allows the advantages of the 5G RAN to be utilized such as mMIMO and mmWave technologies that allow for higher bandwidth throughput to more UEs [5,6,7]. However, by not utilizing the 5G core the network slicing capabilities won’t be available so mMTC and uMTC will not function yet but the high data rates of xMBB will be realized without much change in the latency yet.

The next expected upgrade from 4G to 5G will likely be to continue to use the 4G RAN and the 4G mobile core but only partially for the user plane [5,6,7]. Then use the 5G RAN and 5G mobile core for the user plane and the control plane with network virtualization on the 5G Core. This would essentially be using the 4G network to augment the 5G network when 5G service is not as dependable and available as the 4G service in the given location. So, the 5G network core would now be able to provide the full range of services needed for xMBB, mMTC, and uMTC that is possible due to the virtual network slicing that can provide selective services from the 5G core that support different use cases. However, due to still relying on the 4G network in certain locations the 5G services will be limited to location. In certain locations the full range of 5G functionality may be available for xMBB, mMTC, and uMTC uses but in other locations it may not. This would also make applications significantly dependent on the 5G services such as V2X

(Vehicle-To-Everything) undependable and even dangerous to use considering the safety factor of V2X on the reliability of the 5G network [3].

The next evolution of the 5G network will be when the 5G network can be utilized completely by itself without any support of the 4G network. This would mean that the RAN and the core of the 5G network would run the control plane and the user plane completely by itself. This would allow for the full expectations of 5G to potentially be utilized including xMBB, mMTC, and uMTC as well as newer potential uses still in development or not yet explored.

With the redesign of the network core in 5G there are still several functions of the 4G core that are carried forward into the 5G design. These include: AMF (Access and Mobility Management Function), SMF (Session Management Function), UPF (User Plane Function), PCF (Policy Control Function), AUSF (Authentication Server Function), and UDM (Unified Data Management) [6,7,8].

The 5G core also has several entirely new functions to allow it to provide the expanded and virtual networking functions that it does. These include: NEF (Network Exposure Function), NRF (Network Repository Function), and NSSF (Network Slice Selection Function) [6,7,8]. There are also several other core functions, but these are the most important functions of the core for the majority of use cases [6]. How these service functions of the core network work will be described later when the 5G core is described.

5G Communications Services

The most significant part of 5G that really sets it apart from 4G is the communication services that its design offers.

Through the use of network virtualization within the 5G core many different capabilities, that previously were limited by the monolithic framework of the 4G core, are now possible [5]. This is largely in part due to the use of network slicing as is seen later in Figure 7 from [9]. Network slicing allows for separate instances of the network connections to the 5G core to be created that have different capabilities [5,6]. These can include high data rates, a large number of UEs connected to a single cellular gNB, ultra-low latency, and high reliability as well as many other capabilities. These required services from the network cannot be supported by a single network design such as the 4G cellular network. Therefore, by creating multiple virtual networks within the 5G system a network can be created separately for each use case as needed and due to the virtualization of the network slices the different slices can be scaled up and down as needed [15].

The following use cases encompass the majority of capabilities and uses that will be focused on and include xMBB, mMTC, and uMTC.

Extreme mobile broadband (xMBB) also referred to as enhanced mobile broadband (eMBB) is in essence the evolution of 4G networks capabilities. That is 4Gs capabilities that mainly focus on mobile broadband communications because of the limited ability to provide services for mMTC and uMTC [5,7,14,16,17]. The collection of xMBB services is focused on providing high bandwidth to support business and personal mobile communications as well as provide streaming services and mobile connection to the internet [5,6]. The introduction of the use of the mmWave frequency band, which will utilize the portion of the electromagnetic spectrum from 24 GHz and above, allow for more bandwidth to be transmitted compared to the 4G spectrum [5,6]. Other important features of 5G that are important to some of the xMBB use cases include mobility and area traffic capacity as well [17].

Massive machine type communications (mMTC), which has also been referred to as the IoT (Internet of Things), encompasses another set of network characteristics that were previously not possible to be met by the main design of the 4G cellular network [1]. However, 4G has recently began to provide some limited service for mMTC [5,19]. 5Gs new capabilities make it possible to provide service for mMTC using cellular networks [5]. The service requirements for mMTC are for the cellular network to provide service to a large number of devices within a single cell of the cellular network, up to $1x10^6$ UE within a km^2, while 4G is limited to less than $1x10^4$ connections [5,7,10,17]. There are several other network capabilities categories that are a lesser part of mMTC that include: low transmission sizes, large coverage area, limited mobility, low cost, and low energy use [7]. Not all of these capabilities are necessary for all mMTC use cases as not all IoT devices are stationary or are unplugged, therefore needing low energy use.

Such applications include the smart grid. An example problem with the existing 4G network, even with recent updates, is that when many of the connected machines, IoT devices, like all the smart meters for the electric grid within a specific 4G cellular cell try to send polling information at once in a large metropolitan area with a lot of smart meters. This large number of connections all at once will cause congestion [19,20]. Solutions have been proposed and simulated that could solve this problem that include altering the protocol used to manage the random NOMA (Non-Orthagonal multiple Access) that occurs with mMTC and to try to relieve congestion by managing the large influx of communications all at once, characteristic of mMTC applications, by filtering the incoming communication packets through several different methods [19,20]. However, these solutions have shown to result in increased delay due to increased processing [19]. If the latency of the backhaul network is reduced as it is in 5G then the computation of these types of functions can occur much closer to the RAN connection and reduce the latency involved, another issue with the 4G design is that the mMTC connections limit the existing connections, given the number of connections possible on the 4G RAN [5,7]. By utilizing the 5G RAN, as proposed above in the initial phase of 5G deployment, the increased number of connections possible by the 5G RAN mMIMO antenna arrays would allow for more connections to occur at once within the cellular cell to alleviate the limitations on the number of devices but still leave the issue of not being able to process the connections as quickly as may be needed. For example, the smart grid is becoming more dependent on the information from deployed smart meters that monitor power consumption, and in order to regulate the amount of power on the grid, the power use information needs to have a very low latency for the power regulation to be efficient and to avoid potentially dangerous conditions in which there is too much power produced, too little power produced, or the phase angle is out of sync [21,22]. The latency requirements for critically important applications within the smart power grid require a latency of less than 50ms, otherwise blackouts could occur or damage to critical components of the power grid could occur [22].

