Fiber and Integrated Optics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ufio20

From 1G to 5G

Henri Hodara & Edvin Skaljo

To cite this article: Henri Hodara & Edvin Skaljo (2021): From 1G to 5G, Fiber and Integrated Optics, DOI: 10.1080/01468030.2021.1919358

To link to this article: https://doi.org/10.1080/01468030.2021.1919358

Published online: 20 Dec 2021.

FIBER AND INTEGRATED OPTICS
https://doi.org/10.1080/01468030.2021.1919358

From 1G to 5G

Henri Hodara[a] and Edvin Skaljo[b]

[a]SymbiOptix, Inc., Dana Point, California, USA; [b]Department for Physics & BH Telecom doo Sarajevo, University of Sarajevo, Sarajevo, Bosnia and Herzegovina

ABSTRACT

Wireless Mobile Communications have undergone major "generational" changes since MacDonald of Bell Telephone Laboratories introduced the concept of cellular communications in 1979. Over the last four decades, the technology has evolved from the first generation, **1 G**, followed by **2 G, 3 G**, and **4 G**, to reach today's threshold of the fifth generation, **5 G**.

In order to grasp the magnitude of the impact **5 G** is having on our society (*5 G internet is expected cover up to 65% of the world's population by the end of 2025!*), it is essential to understand the historical development of the previous generations. This review describes and explains the technology from **1 G** to **4 G**, culminating with the current state-of-the-art of **5 G**, and its expected performance over the next ten years.

We cover the major technical features of **5 G** that differentiate them from the previous generations, in particular,the introduction of the **MM-Wave** spectrum combined with **MIMO, M**ultiple **I**nput-**M**ultiple **O**utput antenna arrays. We single out these two features as they provide orders of magnitude increases in channel capacity (20 Gb/s Downlink data transfer rates with reduced latency of 1 ms) and internet connectivity on the order of billions.

We conclude with related **5 G** marketing data, some of it related to the smart phone, a marvel of engineering and a credit to human creativity.

5 G is committed to connecting billions of devices with billions of people, both wirelessly and via the internet. Such a technical advance is a game-changer!

KEYWORDS
Circuit and packet switching; CDMA; OFDMA; SC-FDMA; Mimo antennas; 1G; 2G; 3G; 4G; 5G network architectures; mm-waves; IoT

Introduction

Welcome to the world of acronyms! The **G** stands for **Generation**. It refers to the generational technology advances in mobile telecommunications that took place over the last fifty years. These advances are defined by standards of performance agreed on by national, international and industrial organizations.

Our aim is to guide the reader through the growth of mobile wireless telecommunications beginning in the second half of the 20 th century until now. The growth took place in stages, or generations beginning with the 1st Generation (**1 G**) and ending with the current **5th** Generation (**5 G**), We shall follow the history behind such a spectacular growth, review and call out the major features of each

CONTACT Henri Hodara h2@alum.mit.edu SymbiOptix, Inc., Dana Point, CA 92629, USA

generation and discuss the myriad of applications being introduced into our living space.

Why are we writing about the five generations of Mobile Wireless Communications and more specifically about the 5th generation, 5G? Because their impact on the structure and dynamics of our society has been profoundly altered by this technology. Everyone in the world is interconnected: Person to person, person to machines, machines to machines. This is best illustrated by Figure 32 below taken from the last section of our report, Market Trends that shows the growth of the yearly connections in billions. Interconnections in 2021, according to some survey will exceed 25 billions though the earth population is less than 4 billions!

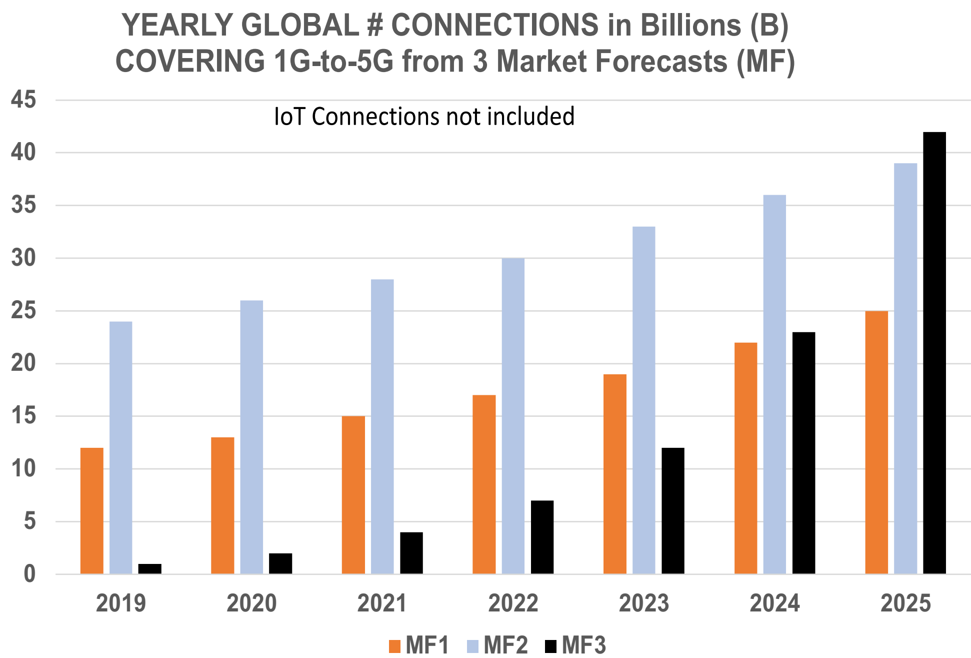


Figure 32. Comparison of 1 G-to-5 G number of connections forecast.

As a preliminary to this journey, it is worth reviewing briefly the technology associated with both "Fixed" communications, often referred to as **POTS, P**lain **O**ld **T**elephone Service and "Mobile" communications, typically represented by Smart Phones such as iPhones, and others.

Fixed communications – Circuit switching

In fixed communications (often called "Land Lines" by the telephone companies), for two parties or subscribers to communicate with each other, they need a dedicated line, referred to as "circuit"."To facilitate this process, a "Local Telephone Exchange" is necessary. Without a Local Exchange, the phone companies would be forced to have a dedicated line or circuit for every pair of possible subscribers. As shown on the left of Figure 1, 9 subscribers would require 9(9–1)/2 = 36 connection pairs. Thousands of subscribers in the same area, with identical area code (three digits) and identical prefix (next three digits in the telephone number) would require an impractical number of connecting wire lines to accommodate every possible telephone call. This is where the concept of Local Exchange situated in a "Central Office" (CO) comes in.

Figure 1, displays on the right how the individual lines from each subscriber converges into a hub, the Local Exchange located in a Central Office (**CO**). Two lines are connected to each other for the duration of a call between two subscribers, by establishing a "circuit" between them.

Circuit switching was carried out manually until near the end of the 19^{th} century when mechanical switches were introduced. In the early 50s, they were replaced by the more effective Ericsson Crossbar Telephone Exchange switches, culminating in the mid 60 s with AT&T installation of the first electronic switching. Electronic switching enabled subscribers to dial long distance across the country without the intervention of an operator.

The Local Exchange described so far enables communications between subscribers. There is another type of exchange, not located in a **C**entral **O**ffice (**CO**) but in a building that links different **CO**s. It is called a "Tandem **O**ffice," and it

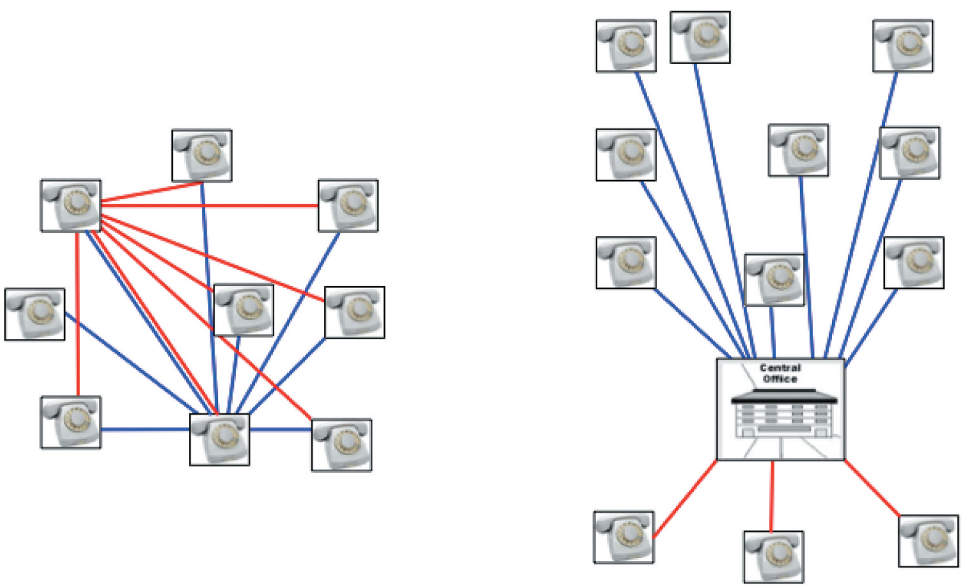


Figure 1. Local telephone exchange. LEFT: Without an exchange, we need a separate line linking every home to every other home. RIGHT: With a central exchange, each home needs only one line linking it to the exchange, which can route calls on to all the other homes.

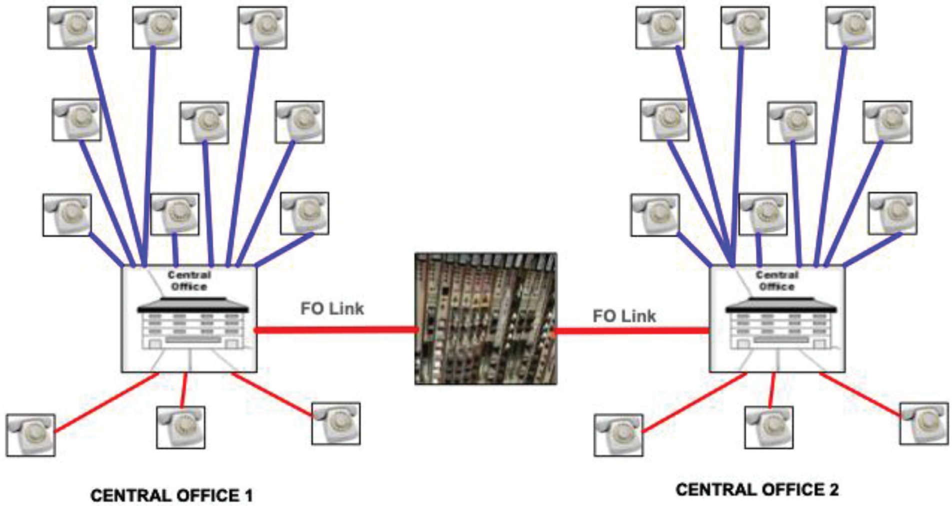


Figure 2. Long distance telephone exchange. LEFT and RIGHT: Local Exchanges (each with different area codes) with switching done in Central Office (CO)MIDDLE: Switches in Tandem Office (TO) connecting COs via Fiber Optic (FO) Cables (Trunks).

houses the electronic switching that routes calls between subscribers across the country with different area codes. It is illustrated in Figure 2, Long Distance Exchange and it is self-explanatory.

We conclude this section by noting that the aggregate of the world's circuit-switched telephone networks is referred to in the literature as **PSTN, P**ubli**c S**witched **T**elephone **N**etwork.

Mobile communications

The two major features that have enabled large-scale mobile communications are packet switching and the cellular network discussed below.

Packet switching

Although circuit switching was still used for voice in the first, **1 G** and second, **2 G** generations of mobile communications, beginning with the latter, voice was digitized. And the ensuing generations, **3 G, 4 G** and **5 G**, converted the voice signals to packets, similar to the ones used for data: **SMS, S**hort **M**essage **S**ervice text messages, and **MMS, M**ultimedia **M**essage **S**ervice, pictures and video.

In Packet Switching, the message is broken up into pieces called Packets that are sent independently, through different paths to their ultimate destinations. To keep track of each piece and reassemble them in the correct order at the receiving end, each packet consists of two major parts, Data and Header.

Breaking communication down into ***packets*** *allows the same data path to be shared among many users in the network.*

For a better understanding of packet switching, we include in this section a detailed description of the process.

In order to communicate between different software applications as well as between elements of networks such as mobile devices, computers and peripherals, the latter three also often referred as "nodes" (an unfortunate choice of words), it has become necessary to:

(1) Break up the communication functions into a hierarchy of smaller, more manageable self-contained sets of tasks called ***Layers*** with minimal coupling between them. Coupler between layers takes place through elementary software operations called ***Services***.
(2) Standardize each layer's communication functions with a set of instructions, that is, algorithms called ***Protocols***.

Together, Layers and Protocols define the packet carrying network architecture.

- Standardization has been developed by various private and governmental organizations, in particular the **I**nternational **S**tandards **O**rganization, **ISO**. These standards embodied in the **O**pen **S**ystems **I**nterconnection, **OSI** allows different equipment and software applications to communicate with each other. The **OSI** model of standards has often been criticized as "bureaucratic" and developed by "committee." Instead, since the bulk of mobile communications travels globally through the internet, a set of standards developed originally for the internet by Cerf and Kahn from **ARPA,** ***A****dvanced* ***R****esearch* ***P****rojects Agency* in the middle 60 s are used. The standards developed for the internet are a simplified version of the **OSI** standards and consist of only four layers (instead of OSI's seven), each with its own set of protocols. The internet standards are called **TCP/IP** in reference to the principal protocols used, ***T****ransport* ***C****ontrol* ***P****rotocol*, **TCP** and ***I****nternet* ***P****rotocol*, **IP**.

Each layer consists of packetized "chunks" of **DATA**, often referred as SEGMENTS, PACKETS, or FRAMES (the name depends on the layer where it is used: Segments is in relation to Transport layer, Packet in relation to Network layer and Frame to LAN layer) to which a **HEADER** is appended, as shown in Figure 3. Key items carried by the header include: Protocol that applies to this layer and the function it performs, Source Address and Destination Address.

As one goes down the layers, other protocols apply, other functions are performed and the appropriate instructions (protocols) are appended in a process known as **Encapsulation**.

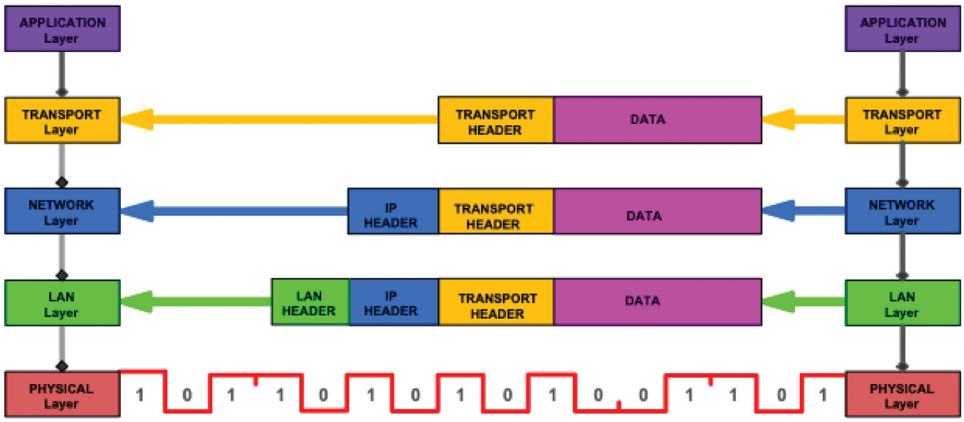


Figure 3. Internet protocol layers.

Figure 3 shows and describes the format and function of each of the four layers. where the network access layer (the lowest layer of the four layers of ARPA's layers) is shown as two layers, one is LAN, and another the Physical Layer.

Application layer

Beginning from the top let is the Application layer, which encompasses a variety of software programs utilized by a mobile subscriber. To name a few:

- E-Mail programs, such as **SMTP,** ***S**imple **M**ail **P**rotocol*
- **HTTP,** ***H**yper **T**ext **T**ransfer **P**rotocol* used by the World Wide Web defines how messages are formatted and transmitted,
- **FTP,** ***F**ile **T**ransfer **P**rotocol* as the name indicates serves to provide transfer of files between two computers (often referred as between Client and Server).

Transport – Layer 4

In order to transmit the data obtained from the Application Layer between Sender and Receiver, a segment with data and a header must be assembled. The data consists of anywhere from around 100 bytes to over 1000 bytes. The header includes the required instructions (protocols). There are two types of protocol in this layer: **TCP** and **UDP**

- **TCP,** ***T**ransmission **C**ontrol **P**rotocol* is the original protocol developed by ARPA, when it was under the influence of the telecommunication industry. It is a *Connection Oriented* protocol much like telephony *Circuit Switching*. It establishes a transmission channel with end-to-end connection between sender and receiver, which guarantees delivery of the message with error correction, and the receiver sending back an *acknowledgment* (**ACK**) packet.

- **UDP, *User* D*atagram* *P*rotocol** is a *Connectionless* transmission. Unlike **TCP**, packets are sent in many directions, so as not to depend on one end-to-end channel. The protocol adds numbers to each packet, so that upon receipt they can be reassembled in the proper order, independently of their arrival time. Unlike **TCP**, there's neither guaranty of delivery, nor error correction, and the lack of *Packet acknowledgment* (**ACK**) makes it speedier and ideal for video streaming.

(UDP is at the root of Packet Switching, developed by ARPA in 1969 under the name ARPANET. Major contributors included Cerf and Kahn from the US, Donald Davies from the UK and Louis Pouzin from France.)

NETWORK (Internet) LAYER 3

Now that the packets include the instructions for successful transmission formulated in Layer 4, a new Header is appended in this layer that includes both, Sender and Receiver Addresses, as well as the protocol being used.
Internet Protocol is the addressing protocol and it serves two functions:

1) it labels every device and interface throughout the network with a unique identifier;
2) it provides a unique address to such devices and interfaces

LAN (Local Area Network, e.g. Ethernet) LAYER 2

The packet in this layer with the appended Header is called a FRAME. The protocols in this packet enable the frame to navigate through the local area network by specifying a new type of address, the **MAC** (***M**edium **A**ccess **C**ontrol*).

MAC, *Medium Access Control.* What is a MAC address and how does it differ from an IP address? MAC addresses are physical addresses that identify a specific device. They often reside in an NIC, Network Identification Card, provided on the device by its manufacturer.

MAC addresses identify machines within the same broadcast network in layer 2, i.e. **LAN**, Local Area Network, while IP addresses used in layer 3 establish a connection between two devices.

MAC. addresses work in the following manner: Assume my device wants to send a packet to some **IP** destination address. If the destination device resides in the same local network as the sending device, then the connection is successfully established between the two through the IP protocol. On the other hand, if the receiving device resides in another network, it will not be found, and the packet must be sent to the closest router in the local network, and that router is identified by its physical and unique MAC address. There are two ways to reach the proper router that can forward the packet to the next network::

- ***BROADCASTING*** to all the routers in the local network their respective **MAC** address. The one that recognizes the destination IP address is selected to be on the correct hop, while the other routers discard the packet;
- **ARP,** *Address **R**esolution **P**rotocol.* It is a protocol capable of establishing which router can process the packet based on the sender's destination IP address.

In either case, the router has two MAC addresses, incoming and outgoing address. If the router can access the destination address, it will send the packet to its IP address. Otherwise, it will identify through its routing table the next routing hop's MAC address based on the packet's source IP address, forward the packet to it, thus repeating the process until it reaches its final destination.

An example of how **MAC** and **IP** addresses are handled in this situation is illustrated in Figure 4 [1]. The following are worthy of note:

- All routers have **IP** and **MAC** addresses on the oncoming and a different set of addresses on the outgoing side.
- The **IP** source and destination addresses on the packet never change through all the routings.
- **MAC** addresses only change when the packet crosses from one local network to another local network on its way to final destination.
- As the packet exits the router, it acquires its outgoing **MAC** address that becomes the Packet source **MAC** address.
- The packet destination **MAC** address is the incoming **MAC** address of the next router, obtained through the routing table of the exiting router.

Thus, we see that the packet is continually changing its destination **MAC** address as it hops through each router, but its final destination **IP** address never changes.

Cellular network

The concept of cellular networking was first introduced by McDonald of the **B**ell **T**elephone **L**aboratories (**BTL**) in his 1979 seminal paper, *The Cellular Concept* [2,3].

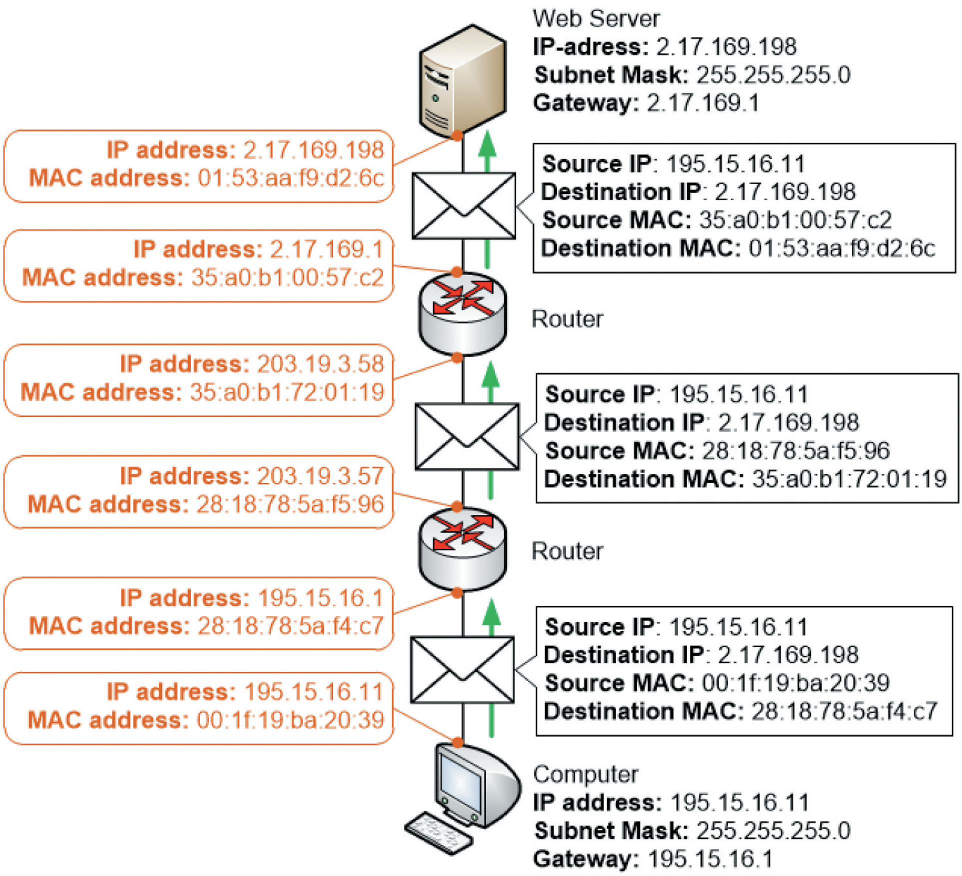


Figure 4. How MAC and IP addresses work (Ref. 1).

The concept consists of dividing an area covered by a large transmitter station into *smaller areas or cells*. Each *cell*, with lower transmitter power communicates wirelessly between telephone users in different cells, on a limited number of frequencies. Two significant results are thus accomplished;

(1) The ***FREQUENCY REUSE*** of a limited number of frequencies allows an area encompassing many cells to handle a number of simultaneous calls much greater than the number of allocated channel frequencies. The same frequencies are used in non-adjacent cells but they are low enough to avoid interference with adjacent cells operating on a set of different *frequencies*.

RESULTS: **System Capacity or number of subscribers per allocated channels is increased**. As we shall see further on, the Capacity is magnified by how many subsets of cells of different frequencies reuse the same set of frequencies.

(2) Moreover, ***CELL SPLITTING*** is a process, such as if the size of the cells is reduced, it allows an increase in simultaneous calls without adding any new frequencies. (See detailed explanations further down, and Figure 8)

RESULTS: **Subscriber Density or number of subscribers per unit area is increased**. The same spectrum allocation is maintained, thus yielding an increase in Spectral Efficiency.

Prior to discussing the two concepts above, let us turn to the choice of cell geometry

Cell geometry

A circular cell is the simplest geometry to implement. Figure 5(a) below, makes it clear what the problem is with such a geometry. The top part of the figure shows tight "packing," but with gaps. Coverage with a set of different frequencies is confined within the boundary of each cell; thus, the gaps between circular cells create dead zones where no communication is possible. The bottom of Figure 5(a) shows how to make the packing tighter by overlapping the circular cells, but that results in interference between cells.

Figure 5(b) shows three polygon configurations inscribed in the circle: Equilateral triangle, square and hexagon, with radii (defined as the line from the center of the figure to any of its vertices) equal to that of the circle. Of the three polygons, the hexagon has the largest area. Selection of higher order polygons to come closer to the surface area of the circular cell reaches a point of diminishing returns.