Ultra-Reliable Machine Type Communication (uMTC), also known as URLLC (Ultra Reliable Low Latency Communications), is the final major category of use that is made possible with cellular networks through the upgrades of 5G [1,5,7]. The main requirements of cellular networks of this use case, that are provided by the advancements of 5G, are to provide high reliability and low latency [1,5,7,23]. The main uses identified for high reliability and low latency include vehicle communications and manufacturing automation [1,3,4,5,7].

Ultra-low latency is important in these cases largely because of the safety factor involved if the devices do not receive and send control and monitoring data fast enough. The devices have a very limited time in which they are expected to respond and adapt to changing conditions and require a latency of around 1ms or less, in the user plane specifically, to avoid potential hazardous situations to devices, products, and human safety [3,4,5,7,23].

Previously ultra-low latency was unachievable in 4G networks that have an end-to-end latency of 30-100ms [17,23]. Note that this includes the latency of not just the user plane, which is as low as 4ms, but also the control plane which significantly increases the latency in 4G networks [17,23]. The 4G RAN backhaul is found to be one the main limiting factors with the ability to achieve low latency. This is because the location of the mobile core for the 4G network is potentially located a large distance away from the UE accessing the network, and the network itself is designed to only service MBB with its network connections [6]. This makes the latency higher than if it was closer. Also, if the 4G network did provide network service specific for the requirements for ultra-low latency and reliability it would significantly impede the service quality for other types of network connections such as for the xMBB and mMTC as all services on that particular 4G network section would need to have the network connection setup to only serve the services of uMTC, which would decrease bandwidth and the number of devices supported [23]. The physical latency is composed of the times for transmitting a packet, for signal propagation, for packet processing, for retransmission, and for packet preprocessing according to [23]. 5G solves this low latency issue in several ways. First the 5G RAN is designed to provide mobile core functionality much closer to the UE than 4G [6,9]. 5G achieves this by physically having hardware that implements the 5G mobile core as close as the gNB base station that the UE is connecting to. Secondly 5G creates a new frame structure through a virtualized network connection that can be virtualized anywhere within the 5G backhaul network, and in this case due to the network virtualization the network connection can be setup and designed to specifically tailor to ultra-low latency and reliability needed for uMTC, which can be seen from Figure 7 from [9] [23]. This still can have an effect on the overall network functionality as a uMTC packet would need to be given priority over the other main use cases of xMBB and mMTC, which has been shown to reduce throughput on xMBB [23].

Reliability is the second factor that 4G networks have not been able to provide at the level necessary for uMTC for the safety and functionality reasons previously discussed. This is because the network connection established through 4G uses a best effort connection approach that doesn’t provide reliable services connections [23]. To solve this shortcoming of 4Gs design we are brought back to the main design change of 5G that allows for separate virtual network connections to be established for each connection depending on the connections service requirements. The key issues that 5Gs virtualized network slices improves service for, with regard to reliability, are that 4Gs channel coding and hybrid automatic repeat requests are improved by 5Gs channel estimation accuracy, which improves channel coding gain, through increases the resources given to the uMTC connection that may include allocating more resource blocks, frequency, antennas, and special domains [23]. This is only possible due to virtual network connection for the specific uMTC connection [5,7,23]. Another aspect of reliability for this use case is to be able to provide reliable service for moving UEs such as vehicles. The uMTC requirements provide support for UEs moving at speeds of up to 500km/hr, but the bandwidth throughput capability decreases as speed increases making this service better suited for uMTC applications than for xMBB applications [11].

The 5G solution to latency and reliability, while reducing services for other service use cases, still results in better service overall for all use case which was not possible in the previous 4G design [11].

4G and 5G Architecture

The 4G and 5G cellular architecture is composed of a cellular access network that is composed of the physical RAN (Radio Access Network) and the Mobile Core. The RAN is composed of UEs (User Equipment) that can include cellular phones and, more recently with 5G, can also include vehicles, monitoring and actuating devices, and the potential for many more devices that have mMTC and uMTC requirements. The UEs connect to NBs (eNBs within 4G RAN and gNBs within 5G RAN) base stations. These NBs then connect to the backhaul network or RAN which then connects to the Mobile Core and then to the internet as seen in Figure 1 from [7]. With 4G and 5G the Mobile Core is virtualized, but with 5G the mobile core can be considered to be located either centrally, regionally, or at the edge of the network, close to the gNB, as can be seen in Figure 7 from [9]. This is what physically reduces the latency in 5G based on where the mobile core is virtualized along the 5G backhaul.