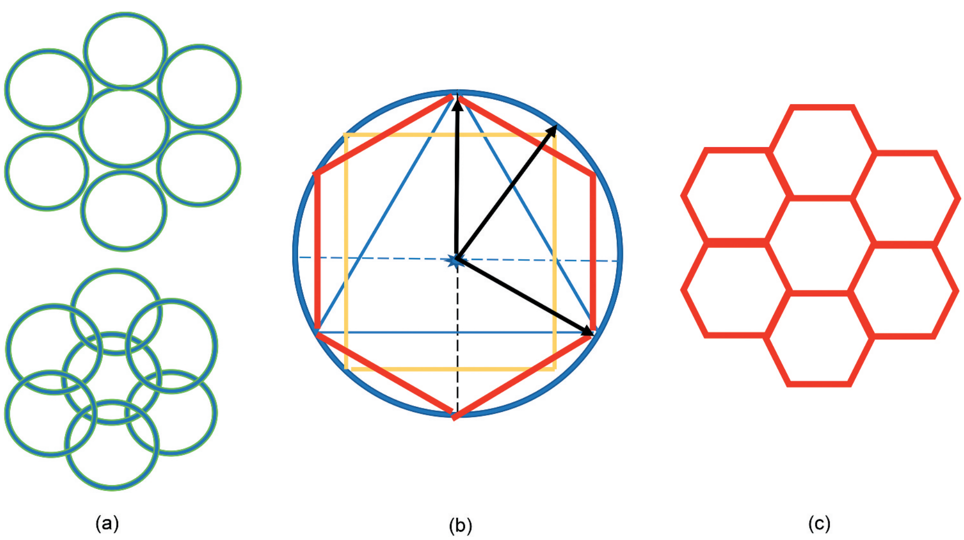


Figure 5. Cell geometry: Packing efficiency of circles and hexagons.

Figure 5(c) shows the packing efficiency of the selected hexagonal shape for communicating cells.

Frequency reuse

To fully comprehend the advantage of frequency reuse, Figure 6 illustrates the difference between a single cell (on the left of the figure) as opposed to a cluster of cells (on the right of the figure) covering the same area. The single cell requires a powerful transmitter located on a tower tall enough to overcome hill obstacles and capable of handling in this example approximately 800 frequencies, or 400 duplex voice channels, each with two frequencies. On the other hand, with a cluster of seven cells, each cell handles a different set of 57 (7 × 57 = 399) frequencies of simplex voice channels, labeled A through G, with much smaller individual transmitter towers and covering an area about one-seventh of the area covered by the single large cell on the left. Given that radiated power decays as the inverse square of the distance, each cell in the cluster needs to radiate an amount of power only 1/49 of the large single-cell transmitter, and it is adjusted so that the overspill in the adjacent cell is beyond the detectable level of cell phones in that cell. The system capacity, as shown in the inset is still 400 duplex channels.

With frequency reuse, the ***cells size in the cluster is kept the same*** as shown in the example of Figure 7. The cluster is replicated a second time with same set of frequencies, labeled A_2 through G_2, such that a transmitter in the cell labeled, say C_1 in the first cluster transmits the same set of 57 frequencies as the

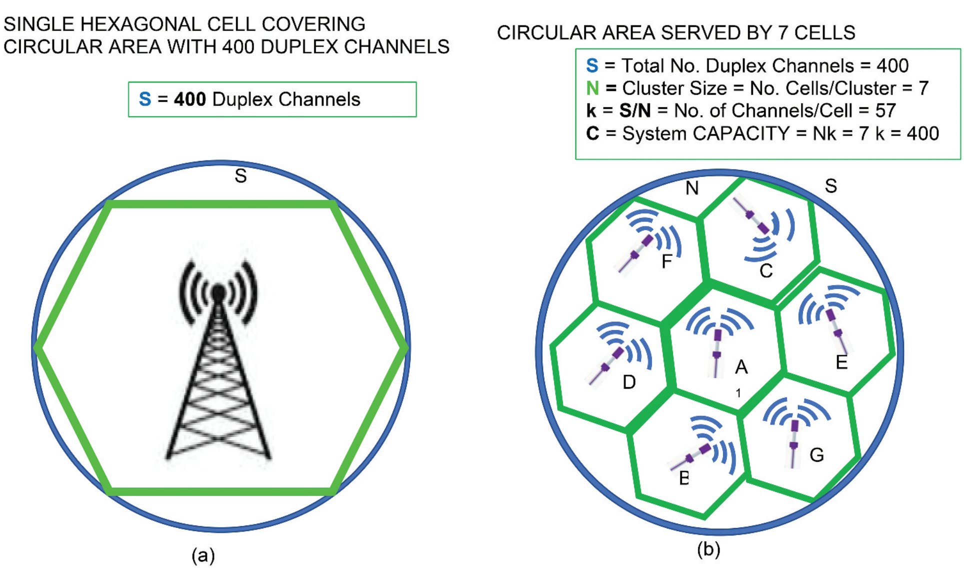


Figure 6. Circular area served by single cell and cluster of 7 cells.

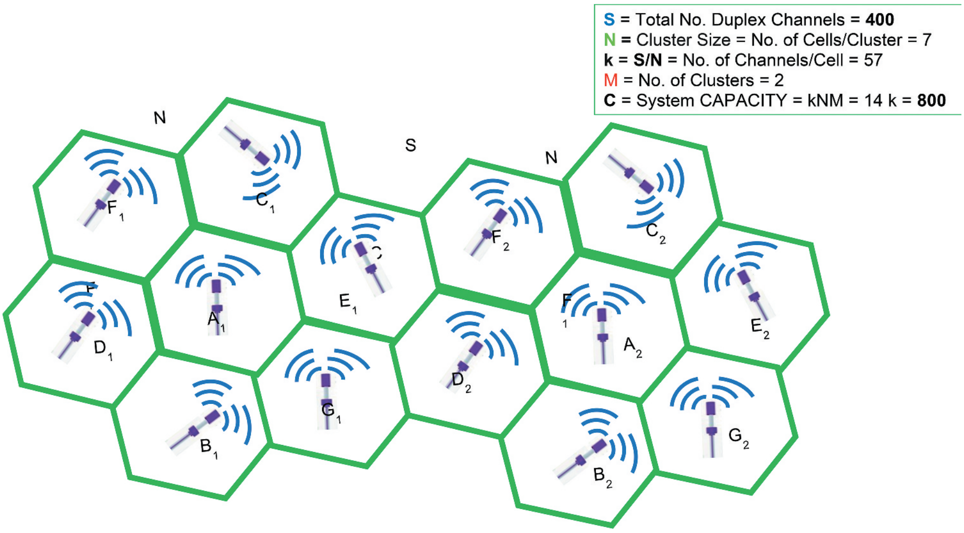


Figure 7. Frequency reuse with 2 clusters.

one labeled C_2 in the second cluster. Note that the cells with the same labels, hence the same set of frequencies are arranged to be separated far enough from each other to prevent interference.

All the frequencies used in the first cluster are reused in the second cluster doubling the system capacity to 800 voice channels, without having to resort adding new voice channel frequencies. Adding more clusters of the same size, say M cluster will increase the system capacity by M. *The gain in additional channels come at the expense of increasing the geographical area*. Such increased System Capacity, labeled **C** in the figures follows from the simple formula:

C = **k**(No Channels/Cell) **x** **N**(Cells/cluster) x **M**(No Clusters)

Cell splitting

When a cluster reaches its full capacity handling a specific number of voice channels, say S = 400 as in Figure 8, and additional voice channels are required to handle the traffic increase, cell splitting allows accommodating the additional channels without adding new frequencies, and unlike frequency splitting with no increase in geographical area. It does this by reducing the cell size and increasing the number of clusters. Figure 8 illustrates how the original single hexagonal cell of 400 channels is converted to 4 clusters of 7 cells per cluster. With each cluster handling 400 channels, and with the same frequencies in each, the total system capacity has increased four-fold to 1600 voice

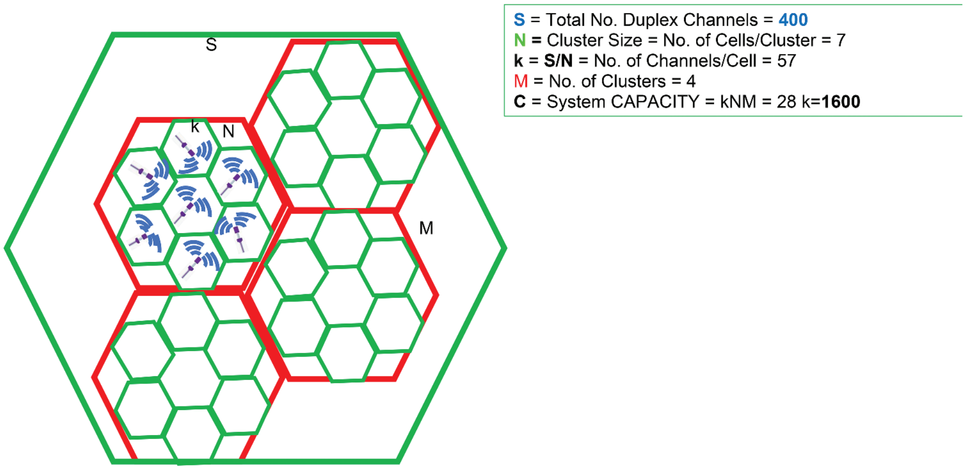


Figure 8. Cell splitting with 4 clusters.

channels. Moreover, the same geographical space is now able to serve 1600 subscribers, thus increasing four-fold the number of subscribers per unit area, that is, the subscriber density.

The cell areas in each cluster are now $4^2 = 16$ times smaller, with correspondingly much smaller transmitting tower and reduced radiated power. Such are the benefits of cell splitting: *Increased system capacity and subscriber density.*

Cluster size tradeoff

We conclude this section with a tradeoff indicating how System Capacity **C** varies as a function of various cluster size for **fixed geographical area, cell size, and channel availability.**

The tradeoff is summarized in Table 1 below with fixed:

Geographical area $\mathbf{A_0} = 500$ Km^2

Cell area $\mathbf{a} = 5$ **km**^{2}

Number of available duplex channels **S** = 400.

Table 1. Cluster size tradeoff.

Geographical Area: $A_0 = 500$ Km^2
Cell Area: a = 5 Km^2
No Duplex Channels: S = 400

No Cells/Cluster N	Channels/Cell k = S/N	Cluster Area (Km^2) A_c = a x N	No Clusters M = A/A_c	System Capacity S = k x N x M
1	400	5	100	40.000
4	100	20	25	10,000
7	35	35	14	5,600

Though System Capacity is highest with the lowest number of clusters, i.e. N =1, co-channel interference is reduced with increased cluster size. Thus, a "sweet spot" striking a compromise leads to a choice of **N** =7.

Having laid out the foundations of mobile communications based on telephone circuit and data packet switching in hexagonal cell networks, we are now in a position to review and discuss the evolution (or revolution) that has taken place over five generational (1 G to 5 G) technologies till now.

Five generations of mobile communications

From 1 G TO 5 G

1 G, the first generation of mobile networks was introduced by **N**ippon **T**elegraph and **T**elephone, **NTT** in Tokyo in 1979. ***It broadcast only voice in an analog format.***

Concurrent with the Japanese introduction, V. H. Mac Donald of **B**ell **T**elephone **L**aboratories (**BTL**) published the same year his concept of "cellular" telephony, described in Section 3, above.

In 1983, the U.S. **F**ederal **C**ommunications **C**ommission, **FCC** approved the **DynaTA, D**ynamic **A**daptive **T**otal **A**rea **C**overage of 8000X telephone, the world's first commercial cellular device demonstrated by Martin Cooper of Motorola. The resultant cellular technology was **AMPS, A**dvanced Mobile **P**hone **S**ystem.

However, before we explain the inner workings of this **A**dvanced **M**obile **P**hone **S**ystem and understand its features and shortcomings, it behooves us to begin with an overview of the **AMPS** network architecture and trace out the path of a mobile communication across such a network. Refer to Figure 9 below.

1 G network architecture

Let us trace a communication between two mobile subscribers, UE1 and UE2 in adjacent cells shown in the network architecture diagram of Figure 9.

Bear in mind that **AMPS** is a *circuit switching* network, much like landline telephones. So, when two users need to communicate with each other, an end-to-end connection (*connection oriented*) must be assigned to them for the duration of the call.

UE1 broadcasts a message on one of the **R**adio **F**requencies, **RF** available in the **AMPS** system. The RF signal is modulated (AM or narrowband FM) by an ***analog signal***. *Analog signals carry both, AM or narrowband FM voice, and digital Signaling information, which uses **F**requency **S**hift **K**eying, **FSK***. The resultant modulated **RF** signal becomes one of the components of the **F**requency **D**ivision **M**ultiple **A**ccess, **FDMA**.

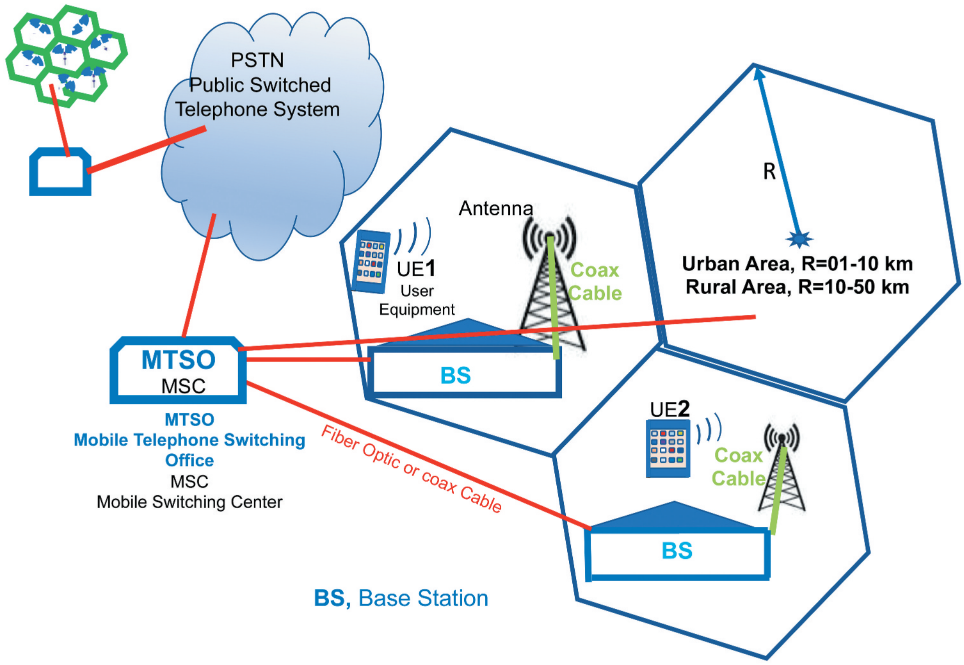


Figure 9. 1 G AMPS network architecture.

The tower antenna receives not only this **RF** signal but many others on different assigned frequencies from other users in the same cell, in short, the composite **FDMA** signal. The **FDMA** signal is relayed down via a coax cable to a shelter on the ground, the **B**ase **S**tation, **BS**, which separates all the **UE**s' frequencies, and send them to the **M**obile **T**elephone **S**witching **O**ffice, **MTSO**. How is this last operation done? Figure 9 shows a fiberoptic cable, but in the early days, the connections between the **MTSO, M**obile **T**elephone **S**witching **O**ffice and the **BS, B**ase **S**tation was done through the so-called **T1** line. Basically, the **RF** signals received at the **B**ase **S**tation, were demodulated and reduced to the voice audio signals, each of approximately 4 kHz bandwidth. Each voice signal was digitized and encoded in **P**ulse **C**ode **M**odulation, **PCM** format, with a resultant data rate of 64 Kb/s. Then, according to T1 (E1 in Europe) standards, 24 (32 in Europe) such signals were **T**ime **D**ivision **M**ultiplexed, **TDM**, bringing the total transmission rate to 1.544 Mb/s (2.048 Mb/s in Europe). Upon reception at the **MTSO**, the individual voice signals were recovered and forwarded as we shall see shortly. Later on, when fiber-optic cable became available, the composite 24 time-division-multiplexed signals were used to modulate an optical carrier and transmitted to the **MTSO**, where upon reception, they were demodulated.

In the **MTSO**, connections between the respective base stations are made according to the frequency of each signal, which is unique for that particular end-to-end connections. Similarly, the others are sent to their respective destinations, some to nearby cells, in the same cluster like the one where **UE**2 is located, or to the core network called **PTSN, P**ublic **S**witched **T**elephone **S**ystem. The latter is an aggregate of the world's circuit-switched telephone networks operated by the telephone companies.

Upon selection of the **UE**1 signal, which contains specific information about the sender's and receiver's addresses, the **MTSO** broadcasts it to all **B**ase **S**tations. When **UE**2 receives it, it replies through the same channel frequency and the **MTSO** assigns a dedicated frequency to both, as well as setting the router switches. (part of the **MTSO)** to provide a dedicated path between the **UE**s.

From the **MTSO**, the process is reversed and the **B**ase **S**tation, **BS** where **UE**2 is located receives all the signals intended for this cell, where they are each impressed on their assigned carrier by **F**requency **D**ivision **M**ultiple **A**ccess, **FDMA**. Prior to this modulation, the audio voice signal undergoes some base-band processing that includes compression, pre-emphasis to boost the high voice frequencies, limiting of the frequency deviation (if the carrier is **FM** modulated), low pass filtering to remove audio signals at the upper range of the voice signal above 3 kHz. In addition, a notch filtering at 6 kHz is included to eliminate the **S**upervisory **A**udio **T**one, **SAT**. The latter helps mobile stations distinguish neighboring **B**ase **S**tations from one another, thus preventing interference.

1 G performance features

Having reviewed the network architecture, we now spell out the features of the **1 G AMPS, A**dvanced Mobile **P**hone **S**ystem.

AMPS was introduced and deployed in the US in the early 80s and remained the technology of choice through the end of the decade, until the advent of the second generation, **2 G** of cell phone technology

AMPS is an **analog** mobile phone system standard originally developed by **B**ell **T**elephone **L**aboratories (**BTL), (**also known as Bell Labs) was later modified in a cooperative effort between Bell Labs and Motorola.

AMPS operates over **two** frequency bands, each 25 MHz wide:

- 824 MHz to 849 MHz
- 869 MHz to 894 MHz

In each band, half of the frequencies are allocated to the Sending or *Forward channel* and the other half to the Receiving or *Reverse* channel. The bands are in turn subdivided in 30-kHz wide channels, and that subdivision is accomplished with **FDMA, F**requency **D**ivision **M**ultiple **A**ccess. It results in 832 multiplexed channels for each band (25 MHz/30 kHz).

FDMA, Frequency **D**ivision **M**ultiple **A**ccess is a multiple access protocol which allows each user (subscriber) to send their information by frequency modulating a different and non-overlapping frequency in the two bands mentioned above. One major problem inherent to FDMA is low spectral efficiency because 1) if the transmitting station is idle, that frequency cannot be re-assigned to another station, which results in wasted spectrum; 2) spectral band guards do not carry information; but they are necessary to ensure non-overlap of frequencies, thus, prevent interference.

*Circuit **S**witching **T**echnology* is used by AMPS to transmit voice only but no data. *Analog **B**andwidth* of each, voice and data is 0 kHz.

Finally, all the transmitted calls are routed to a Core or Backbone network:

PSTN, Public **S**witched **T**elephone **N**etwork which is the *aggregate of the world's circuit-switched telephone networks* that are operated by national, regional, or local telephone operators, providing infrastructure and services for public telecommunication.

We have summarized in Table 2 the key performance features of **1 G**.

1 G technology was plagued with many other problems such as small coverage, no encryption, inherent poor sound quality associated with analog signals and ***no roaming***. Thus, it opened the door to improvements that the second generation, **2 G** of mobile telephony would bring about

2 G, the second generation of wireless telephony originated in Finland, in 1991, under the **G**lobal **M**obile Communications **S**ystem, **GMS** standards. The standards developed by the **E**uropean **T**elecommunications **S**tandards **I**nstitute, **ETSI** contributed to a major leapfrog in wireless mobile technology. In addition to enabling encrypted calls, digital signaling was substituted for analog, with ensuing higher quality of transmission. Communication was no longer restricted to voice, allowing telephone users to send and receive text messages up to 160 character data packets, known as **S**hort **M**essage **S**ervice (**SMS**). Moreover, pictures, video, and audio contents could also be transmitted between telephone users; the data packets became known as **M**ultimedia **M**essaging **S**ervice, **MMS.**

Three major features distinguish **2 G GMS** from **1 G AMPS**:

- Voice signals are no longer analog, but digital with a corresponding improvement in voice quality. ***GMS** is still a circuit switching system.*
- Added Text Messaging with **SMS**, plus eventually pictures and video streaming with **MMS**.
- Data Encryption, which enhanced transmission security

The second generation underwent a lot of improvements over a decade, with increasingly better performance from the original **2 G GMS** standards. The original **2 G GMS** was a slow circuit switching system capable of handling a limited amount of data. A packet switching feature was introduced with the

Table 2. Summary 1G performance features

KEY FEATURES	1G
Lifetime	1980-1990
First Location	USA
Technology	**AMPS** **A**dvanced **M**obile **P**hone **S**ystem.
Frequency Spectrum - Bands	824 MHz-849 MHz - 25 MHz 869 MHz-894 MHz - 25 MHz
Channel Spacing No. Channels/Band	30 kHz 833
Access Protocol	**FDMA** **F**requency **D**ivision **M**ultiple **A**ccess
Switching Technology	CIRCUIT SWITCHING Voice Only
Analog Bandwidth Data Rate	10 kHz
Backbone (Core)Network	**PSTN** **P**ublic **Switched** **T**elephone **N**etwork

next improvement through **GPRS, G**eneral **P**acket **R**adio **S**ervices often referred as **2.5 G**. Further improvement followed with **EDGE, E**nhanced **D**ata **R**ate **E**volution for **G**SM **E**volution, named **2.75 G**! *As an example, the transmission data rate, beginning with the original* ***GMS*** *standards went from 9.6 Kb/s to a range of 35 Kb/s-171 Kb/s with* ***GPRS****. By the end of the century, prior to the introduction of* ***3 G****, data rate with* ***EDGE*** *reached the range 126–384 Kb/s.*

We omit discussions of advanced **2 G** systems, **GPRS** and **EDGE**. Their performance is closer to **3 G**, which will be treated further on. Focusing on the early **2 G GMS** helps appreciate the improvements brought in eventually by **3, 4** and **5 G** systems.

Before we review the inner workings of **GSM**, we first describe its networks architecture.

2 G network architecture

The many functions performed by **GSM** are carried out by the four major subsystems shown in the network architecture of Figure 10:

- The **M**obile **S**tation, **MS** is the equipment the subscriber utilizes to communicate, such as a smartphone. It contains a special electronic chip, the **SIM, S**ubscriber **I**dentification **M**odule which identifies and authenticates the subscriber and allows communication with different service providers on an international scale. In addition, the **SIM** card provides security to the user.

- The **B**ase **S**tation **S**ubsystem, **BSS** consists of two elements:
  - The **B**ase **T**ransceiver **S**tation, **BTS**. It comprises the antenna mounted on top of a tower, which relays transmitted and received signals to the radio equipment housed on the ground. The radio performs several key functions, not only modulation but multiplexing, both **TDMA, T**ime **D**ivision **M**ultiple **A**ccess and **FDMA, F**requency **D**ivision **M**ultiple **A**ccess as required by **GSM**, and described further on in the section, **GSM** Mode of Operation
  - The **BCS, B**ase **C**ontroller **S**tation serves several **BTS, B**ase **T**ransceiver **S**tations, A key function is allocation of time slots required by **TDMA** transmission among the various **MS, M**obile **S**tations, within the **BTS, B**ase **T**ransceiver **S**tations under control.
- The **NSS, N**etwork **S**witching **S**ubsystem is the repository of the database and control functions required to set up calls and establishes connections with the **MS**, Mobile **U**sers. It consists of five elements:
  - The **M**obile service **S**witching **C**enter, **MSC**, often referred simply as **M**obile **S**witching **C**enter. It is the "*brain*" of the network subsystem. It functions much like the land lines **C**entral **O**ffice, **CO** described earlier, and works as an "exchange" that sets up connection-oriented circuits between callers through appropriate routing. It operates using data stored in the following four elements:

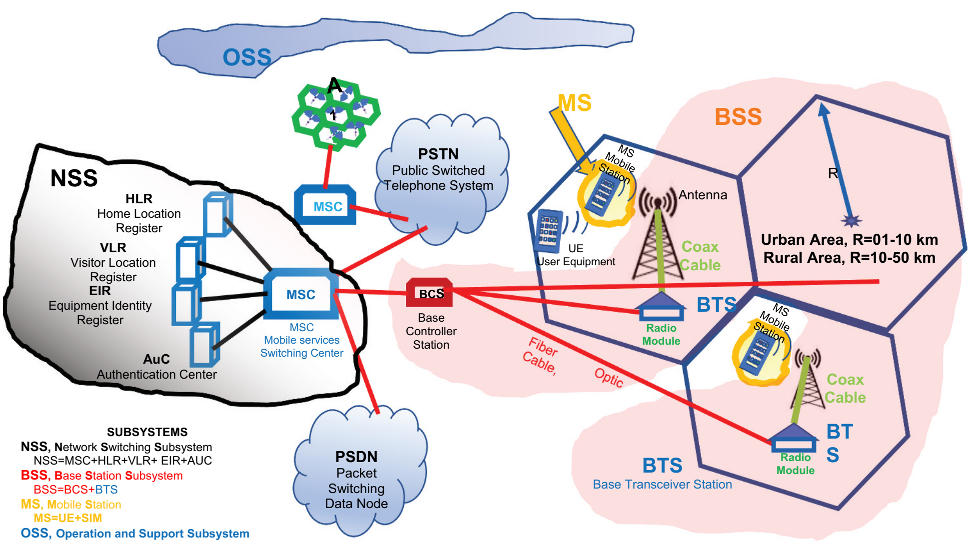


Figure 10. 2 G GSM network architecture.