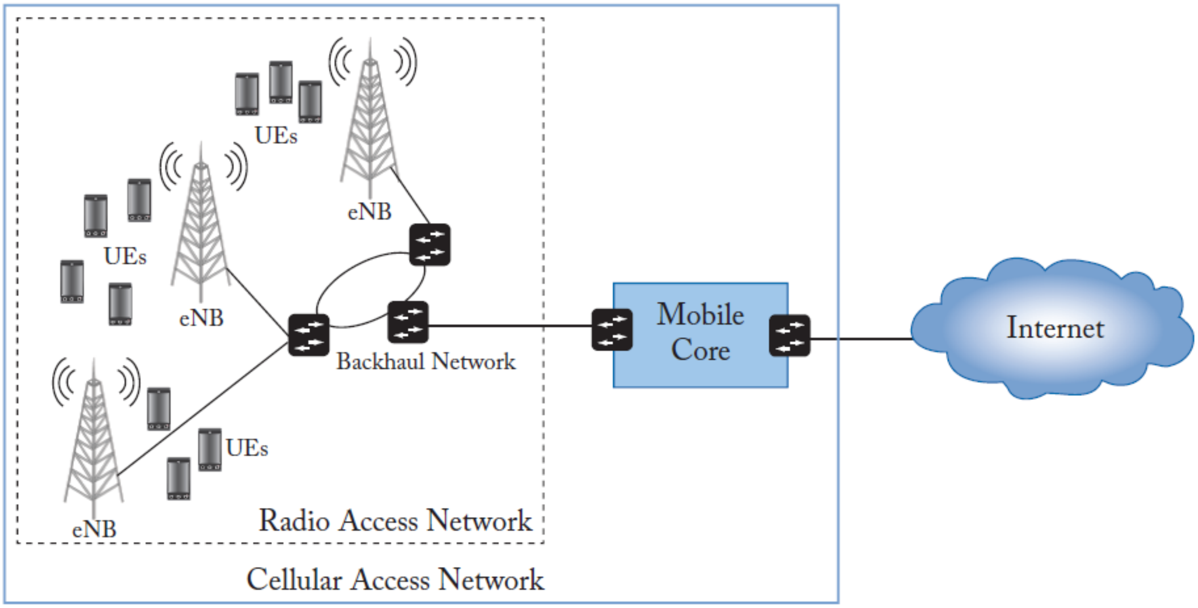

Figure 1 - 4G Network. From [7]

<u>*RAN (Radio Access Network)*</u>

The RAN provides the physical radio signal network through cellular bases stations that provide cellular communication services over several frequency bands of the electromagnetic spectrum [5,7]. Each frequency band has different capabilities and limitations based on the inherent physical properties of that portion of the electromagnetic spectrum. 5G supports all the frequency bands of 4G and more. These include 600 MHz to 2.5 GHz, which 4G supports, and also includes 2.5GHz to 6GHz, which 4G also partially supports, and 24GHz and above, which is only supported by 5G as can be seen in Figure 2 From [6] below.

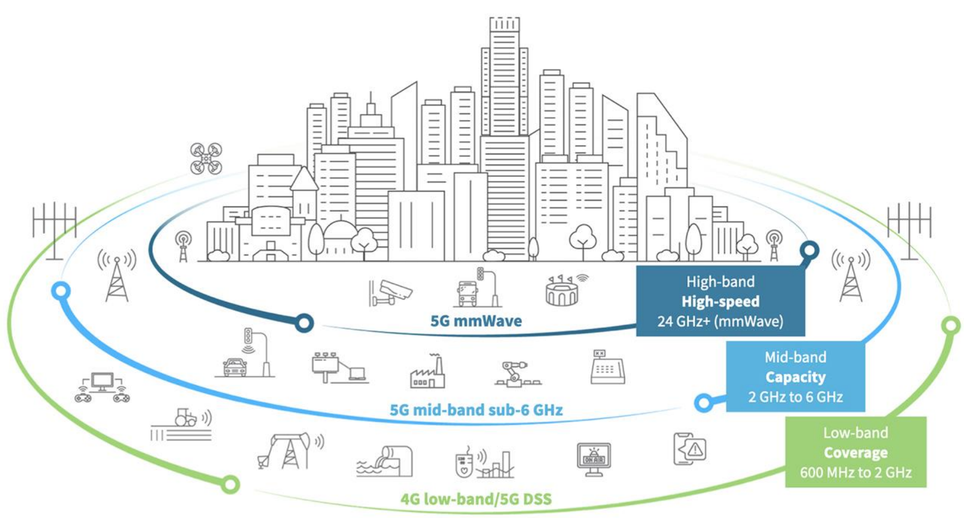

Figure 2 – 4G & 5G Frequency Bands. From [6]

These frequency bands are specific to the Americas geographical region [29]. From region to region across the globe these frequency bands are licensed and need approval to broadcast, which is typically granted from the residing countries governing body having jurisdiction in that area [29].

The 600 MHz to 2.5 GHz frequency band can support cellular communications for up to 45 miles and can pass through obstacles such as buildings making it the most ubiquitous frequency spectrum coverage across the globe [11,12]. The lower frequency of this band results in less bandwidth being able to be supported at this frequency compared to the other higher frequency bands supported. The latency is also increased due to the potential distance the cellular signal may have to travel to and from the gNBs. The gNBs are setup to provide service to a limited number of UEs that can be supported on the available frequency band that can be offered within the cell area of the gNBs. This is also designed in a way so that neighboring gNBs do not overlap frequency bands with each other so one neighbor may not utilize the same frequencies within the 600 MHz to 2.5GHz frequency range that its neighbor gNB/s utilize [5,7]. This same frequency band range partial utilization is also used on the other higher frequency bands to avoid having a connection interfere with one another. We will later see that there are several ways to alter the frequency band delivery to still allow utilization of the same frequency of different devices within the same area by methods that include utilizing OFDM [5]. The wide coverage area that this frequency band offers makes it the best option for the mMTC use case for devices such as wireless sensor and actuator nodes spread across acres of agricultural land [5,23]. High availability of this frequency band also makes it better suited for uMTC uses as high bandwidth is not needed [5].

The next higher range of frequencies between 2.5 and 6 GHz are sometimes grouped with the previous lower frequency band, but are more recently utilized in cellular service than the previous frequency range [5]. These frequencies are supported by 5G and somewhat supported by 4G cellular services [29]. This frequency range provides better throughput than the previous lower frequency band but does not provide as much area coverage because higher electromagnetic frequencies don’t travel as far and are more easily absorbed or reflected by solid objects. This frequency band is also a good option for mMTC but bellow 2.5 GHz is better [5].

The final frequency band, that is only supported by 5G, is the 24 GHz and above frequency band [5,29]. This frequency band provides the least amount of area coverage and needs line of sight, due to these higher frequencies being obstructed by obstacles, but it provides the best capacity for higher bandwidth applications [5,7]. The deployment of this frequency band support is more costly to deploy than the other frequency bands because it requires more gNBs to be deployed in a smaller area than the previous frequency bands in order to provide uninterrupted service across an area [5]. This frequency band is suited for all three different applications in certain aspects of their intended use but not all. This frequency band provides a higher bandwidth than is possible by the other lower frequency bands so it can meet the need of xMBB but it does not provide as much coverage so it would not be as well suited for uMTC that need large coverage and high mobility [5]. This low mobility and coverage means that in the sense of moving UEs it does not provide as much reliability for mMTC or uMTC that the lower frequency bands provide [5]. It can however provide more reliability with certain applications that don't need high mobility and large area coverage. This is because there is not a lot of use of this higher frequency band so more spectral space can be dedicated to providing service, as well as the UEs being closer to the gNBs which also reduces latency and increases reliability [7]. This higher frequency band also allows for smaller scheduling blocks of time that increases the reliability of the service at rates as low as 0.125ms so more can happen in a shorter amount of time which increases reliability for such applications such as uMTC applications that need this type of reliability such as stationary UEs such as industrial controls and sensors that can be located in a smaller area and have constant uninterrupted service [7]

In order to deploy this higher frequency band, that has a wavelength that is only millimeters wide, is to deploy new physical RAN technology called mmWave [5,7]. The hardware required to produce this frequency range only needs much smaller antennas then the previous lower frequency bands that require antennas that are much larger [5]. This high frequency band with millimeter wide wavelength has a much higher loss over distance than the lower frequency bands [5,7,25]. A simplified view of the loss of signal over distance can be calculated from a free space loss equation with relation to the frequency and wavelength of the signal as is seen in the equation below from [25].

Free space path loss: $FSPL = \left(\frac{4\pi df}{c}\right)^2 = \left(\frac{4\pi d}{\lambda}\right)^2$ (1)

d = distance from gNB to UE

f = frequency of cellular signal

c = speed of light

λ = wavelength size of the cellular signal

In this equation it can be seen that the distance and frequency are inversely proportional to the wavelength of the signal. So, if the distance or the frequency increases, irrespective of each other, the signal is reduced and if the wavelength increases the signal strength is increased. Therefore, the higher frequency bands have much higher signal loss than the lower frequency bands do over that same distance. What happens is that the signal will spread out omnidirectionally as it travels outward as can be seen in Figure 3 from [25] below.

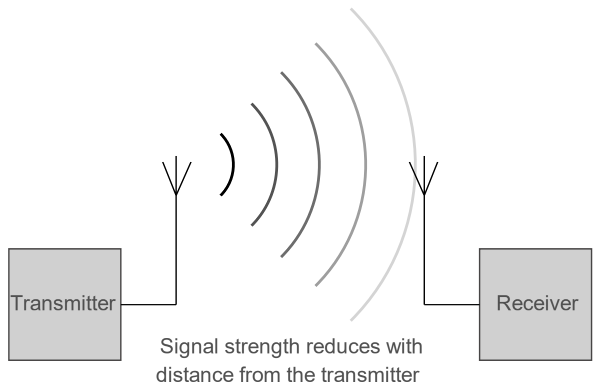

Figure 3 – Radio signal reduction over distance. From [25]

This phenomenon can also be examined with reguard to the antennas themselves using Friis propogation formula that, assuming the Rx and Tx antennas are isotropic, calculates the signal strength at the receiving antenna, based on the receiving antennas area, and the transmitting signal power [18].

Friis transmission formula from [18]:

Antenna receiving signal power = $P_r\ (d) = P_t G_t G_t \left(\frac{\lambda}{4\pi}\right)^2 d^{-n}$ (2)

P_r = Transmitting antenna signal power

G_t = Transmitting antenna gain

G_r = Receiving antenna gain

λ = wavelength size of the cellular signal

d = distance from gNB to UE

n = 2 in free space (no obstructions/line-of-sight)

With Friis tramsission formula it can be seen that the signal at the receiving antenna is similarly reduced by the size of the antenna aperture, the disance, and the increasing of frequency or inversly the decreasing of wavelength.

To help overcom this drawback of omnidirectional signal propogation a radio technology known as beamstearing also known as beamforming is used [2,5,16].

Beamforming is a way of altering the radio signal in a way that a larger portion of the signal is propagated in a specific direction [2,5]. This is accomplished by using multiple antennas that create multiple signals that when combined will cause parts of the signal to be reduced and parts of the signal to be increased based off how the signals interact with each other in either a constructive or destructive way to the signals from each antenna [2,5]. This use of multiple antennas is known as MIMO (Multiple Input Multiple Output) and will be discussed in more detail next with the introduction of mMIMO (Massive Multiple Input Multiple Output) in 5G. The antennas can adapt the signals to follow UE that is connected to the antennas by scanning for the connecting UEs connected [2,5]. The use of digital antennas in the MIMO array of antennas provides for more adaptability than just using analog or a hybrid form of analog and digital antenna arrays [2]. The use of beamforming allows for an increase in efficiency and a reduction in interference that allows for the shortcomings of higher frequency bands to be utilized [2].

The use of multiple antennas is known as MIMO and increases the number of UEs that can be connected within a single cellular cell, the number of streams that can be connected to a single UE, which in turn increases the bandwidth to a single UE, and the ability to create constructive directional radio beams to different UEs to supplement the shortcomings of mmWave while maintaining the benefits of the mmWave frequency band which supports much higher bandwidth due to the increased frequency and the uncongested frequency band with relation to the lower frequency bands also utilized by 5G [5,27]. MIMO is already used in 4G and also in WiFi (IEEE 802.11n) but is enhanced in 5G to support more antennas and therefore increase all the previously listed capabilities of MIMO to result in what is referred to as mMIMO in 5G deployment [27]. MIMO functions by several methods it uses multiple antennas to transmit data to a single antenna on a UE or multiple antennas on a UE as is seen in Figure 4 from [27] below.