- ○ **H**ome **L**ocation **R**egister, **HLR** as its name indicates, stores and administers a large database pertinent to subscribers' operation.
- ○ **VLR,** **V**isitor **L**ocation **R**egister, functions like **HLR**, but is dynamic inasmuch that it provides similar information temporarily while the subscriber is roaming across different cells.
- ○ **EIR,** **E**quipment **I**dentity **R**egister, identifies all mobile stations, by storing an International Mobile Equipment Identity (IMEI) in the form of number unique to each device.
- ○ **AuC,** **Au**thentication **C**enter, as the names indicates authenticates a subscriber initiating a call by verifying the validity its mobile device's **S**ubscriber **I**dentification **M**odule, **SIM** card
- **O**peration and **S**upport **S**ubsystems, **OSS** supports, both the **N**etwork **S**witching **S**ubsystem, **NSS** and the **B**ase **S**tation **S**ubsystem, **BSS** by monitoring and controlling their operation, in particular the traffic loading of the latter.

GSM mode of operation

In order to accommodate communication within a cell between many **M**obile **S**tations, **MS** and a single base transceiver station (**BTS**). **GSM** utilizes **TDMA**, Time Division Multiple Access, a form of time division multiplexing; and **FDMA**, frequency division multiple access, much like frequency division multiplexing. Furthermore, since information is digital, it lends itself to **RF** pulses or bursts transmission by each device. The resultant combination of **RF** bursting with TDMA and FDMA contributes to an increase in information capacity. How it comes about is shown by observing Figure 11 graph, which follows.

A common frequency named a "***channel***" is assigned to eight **MS** transmitters in time sequence as shown for instance by the third row of blocks along the time axis. Each block corresponds to a "***time slot***" of 577 μs duration, during which the respective transmitters emit a 156-bit packet, as shown. Following transmission of the 8th slot, the sequence is repeated in time. The eight time slots form a group of 4.615 ms duration, called a **TDMA** "***frame***","-which defines a Logical Channel.

The Physical Channel corresponds to the burst period allocated to each of the eight time slots in the frame. The transmitter data rate or time slot data rate is 156 bits/4.615 ms = 33.8b/s, the theoretical limit. Actual transmitter date rates for early **GSM** did not exceed 9.6Kb/s. The "***channel rate*** "is defined" as the number of packet bits transmitted during a time slot duration. It is 156 bits/577 μs or 270.36Kb/s.

Refer now to the frequency axis of Figure 11, based on GSM specification of allocated frequencies. GSM operates in two frequency bands: one in the

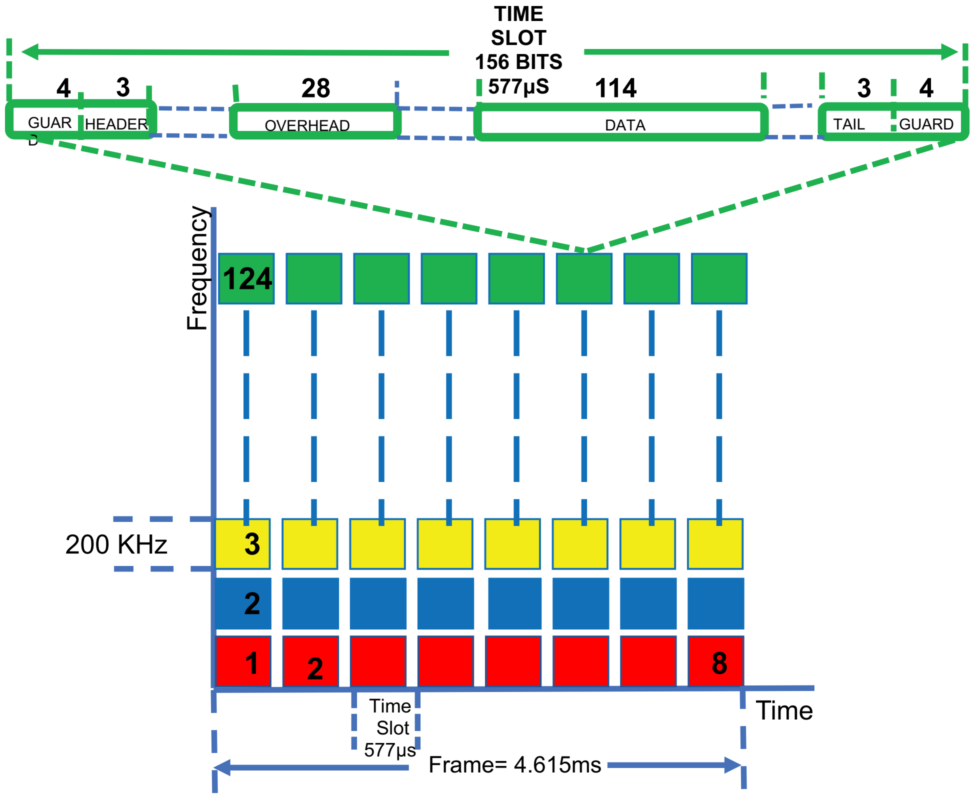


Figure 11. TDMA-FDMA channel allocation in GSM.

900 MHz band and the other in the 1900 MHz Each band is broken up into two sub-bands:

- 890–915 uplink/935-960 MHz downlink.

Uplink or Reverse channel is transmission from **MS** to **BTS**
Downlink or Forward channel is from **BTS** to **MS**
Each band is 25 MHz wide, broken in 200 kHz frequency slots
Thus, 25 MHz/200 kHz provides 125 channels, but only 124 as shown in Figure 11 are available. The 125th is a guard band.

- 1850–1910 uplink/1930-1990 MHz downlink, 60 MHz wide.

The **TDMA-FDMA** system requires accurate synchronization. It is accomplished by having the Base Station sends commands to the mobile phone **MS** instructing timing adjustment of the bursts to account for propagation delay. The adjustment takes place in the guard portions between slots (shown in

Figure 11) where no data is present. When a call is initiated, the transmitter has no knowledge of the time adjustment required. So, there is a whole-frequency channel reserved for such a transaction. It is called **RACH, R**andom **A**ccess **Ch**annel.

Finally, one word about modulation. We would want a scheme that is not amplitude dependent like **A**mplitude **S**hit Keying, **ASK**, because it introduces noise in the system. So, a logical approach is to use a phase shifting scheme as opposed to an amplitude varying scheme. One candidate is **M**inimum phase **S**hift **K**eying, **MSK**. GSM uses a variant of MSK, **G**aussian **M**inimum phase **S**hift **K**eying, **GMSK**. It shapes the incoming bit stream with a Gaussian filter (a filter whose impulse response is a Gaussian function). Its major advantage compared to **MSK** is a reduction in sideband power, thus lower interference with adjacent frequency channels.

In this **2 G** section, we do not discuss cdmaOne (IS-95), a **CDMA** access protocol introduced by Qualcom in the US in the 90s, because GSM was much more prevalent than **CDMA**. In 1996, there were 50 million GSMA subscribers, globally compared with 1 million CDMA subscribers. **CDMA** will be treated in the succeeding generations when it became for a while the protocol of choice. However, we added Appendix A, "Explaining **CDMA**","to help the reader understand **C**ode **D**ivision **M**ultiple **A**ccess, **CDMA** in the generations that follow, **3 G, 4 G**, and **5 G.**

As **GSM** upgraded, first to **G2.25 GPR, G**eneral **P**acket **S**ervice, that made it possible to deliver packets, and up to **G2.5 EDGE, E**nhanced **D**ata Rates for **G**SM **E**volution, **CDMA** access protocol developed into **CDMA-2000** that became one of the standard access protocols in **3 G**, as we shall see in the next section.

We conclude this section on **2 G GSM** with Table 3 summarizing the key performance parameters of that generation.

3 G, the third generation of mobile communications heralded the dawn of the 21^{st} century. It raised the frequency spectrum to two bands: 1885–2025 MHz and 2110–2200 MHz, and data rates to 2 Mb/s and beyond. Its development period occurred during the first decade of this century, from the year 2000 to 2010. Upgrades continued beyond 2010 with demonstrated data rates in the 20 Mb/s, following the introduction of **4 G**, the fourth generation of wireless communications technology.

3 G technology was first introduced by the Japanese company **N**ippon **T**elegraph and **T**elephone (**NTT**) in 2001. It made international communications possible with additional services such as video streaming and conferencing, and **V**oice over **IP** (**I**nternet **p**rotocol) **VoIP.**

Main features of **3 G** technology are:

- It supports greater voice & data capacity and lower cost data transmission at rates of 2 Mb/s and higher.

- **It provides transmission security through encryption**.
- Unlike **2 G** where voice is analog while data transmission is rudimentary, **3 G** provides **digitized voice and data**.

The **3 G** specifications were developed by **ITU**, the **I**nternational **T**elecommunication **U**nion and made public as a standard under the label **IMT, I**nternational **M**obile **C**ommunications. Two competing systems were developed:

- **UMTS, U**niversal **M**obile **T**elecommunicatio**ns S**ervice developed in Europe and prevalent in most of the world, and
- **CDMA**2000 developed in the US and originally deployed in South Korea.

Yet, the two competing systems **UMTS** and **CDMA**2000 use, the same access multiplexing technology: **CDMA, C**ode **Di**vision **M**ultiple **A**ccess, a spread spectrum technique developed by Qualcom in the US. **CDMA**2000 uses a narrow band **CDMA** with 1.25 MHz channels while **UMTS** uses a **W**ideband **CDMA**, referred to as **WCDMA**, with 5 MHz channel bandwidth.

What is CDMA, Code division multiple access?

We refer the reader to Appendix A, **CDMA** devoted to a simplified explanation of its concept and operation. However, it behooves us at this juncture to explain briefly how **CDMA** works.

CDMA is a Spread Spectrum technique. It takes the signal and codes every data bit with specified number of "code bits" named "chips" that may vary from as little as 4 to as high as 64 and 128. That number of chips per data bit is called the **S**preading **F**actor, **SF;** or better stated,

Table 3. Summary 2 G performance features.

KEY FEATURES	2G
Lifetime	1991-2000
First Location	FINLAND
MOBILE SYSTEM	**GSM, G**lobal **S**ystem for **M**obile communications **cdma**One (IS-95) **C**ode **D**ivision **M**ultiple **A**ccess
Frequency Spectrum (MHz)	Uplink:890-915 Downlink: 935-960 Uplink:1850-1910 Downlink: 1930-1990
Channel Spacing No. Channels/Band	200 kHz 124
Access Protocol	**TDMA/FDMA** Time Division Multiple Access/Frequency Division Multiple Access (IS-95) **cdma**One, **C**ode **D**ivision **M**ultiple **A**ccess

SF = Code Chip Rate/Baseband Information Rate = Mc(hip)/s/Mb/s

To transmit the much longer temporal signal in a reasonable time, requires much greater bandwidth, and the **S**preading **F**actor, **SF** is a measure of the extent of the signal "spread" over the entire spectrum.

The same treatment applies to all signals, which now share the same frequency channel. How does the receiver distinguish and recover a specific signal? That's where the coding comes in. Every sender, that is, user has an assigned code, different from all the other sender codes that is used to "***spread***" the signal according to that code over the entire frequency spectrum assigned to the users in a cell (see Appendix A: How does CDMA work?). The codes form a so-called "orthogonal" set with the property that if the received signal, which is a linear superposition of all the different users signals in one cell is multiplied by the user's assigned code, the transmitted signal from that specific **user-sender** is recovered. Multiplication of the received composite signal by any other orthogonal code will recover the signal sent by the user with this assigned code. Furthermore, code orthogonality ensures that the cross-correlation between any two codes is near zero,

This technique spreads the spectral channel bandwidth and allows every user to share that same "spread" channel. The advantages that ensue are a significant simplification of the multiple access protocol, plus interference immunity (Appendix A: How does CDMA work?).

3 G brought to wireless communications a standardization of network protocols. Furthermore, in addition to providing still **2 G** telephone circuit switching which ties a channel between two users for the duration of the call, both **3 G UMTS** and **3 G CDMA2000** are able to share the same channel between many users by breaking the respective messages into packets of an agreed size, that once received by the specific user could be reconstructed into the original message. This method of transmission, **packet switching** became the norm in **3 G** wireless communications and succeeding generations.

There are two types of packet switching illustrated in Figure 12(a) **connection-oriented** also known as virtual **circuit switching**, and **connectionless or datagram switching**. The difference is how address information is conveyed.

Connection-oriented packet switching used in the 3 G mobile networks requires a virtual circuit, much like the physical analog of circuit-switching telephony. This method lightens the packet header load of addresses but demands a hand-shaking protocol at the initiation of transmission, making it equivalent to a physical circuit-switching. Connection-oriented packet switching operates under the Transmission Control Protocol (TCP) at the Transport Layer level shown in Figure 3, Internet Protocol Layers, and is reproduced below for ease of reading as Figure 12(a) Right.

On the other hand, connectionless packet switching, the method of choice in 3 G, uses the Internet Protocol (IP) to assign sending and destination addresses, at the Network Layer level shown in Figure 12(b). Connectionless packet switching, as we shall see further on, is the method of choice in 3 G. It uses the Internet Protocol (IP), both on the Ethernet and the Internet.

The difference between these two types of switching is best understood by looking at Figure 12(a), where packets travel from Sending node A to Destination node B.

The figure on the left illustrates Connection-oriented switching of three packets (1, 2 and 3), which travel along the established routes A, D, C, and B. During transmission, the order of packets must be kept. As a result, when communication is initiated, a point-to-point connection is established by exchanging packets between nodes A and B, so that as each sent packet reaches the destination node B, an ACKnowledgment (ACK) packet is sent back from node B to node A. Once communication is completed, a connection release packet is sent by Sender to free the network for another connection.

In connectionless transmission (right side of Figure 12(a)), packets are allowed to travel on different routes, as one can see: Packets 1 and 2 are

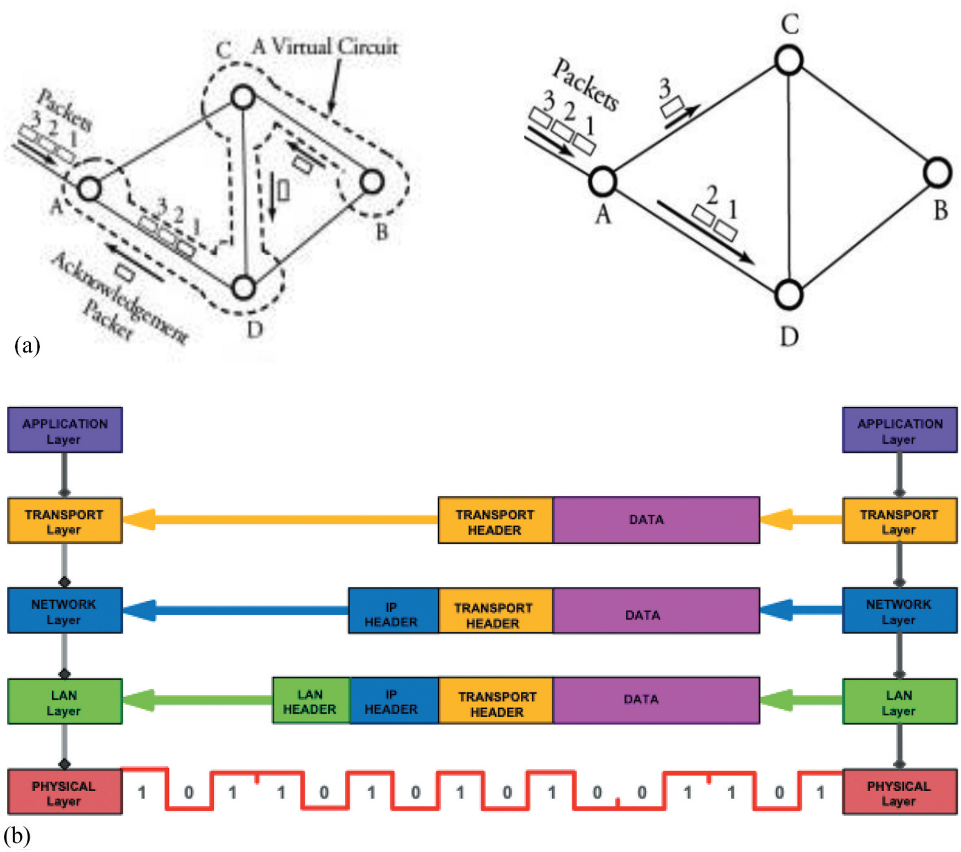


Figure 12. (a) Left: Connection Oriented or Virtual Switching. (a) Right: Connectionless or Datagram Switching. (b) Internet Protocol Layer.

using route A, D, C and B, while packet 3 is using route A, C and B. Their arrival time does not need to be in the order they were sent, because each packet is assigned an order number (as part of its sending address) that allows the receiver to put them in the correct order. Since there is no acknowledgment packet from node B, it yields a simpler and faster connection. Although compared to connection-oriented transmission, connectionless transmission has lower reliability, it is much less costly.

Ref: Nadir F. Mir, *Computer and Communication Networks, 2nd Edition*, Pearson 2014. (out of print)

We shall now look at the network architectures used in **3 G**, beginning with the **UMTS** architecture, which is completely different from **2 G GSM**. Then we shall briefly review **CDMA2000** architecture which has kept most of the **2 G GSM** features.

3 G UMTS network architecture

As shown from Figure 13, **UMTS** architecture comprises three elements or subsystems:

- **UE, U**ser **E**quipment, the counterpart of **GSM MS, M**obile **S**tation. It consists of: It consists of:
- **ME, M**obile **E**quipment (i.e. smart phone, smart pad) renamed **UE(!)** that contains the **SIM, S**ubscriber **I**dentity **M**odule, which is the second sub-element of the **UE,**
- **USIM**, second sub-element, stands for **U**MTS **SIM** to tax the reader's memory of acronyms!

- **UTRAN, UMTS R**adio **A**ccess **N**etwork with two sub-elements:
  - Node B
  - **RNC, R**adio **N**etwork **C**ontroller

- **CN, C**ore **N**etwork that contains:
  - **CS, C**ircuit **S**witched network'
  - **PS, P**acket **S**witched network

There is no sufficient difference between **3 G UE, U**ser **E**quipment and **2 G ME, M**obile **E**quipment to warrant any further discussion, so we turn our attention to **UTRAN, U**MTS **R**adio **A**ccess **N**etwork and **CN, C**ore **N**etwork elements.

UMTS Radio **A**ccess **N**etwork's **Node B** replaces the **BTS, B**ase **T**ransceiver **S**tation of the previous **2 G** architecture. Both, **BTS** and **UMTS**, are elements that process signals transmitted or received by the cell antenna. **UMTS** includes not only the radio module of **BTS** responsible for modulation,

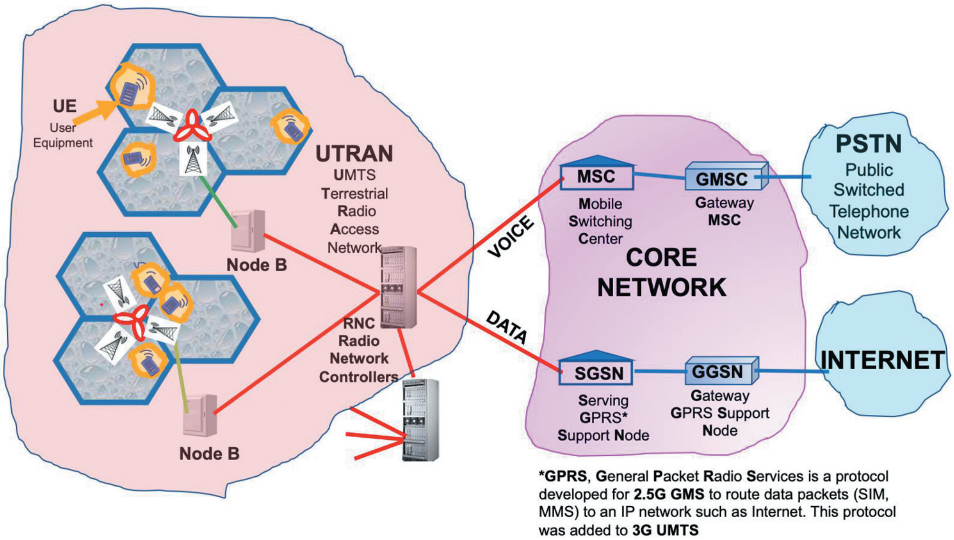


Figure 13. 3 G UMTS network architecture.

multiplexing and all the functions associated with RF transceiving, but also Baseband processing.

However, **Node B** is not located in every cell. As shown in Figure 13, the cells are arranged in groups of three; within each group, the three antennas are situated in a corner common to all three cells, and each antenna beam is "sectored" to create a 120°fan pattern (shown in red). The resultant three beams give a 360°coverage. ***In one of the three cells, the antenna is connected to Node B, and that cell is referred to as the Serving Cell***. Thus, a transmitting **UE,** User **E**quipment such as a moving smart phone that falls within the beam coverage of one of the three antennas communicates directly with it; that information is then transmitted to the Server Cell antenna and relayed by cable to the connected **Node B.**

The other sub-element of **UTRAN, UMTS R**adio **A**ccess **N**etwork is the **RNC, R**adio **N**etwork **C**ontroller. Note whereas one **Node B** only supports a group of three cells, the RNC supports many **Node B**s as well as other **RNC**s. The **RNC** performs two important functions:

- A Control function as its name indicates, under the overall category of Resource Management that includes allocation of channels and scheduling or load control of transmitted packets.
- A Mobile Management function related to the location of the **UE, U**ser **E**quipment (mobile phone such as subscriber) which is moving not only from cell to cell but also outside of its subscriber's local area network to

a faraway geographical location, as in “roaming”. And, in addition, ***encryption of the signal to be transmitted to mobile phone.***

The third major element, the **C**ore **N**etwork (**CN**) is fundamentally different from the **2 G GSM element**. Note that there are two distinct paths coming out of the **RNC, R**adio **N**etwork **C**ontroller, one for voice, and one for data.

Voice

- Voice is circuit switched and sent to an **MSC, M**obile **S**witching **C**enter, much as in **2 G GSM**, which one of its functions is to set up and terminate end-to-end calls between subscribers; another key function of the MSC is to send the signal through the gateway described below.
- **GMSC, G**ateway **MSC** is the inter-network route that enables the subscriber’s signal coming out of the **MSC** to reach the final network, the **PSTN, P**ublic **S**witched **T**elephone **N**etwork, (or vice versa coming out of the **PSTN** and reaching the **MSC**).

In order not to clutter (Figure 13), we omit showing the various Register modules identical to the ones used in **2 G GMS** (**H**ome **L**ocation Register, **HR; V**isitor **L**ocation **R**egister, **VLR; E**quipment **I**dentity **R**egister, **EIR** and **Au**thorization **C**enter, **AuC**) in the Core Network voice path.

Data

The data switching path through the Core Network of Figure 13 goes through two elements:

- **SGSN, S**erving **GPRS S**upport **N**ode. **GPRS, xc** is a protocol developed specifically under a **GSM** upgrade labeled **2.5 G**, which allows text messages transmission limited to 160 characters via **SMS, S**hort **M**essage **S**ervice, packet switching.
- **GGSN, G**ateway **GPRS S**upport **N**ode which is followed by an internetwork router, much like the one in the voice path. It receives the packet from the **SGSN, S**erving **GPRS S**upport **N**ode, which carries the **GPRS** protocol and converts the data into a protocol required by the IP network, Internet in Figure 13

We conclude our review of 3 G networks by describing an architecture that applies to CDMA2000, shown in Figure 14.

During its lifetime **3 G CDMA** 2000 went through many updates, and we have selected in Figure 14, the architecture **1xEV-DV** introduced in 2004. **EV** stands for **EV**olution and **DV** for **D**ata and **V**oice. Indeed, a major feature of this update is Integrated Voice and simultaneous high-speed Data on a single carrier. The network is capable of delivering in the Forward Link (also called Down Link from Tower to subscriber phone) data rates up to 3.09 Mb/s.

Referring to the network architecture of Figure 14, CDMA 2000 1xEV-DV consists of three major elements: A **M**obile **S**tation, **MS, a R**adio **A**ccess **N**etwork, **RAN** and two **C**ore **N**etworks (**CN**), one for voice and one for data.

- **M**obile **S**tation, **MS.**

It is the counterpart of the **UE, U**ser **E**quipment of **UMTS, U**niversal **M**obile **T**elecommunicatio**ns S**ervice described. It performs similar functions, which will not be repeated here.

- **R**adio **A**ccess **N**etwork, **RAN**

The **RAN** is the "heart" of the network that links communications to and from the **MS, M**obile **S**tation, to the Voice and Data core networks. It includes two sub-elements, the **B**ase **T**ransceiver **S**tation, **BTS**: and the combined **B**ase **S**tation **C**ontroller, **BST**/**S**elector and **D**istribution **U**nit, **SDU**.