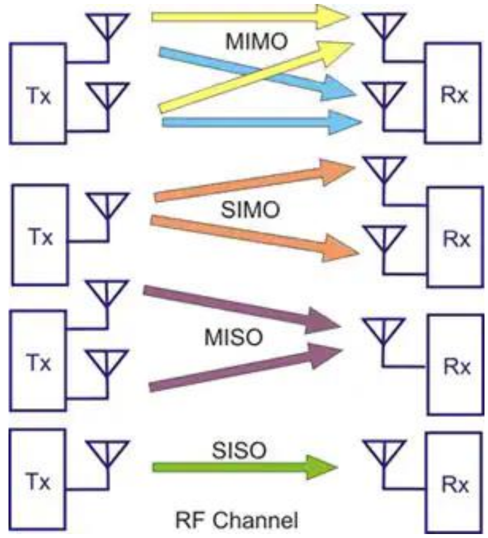

Figure 4 –Different types of antenna input and output options. From [27]

These multiple signals may take different paths as a result of signal reflection to reach the UE as is shown in Figure 5 from [26] below. These signals will arrive at different times and can be taken advantage of by using what is known as spatial multiplexing in which multiple signals are sent with spatially different information that travels along different paths but are sent from the same antenna on the same frequency, bu each signal is orthogonal to one another so they don't impede with one another [5]. This results in the receiving UE getting more bandwidth when the different signals are received. This leads us into how the signal is orthogonally separated through a process known as OFDM.

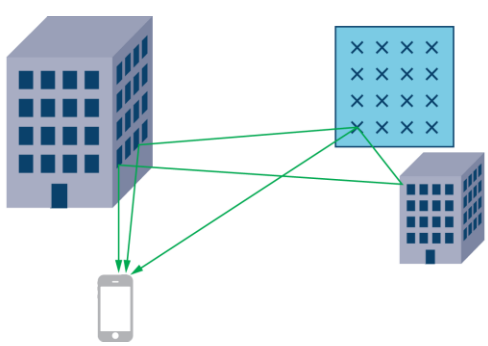

Figure 5 – Radio signal multipath reflections. From [26]

Using these orthogonal signals are what is known as OFDM (Orthogonal Frequency-Division Multiplexing), and is used in 4G and 5G, which divides the bandwidth within a given frequency to sub-carriers that are orthogonal to each other and that contain a portion of the data stream to the perspective UE or gNB [5]. This allows for more bandwidth to be sent at once between a UE and a gNB [5,7]. Combined this allows mMIMO to provide more bandwidth, more UE capacity, and more reliability within the given cell that it operates, and currently in more recent cellular devices there are about 2 antennas so these UEs can take advantage of a 2 x 2 MIMO signal that makes xMBB possible in these devices [26,27]. The processing necessary to perform these complex computations to split signals and adjust the network was previously not capable of being supported on a large scale due to the limited computational capabilities of the UEs and the mobile cores of the cellular networks [26,27]. The actual introduction of SDN (Software Defined Networking) usage in 4G networks changed this, and of which is further enhanced in the 5G network, allowing for the cellular network and the UEs to simplify and redistribute the processing necessary to provide mMIMO through OFDM on a globally accessible scale [5,7,27].

5G Mobile Core

With the advent of SDN usage in 4G to provide NFV (Network Function Virtualization) it became possible to provide a separation between the user plane data sent and the network control data sent that manages the different connections and data flows which significantly decreased the computational power and communication overhead to manage connections by the UEs and the main cellular mobile core which can be seen in Figure 6 from [7] [28]. The 5G mobile core further advanced the virtualization of the mobile cores functionality by not only utilizing the NFV introduced in the 4G network, within the 4G mobile core, but also to completely virtualize the mobile core itself to make it possible to provide the mobile cores functionality anywhere a cloud server is available to support it, through virtualization of the mobile core as can be seen in Figure 7 from [9]. Here in the 5G mobile core it can be seen that the mobile core can have its user plane and data plane functions virtually separated to be processed in different physical locations on virtual servers of the mobile core geographically. This was limited to only happen on a dedicated server for the 4G networks monolithic mobile cores design which can better be seen in Figure 7 from [9] as the xMBB slicing section but with all the analytics processing also only happening in the central data center location of the cloud. What 5Gs mobile core introduces is the ability to geographically change the location of the core functions of the mobile core allowing for much less latency for communications that need low latency [7]. 5G also introduces the ability to provide separate network connection functionality tailored to different needs as outlined earlier that include the needs of xMBB, mMTC, and uMTC through the use of network slicing provided by separately definable NFV sessions known as network slices [7].

The mobile core itself separates the user plane and the control plane data and processes these data streams separately [7]. The 4G and 5G mobile cores use an SBA (Service Based Architecture) that is designed to provide separate services at different network levels depending on the needs of the network connection [6,7]. The user plane is mainly the connection stream that contains the main data flow between the internet and the UEs such as video or audio data steams while the control plane is the connection stream that supports the management of the connected streams and the services that the mobile core offers that support the connection and the user plane [6,7,8]. In the 4G mobile core the separate services that are also a part of the 5G mobile core are: AMF(Access and Mobility Management Function) which provides connection and security management, SMF (Session Management Function) which provides management for specific connection sessions, UPF (User Plane Function) which provides routing and QoS support for the user plane packets which are then routed from the UE to the internet under the management of the control plane, PCF (Policy Control Function) which provides rules for the functions of the control plane, AUSF (Authentication Server Function) which provides authentication functions in the control plane, UDM (Unified Data Management) which provides more access management security [8]. The 5G mobile core further expands its capacities by introducing the following services: NEF (Network Exposure Function) which provides communication between the internal and external processes, NRF (Network Repository Function) which provides support to keep track of network service functions, NSSF (Network Slice Selection Function) which provides management of 5Gs introduction of network slices [8]. All of these services can be viewed as a logical distribution of services and functional connections within the mobile core in Figure 6 from [7] below.