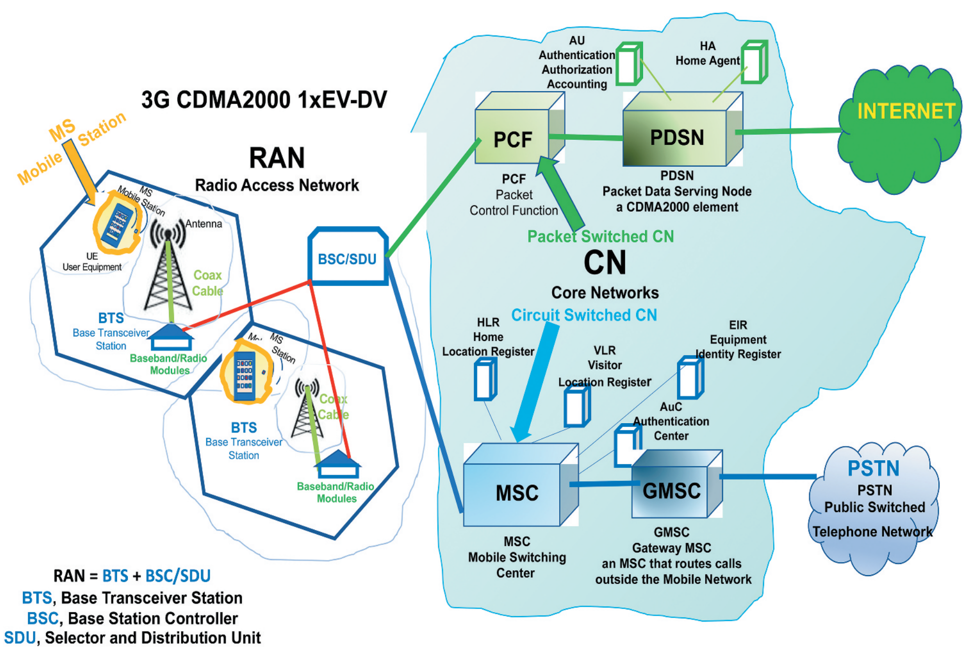


Figure 14. 3 G CDMA2000 1XEV-DV Network Architecture.

- The **B**ase **T**ransceiver **S**tation, **BTS** includes the antenna mounted on top of the tower, and the housed equipment on the ground. The latter consists of two sections that handle both, the forward and reverse links:
  - A Baseband processor that comprises of a baseband adaptive modulation/demodulation, such as 16-QAM, encryption, decryption, and Multiplexing/Demultiplexing Access using CDMA, Code Division Multiple Access. A Block Diagram of this function is shown in Figure 15 taken from the excellent review by Agrawal et al.[3]
  - An **RF, R**adio **F**requency section that consists of RF modulation/demodulation, up and down frequency conversion, and amplifiers/filters. Note, that all the RF equipment will be placed on top of the antenna tower in future generations, **4G** and **5G**.
- The **B**ase **S**tation **C**ontroller, **BSC/SDU**, **S**elector and **D**istribution Unit.
  - The **BSC, B**ase **S**tation **C**ontroller performs the functions of routing voice and data to the core networks. A **BSC** serves several cells, each with their respective **BTS**. In addition, the **BSC** carries out the Mobility Management tasks which include keeping track of the location of the **MS, M**obile **S**tation and carrying out "handoffs" as it goes from one cell into another.
  - The **SDU, S**elector and **D**istribution **U**nit is a special switch capable of identifying the destination of packets and frames by their addresses: A packet with a TCP address header is sent to the voice core network, while one with a **UDP** address is sent to the data network.

Core network CN

The Core network consists of two sub-networks, the Circuit Switched Voice and the Packet Switched Data.

- The Circuit Switched Voice is identical to the **UMTS, U**niversal **M**obile **T**elecommunicatio**ns S**ervice described above. It includes the same two elements, the **M**obile **S**witching **C**enter, **MSC** and the **GMSC**, the **G**ateway from the **MSC** to the **PSTN, P**ublic **S**witched **T**elephone **N**etwork used in the **UMTS, U**niversal **M**obile **T**elecommunicatio**ns S**ervice. The reader is referred to the previous section on **UMTS** for explanations, which includes the four "registry" elements shown in Figure 14
- The Packet Switched Data sub-network consists of two major sub-elements: **PCF, P**acket **C**ontrol **F**unction and **PDSN, P**acket **D**ata **S**erving **N**ode discussed below:
  - The **PCF, P**acket **C**ontrol **F**unction is a router and switch that performs a function equivalent to the **MSC, M**obile **S**witching **C**enter in the Voice sub-network. The **PCF** takes the packets received from the **RAN, R**adio **A**ccess **N**etwork and sends them on to a Gateway, the

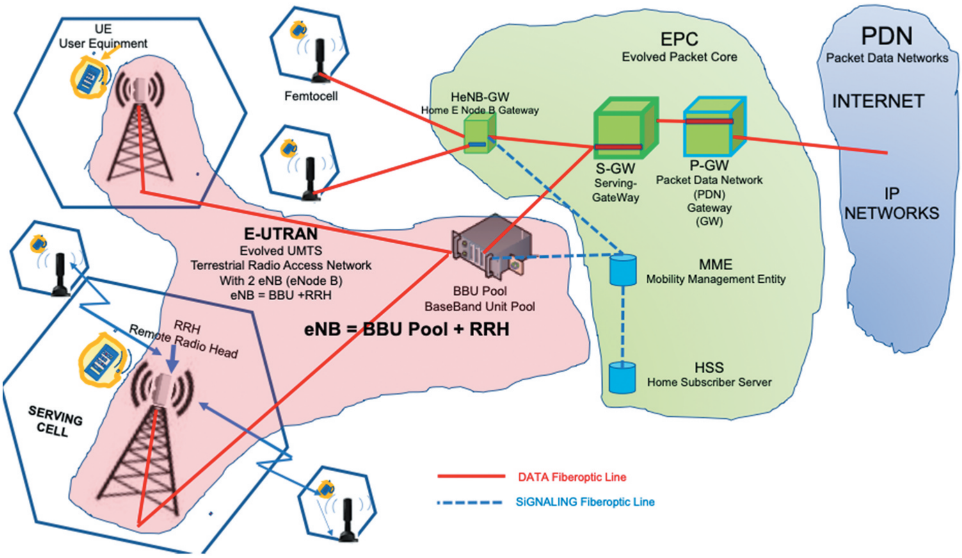


Figure 15. BTS, Base transceiver station: Baseband/RF.

PDSN or **P**acket **D**ata **S**erving **N**ode that forward them to the **IP** (**I**nternet **P**rotocol) or private data networks, such as the Internet shown in Figure 14.

- The **PDSN** or **P**acket **D**ata **S**erving **N**ode as stated just above is the gateway to the **IP** networks. It includes "registry" elements similar to the ones used in **UMTS, U**niversal **M**obile **T**elecommunicatio**ns** **S**ervice (HLR, VLR, EIR and AuC) but bear different Names: **AAA, A**uthentication, **A**uthorization and **A**ccounting and **HA, H**ome **A**gent.

As its name indicates, the **AAA, A**uthentication, **A**uthorization and **A**ccounting elements authenticate the subscribers for access to the Core Network as well as storing statistics for billing and invoicing. The **HA, H**ome **A**gent, among its many functions keeps track of the subscriber's registration and enables the packets access to the **PDSN, P**acket **D**ata **S**erving **N**ode.

We conclude this section with Table 4 that summarizes the key performance parameters of **3 G**, in particular **UMTS** WCDMA and **CDMA** 2000 1xEV-DV. Note from the Table, third column, rows 9 and 10 that the Downlink Speed is 3.072 Mb/s while the Chip Rate is only 1.2288 Mc/s. How could that be? This is explained as follows: First, 16-QAM adaptive modulation is used, thus reducing transmission rate by a factor of 4 (Log_2M, M = 16) to a symbol rate 3072 Kb/s/4 = 768 Kb/s. Assuming 20% overhead in each bit [4], the net symbol rate R_s = 614 Kb/s. If the signal is Time Division Multiplexed (TDM) between two subchannels* with their respective orthogonal codes, the symbol rate per each subchannel is 307 Kb/s, yielding a code with **S**preading **F**actor,

SF = Chip Rate/Symbol Rate = (1,228.8 Kb/s)/(307.2 Kb/s) = 4 (4 chips/symbol)

The reader is referred to [5] for a thorough discussion of this topic.

4 G long-term evolution architecture

4 G, the fourth generation in the development of wireless communications was a major step forward that required new devices. One of such devices, Apple's iPhone has been a major contributor to the growth of **4 G** technology.

This new technology was introduced in 2009 in Sweden and Norway, as the **L**ong-Term **E**volution, **LTE** based on standards specified by the **I**nternational **T**elecommunications **U**nion-**R**adio, **ITU-R** organization under the **3**rd **G**eneration **P**artnership **P**roject, **3 GPP**.

Table 4. Summary 3 G performance features.

KEY FEATURES	3G	
Lifetime	2000 - 2010	
First Location	JAPAN	
MOBILE SYSTEM	UMTS WCDMA Universal Mobile Telecommunications Service	CDMA 2000 1xEV-DV Code Division Multiple Access
Frequency Spectrum (MHz)	1885-2025 /2110-2200	
Channel Bandwidth	5 MHz	1.25 MHz; 3x1.25 MHz
Access Protocol	WCDMA/TDMA	CDMA
Switching Technology	Digitized Voice and Data Circuit and Packet Switching	
Peak Data Rates	Forward (Downlink) 2.072 Mb/s Reverse (Uplink) Variable	Forward (Downlink) 3.072 Mb/s Reverse (Uplink) 451.2 Kb/s
Chip Rate	3.84 Mc/s	1.2288 Mc/s
Core Network	Packet Switched SGSN Serving GPRS Support Node Circuit Switched MSC Mobile Switching Center	Packet Switched PDSN Packet Data Serving Node Circuit Switched MSC Mobile Switching Center

Concurrently, a competitive standard, 802.16, mobile **WIMAX, W**orldwide **I**nteroperability for **M**icrowave **Access** was introduced in the US by the **IEEE, I**nstitute of **E**lectrical and **E**lectronics **E**ngineers. Neither **LTE** nor **WIMAX** was able to meet the **4 G** standards, and they were considered stopgaps, or updated versions of the **3 G** standards.

Finally, in March 2011 an updated version of **LTE**, named **LTE**-**A**dvanced, for short **LTE-A** was approved by the **3**rd **G**eneration **P**artnership **P**roject, **3 GPP**, as it met the 4 G standards established by the **I**nternational **T**elecommunications **U**nion-**R**adio (**ITU-R**) organization.

At about the same time frame, an updated version **WIMAX2** was released by the **IEEE** that was competitive with the **LTE-A** version. Unfortunately, **LTE-A** was already firmly established throughout the world, and **WIMAX2** was unable to displace it.

Compared to the previous generations of mobile wireless communications, the new standards of **4 G** were to some extent more "revolutionary" than the long-term evolution implied in its name, **LTE.**

Key features of **LTE-A** network are summarized below and will be discussed in greater details further on:

- *Network Cells*: Mix of Macrocells, Microcells, Picocells and Femtocells with the following Emitted Power (in Watts)/Range (in meters).
  - Macrocells: 20 W/35 Km
  - Microcells: 5 W/2 Km
  - Picocells: 1 W/200 m
  - Femtocells: 0.1 W/10 m
- Frequency Spectrum:
  - Over 40 bands operating globally between low frequency, 700 MHz and high frequency 2,800 MHz
  - Each band has both Uplink (Reverse, from Mobile Station to Base) and Downlink (Forward, from Base to Mobile Station)
- The four mostly Deployed Networks in their respective Bands are:
  - 327 Networks/1800 MHz
  - 183 Networks/800 MHz
  - 164 Networks/2800 MHz
  - 128 Networks/700 MHz
- Channel Bandwidths:
  - 1.4 MHz, 3 MHz, 5 MHz, 10 MHz, 15 MHz, and 20 MHz
- Carrier Aggregation:
  - Effective bandwidth, hence, data rate can be increased **fivefold** by spreading the data over a maximum of 5 different frequency carriers.
  - Thus, an increase from 20 MHz to 5 × 20 MHz = 100 MHz
- Subcarrier Spacing: 15 kHz
- Access Protocol:

  - Downlink: **OFDMA, O**rthogonal **F**requency **D**ivision Multiple

- Access
  - Uplink: SC-FDMA, Single Carrier – Frequency Division Multiple Access (SC-FDMA has lower Peak-to-Average Power than OFDMA)
- *Switching Technology*: Integrated Voice and Data, **Packet Switching**
- Low Mobility Download/Upload Peak Data Rates: 1 Gb/s/500 Mb/s
- High Mobility (500 Km/hr) Download Peak Data Rates: 100 Mb/s
- Latency < 50 ms
- Downlink Baseband Coding Modulation, up to 256-QAM
- *Downlink Directional Antennas*, up to **8 × 8 MIMO** (**M**ultiple **I**nput **M**ultiple **O**utput)

At this junction, the best way to discuss the above features is in the context of the **4 G LTE-A** network architecture. A representation of the network is shown below, Figure 16: **4 G LTE-A** Network Architecture.

The network is composed of three basic elements:

- The **U**ser **E**quipment, **UE**. It is the same element described in the 3 G **UMTS, U**niversal, **M**obile **T**elecommunications **S**ystem, architecture (Figure 13) and serves identical functions. In short, the **UE** is nothing more than the smart phone or smart pad that one uses for cellular communications. It is pictured in Figure 16 as a little blue phone, surrounded by an orange "halo".
- The second element, the heart of the system, **E-UTRAN**, an abbreviation for **E**volved- UMTS **T**errestrial **R**adio **A**ccess **N**etwork, handles all communications between the **UE** and the core network (the third element of this architecture), re-labeled **E**volved **P**acket **C**ore, **EPC**. It is a simplified version the 3 G **UTRAN** architecture of Figure 13, whereby the two original elements, Node **B** and **R**adio **N**etwork **C**ontroller, **RNC** are now combined into a single entity, **e**volved **N**ode **B** or **eNB.**

Two major features of **eNB** warrant mentioning:

- The old **3 G Node B** of Figure 13 that combined the radio equipment and the baseband band processing and was located on the ground is now split in the **eNB**: The radio part is mounted on top of the antenna tower and re-labeled **RRH, R**emote **R**adio **H**ead; and the **B**ase**B**and processing **U**nit, **BBU** is now distributed, so as to serve many cells and shown as the **BBU** Pool.
- The combined **BBU** Pool and **RRH** are what constitute the **eNB**.
  - Distributed nature of **eNB**. It is no longer associated with one cell and serves many cells. For instance, a caller in one of the small femtocell shown

in Figure 16 has two ways of communicating through a **PDN, P**acket **Data N**etwork such as the internet: 1) Directly through the air to the small femtocell antenna, which relays if in turn to a major cell called the Server Cell. The Server Cell then relays the data through a fiberoptic line to the **B**ase**B**and Unit Pool, **BBU** Pool, through the core network, **E**volved **P**acket **C**ore, **EPC** and onto the **PDN, P**acket **Data N**etwork; 2) or through the air to the small femtocell antenna, which bypasses the server cell, but instead sends the data via a fiberoptic line directly to a special relay node known as **HeNB, H**ome **e**volved **N**ode **B** which in turn sends it through **E**volved **P**acket **C**ore, **EPC** and onto the **PDN, P**acket **Data N**etwork.

One final but important remark regarding the splitting of functions between the two constituents of **eNB**, the **BBU, B**ase**B**and **U**nit pool and the **RRH, R**emote **R**adio **H**ead:

- The **BBU** comprises a **digital signal processor** to process forward voice signals for transmission to a mobile unit and to process reverse voice signals received from the mobile unit. Major functions for sending and receiving:
  - Adaptive base band modulation/demodulation up to 258-QAM

- Encyphering/Decyphering
- Multiplexing/Demultiplexing Access (which are discussed further on)

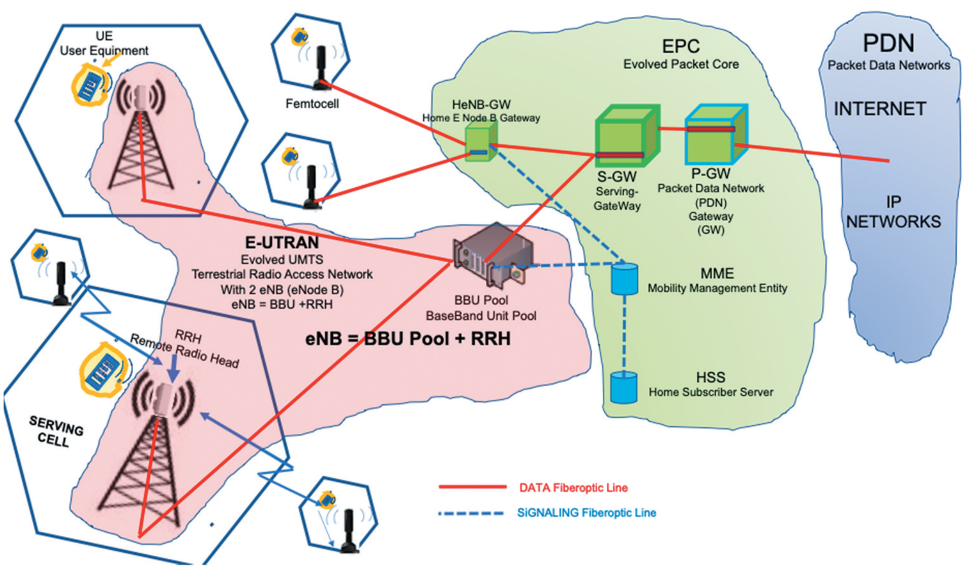


Figure 16. LTE-Advanced network architecture.

- The third element of the **4 G LTE-A, L**ong **T**erm **E**volution-**A**dvanced network is the core network labeled **EPC, E**volved **P**acket **C**ore network. There are two major sub-elements:
  - A router, labeled **S-GW, S**erving **G**ate**W**ay. It takes the fiberoptic line Data Packets output coming out of the big cells through the **BBU** pool as well as the output from the femtocells through the **HeNB** and route them through the **S-GW**. The router ensures correct convergence and timely delivery of the voice and data packets.
  - The final sub-element, ***P*-GW** or **PDN-GW** is the **P**acket **D**ata **N**etwork **G**ate**W**ay. Its major function, that of an **IP, I**nternet **P**rotocol router that connects the voice and data packets to the intended **P**acket **D**ata **N**etwork, **PDN**, whether Internet of private (VPN) IP network.

In addition to the aforementioned two sub-elements, there are two other sub-elements that perform "housekeeping" functions. They are blue-colored in Figure 16 and carry the signaling functions (shown by dotted fiberoptic lines), as opposed to the data functions shown with the red fiberoptic lines. The signaling function is often referred to as the "control plane" function while the data function is called the "user plane" function. These two sub-elements are the **MME, M**obility **M**anagement **E**ntity and **HSS, H**ome **S**ubscriber **S**erver.

- The **MME, M**obility **M**anagement **E**ntity, controls all the signaling functions required by the mobile user. It is a critical control sub-element dealing with the **EPC, E**volved **P**acket **C**ore network control plane. It manages sessions, authentication, paging, mobility, bearers and roaming.
- The **HSS, H**ome **S**ubscriber **S**erver, is a database server that contains all the pertinent information about the user's mobile phone, in particular all its identifying information as well as the pertinent subscriber information.

The quest for higher data rates:

- MULTIPLEXED MODULATION
- MULTIPLE ANTENNAS

Multiplexed modulation

We begin with **OFDM, O**rthogonal **F**requency-**D**ivision **M**ultiplexing, as it is the foundation of the two multiple-access techniques used in **4 G LTE-A** (4th **G**eneration, **L**ong **T**erm **E**volution-**A**dvance):

OFDMA, Orthogonal **F**requency-**D**ivision **M**ultiple **A**ccess in the Downlink

SC-FDMA Single **C**arrier (Orthogonal) Frequency-Division **M**ultiple **A**ccess in the uplink

OFDM

The worldly proliferation of digital wireless communications demands insatiably higher bit rates and consequently greater bandwidths. Unfortunately, multiple-path delays are responsible for errors, as transmitted bits run into each other. The result is the so-called Intersymbol Interference.

The solution is to break up the frequency spectrum of the transmitted signal into many spectral bands, narrow enough such that the channel characteristics in each band are essentially constant. This is called flat fading. Flat fading, when considered in the time domain, is equivalent to having an impulse response that is much shorter than the bit length.

The resultant narrow spectral band signals are implemented with a technique known as **O**rthogonal **F**requency **D**ivision **M**ultiplexing, **OFDM**. The meaning of orthogonal will be clear after we describe the key advantages of **OFDM**.

Refer to Figure 17: **OFDM** FEATURES. Without **OFDM**, Figure 17(a) compares the difference: in the presence of channel interference between a broad band signal of bandwidth B (Figure 17(a) left) with the data spread into MULTI-CARRIER frequency bins (Figure 17(b) right). With large bandwidth, the signal data bits are of short duration, so that frequency selective transmission channel fluctuations create large amplitude changes in those data bits, with consequent high **B**it **E**rror **R**ate, **BER**. This is called **Selective fading.**

The situation shown in Figure 17(a) is actually oversimplified. The channel frequency response is characterized not only by the amplitude spectrum shown in Figure 17(a) but also by a phase spectrum that reflects the time

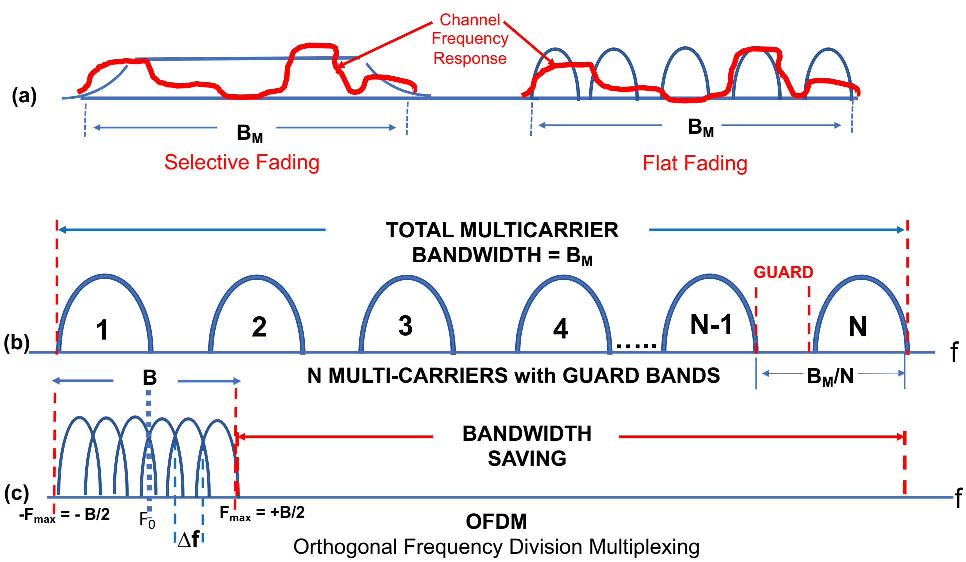


Figure 17. OFDM features.

delay incurred by different transmission paths. So, even if the channel amplitude spectrum is constant, the variation in phase caused by the time delay spread could result as well in varying amplitudes of the data pulses (i. e. in-phase and out-of-phase interference) but also in symbol interference. As a rule, if the average channel transmission path delay is greater than 20% of the pulse duration, one data bit will interfere with the succeeding one, bringing on symbol interference. Typical delays range from a fraction of one microsecond on the order of a few microseconds, sufficient to bring about symbol interference in signals having data rates 10 Mb/s or higher.

On the other hand, when MultiCarriers (not necessarily orthogonal) are used, the data are spread in N narrow frequency bins as shown in Figure 17(a) right. The channel amplitude frequency fluctuations in each bin are nearly constant; if the channel-phase spectrum is also flat within those bands, then bit delays as well amplitude fluctuations will be significantly lower yielding lower symbol interference and consequently lower **BER, B**it **Error rate.**

This operating regime is referred to as ***Flat Fading***.

Bandwidth inefficiency occurs with non-orthogonal multicarriers as shown in Figure 17(b). This is due to the use of Guard bands needed to avoid interference between the various frequency carriers. The situation is remedied in two ways with **OFDM**:

- The frequency bins, shown in Figure 17(c) are all squeezed together in an overlapping pattern that provides a significant bandwidth saving and higher spectral efficiency.

How does one deal with the carrier interference created by the overlap? The answer follows:

- The frequency carriers are orthogonal, which amounts to say they do not interfere with each other. How is that accomplished is explained Appendix B: How to make carriers orthogonal in OFDM? In a few words, orthogonality holds when the modulated subcarriers interval is equal to the Nyquist rate, $2\ F_{max} = B$ (shown in Figure 17(c))

Two such subcarriers whose spectra are modulated by a square pulse representing a transmitted binary symbol are illustrated in Figure 18 (Refer to Appendix B). Note that the spectrum major lobe of peak frequency F_0 has its nulls occurring at $\pm 1/T$, $\pm 2/T$, $\pm 3/T$ and so on, where $T = 1/\Delta f$, and $\Delta f = B/N$, i. e. the total bandwidth divided by the number of subcarriers, N. On the other hand, the major lobe of the modulated frequency subcarrier with peak frequency $F_0 + 1/T$ has a null at the peak of the spectrum of center frequency F_0, thus eliminating the bulk of the interference where the energy is concentrated, that is, in the major lobes.

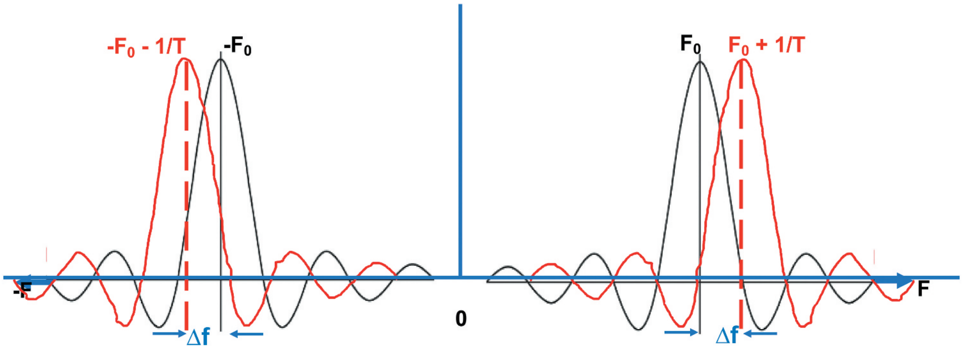


Figure 18. Two OFDM orthogonal carriers (two-sided spectrum).

The lack of interference between the two modulated carriers whose peak frequencies occur at $F_{0,,}$ F_0 ±1/T, F_0 ±2/T, F_0 ±3/T is the result of the orthogonality explained in Appendix B.

A major benefit of the bandwidth reduction brought about with **OFDM** is a greater spectral efficiency. Recall that the spectral efficiency is the ratio of the bit rate, R_b divided by the total bandwidth, B. As the latter is made smaller, as shown in Figure 17(c), the spectral efficiency is increased. There has been a significant improvement going from the first generation, **1 G** where the spectral efficiency was less than 1bit/(s.Hz) to the fourth generation of wireless communications, **4 G** that attained spectral efficiencies of 6 bits/(s.Hz) and higher.

Note how the peak of modulated carrier spectrum of center frequency F_0 coincides with the null point of modulated carrier spectrum whose center frequency is F_1 + 1/T, thus eliminating carrier interference.

Let us now see how transmission of **OFDM** is implemented. When the concept was originally proposed by Robert Chang [6] of the Bell Telephone Laboratories in 1966, it involved a lot of hardware, banks of subcarrier oscillators and coherent detectors. It did not take off until the 1971 when Weinstein and Erbert [7,8] substituted Fast Fourier Transforms, **FFT** to accomplish the same results with algorithmic software.

The original **OFDM** transmitter and receiver hardware implementations are shown, respectively, in Figures 19 and 20.

Let us start with the transmitter, Figure 19 where the serial data stream inputs the transmitter on the left at the bit rate R_b, each bit of duration T_b = 1/R_b. We first describe its operation in a simple manner by omitting baseband adaptive modulation.

As the serial stream enters the **S**erial-to-**P**arallel (**S/P**) converter, each incoming bit is assigned to one of the parallel outputs in succession, whereby it modulates a subcarrier frequency, from f_0 to f_{N-1}. Thus, the N bits undergo

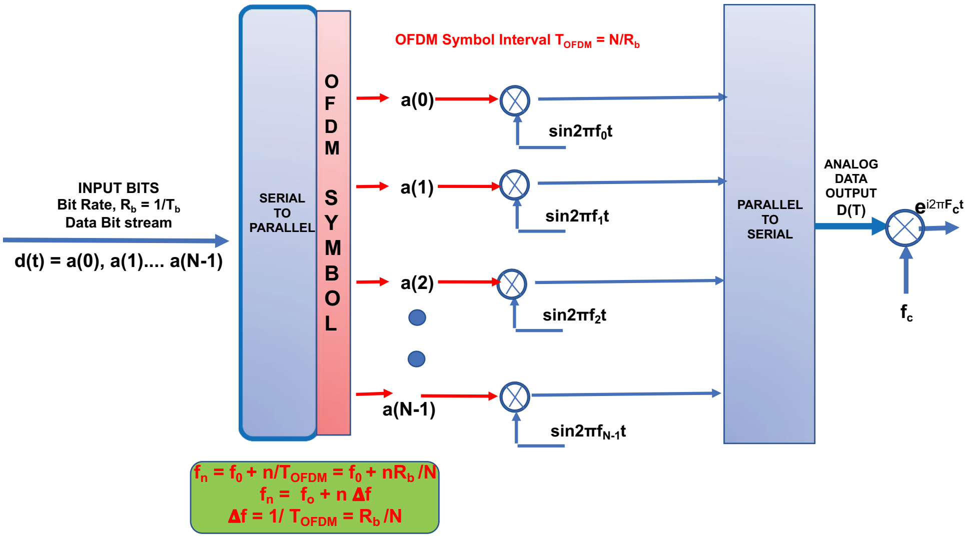


Figure 19. OFDM transmitter, hardware implementation.

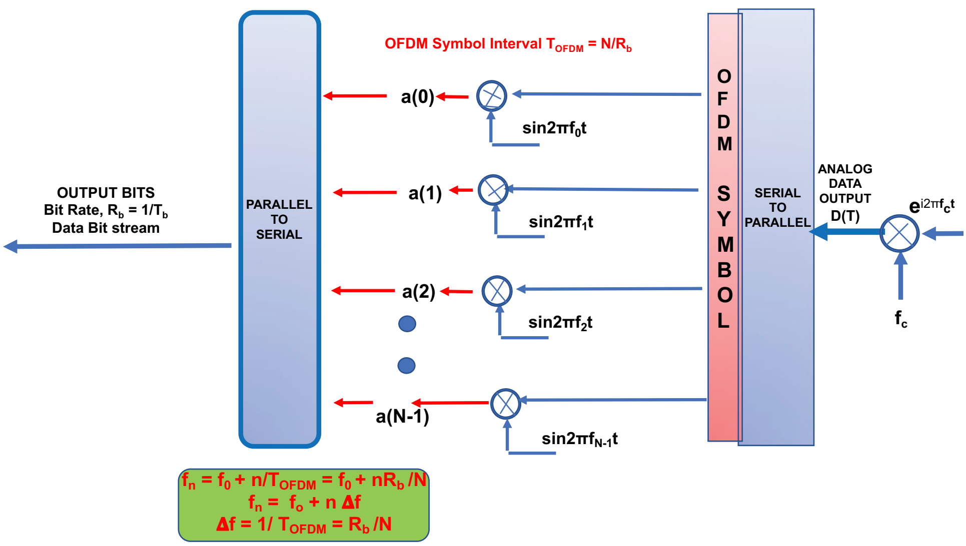


Figure 20. OFDM receiver, hardware implementation.

a delay of $NT_b = N/R_b$ going through the N shift-registers. As a result, the time interval between succeeding groups of N bits is NT_b; we call such a group of N parallel bits at an instant of time an **OFDM** Symbol. The N bits that have now been converted to analog signals are then demultiplexed through a **Parallel-to-Serial**, **P/S** converter, whereby they modulate an **RF** (**R**adio Frequency) $\mathbf{f_c}$ carrier in one of the GHz bands assigned to **LTE-A**. Note that the $\mathbf{f_c}$ carrier is nothing more than one of the N assigned subcarriers after a change of coordinates shown as F_0 in Figure 17(c)

For the sake of simplicity, we have omitted showing, after the **P/S** converter, **CP, C**yclic **P**refix module designed to create a guard band that eliminates **I**nter**S**ymbol **i**nterference, **ISI** as described further down.

The hardware implementation of the Receiver is shown in Figure 20. It performs the reversed operations of the transmitter and the figure is self-explanatory.

Discrete **F**ourier **T**ransforms (**DTF**) play a very important role in wireless transmission of signals, particularly between base stations (i.e. Cell Towers) and mobile users (i.e. smart phones). The use of **DTF** simplifies greatly transmitter and receiver implementation. With N parallel transmitted signals in parallel, as is the case, each signal modulates an **RF** (**R**adio Frequency around 2 GHz e.g.) subcarrier at one of the N frequencies. At the receiving end, N local oscillators at N corresponding frequencies are required for coherent detection. N can be a large number, anywhere from 48 to 1024 or higher, thus requiring such a large amount of "hardware" that renders the scheme impractical.

DTF, Discrete **F**ourier **T**ransforms, comes to the rescue by replacing this unwieldy mass of hardware with an algorithmic software. The adoption of **OFDMA** (**O**ptical **F**requency-**D**ivision **M**ultiple **A**ccess) in 4 G LTE-Advanced wireless communications (to be discussed further on) owes it primarily to **DTF**, a simple and cost-effective Access Technology.

The resultant algorithm developed to carry out the **DTF, D**iscrete **F**ourier **T**ransform is a very rapid one, computationally; it is called, **F**ast **F**ourier **T**ransform, or **FFT**. From now on, we shall use the terminology, **FFT** instead of **DFT**, and **IFFT** for the **I**nverse **F**ast **F**ourier **T**ransform.

Refer now to Figure 21, the transmitter block diagram where all the hardware has been replaced by an **IFFT** processor. Note the following:

- The QAM symbol carried as I and Q components by each of the N subcarriers, is expressed as a complex number, a + jb in rectangular coordinates. A short tutorial on QAM is added as Appendix C.
- The QAM symbol values are converted to polar form in terms of amplitude $|\mathbf{A_k}|$ and phase, $e^{i\phi}k$. and loaded in the **IFFT** processor bins, with each bin corresponding to a different subcarrier frequency.
- Amplitude and phase modulation of the respective subcarriers, f_k processed through the **IFFT** turns out as shown in Appendix D, nothing more than the discrete **I**nverse **F**ourier **T**ransform, **IFFT** (**I**nverse **F**ast **F**ourier **T**ransform) of the signal sampled at the rate of the full bandwidth, **B**

The set of N numbers or symbol values located in their respective frequency bins are then transmitted by modulating an RF carrier in one of the 4 G bands, usually around a few GHz. Note that every T_{OFDM}, a new constellation symbol is assigned to each channel.

We omit a block diagram of the receiver as its operation is the inverse of the one shown in Figure 21, **IFFT** Transmitter. Note that the Fourier transform performed at the receiver is the inverse of the transform of the one done on the transmitter. Both transforms are identical except for a normalizing factor and sign change in the exponential functions described in Appendix D.

We have also omitted for lack of space the important topic of **C**yclic **P**refix, **CP**, powerful coding technique that reduces **ISI, I**nter**S**ymbol **I**nterference, in dispersive channels whereby the delays caused by different propagation paths are responsible bit errors. The situation is remedied by adding the tailed portion of each OFDM symbol to the header so as to create a guard band of duration equal to the expected delay. The technique has been described extensively in the literature and the "World Wide Web" is full of videos with varying degrees of explanation. In calculations of data rates, as we shall discuss further on, **C**yclic **P**refix, **CP** adds overhead to each symbol.

If channel dispersion time delay is T_d, the overhead is simply T_d/T_{OFDM} .

We now work out an example of an OFDM system, taken from Ref [6].

Consider and OFDM system with the following parameters:

Calculate System Data Rate, R given the three parameters below:

Total Bandwidth B = 1.0 MHz

Delay Spread, $T_d = 25\mu s$

Jitter Roll-off Factor r = 0.25, thus B = R (1 + r) = 1.25 R

Number of Channels, N =128

To avoid **ISI** (**I**nter**S**ymbol **I**nterference) caused by Delay Spread, choose an OFDM symbol interval, $T \geq 5\, T_d = 125\ \mu s$

Thus, Channel Bandwidth $\Delta f \leq 1/T = 8$ kHz.

Using N = 128, $\Delta f = B/N = 7.8$ kHz, which satisfies inequality above.

Channel Rate, R_N with roll-off factor of .25: $R_N = 7.8$ kHz/1.25 = 6.24 Kb/s

System Data rate based on 64-QAM:

$\mathbf{R} = N.R_{N.}Log_2M = 128 \times 6.24$ Kb/sx6 = **4.8 Mb/s**

Spectral Efficiency = R/B = 4.8 b/Hz

We conclude our discussion of **OFDM** with some important definitions, which the reader will encounter in the literature plus a simple example.

*A **Resource Block** (**RB**) is the smallest unit of **resources** that can be allocated to a user*. A typical **Resource Block, RB** is a frequency-time element defined in an **OFDM** two-dimensional block allocation, Figure 22.

$RB = 12\ \Delta f \times 7\ T_{ofdm}$

In **LTE-A**, the subcarriers frequency separation is 15 kHz. The Resource Block, RB 12 × 15 kHz subcarriers times seven OFDM intervals.

An additional metric is the "**Slot Time**" = 7 $\mathbf{T_{ofdm}}$ = 0.5 ms in **LTE-A.**

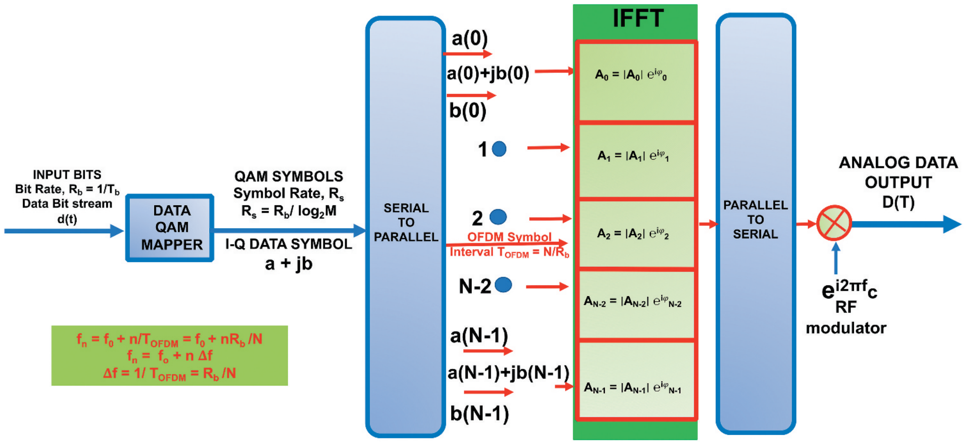


Figure 21. IFFT transmitter.

With B = 5 MHz, after guard band the effective **B** = 4.5 MHz. Without OFDM, the incoming data bit duration is $\Delta t = 1/\mathbf{B} = 0.222$us. With OFDM, the number of subcarriers $N = \mathbf{B}/\Delta f = 4.5$ MHz/15 kHz yields = 300. The resultant pulse duration after Serial to Parallel conversion (Figure 22) is N/B = (300) x(1/B) = 300 × 0.222 = **66.7 µs** $=\mathbf{T_{ofdm}}$. We call $\mathbf{T_{ofdm}}$ the OFDM Symbol Interval

The number of subcarriers per user (without carrier aggregation) is determined by the bandwidth. LTE-A provides the following bandwidths in MHz: 1.4,3, 5, 10, 15 and 20 MHz

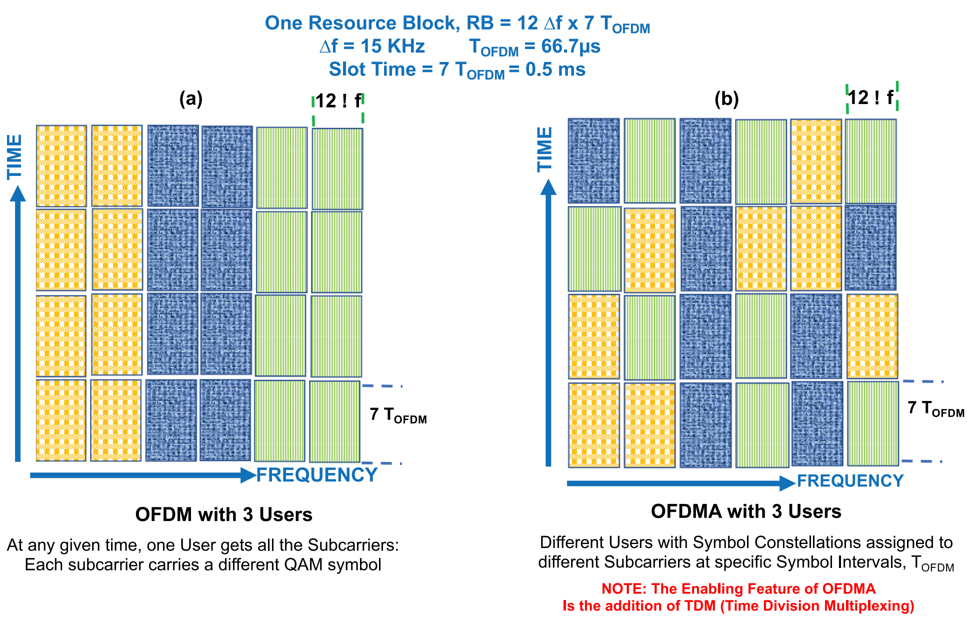


Figure 22. OFDM – OFDMA frequency-time grid.

Example: A 5 MHz bandwidth with 10% guard interval amounts to an effective Bandwidth of **B** = **4**.5 MHz. Each subcarrier is 15 kHz wide, and the total number of subcarriers is N = 300. Because all the modulations are done with an N-**IFFT** (**I**nverse **F**ast **F**ourier **T**ransform) it demands N to be an exponent of 2. The exponent that gives the closest number to 300, is 9, so $N = 2^9 = 512$,

There are two major drawbacks to OFDM.

- In order to mitigate InterSymbol Interferences caused by multipath delays, we discussed previously how the OFDM symbol interval T_{OFDM} should be as long as feasible (much longer than the average channel delay), which means a large number N of subcarriers. Recall that the OFDM symbol interval increases directly with the number N of subcarriers.

The LTE-A standard demands that all N channels be allocated to one user at a particular time, as illustrated in Figure 22(a). In short, when a User has completed its transmission, another User can take over at a succeeding time, but "inherits" all the N channels. That's why OFDM is not used in LTE-A, and instead **OFDMA, O**rthogonal **F**requency **D**ivision **M**ultiple **A**ccess, is employed, whose discussion follows shortly.

- The other problem with **OFDM** is its high **P**eak-to-**A**verage **P**ower **R**atio, which makes it unsuitable for the uplink, given the small size of the mobile **U**ser **E**quipment, **UE** (Smart phone, Smart Pad, etc.). **OFDMA** has also this problem, so we discuss it below.

OFDMA

A major problem with **OFDM**, in addition to high **P**eak-to-**A**verage **P**ower **R**atio, **PAPR** is its lack of flexibility. This is illustrated in the Frequency-Time Grid of Figure 22(a) which shows transmission by 3 Users, in terms of the **R**esource **B**locks, **RB**. Once a series of subcarriers is assigned to a User, like the one on the far left of the diagram, should the channels become noisy or exhibit too many delays, the User has no possibility of switching to another set of frequencies. It is stuck with the original allocated frequencies. This lack of versatility is a serious drawback.

A solution to the above drawback is **OFDMA, O**rthogonal **F**requency **D**ivision **M**ultiple **A**ccess, which enables "Access" to another frequency via **T**ime **Di**vision **M**ultiplexing, **TDM** as shown in Figure 22(b) for the same 3 Users. In that figure, the availability of **TDM** allows a User, such as the one on the far left, switching to another set of frequencies after one **RB, R**esource **B**lock has elapsed in time. Recall that an **RB** is the smallest unit of **resources** that can be allocated to a user. Thanks to **TDM**, a User can now "access" another set of available subcarriers, once the minimum required time of an **RB** duration, has elapsed. This is the key feature of **OFDMA**, namely Access in time of another set of subcarriers when the channel conditions demand it.

OFDMA, much like **OFDM** suffers of high **P**eak-to-**A**verage **P**ower **R**atio, **PAPR**. The high peak power occurs when the normally *out of phase* subcarriers fall *in phase* (constructive interference) at one instant of time, thus making the signal envelope shoot up to a maximum value. As shown in Appendix E, PAPR increases directly with the number of carriers N.

OFDMA is only used in the Downlink, from the Base Station to the various mobile Users, the reason being that its high **P**eak-to-**A**verage **P**ower **R**atio, **PAPR** can be mitigated with large dynamic range power amplifiers that can be accommodated in larger ground housings. The resultant bulky hardware cannot be accommodated in the small Mobile User devices. As a result, we need to turn to another approach for the Uplink, **SC-FDMA, S**ingle **C**arrier-**F**requency **D**ivision **M**ultiple **A**ccess.

SC-FDMA

SC-FDMA differs from **OFDMA** in one significant respect:

OFDMA transmits all symbols in parallel, say in M adjacent subcarriers of 15 KHz bandwidth each, while each symbol remains fixed during the **OFDMA** symbol interval of duration 66.7 μs (equal to $1/\Delta f = 1/15$ kHz).

On the other hand, **SC-FDMA** transmits ***those M symbols in series, but M times as fast, over the same 66.7 μs interval*** called SC-FDMA interval, but it does it by spreading each symbol over M subcarriers, while holding it constant for only 1/Mth of the **SC-FDMA** interval, or (1/M)) x 66.7 μs. Thus, **SC-FDMA** achieves the same symbol rate at **ODFMA**.

How are each of the M = 4 symbols spread across 4 distinct adjacent channels? Refer to Figure 23, an overall block diagram of the **SC-FDMA** transmitter and receiver. It differs from the **OFDM** transmitter block diagram of Figure 21, by the addition of a QAM mapper that maps out the M symbols to a much larger set N of subcarriers followed by a **DTF, Discrete Fourier Transform**, referred to as an **FFT, Fast Fourier Transform**.

Figure 23 shows on the upper left the symbols time sequence, $\mathbf{x_i}$ going through a **S**erial-**to**-**P**arallel, **S-to-P** converter and being converted by the **FFT** into a parallel sequence of corresponding complex frequency coefficients, $\mathbf{X_k}$. The mathematical derivation is given in Appendix F, where it is noted that unlike **OFDMA**, the complex coefficients, $\mathbf{X_k}$ are no longer directly related to the original symbols, $\mathbf{x_i}$ **.**

The next step in the block diagram is a most important feature of **SC-FDMA** implementation. It is a **Mapping** operation, which spreads each of the M complex frequency coefficients, $\mathbf{X_o}$, $\mathbf{X_1}$, … … $\mathbf{X_{M\text{-}1}}$ across M adjacent channels called ***Localized Subcarrier Mode***. Another configuration,

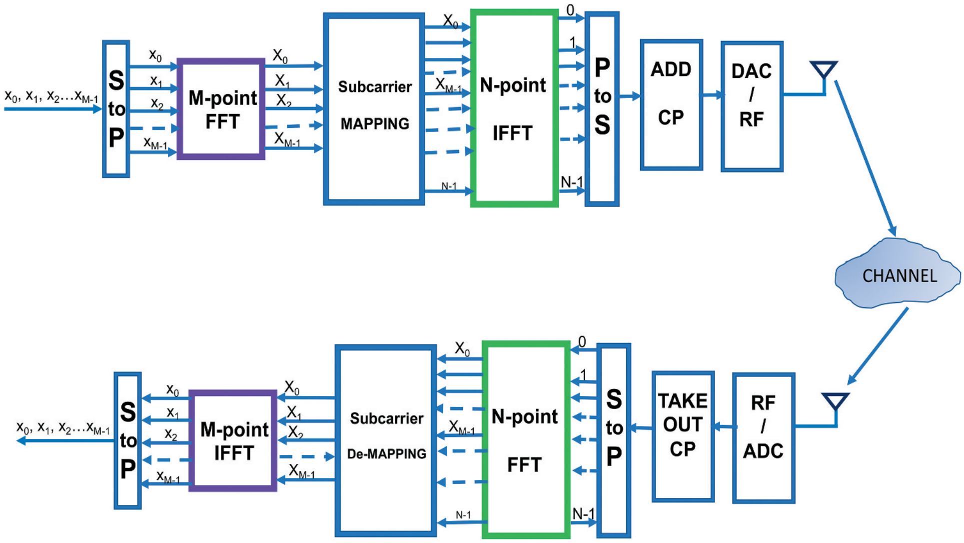


Figure 23. SC-FDMA transmitter receiver block diagram.

Distributed Subcarrier Mode, spreads the $\mathbf{X_m}$ in non-adjacent ones, evenly distributed across the N available channels, and yields identical results.

Going back to the ***Localized Subcarrier Mode***, the M complex coefficient, $\mathbf{X_m}$, is assigned to M adjacent frequency channels and fed into an **I**nverse **D**iscrete **F**ourier **T**ransformer, **IFFT**. From there on, they follow the identical route of **OFDMA** symbols: Conversion to time domain, summing, that is, **P**arallel-**to**-**S**erial ***P*-to-S** conversion, **CP** or **C**ycling **P**refix addition to cancel out channel time delays as discussed earlier, **DAC, D**igital-to-**A**nalog **C**onversion, and modulation of an RF carrier in one of the LTE-A assigned bands. The receiver in Figure 23 block diagram performs the inverse operations shown in the transmitter and does not need any further explanations.

To sum up: The output of the **IFFT** is an exact replica of the input data symbols, in both, Localized and Distributed Subcarrier Modes. Furthermore:

- The **FFT** output is a sequence of M complex coefficients in the frequency domain. Each coefficient is a linear combination of the M constellation symbols, with a frequency spectrum encompassing a bandwidth MΔf where $\Delta f = 15$ kHz in LTE-A.
- The FFT parallel output of the M complex coefficients gives rise to an SC-FDMA symbol of same duration as the equivalent OFDMA symbol as shown in Figure 24. However, in the SC-FDMA case, those M coefficients are sent serially over the SC-FDMA symbol duration, consequently at a rate 1/MΔf.

Furthermore, in order to modulate an RF carrier in one of the **LTE-A** allotted frequency bands, around 2 GHz, all the baseband modulated symbols can be upconverted to this high frequency by reassigning them to corresponding bins in the **IFFT**

This is nicely illustrated in Figure 24, with a simple QPSK modulation example of M = 4 Subcarriers, taken from Rumney [6]. It shows in the upper left corner a line tracing in time the succession of the 4 symbols being fed to the **FFT** after sampling the waveform at 4 times the **SC-FDMA** symbol interval. The resultant frequency spectrum is derived and shown in Appendix F.