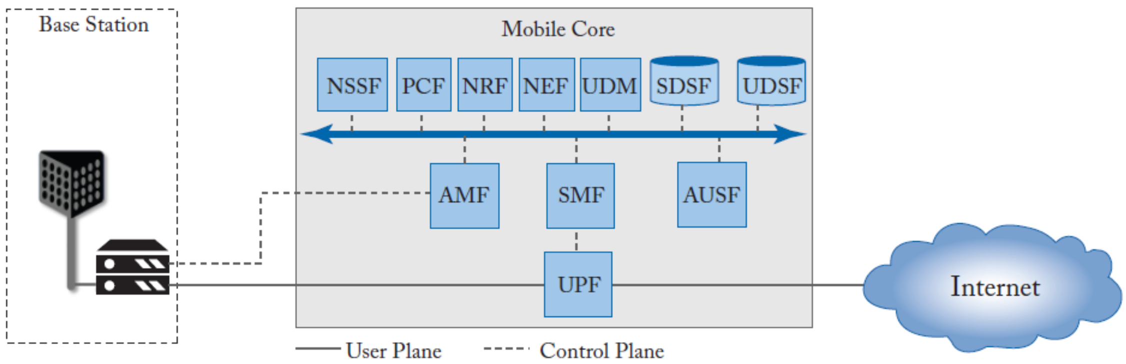

Figure 6 – 5G Mobile Core. From [7]

As can be seen through the description of the main functions of the mobile core that the user planes processing is significantly simplified as it is managed mainly by the control plane allowing the user plane data streams to be much faster due to less computation overhead which results in less latency and therefore more throughput ever increasing the bandwidth capabilities of cellular networks to provide ever increasing bandwidth and lower latencies.

The 5G mobile core design offers the capability to provide even finer grain control of the NFV than the previous 4G mobile core. With the increased services added to the 5G mobile core 5G can provide mobile core virtualization on a service level that allows for the separate services provided by the mobile core to be separated geographically on different cloud servers due to the complete virtualization of the mobile core into separate services. As can be seen from Figure 7 from [9] the separate services, mainly the user plane and the control plane services can be separated and provided at different locations to improve service. Also, arguably the largest improvement of 5G over 4G is the ability to provide separate service functions to separate connections that run on the same 5G mobile core that was not possible before with 4Gs mobile core. This separation of network functions, due to virtualization of the network is known as network slicing [5,7,9].

Network slicing is in essence providing separate service levels to different network connections on the same network by managing the scheduling of packets on the RAN, such as was previously descripted to support applications such as lower latency application connections that have uMTC needs, and largely by virtually setting up mobile core instances that only provide the needed services for the specific network slice connection based on the connections needs that have largely been laid out previously as the needs for xMBB, mMTC, and uMTC. In Figure 7 from [9] below several possible network splices can be seen that provide the needs of xMBB, mMTC, and uMTC.

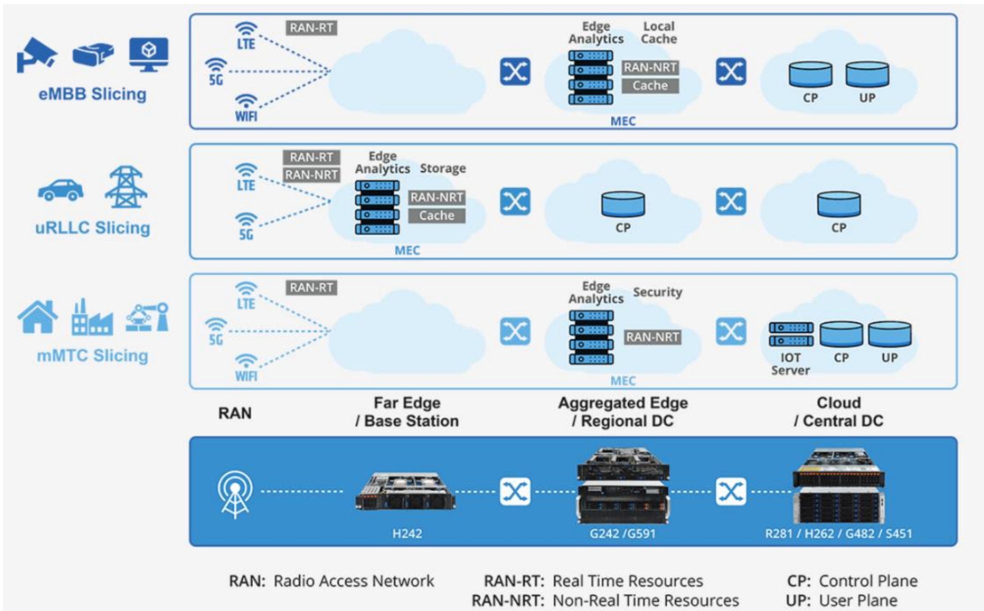

Figure 7 – Network Slicing. From [9]

The xMBB slice is shown to support the needs of extreme mobile broadband by providing high levels of bandwidth through the use of providing several connections at once that can all stream data along the user plane to the UE. The data can even be cached in a server closer to the connected UE to reduce the amount of traffic across the internet if the data is redundant such as a live video broadcast that is viewed online by several users in a nearby cellular cell location such as several students or employees viewing a teleconference on phones, tablets, or laptops connected to cellular service and possibly Wi-Fi service as well at the same time. In the uMTC scenario it can be seen that the services needed that include real time latency can be provided by creating a network slice that provides the user plane service as close to the physical UE connection that offers the lowest latency. This type of virtual mobile core functionality provided is referred to as edge computing and can even be located within the cellular tower providing the RAN connection for the UE. The combination of the reduction in latency due to the computation needed for the user planes functionality being reduced, that is managed by the control plane, and the reduction in distance the signal has to travel by using edge cloud computing results in 5G being able to offer uMTC that was previously restricted by 4Gs main design limitations [5,7,9]. Therefore, network slicing can be seen as the main key to 5Gs ability to provide a much more diverse range of services than 4G is capable of offering.