What makes **SC-FDMA** so different from **OFDMA**? In **OFDMA** N symbols transmitted *at the same time*, each on a different 15 kHz subcarrier amount to a multi-carrier transmission. This simultaneous time transmission of the symbols generates "beating" and resulting constructive- and destructive-phase interferences. Such a phase interference is what accounts for the highs and lows in signal strength, the high **PAPR, P**eak-to-**A**verage **P**ower **R**atio of OFDMA.

On the other hand, **SC-FDMA** symbols are transmitted sequentially, hence, exhibit a much lower **PAPR**. The lower **PAPR of SC-FDMA** accounts for its use in the Uplink, where the size and dynamic range of **M**obile **U**sers equipment is limited.

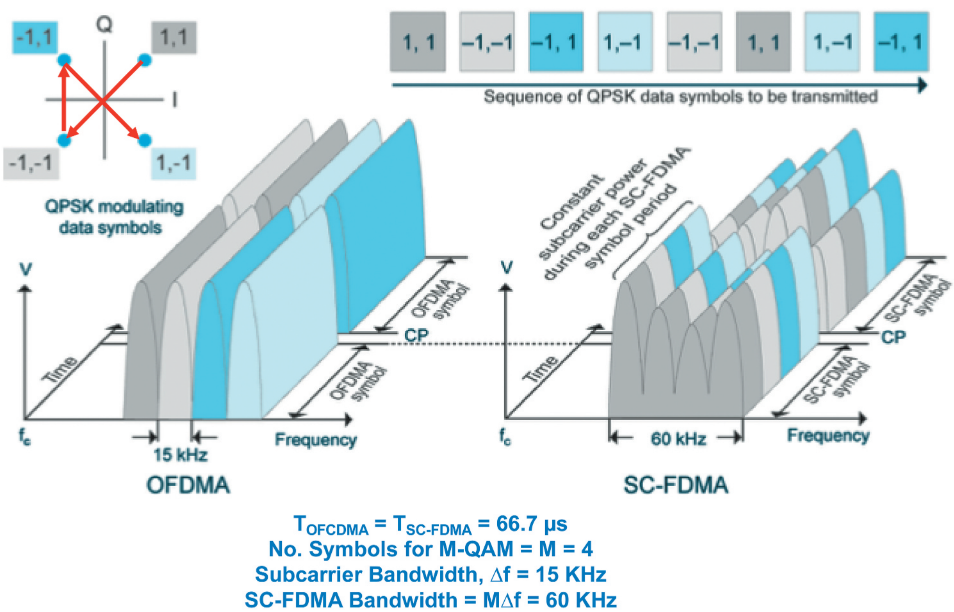


Figure 24. Symbol transmission with OFDMA and SC-FDMA.

PAPR, Peak-Average Power Ratio

A major problem with **OFDM** and **OFDMA** is their high **P**eak-to-**A**verage **P**ower **R**atio, **PAPR**, which makes it unsuitable for the uplink, given the small size of the mobile **U**ser **E**quipment, **UE** (Smart phone, Smart Pad). For instance, a typical **PAPR, P**eak-to-**A**verage **P**ower **R**atio of 10 dB would increase the battery consumption so as to an impractical size. In addition, it would drive the RF power amplifier to saturation, outside of the linear regime. This is illustrated in Figure 25 **P**eak-to-**A**verage **P**ower **R**atio, **PAPR**. A heuristic explanation given in Appendix E shows that the **PAPR** increase as N, the number of subcarriers. Such swings between peak power and average power are understandable; the subcarriers, being sinusoidal, when added together for transmission can reinforce or cancel in unison, as shown clearly in Figure 25. For instance, a transmitter of average power 0.2 watt is "stressed" to put out 2 watts if = **PAPR** = 10 dB.

The way around this problem, as discussed in the previous section is to use **SC-FDMA** or **Si**ngle **C**arrier (orthogonal) **F**requency **D**ivision **M**ultiple **A**ccess in the uplink.

The comparative performance between **OFDMA** and **SC-FDMA** is described in the literature in terms of a **CCDF, C**omplementary **C**umulative **Di**stribution **F**unction. Figure 26 provides an example for a variety of cases, taken from Reference [7], and are self-explanatory. In this figure, **IFDMA** is the Distributed Subcarrier Mode, where the "**I**" of **IFDMA** refers to Interlace mode; and **LFDMA** refers to the Localized case. Note that the vertical axis label is the **C**omplementary **C**umulative **Di**stribution **F**unction **Pro**bability of the ratio of **PAPR** to a reference $\mathbf{PAPR_O}$.

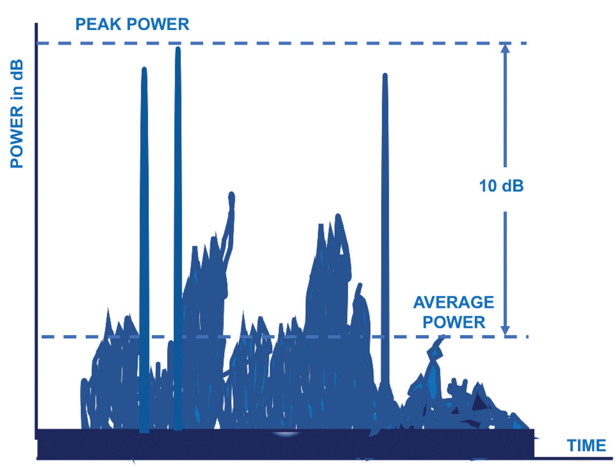


Figure 25. PAPR, Peak-to-Average Power ratio in OFDMA transmission.

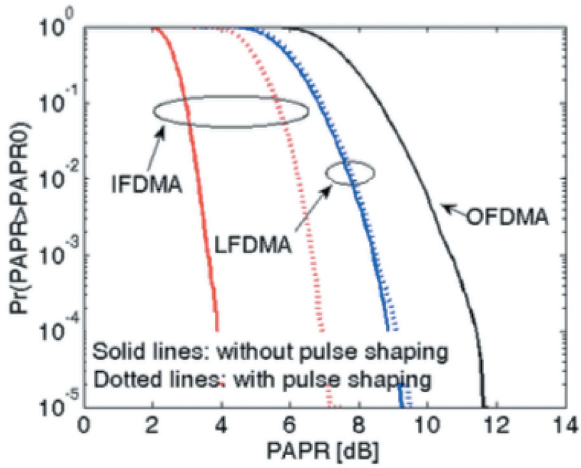


Figure 26. Comparison of CCDF of PAPR for IFDMA, LFDMA, and OFDMA. M = 256, N = 64, and Pulse Shaping Parameter, α = 0.5. (a) QPSK. (b) 16-QAM

Multiplexed antennas

Communications through the atmosphere are plagued by multi-paths. Objects on the ground, as well as atmospheric scatterers, contribute to this problem. They result in a degradation of signal quality that contribute to increased error rate of the digital data.

One way around the problem is through space diversity with multiple antennas operating at the same frequency. This technique called **MIMO,** **M**ultiple **I**nput-**M**ultiple **O**utput antennas. Is illustrated in Figure 27 with two pairs of antennas.

The two transmitting antennas on the left can be located in the same mobile user or be in two different smart phones. The pair of receivers on the right are, as a rule, located in the same base station. The intervening atmospheric path in the figure illustrates multipaths created by both, a small building and an atmospheric scatterer.

The technique is simple and well known, and it has been used in a variety of microwave communication systems since the 1950s. It consists of characterizing the intervening path in terms of a transfer function, defined by the coefficient, h_{11}, h_{12}, h_{21} and h_{22} shown in the figure which are measured by sending probe signals, together with the data.

In the figure, the x's and y's refer to either the data signals or the probe signals. Since the two transmitted and the two received probe signals are known, a total of 4 measured quantities, the equation in the figure allows to solve for the 4 unknown parameters of the medium, h_{11}, h_{12}, h_{21} and h_{22} . Once these 4 parameters are known, one can calculate the senders' data

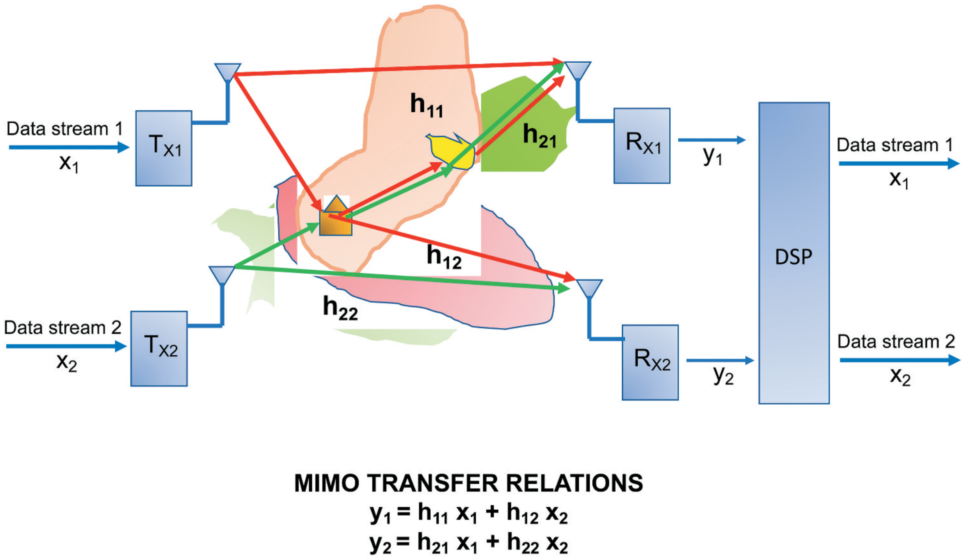


Figure 27. MIMO mitigation of multipaths reception.

streams from the two received data stream and the four h parameters derived from the probe measurements,

This technique is not limited to two pairs of antennas. Variants include **SISO, S**ingle **I**nput (one transmitting antenna) with **S**ingle **O**utput (one receiving antenna), **SIMO, S**ingle **I**nput, **S**ingle **O**utput, and in its more general form **MIMO**, which can have more than two transmitting and receiving antennas.

We conclude this section with Table 5, a summary of **LTE-A** features.

5 G – A Game changer

Introduction

The first **5 G** network was deployed in South Korea, in April 2019. Ericsson of Sweden predicted that **5 G** internet would cover up to 65% of the world's population by the end of 2025! At present, **5 G** networks and technologies are in continued state of development, with standards changing from year to year. It is expected development will continue at least until 2020.

The 5th generation of wireless networks is a game changer! It goes beyond the **4 G L**ong **T**erm **E**volution **(LTE)** that connects mobile devices to the Internet (often referred as Mobile Internet). 5 G is committed to connecting "billions" of devices communicating through the Internet between people and devices such as mobile phones, tablets, computers, sensors and actuators. Such massive connectivity demands new standards of data transfer, in particular:

Table 5.

KEY FEATURES	4G LTE-A
Lifetime	2010 - PRESENT
First Deployment	2009 STOCKHOLM, SWEDEN
Mobile System	Frequency Domain Multiplexed Modulation MIMO Antennas
Frequency Spectrum (MHz)	Nearly 80 Frequency Bands with Range: 600 MHz to 2300 MHz
Channel Bandwidth Occupied Bandwidth	1.4 MHz, 3MHz, 5 MHz, 10 MHz, 15 MHz, 20 MHz 1.1 MHz, 2.7 MHz, 4.5 MHz, 9.0 MHz, 13.5 MHz, 18 MHz
Access Protocol	Downlink: OFDMA/Uplink: SC-FDMA
Switching Technology	Internet Protocol (IP) Packet Switching (Voice and Data)
Specified Peak Data Rates	Downlink: 1000 Mb/s - Uplink: 500 Mb/s
Core Network	EPC, Evolved Packet Core Digitized Voice (TCP/IP Protocol)/Data Packets (IP Protocol)

- Very high speeds up to 10 Gb/s in the Uplink, and 20 Gb/s in the Downlink
- Very low latency, from 110 ms down to 1 ms
- Connections via the Internet on the order of many billions
- EE, Energy Efficiency, increase from 1 Megabit/Joule to 1 Gigabit/Joule in the next ten years

The latter makes it necessary to open up the frequency spectrum by bringing in **M**illi**M**eter **Waves** (**mmWaves**).

mmWaves require new mobile phones with low power, miniature directional antennas (MiMO, Multiple input/Multiple Output) and built-in modems (Modulator/Demodulator) capable of up or down conversion of **mmWaves** frequencies between 30 and 300 GHz.

The enormous area coverage that results from the expected billion connections between humans and machines through the internet is referred to as **I**nternet **o**f **T**hings (**IoT**), a term coined around 1999 by Kevin Ashton, a British manager at Proctor and Gamble who developed the **RFID, R**adio **F**requency **I**dentification **T**ag, a tag used to identify and track merchandise in various businesses such as department stores.

In short, 5 G wireless connectivity takes place between:

- Cell towers, WLAN (Wireless Local Area Network), Internet, etc.

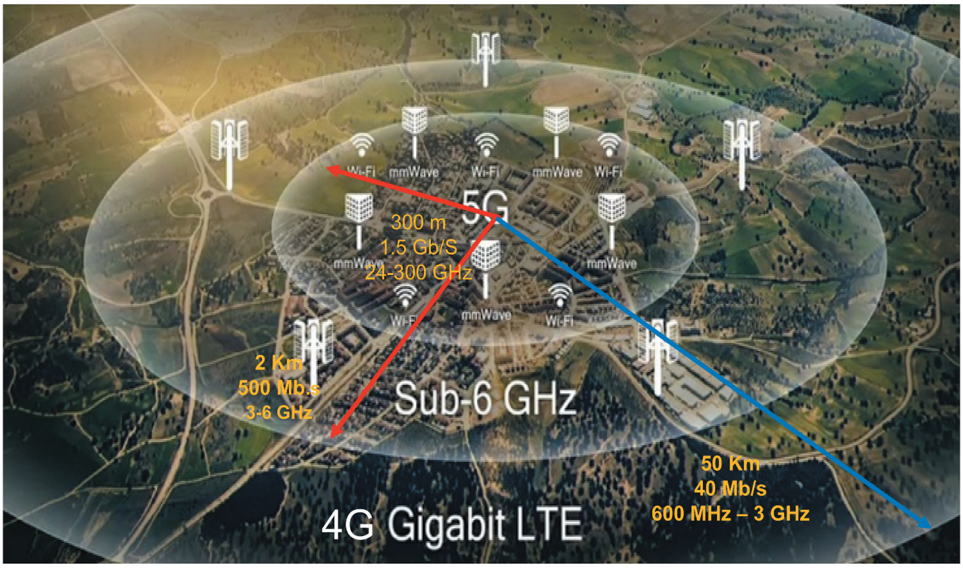


Figure 28. 5 G data rates, frequency range and covered radial range.

- People-to-People Communications (smart phones, Tablets, Personal computers), including multi-persons video calls
- Machine-to-Machine Communications such as Monitoring of sensor networks, Industrial Automation, Self-Driving cars

How does **5 G** cope with such an increased capacity and connectivity? By the addition of **M**illi**m**eter-**W**ave spectrum, **mmWave** combined with smaller antennas. In essence,

- Shifting to the **mmWave** spectrum increases *Data Capacity* through the resultant greater bandwidth that is available at higher frequencies.
- Smaller antennas, that is, 3 mm at 100 GHz (!), with **MIMO, M**ultiple **I**nput-**M**ultiple **O**utput antenna arrays that provide greater *Connectivity*
- Both of the above lead to the formation of **Micro-Cells** a few hundred meters apart and Femto-**C**ells less than hundred meters apart.

The overall frequency spectrum covered by **5 G** extends in principle from 3 to 300 GHz! Since **5 G** operation must support old **4 G** devices, we include the frequency range of **4 G LTE**, below 3 GHz, down to 600 MHz

The following Figure 28: 5 G Data Rates, Frequency range and Covered Radial Range summarizes the situation. Note the three major frequency bands:

- High Band, **mmWave**, covers the range from 24 to 300 GHz (12.5 to 1.0 mm)

- Middle band, Sub-6 GHz covers a range from 3 GHz to 6 GHz
- Low band, the **4 G** spectrum spanning between 600 MHz up to 6 GHz

The data rates shown in the figure are not the 5 G specified goals but rather an achieved average.

Network architecture

The **5 G** network architecture differs from the **4 G** architecture in one significant respect: Introduction of distributed computing, where the central servers reside in the "cloud," hundred or thousand miles away from the cellular network, with the addition of smaller cluster of servers located close to the network in the so-called **E**dge **N**etwork. The advantage of the **E**dge Network is its proximity to the cellular network, thus reducing **Latency**, an important consideration in **IoT** (**I**nternet **of** **T**hings) where communications between a multitude of devices and people takes place. The existence of the **E**dge **N**etwork is made possible, for example, by storing application programs much smaller than those found in the cloud core network, such as turning on and off sensors in one's home or low-latency automobile collision avoidance systems.

This is shown in Figure 29, 5G Generic Network Architecture. We call it, "Generic" because the industry has not settled on one or two specific architectures. In fact, half-a-dozen network architectures have been proposed with an "alphabet soup": C-RAN, D-RAN F-RAN, H-RAN, O-RAN, V-RAN, based

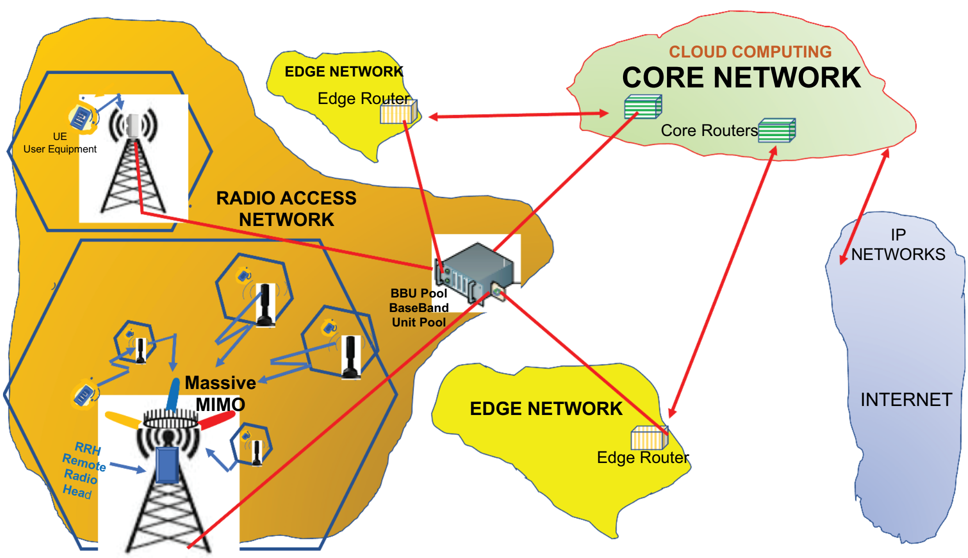


Figure 29. 5 G generic network architecture.

on different features, the most salient being whether it is a centralized or distributed architecture. An excellent comparison of some of these architectures can be found in the seminal review article listed in Reference [9].

All **5 G** networks consist of the two key elements shown in Figure 29:

- **R**adio **A**ccess **N**etwork, **RAN**
- Cloud residing **C**ore **N**etwork, **CN** with satellite **E**dge Computer **N**etworks near the cellular network.

As seen in the figure, **the R**adio **A**ccess **N**etwork, **RAN** consists of a variety of cells and base stations (shown as masts with antennas) ranging in size from Microcells (Femtocells omitted in order not to burden the figure) to Macrocells, with large antenna masts topped with directional antenna arrays, the **MIMO**s (**M**ultiple **I**nputs-**M**ultiple **O**utputs) discussed at the end of the **4 G** section. **MIMO** are capable of receiving more data simultaneously. Three directional lobes of the antenna pattern are illustrated in Figure 29, at the top of the big tower mast.

The Microcells and the smallest ones called Femtocells, a unique feature of **5 G** technology, utilize **mmW**aves that greatly increase coverage and provide data rates an order of magnitude faster than in **4 G** because of their location in a much higher portion of the frequency spectrum.

Another major difference with **4 G** networks is the splitting of functions between baseband processing **(BBU, B**ase**Ba**nd **Un**it) and RF (**R**emote **R**adio **H**ead, **RRH**) signal processing. In **4 G**, these two functions were integrated as one in a housing at the base of the transmitting tower. A more efficient distribution of these two functions is accomplished in distributed **5 G** networks:

The **BBU, B**ase**B**and **U**nit whose sending and receiving functions consist of,

- Adaptive base band modulation/demodulation up to 258-QAM
- Enciphering/Deciphering
- Multiplexing/Demultiplexing Access
- Digital Signal Processing, DSP

no longer resides at the base of the cell tower. Instead, it is located away from it, in order to serve several Macrocells, as shown in the figure. It is now part of a pool serving many Macrocells, hence its name **BBU P**ool, BaseBand unit pool

The **BBU P**ool, **BaseBand Unit P**ool communicates with Macrocell Towers via fiberoptic lines by transmitting data packets from the Core and Edge Networks (known as *Fronthaul*) as well as routing the received Baseband signals to these networks. It also communicates with the Core and Edge Networks via fiberoptic lines: Receiving data packets originating in Packet Data Networks such as Internet and forwarding BaseBand signals to these Networks (referred to as Backhaul). Figure 29 illustrates these transactions.

The second function in **5 G** networks is carried out by the **RRH, R**emote **R**adio **H**ead displayed in the figure. It is now deployed on top of the cell tower mast to perform the following functions

- **DAC, D**igital to **A**nalog **C**onversion before transmission and **ADC, A**nalog-to-**Di**gital **C**onversion upon reception . . .
- Optical to electrical signal conversion, and vice versa
- Amplification of transmitted and received signals
- Frequency Up conversion before transmission and Down conversion upon reception.

The **C**ore **N**etwork, **CN** is an exchange network that manages all the mobile voice, data and internet connections. It is distributed with major servers located in the Cloud, and local smaller servers positioned in satellite or **Ed**ge **N**etworks that act as providers to nearby users who require low latency. Both Core and Edge are shown in Figure 29.

We now review some of the significant technologies that have made 5 G viable.

(1) **M**illimeter **Waves, mmW**aves: Advantages are two-fold: Extended Coverage and Higher Data Rates.

Increased data rates are the direct results of operating in a higher portion of the frequency spectrum, from about 30 GHz, and theoretically up to 300 GHz where data rates as high as 10 Gb/s occupying a bandwidth of approximately 10 GHz constitute a small portion of the available spectrum.

Extended Coverage is obtained by splitting the cell size to micro- and ffemto-cells thus allowing multiple re-use of the same set of frequencies used in a macro-cell.

(2) Massive **MIMO** (**M**ultiple **I**nput, **M**ultiple **O**utput) Antennas.

MIMO, one of the **5 G** breakthrough technologies, reduces channel interference caused by transmission multipaths. In addition, the small size antennas fit in a user smart phone thus provide gain in the **mmW**aves band.

Furthermore, "massive" **MIMO** antennas located on top of macro-cell towers have directional beams, as seen in Figure 29 that enable communication with multiple users on the same frequency.

(3) **N**etwork **F**unction **V**isualization, **NFV** decouples network functions, such as routing, packet processing, security and many others, from

proprietary vendor hardware, and replaces them with software running on so-called **V**irtual **M**achines, **VM**s. The latter operate on software, instead of a physical computer, that runs programs and applications. **VM**s, though, reside on physical computers named “Host Machines.” Thus, all network functions are reduced to software programs now “virtualized,” that are placed on commercial physical servers, the Host Machines. Such a virtualization allows each **V**irtual **M**achines, **VM** to be assigned a portion of the network functions, thus providing a flexibility, that will be used in **Network Slicing**, as described further on.

A better understanding of **N**etwork **F**unction **V**isualization, **NFV** can be illustrated by reviewing briefly an older technology, **S**oftware **D**efined **N**etwork, **SDN** that is also utilized in **5 G**.

Software **D**efined **N**etworking, **SDN** replaces the proprietary commercial routers (more exactly their firmware) with software that provides two functions:

- The Control Plane (a Plane is a software layer)) consists of the signals that determine how packets are forwarded. It is responsible for building the routing or Lookup table.
- The Data Plane, also known as User Plane or Forwarding Plane, moves the packets from input to output, and by referring to the Lookup table forward those packets to their intended destination.

The decoupling of these two functions allows the control signals to be generated in a remote location, away from the routers, so that they can now control many routers instead of just one. In short, the software implementation of these two functions does away with proprietary routers hardware and reduces cost.

To recap:

Network **F**unction **V**isualization, **NFV** breakdowns the functions of large physical servers to its software program elements, making them independent of the physical platform and virtualizing them on a **Vi**rtual **M**achine, **VM**.

Software **D**efined **N**etworks, **SDN** convert physical networks to virtual networks. Virtual networks behave like physical networks and have all the advantages of dealing with software programs (easily modified) compared to hardware that does lend itself to major modifications.

NFV and **SDN** technologies are enabler of:

- Efficient infrastructure scaling
- Lower computational redundancy
- Less hardware systems
- Energy consumption reduction

(4) **N**etwork **S**licing provides a method of separating the multiplicity of virtual **networks** residing on a physical host into different virtual

networks that assign a specific machine or device to a corresponding user application. The result allows service providers to dedicate specific portions of their networks to a particular machine-user pair, as illustrated in the self-explanatory Figure 30: Network Slicing.

Summary

We conclude with the Summary Table 6 that lists major Performance Features, both currently achieved and future goals. Bear in mind that **5 G** is not fully deployed yet. As time goes on with ongoing developments until 2030, expect to witness significant developments in this decade, 2020 to 2030.

By the end of the decade in 2030, **5 G** will have achieved significant advances in the following performance parameters not listed in Table 6:

- 100 Billion Total machine-machine Connectivity!
- 1000-fold increase in System Capacity
- 1000-fold Battery Life lower energy consumption
- threefold increase in Spectral Efficiency, from 5 b/s/Hz to 15 b/s/Hz
- 99.99% availability
- 100% coverage!

Market trends

The deployment of **5 G** networks is at best spotty, because the majority of the providers only provide limited coverage in the **mil**li**me**ter **wave**, **mmWave** band of the spectrum, the key feature of the **5 G** technology.

mmWave touted for its rapid data rate, 20 Gb/s and very low latency, 1 ms,

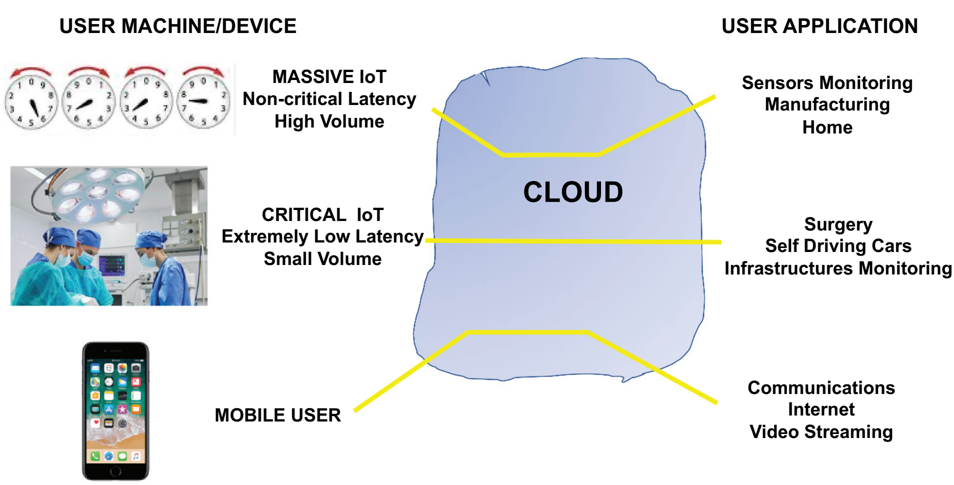


Figure 30. Network slicing.

both more than an order of magnitude better than **4 G**, is limited in range to less than a few hundred meters. So far, it has found applications mostly in large, but limited spaces such as stadiums and amphitheaters. As a result, the wireless communication industry is promoting a throttled down version of **mmWaves**, emphasizing instead the Sub-6 GHz band (see Figure 28) where latencies approaching 1 ms are still achievable.

In addition, there are not too many smart phones that have modems capable of handling millimeter waves. Samsung was one of the first companies to come up with the Galaxy 10 for 5 G, but unfortunately at a price of 850 USD! Several companies followed suits, except for Apple, which finally announced a **5 G** iPhone in September 2020. (More description of 5 G phones further on in this section).

While there were only a handful of **5 G** network deployments in 2019, the number has grown significantly in 2020, albeit with emphasis in the Sub-6 GHz band, but with limited coverage in the **mmWave** band. An interactive coverage of **5 G** can be found in the interactive map provided by Ookla, available at the following website: https://www.speedtest.net/ookla-5g-map

WIreless market revenue forecast

The forecast of the wireless communication market has been investigated by a multitude of market survey organizations. The results vary widely between estimates, both with respect to revenue and **C**ompound **A**nnual **G**rowth **R**ate, **CAGR**. We can at best provide an indication of the growth based on an

Table 6. Summary of 5 G performances features.

KEY FEATURES	5G
Lifetime	2020 – DEVELOPMENT UNTIL 2030
First Deployment	APRIL 2019 SOUTH KOREA
Mobile System	Base Band Multiplexed Modulation MASSIVE MIMO Antennas NFV, Network Function Visualization NS, Network Slicing
ACHIEVED Frequency Spectrum	Sub-6 GHz: 450 MHz – 6 GHz mmWave: 24.25 GHz to 52.6 GHz
SUBCARRIER SPACING	15 kHz, 30 kHz, 60 kHz, 120 kHz, 240 kHz
Access Protocol	Downlink: OFDMA/Uplink: SC-FDMA
Switching Technology	Internet Protocol (IP) Packet Switching: Voice and Data
Specified Peak Data Rates	Downlink: 20 Gb/s - Uplink: 10 Gb/s
Core Network	CLOUD DISTRIBUTED CORE NETWORK + EDGE NETWORKS
2030 GOALS	Latency: 1ms - < 10 ms / Mobility: Up to 500 Km/h Broadband Dense Areas Connectivity: 2000 Users/Km^2 Broadband Indoor Areas Connectivity: 75,000 Users/Km^2 Broadband Crowd Access Connectivity: 150,000 Users/ Km^2

average of all these estimates. But first, we must define how revenues are calculated. Revenues for 5 G Network infrastructures consist of three major components (refer to Figure 29):

- **R**adio **A**ccess **N**etwork, **RAN**
- **C**ore **N**etwork, **CN** and **E**dge **N**etwork, **EN**
- Transport **Ne**twork (Coax and Fiberoptic cables)

Following is a Pie chart, Figure 31(a) (www.mordorintelligence.com) indicating the relative revenue contributions of the above three components to the total network infrastructure market.

We tried to project the infrastructure revenue growth over the next five years 2020– 2025), an impossible task! Combining the web literature, the variability blows one's mind, with projections as high as 300 USD **B**illions and **C**ompounded **A**nnual **G**rowth **R**ates, **CAGR** reaching 130% We selected the ones that cluster together, as an indication of some rationality in the process and came up with the chart of Figure 31(b), which shows an estimate of 15 $B by 2024, at a CAGR of 63%. We believe that all these numbers were churned out before the onset of the current Covid-19 pandemic, and one will have to wait until late in 2021 to verify whether these projections hold up or as it is likely need to be revised down.

*Wireless mobile devices *connectivity*

In the ffollowing,we begin our projections of connectivity without including IoT connections, as they will be taken up, separately in the next section. The projections are not very reliable, given that they are all made prior to the serious worldwide economic slowdown brought about by the Covid-19 pandemic.

In 2017, GSMA (gsma.com) estimated the worldly number of mobile subscribers at 5**B** (**B**illion). Given that a subscriber could have multiple **SIM, S**ubscriber **I**dentity **M**odule cards, the *total number of connected devices*, such as smart phones, tablets, eand so onis increased by approximately 54%, from 5B to **7.7B** in 2017. Assuming a conservative **CAGR, C**ompound **A**nnual **G**rowth **R**ate of 25%, we obtain the prediction, labeled MF1, shown in Figure 32, where we have added two other projections, MF2 and MF3 for comparison purposes. However, given the current economic slowdown induced by the pandemic, we adopt the more conservative projection of MF1, which predicts 17B connection by 2022 (IoT connections not included).

**Connectivity is a generic term describing the number of connections or its equivalent, the number of connected devices.*

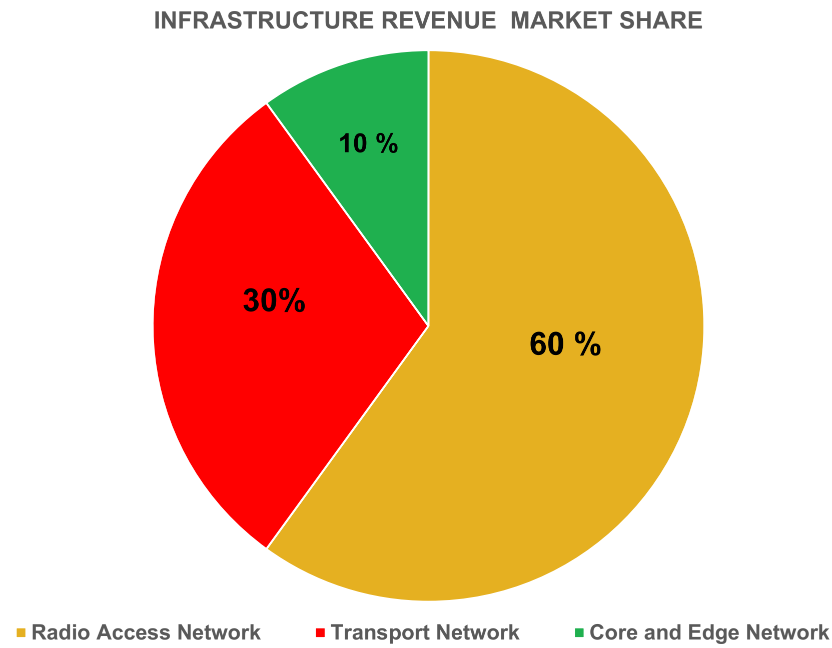


Figure 31. (a): Infrastructure networks market share. (b): Network infrastructure revenue forecast.

The additional Figure 33 displays the growth of connected wireless devices, both, total and **5 G** only. Since all these projections were calculated without taking into account the Covid-19 economic slowdown, they should be taken

with a grain of salt! In particular, the growth of **5 G** connections in 2020 should experience a significant slowdown, whose consequences will not be known until the end of that year.

Having described forecasts of both, Infrastructure revenue and Number of Connections of Wireless Mobile Devices without including **IoT** (**I**nternet **o**f **T**hings) devices, we now turn to the forecast of Number of connections and Revenue of **IoT** devices in the following section titled, ***IoT, INTERNET OF* THINGS CONNECTIVITY**

However, we preface this section, prior to defining the concept of **IoT**, with an illustrative example that should give the reader a better understanding of that concept.

IoT, Internet of things connectivity

Before defining IoT, **I**nternet of **T**hings, we introduce the concept by giving an example of an application. It is a simplified version of a case study described in Sidker's et al. excellent paper, *IoT-enabled Smart Lighting Systems for Smart Cities* [10], and we have summarized the concept in the following Figure 34: IoT Smart City Application.

On the left of the figure are shown three "streets" with lampposts. Each lamppost is equipped with transducers, capable of transmitting and receiving control signals using the local communication protocols listed on the right of the lampposts. The transducers, low-power complex devices referred as *IoT devices*, are capable of transmitting diagnostic signals indicating the status of each lighting post-performance parameters, such as luminosity of the lamp, wattage consumed, On-and-Off timing, etc. The IoT devices upon reception of control signals are designed to "act" upon the received signals and make corrective adjustment of the parameters, if necessary.

Further, to the right in the figure, the Local Hub Gateways have two functions:

- They provide connecting wireless path between the IoT devices and the Internet by converting the local protocols to an IP protocol or vice-versa.
- They also act as "Local Controllers" by communicating diagnostic information among themselves and taking corrective action. As an example, if the monitored local weather indicates a dark sky caused by a storm, the local controllers have the ability to turn on the streetlights earlier, without connecting with the internet.

Finally, the figure on the right illustrates the signal paths through the internet to the "Cloud" (Refer to Appendix G for an explanation of "How the Cloud Works). The applications containing the database and all pertinent information and actions required to operate and maintain the street lighting are stored in the virtual cloud servers.

What is IoT?

The **I**nternet **o**f **T**hings, **IoT** was invented by Kevin Ashton of Proctor & Gamble, UK in 1999. It was specifically **R**adio **F**requency **id**entity **(RFID)**, a technique to identify and track tagged objects, in particular merchandise in department stores. What Ashton called "things" is nothing more than the tagged objects, now succeeded are by the so-called as **IoT** devices.

IoT, as we know it today, took off a little less than a decade ago and has grown so rapidly as to permeate every aspect of our lives. There are a multitude of definitions of **IoT** and here is our version:

"**I**nternet **of** **T**hings, **IoT**, where Things refer to any objects, are devices with unique identifiers capable of communicating and interacting with the cloud through the internet, and <u>without human interaction</u>. **IoT** enables the connecting and exchanging of data between multiple, alike and distinct devices and systems"

IoT distribution and connectivity

We took a look at worldwide forecast of the number of IoT-connected devices until 2025. As always, they vary widely, and since they all fail to predict the effects of Covid-19, they are not very reliable. We took the ones that clustered together and yielded more conservative predictions; the results are shown in Figure 35: Yearly number of Connected IoT Devices. Although they may not be accurate predictions, they serve to illustrate the order of magnitude of the number of connections and provide an indication of their growth. It predicts 11.5B worldwide connected devices for the year 2020 and above 23B for 2030!

It is interesting to compare Figure 35 with the estimated number of IoT devices that are 5 G compatible. Figure 36: Yearly Global # IoT 5 G Connections, which follows, indicates a 5 G contribution of 3.6B by 2025, a little more than 15% of the total.

We conclude this discussion with Figure 37, which elaborates on the forecast of the 5 G contribution to the total # of connected devices; it is self-explanatory. It shows the relative contribution of the three categories of IoT devices, with the bulk of them among the so-called Massive IoT devices.

Following our coverage of **IoT** Connectivity, we now turn to a revenue forecast.

IoT, Internet of things revenue forecast

Figure 38 shows an estimate of **IoT** devices revenue, beginning with the year 2018 and extending to 2025. It is based on a **CAGR, C**ompounded **A**nnual **G**rowth **R**ate of 26.8%. This growth rate is a composite based on various scattered piece of data.

The resulting revenue growth should be taken only as an indication of the market trend, while the numbers themselves are at best order of magnitude estimates. Not shown in the graph are the following observations that apply to this estimate:

- Both, **NA, N**orth **A**merica and, **EU, Eu**rope are about equal in both growth rate and fraction of the total revenue.
- **APAC, A**sia **Pac**ific, which includes Japan, China, South Korea, Taiwan and India, are major contributors to the global revenue growth. **APAC** is expected to capture a fraction of the global market, larger than **NA, N**orth **A**merica or **EU, Eu**rope by 2020. However, once the impact of Covid-19 on the economy is assessed at the end of 2020, all these forecast may need significant revisions.

We conclude this Market Trend section with a forecast of smartphone revenues, both, Global and in the US. Smartphones being one the key components

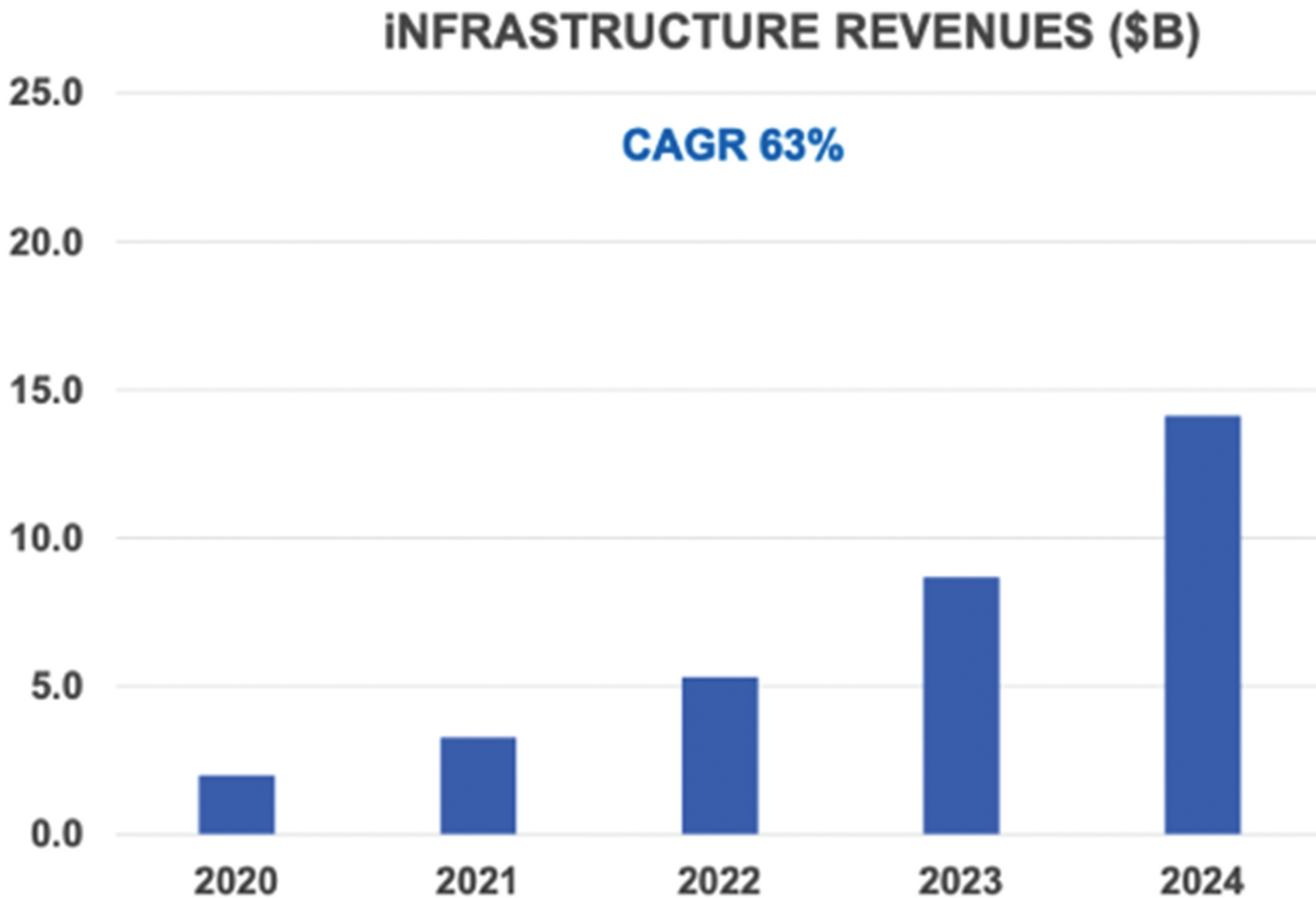


Figure 31. Continued.

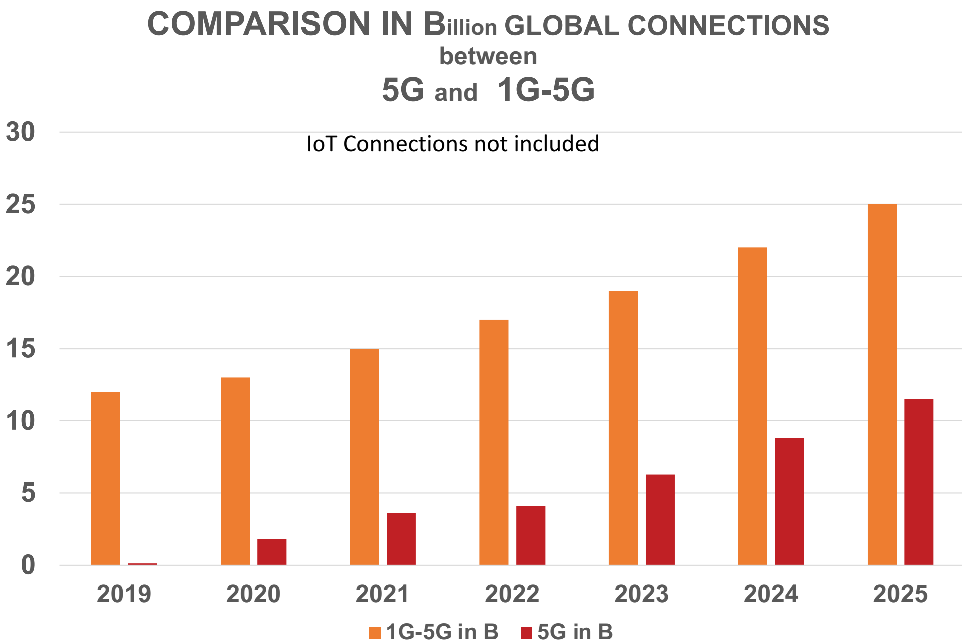


Figure 33. Connectivity forecast between total and 5 G only.

of **5 G** implementation, this section will be followed by a review of Smartphones technology and their availability in the market.

Smartphone revenue forecast

As an aside, before tackling the revenue growth, here is some interesting observation.

At the end of 2019, the total worldwide number of smartphone users was around 3.2Billions! It is estimated that by the end of 2020, in spite of the economic impact of Covid-19 on the economy, the number of smartphone users will reach 3.5B which is equal to 45% of the world population of 7.8B.

Based on published growth rates for both, number of smart phones and number of people on earth, we derived Figure 39: World Growth in Number of Smartphones vs Population. It shows a crossover point in the year 2030 where both numbers, smartphones and people, are equal. Beyond 2030, the number of smartphones in the world will exceed the number of people on earth and continue growing faster!

An estimated Revenue forecast over the decade spanning the period 2015 to 2025 is shown in Figure 40: Worldwide Smartphone Sales in $B (Billion US dollars). It shows revenue from sales estimated to reach US 800 USD billion by

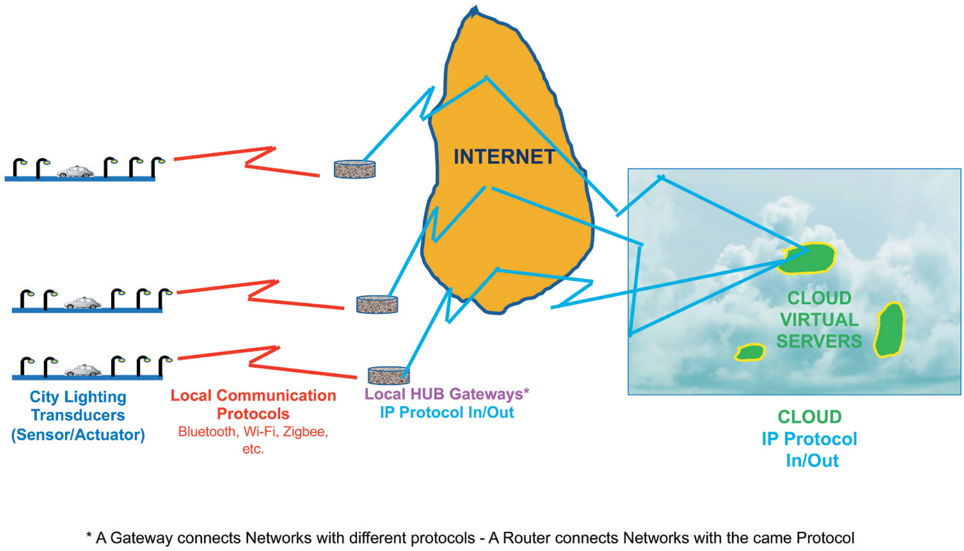


Figure 34. IoT Smart city application.

2025, just about twice the revenue of 2015, a **CAGR, A**nnual **C**ompound **G**rowth **R**ate of 7.2%.

Market capture among principal vendors is displayed in Figure 41: 2019 Percentage Smartphone Sales according to Vendors. It shows that worldwide, the leader is Huawei with a market capture of about 38% followed very closely by Samsung, while Apple, the US giant is "down in the noise". On the other hand, when it comes to the US market, Apple is undisputedly the leader with nearly 45% capture, dwarfing all the other vendors!

The smartphone and millimeter waves

The fifth generation of wireless mobile communications, **5 G** has been called a game changer. WHY? Most of the features of **5 G** are either

- Same as 4 G such as operating frequencies, packetized voice and data, accessprotocols (OFDMA, Orthogonal Frequency Division Multiple Access and SC-FDMA, Single Carrier-Frequency Division Multiple Access); or
- Improved significantly by an order of magnitude with respect to 4 G through advances in technology such as increased date of 10x4G, or reduced latency (1/10)x4G

What sets **5 G** apart from the previous generation of wireless mobile communications is brought about by a *giant step forward*, the introduction of

mmWave, milli**m**eter wave**s** coupled with **MIMO, M**ultiple Input-**M**ultiple **O**utput antennas. Opening the frequency spectrum to a much high frequency band, that is, smaller wavelengths, allows much greater bandwidths with consequently much higher data rates. Other benefits covered in prior sections of this report include reduced latency, much greater connectivity, not only between people but between IoT devices.

Unfortunately, this *giant step* has proved to be costly, which accounts for the slow worldwide deployment of **5 G** networks. One major application of **5 G** in confined areas with large populations, is in concert halls or stadiums where everybody communicates with each other using short wavelengths, **mmWave** whose limited range (300 m) prevents spillover into adjacent cellular areas

It's only a little over a year since South Korea launched the first deployment of **5 G**. Operating carriers have been slow following suit throughout the world. The current Covid-19 is undoubtedly one of the causes, since application of **5 G** networks to concert halls, stadiums and all venues of confined space is no longer feasible as long as the pandemic endures. Even South Korea has joined the ranks of the world stragglers!

To make the situation more confusing, many operating carriers claim large deployments of **5 G** network, but what they call **5 G** does not include **mmWave**! It only includes the **Sub-6** frequencies below 6 GHz, an extension of the **4 G-LTE** spectrum.

Curiously enough, when one looks at worldwide current and contemplated **5 G** networks that include **mmWave**, the US is the leader, led by the carrier **Verizon**! Verizon claims it will have deployed **5 G** in 60 cities by the end of 2020.

AT&T on the other hand announced, at the end of 2019, the deployment of "**5 G** Plus" networks in 25 cities. This is a misnomer because the so-called "**5 G** Plus" does not include **mmWave**!

T-Mobile, which acquired Spring in April of 2020 announced in June 2019 that it had deployed 5 G networks (with **mmWave** of course) in 5 cities!