Comparison

Based on the study of the advances that 5G provides over 4G a table can be created that showcases the main improvement metrics of 5G over 4G. The separate use cases introduced are also shown with relation to the support they need from the enhanced metrics that 5G is capable of providing, shown in Table 1 below. I have then consolidated this literary study of 5Gs technology improvements over 4G that make the improved metrics, that 5G is designed to be capable of, possible as can be seen in Table 1 as well.

Design Specification Metrics	Cellular Network Generations		Use Cases supported			Supporting 5G Technology
	4G	*5G*	*xMBB*	*mMTC*	*uMTC*	
Peak Data Rate (Gbits/s)	1	20	X			mMIMO, mmWave, Connection Aggregation, Network Slicing, Enhanced Mobile Core Design
User Experienced Data Rates (Mbits/s)	10	100	X			mMIMO, mmWave, Connection Aggregation, Network Slicing, Enhanced Mobile Core Design
Connection Density (Devices/km^2)	10^4	10^6		X		mMIMO, Network Slicing, Enhanced Mobile Core Design
Latency (in user plane) (ms)	10	1	x		X	Network Slicing, Enhanced Mobile Core Design
Mobility (km/h)	350	500	x		x	mMIMO, Network Slicing
Network Energy Efficiency	1x	100x	x	x		Network Slicing, Enhanced Mobile Core Design
Reliability					X	mMIMO, Network Slicing, Enhanced Mobile Core Design

Table 1 - 4G vs 5G, Use Case Support, Supporting 5G Technology [11]

4G has a single network design which limits what capabilities the 4G network can support as was discussed in detail earlier through this study. For example, providing support for certain properties of the network such as high data rates causes a reduction in the possible latency, efficiency, and reliability. The most significant change that 5G offers over 4G is the possibility of supporting multiple networks with differing capabilities at the same time on the same physical RAN through NFV and then creating virtual network slices with different capabilities that can be supported on these different slices as was seen in Figure 7 from [9]. In 4G when a lot of devices were connected at once mMTC could potentially function but in a single cell there would be a lot of latency created that would make it so other services like xMBB and especially uMTC type connections would be limited or impossible due to high latency. 5G solves this by having a mobile core that can process the user plane closer to the UEs through virtualization and further solves this issue by supporting separate virtual network slices that can individually support higher connection density while also supporting reliability, lower latency, and high bandwidth. Massive MIMO also helps support this by providing for more connections and more separate streams to a single device allowing for higher bandwidth and more connections, more reliability as there are more fallback connections if any fail or become degraded. Massive MIMO has also enhanced the ability for beamforming in 5G as it provides for more antennas to provide better beamforming response and scale for increased numbers of UEs in a cell, which makes mmWave frequencies more efficient to use as the radio signal can be constructively increased in the direction of many UE connections at once. The network slices provide the possibility to also setup network connections that have less overhead and are more energy efficient because they require less computations by the UEs so mMTC applications are better supported while still supporting other network functionality. The introduction of support for frequencies of 24GHz and above through the use of mmWave antennas also has helped support more bandwidth in 5G than previously in 4G. Another aspect of 5G that is a part of 4G but is enhanced by 5G is connection aggregation in which multiple connections are possible at once to a single device, increasing bandwidth, reliability, and mobility. The full virtualization of the 5G mobile core in combination with NVFs and network slices has made 5G capable of scaling much more than 4G as the 5G core can have instances of it spun up on edge cloud servers that are much closer to the UEs than was previously possible with 4Gs design significantly reducing latency. This all in all adds up to 5G having many enhanced capabilities over 4G as well as providing for new opportunities than 4G previously couldn’t support.

Conclusion

Cellular service has become one of the leading forms of communication technology that is utilized across the world for personal and professional communications, this has made it important for the cellular services to evolve to meet the high-performance communication networking needs of society [12]. In order to meet the needs set out by high-performance communication networks 5G cellular networks need to utilize the existing communication networks to their full capacity and expand them to meet new network communication needs that are limited by the previous technologies. The main things needed for this high level of performance is high rates of operation, multi-service integration, and the quality of service for these multiple types of services. 5G has been designed to provide increased rates of operation through enhancements to its mobile core, and its RAN. The key defining addition that 5G provides that 4G was lacking is the ability to provide multi-service integration through the use of virtual network slicing to provide support for xMBB, mMTC, and uMTC that 4G has been limited in its ability to provide. The quality of service provided for the network connections offered by 5G is also significantly improved over 4G through 5Gs virtual network slicing. The virtual mobile core and the SDN slices that 5G provides has set it up to not only provide high performance communication services for the current needs of cellular networks but for future needs as well with its ability to easily scale and provide new and unique network connections as new use cases are explored.

Abbreviations

3GPP	3rd Generation Partnership Project
4G	4th Generation Cellular Network
5G	5th Generation Cellular Network
eNB	Evolved Node B (4G Base station)
gNB	Next Generation Node B (5G Base Station)
MBB	Mobile Broadband
MIMO	Multiple Input Multiple Output
mMIMO	Massive Multiple Input Multiple Output
mMTC	Massive Machine Type Communications [aka] IoT (Internet of Things)
mmWave	Milimeter Wave
NFV	Network Function Virtualization
NOMA	Non-Orthogonal Multiple Access
RAN	Radio Access Network
SBA	Service Based Architecure
SDN	Software Defined Networking
UE	User Equipment
uMTC	Ultra-Reliable Machine Type Communications [aka] URLLC (Ultra Reliable Low Latency Communications)
xMBB	Extreme Mobile Broadband [aka] eMBB (enhanced Mobile Broadband)
(Mobile Core)	
AMF	Access and Mobility Management Function
AUSF	Authentication Server Function
NEF	Network Exposure Function
NRF	Network Repository Function
NSSF	Network Slice Selection Function
PCF	Policy Control Function
SMF	Session Management Function
UDM	Unified Data Management
UPF	User Plane Function

References

[1] 5G Explained - How 5G Works. EMF Explained 2.0. (2020, August 23). Retrieved 2021, from http://www.emfexplained.info/?ID=25916.

[2] Riina. (2021, August 20). Beamforming antennas –how they work and are tested? Verkotan. Retrieved 2021, from https://verkotan.com/2021/beamforming-antennas-how-they-work-and-are-tested/.

[3] S. A. Ashraf, R. Blasco, H. Do, G. Fodor, C. Zhang and W. Sun, "Supporting Vehicle-to-Everything Services by 5G New Radio Release-16 Systems," in IEEE Communications Standards Magazine, vol. 4, no. 1, pp. 26-32, March 2020, doi: 10.1109/MCOMSTD.001.1900047.

[4] V. -L. Nguyen, P. -C. Lin and R. -H. Hwang, "Enhancing Misbehavior Detection in 5G Vehicle-to-Vehicle Communications," in IEEE Transactions on Vehicular Technology, vol. 69, no. 9, pp. 9417-9430, Sept. 2020, doi: 10.1109/TVT.2020.2975822.

[5] Osseiran, Afif. 5G Mobile and Wireless Communications Technology / Edited by Afif Osseiran, Ericsson, Jose F. Monserrat, Universitat Politècnica De València, Patrick Marsch, Nokia. 2016. Print.

[6] Remmert, H. (2021, March 19). What is 5G network architecture? Digi International. Retrieved 2021, from https://www.digi.com/blog/post/5g-network-architecture.

[7] Peterson, Larry, and Oğuz Sunay. Synthesis Lectures on Network Systems. 1st ed. Vol. 1. San Rafael: Morgan & Claypool, 2020. Synthesis Lectures on Network Systems. Web.

[8] Koksal, S. (2019, November 12). Core network evolution(3G vs. 4G vs. 5G). Medium. Retrieved 2021, from https://medium.com/@sarpkoksal/core-network-evolution-3g-vs-4g-vs-5g-7738267503c7.

[9] GIGA-BYTE Technology Co., Ltd. (n.d.). [5G-mmtc]smart city solution: Solution - gigabyte global. GIGABYTE. Retrieved November 2021, from https://www.gigabyte.com/Solutions/Networking/mmtc.

[10] 3GPP. (2021). Release 17. Retrieved 2021, from https://www.3gpp.org/release-17.

[11] NOKIA. (2016). Minimum technical performance requirements for IMT-2020 ... ITU. Retrieved 2021, from https://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2020/Documents/S01-1_Requirements%20for%20IMT-2020_Rev.pdf.

[12] Delaporte, A., & Bahi, K. (2021, September). The State of Mobile Internet Connectivity 2021. GSMA. Retrieved 2021.

[13] Lema, Maria A., et al. "Business case and technology analysis for 5G low latency applications." IEEE Access 5 (2017): 5917-5935.

[14] Samsung Electronics Co., Ltd. (2019). 5G Core Vision Technical Whitepaper

[15] Rost, P., Mannweiler, C., Michalopoulos, D. S., Sartori, C., Sciancalepore, V., Sastry, N., ... & Bakker, H. (2017). Network slicing to enable scalability and flexibility in 5G mobile networks. IEEE Communications magazine, 55(5), 72-79.

[16] Marsch, P., Da Silva, I., Bulakci, O., Tesanovic, M., El Ayoubi, S. E., Rosowski, T., ... & Boldi, M. (2016). 5G radio access network architecture: Design guidelines and key considerations. IEEE Communications Magazine, 54(11), 24-32

[17] ITU. (2021, February). Recommendation ITU-R M.2083-0, Detailed specifications of the terrestrial radio interfaces of International Mobile Telecommunications-2020 (IMT-2020).

[18] Xiao, Ming, et al. "Millimeter wave communications for future mobile networks." IEEE Journal on Selected Areas in Communications 35.9 (2017): 1909-1935.

[19] Leyva-Mayorga, Israel, et al. "Adaptive access class barring for efficient mMTC." *Computer Networks* 149 (2019): 252-264.

[20] Ye, Neng, et al. "A random non-orthogonal multiple access scheme for mMTC." 2017 IEEE 85th Vehicular Technology Conference (VTC Spring). IEEE, 2017.

[21] Hui, Hongxun, et al. "Modeling and control of flexible loads for frequency regulation services considering compensation of communication latency and detection error." Applied energy 250 (2019): 161-174.

[22] Hu, Y., Donnelly, M., Helmer, T., Tram, H., Martin, K., Govindarasu, M., ... & Cioni, M. (2010, January). NASPInet specification-an important step toward its implementation. In 2010 43rd Hawaii International Conference on System Sciences (pp. 1-9). IEEE.

[23] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee and B. Shim, "Ultra-Reliable and Low-Latency Communications in 5G Downlink: Physical Layer Aspects," in IEEE Wireless Communications, vol. 25, no. 3, pp. 124-130, JUNE 2018, doi: 10.1109/MWC.2018.1700294.

[24] Triantafyllou, Anna, Panagiotis Sarigiannidis, and Stamatia Bibi. "Precision agriculture: A remote sensing monitoring system architecture." Information 10.11 (2019): 348.

[25] Free space path loss: Details & calculator. Electronics Notes. (n.d.). Retrieved 2021, from https://www.electronics-notes.com/articles/antennas-propagation/propagation-overview/free-space-path-loss.php.

[26] Masterson, Claire. "Massive MIMO and beamforming: The signal processing behind the 5G buzzwords." 10 Massive MIMO and Beamforming: The Signal Processing Behind the 5G Buzzwords (2017): 10.

[27] Chatterjee, P. (2012, December 20). Wireless Mimo Driving RF Challenges. DigiKey. Retrieved 2021, from https://www.digikey.com/en/articles/wireless-mimo-driving-rf-challenges.

[28] Fulton, S. (2018, May 22). What is SDN? how software-defined networking changed everything. ZDNet. Retrieved 2021, from https://www.zdnet.com/article/software-defined-networking-101-what-sdn-is-and-where-its-going/.

[29] Wikimedia Foundation. (2021). List of 5G nr networks. Wikipedia. Retrieved 2021, from https://en.wikipedia.org/wiki/List_of_5G_NR_networks.