As of May 2020, T-Mobile advertised **5 G** deployments in more than 6000 cities in the US. This feat is accomplished by operating in the 600 MHz band, which is cost-effective, but no **mmWave** is used!

T-Mobile, though, is proceeding cautiously building up "true" 5 G systems using **mmWave** in the 28 GHz (10.7 mm)** and 39 GHz (7.7 mm) bands in many cities, such New York, Atlanta and Las Vegas.

The morale of the above is that one must be careful interpreting the claims, because too many carriers with a large number of installations refer to them as **5 G**, though they do not use **mmWave**.

A key component of a **5 G** network is the smartphone; the jump to the **mmWave** spectrum requires major changes in its circuitry. **4 G** phones are incompatible with **5 G** and they cannot be modified. So, the industry had to come up with new models, which unfortunately are pricey, thus generating a lack of enthusiasm among potential users, which accounts partly for the slow growth

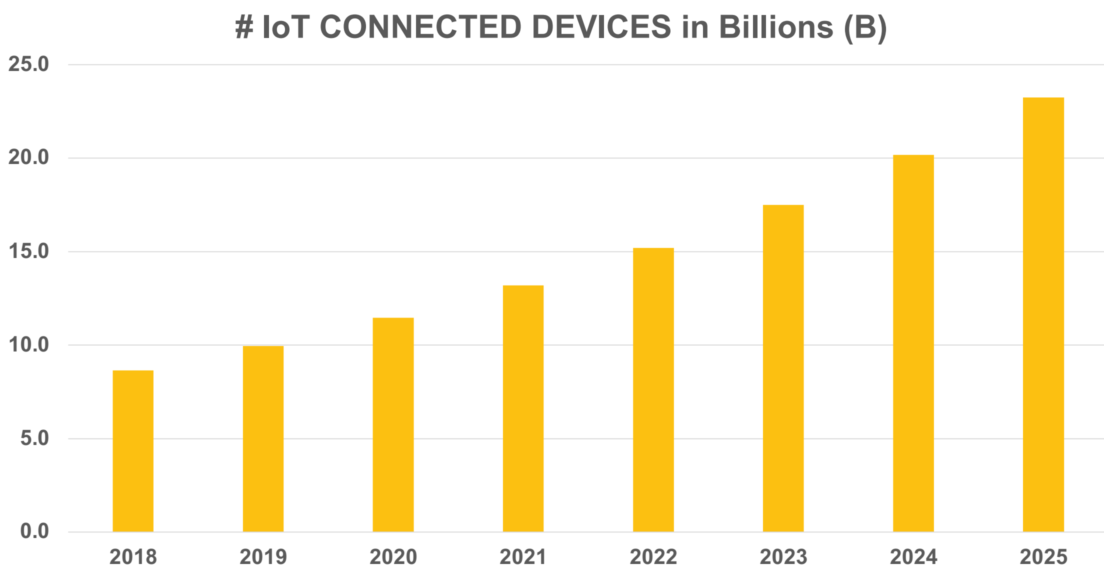


Figure 35. Yearly number of connected IoT devices.

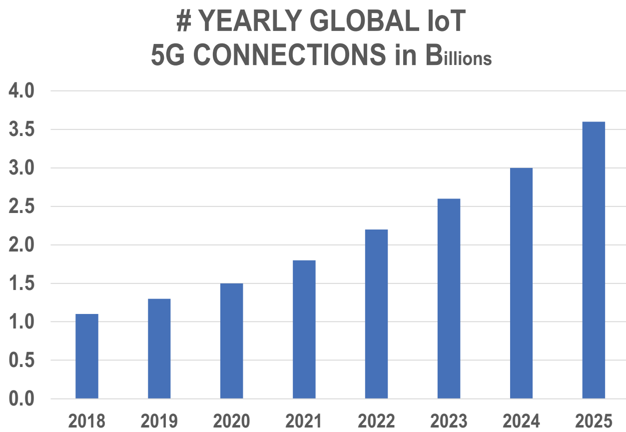


Figure 36. Yearly GLOBAL # IoT 5 G connections.

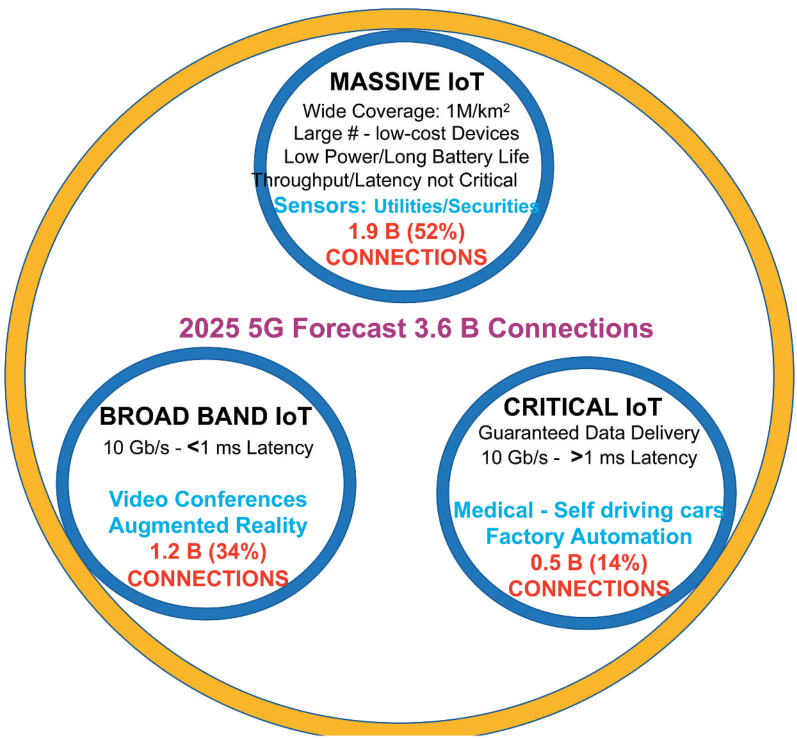


Figure 37. 2025 Prediction of the three categories of IoT devices compatible with 5 G.

of **5 G** networks. As a result, there are few smartphones today that compatible with **5 G**. The US giant Apple that was waiting in the wings for a year, finally announced in September 2020 the introduction of a **5 G** iPhone in the market.

Mention of **5 G always include **mmWave****Note that the wavelength of 6 GHz is 50 mm, thus would need antennas 5 times longer than at 28 GHz.*

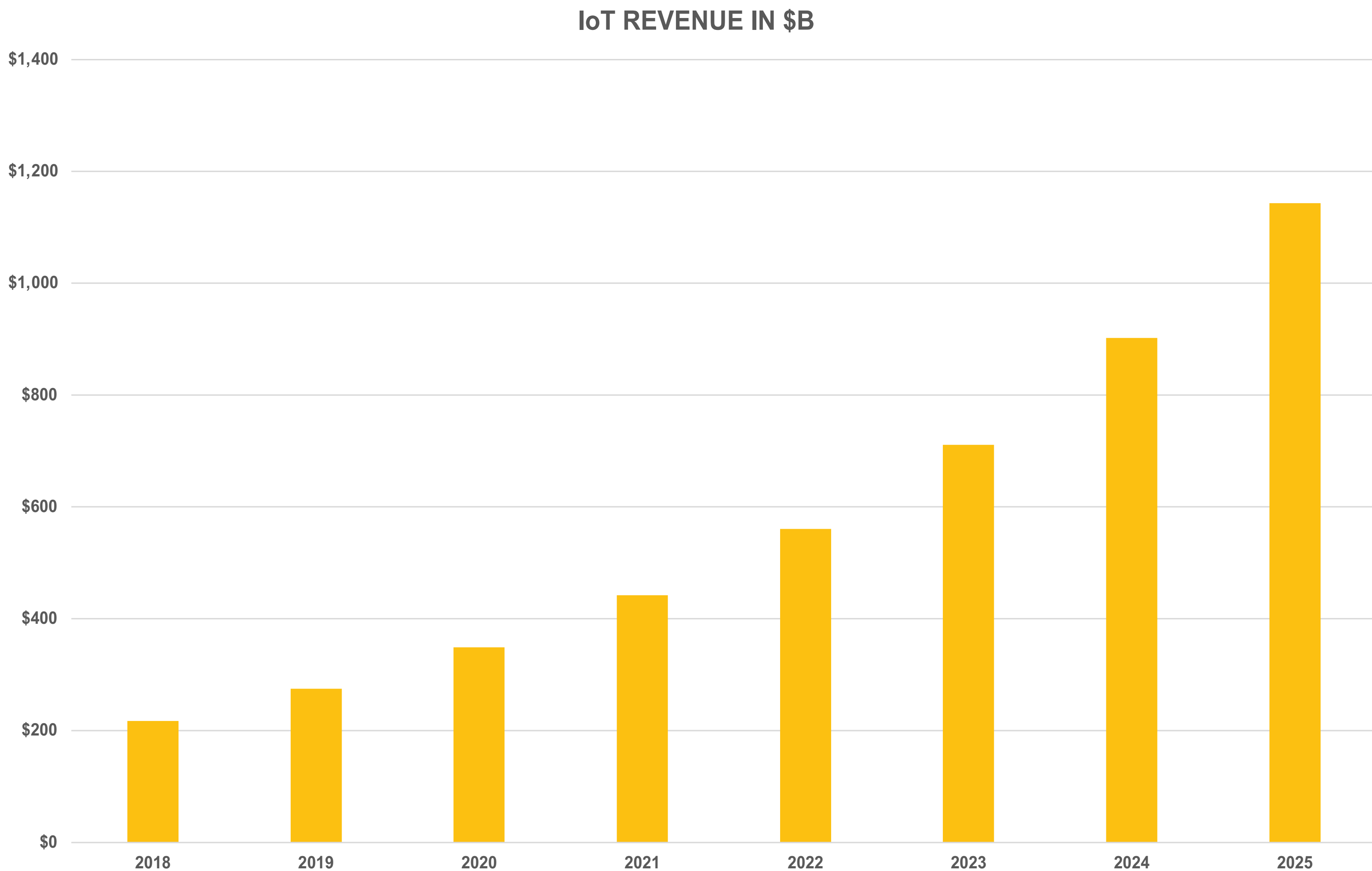


Figure 38. IoT, Internet of things revenue in $Billions.

We conclude this discussion by reviewing briefly the availability of smartphones compatible with the millimeter wave band of **5 G** and describe some of their major engineering features.

The major frequency 5 G bands used in the US are listed in Table 7: US **5 G** Frequency Bands. The three major operating carriers in the US, in order of the number of **mmWave** deployed networks are: **Verizon, AT&T** and **T-Mobile**.

Note from the figure that the frequencies used are either below the Sub-6 GHz band or in the millimeter band. None of the carriers operate in the Sub-6 GHz band (3-to-6 GHz) shown in Figure 28.

As to smartphones that support operations in these bands, only the few listed in Table 8. **mmWave** are just being introduced in the US and the situation changes almost day-to-day. Undoubtedly, there will be more smartphones available in 2021 that support the frequencies of Table 7, particularly in the **mmWave** bands. These phones are still high price, around 1000, USD but as usage increase, they should come down in price.

The most popular and advanced smartphone is the Samsung Galaxy S20 Ultra. It is truly a marvel of engineering as noted from Figure 42: Block Diagram of S20 Ultra Transmitter and Related Hardware. In this figure, we illustrate only the smartphone transmitter block diagram, a simplified version of Figure 23: *SC-FDMA Transmitter Diagram*. We have omitted the receiver, a mirror image of the transmitter, in order to keep the figure to a reasonable size. To understand the block diagram of Figure 42, the reader is referred to the section containing Figure 23 The related hardware shown in Figure 42 is taken from Ref [11].

Note that Apple is missing from Table 8. Until recently, Apple had stayed away from entering the US **5 G** market with an iPhone that would support **mmWave**. Apple did finally announce its entry in the **mmWave** in September 2020, using some variant of the same basic "Snapdragon" Qualcom modem, but we have had no time to review it.

The race for mmWave is on and may become exciting. **STAY TUNED!**

Concluding remarks

This review of Wireless Mobile Communications has been a long, yet hopeful fruitful journey.

It began a little more than forty years ago, in 1979 when MacDonald of the Bell Telephone Laboratories in the USA introduced the concept of cellular network communications, which has now been adopted throughout the world.

The implementation of cellular mobile networks evolved over the following 40 years through five generations of technology known as **1 G, 2 G, 3 G, 4 G and 5 G**. Standards and Specifications for these technologies were accomplished with the cooperation of worldwide committees and technical groups.

The first generation, **1 G** was limited to circuit switched analog voice, much like old telephony. In the second generation, **2 G** voice was digitized, though

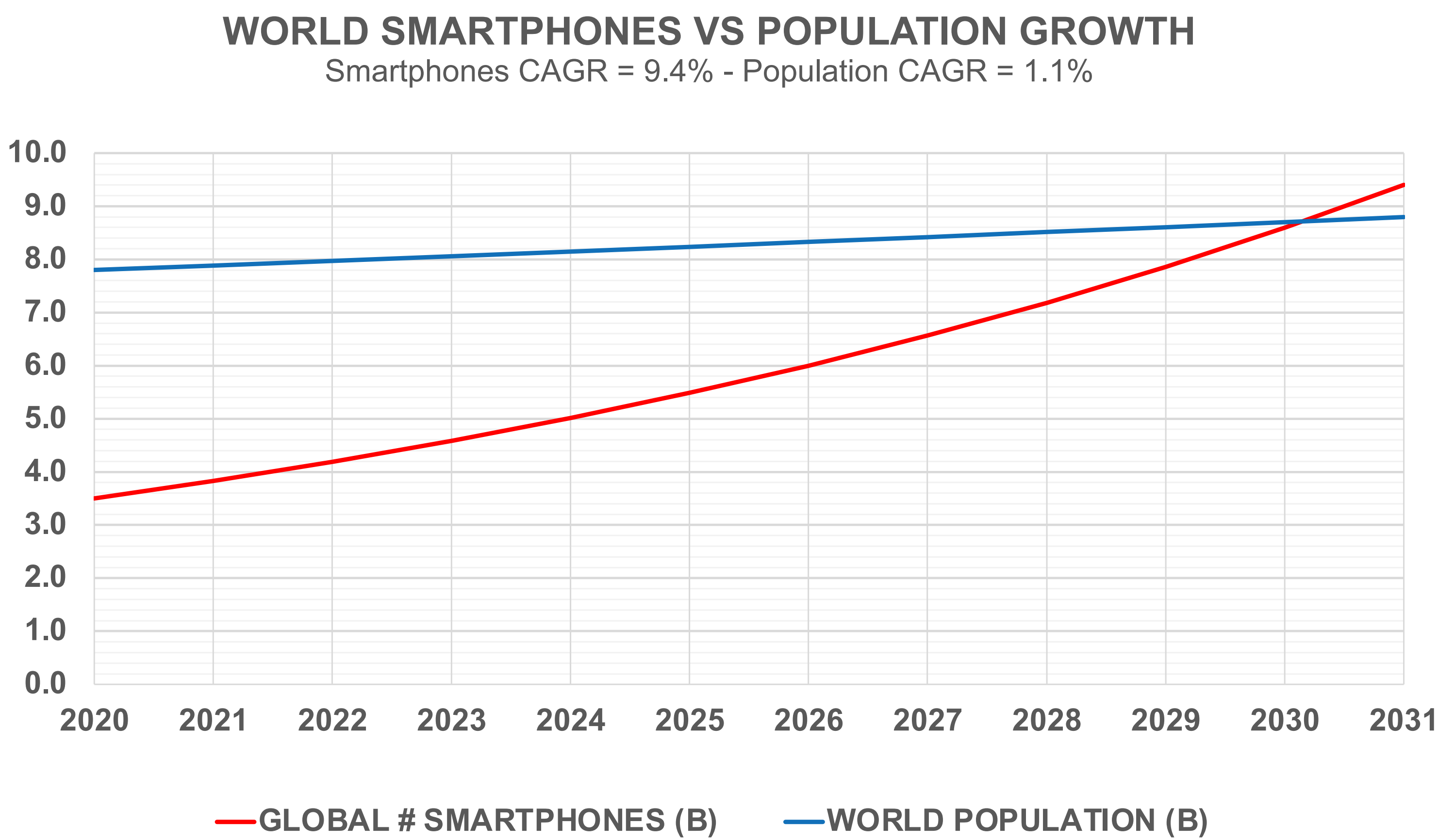


Figure 39. World growth in number of smartphones vs population.

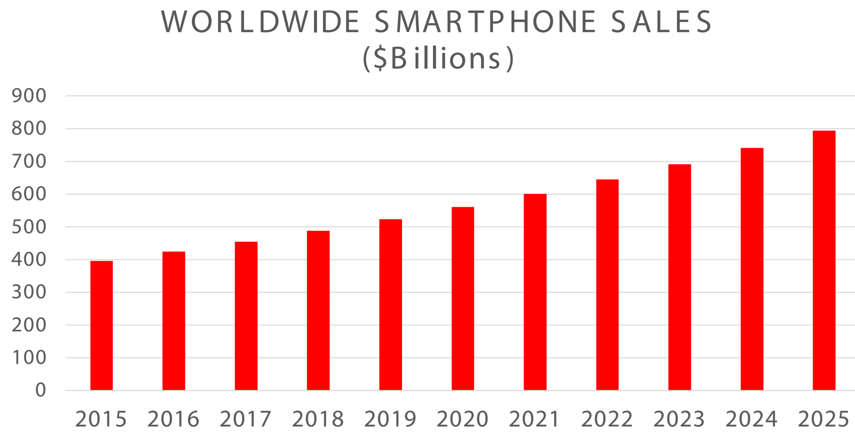


Figure 40. Worldwide smartphone sales in $B.

still circuit switched; short text messaging was added, and downlink data rates were up 300 KHz, compared to the puny 10 KHz analog bandwidth of **1 G**. The third generation, **3 G** was a big step forward with introduction of packet switched data at rates of 3.0 Mb/s for both, data and digitized voice, though the latter was still circuit switched. It took a fourth generation, **4 G** to treat both, voice and data as switched packets with downlink speeds of 1.0 Gb/s.

Then came the game changing fifth generation **5 G** with a giant leap forward: *the introduction of mm-Waves* with key parameters improvement of one or two orders of magnitude. Peak frequency spectrum having languished for several generations around 2000 MHz increased to 6 GHz in the span of one generation. This was accompanied by downlink data rates going from 1.0 Gb/s to 20 Gb/s thanks to the advent of mm-Waves. Other significant improvements include massive **MIMO** antennas; latency reduction from less than 50 ms in **4 G** to less than 10 ms in **5 G**, reaching 1 ms together with 100 billion device-to-device connections, both expected by the end of this decade. Table 9 *summarizes 1 G to 5 G Comparative Performance.*

In short, our world dominated by the advances of communication technologies will never be the same. Hopefully, a better world is dawning, if humankind can control the technological advances it unleashed and apply them to the benefit of society.

Abbreviations

G: Wireless Mobile Technology Generation; 1G:1st Wireless Mobile Technology Generation; 2G: 2nd Wireless Mobile Technology Generation; 3G: 3rd Wireless Mobile Technology Generation; 4G: 4th Wireless Mobile Technology Generation; 4G LTE-A: 4G Long Term Evolution-Advance; 5G: 5th Wireless Mobile Technology Generation; AAA: Authentication, Authorization and Accounting; ACK: Acknowledgment; ADC: Analog-to- Digital Converter; AM: Amplitude Modulation; AMPS: Advanced Mobile Phone System; ARP: Address Resolution

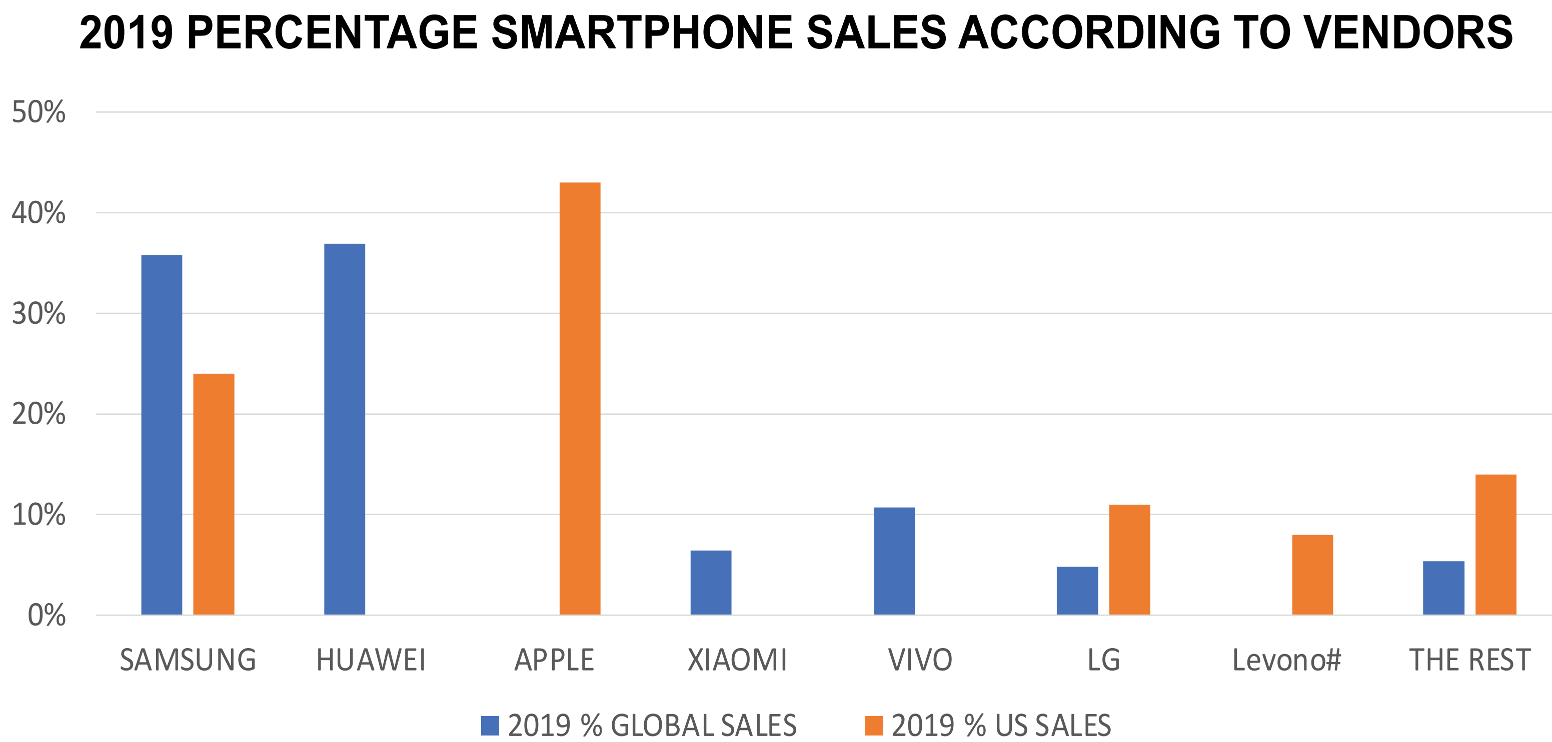


Figure 41. 2019 percentage smartphone sales according to vendors.

Table 7. US 5 G frequency bands.

5 G Sub-6 GHz BAND	FREQUENCY (MHz)	UPLINK (MHz)	DOWNLINK (MHz)	CHANNEL BANDWIDTH (MHz)
n2	1900	1850– 1910	1930– 1990	5, 10, 15, 20
n4	850	824– 849	869– 894	5, 10, 15, 20
n41	2500	2496– 2690		10, 15, 20, 30, 40, 50, 60, 80, 90, 100
n66	1700	1710–1780	2110-2200	5, 10, 15, 20, 25, 30, 40
n71	600	663– 698	617– 652	5, 10, 15, 20
5 G mmWave BAND	FREQUENCY (GHz)	UPLINK (GHz)	DOWNLINK (GHz)	
n260	39	37	40	50, 100, 200, 400
n261	28	27.50	27.35	50, 100, 200, 400

Protocol; ARPA: Advanced Research Projects Agency; AuC: Authentication Center; BBU: BaseBand processing Unit; BCS:Base Controller Station; BER: Bit Error Rate; BS: Base Station; BSS: Base Station Subsystem; BTL: Bell Telephone Laboratories; BTS: Base Transceiver Station; CAGR: Compound Annual Growth Rate; CCDF: Complementary Cumulative Distribution Function; CDMA: Code Division Multiple Access; CN: Core Network; CO: Central Office; CP: Cyclic Prefix; CS: Circuit Switched network'; DAC: Digital-to-Analog Converter; DSSS: Direct-Sequence Spread Spectrum; DTF: Discrete Fourier Transform; DV: Data and Voice; DynaTAC: Dynamic Adaptive Total Area Coverage; E-UTRAN: Evolved- UMTS Terrestrial Radio Access Network; EDGE: Enhanced Data Rate Evolution; EIR:Equipment Identity Register; EN: Edge Network; eNB: evolved Node B; EPC: Evolved Packet Core; EPC: Evolved Packet Core; ETSI: European Telecommunications Standards Institute; EV: Evolution; FCC: Federal Communications Commission; FDMA: Frequency Division Multiple Access; FFT: Fast Fourier Transform; FM: Frequency Modulation; FSK: Frequency Shift Keying; FTP: File Transfer Protocol; GGSN: Gateway GPRS Support Node; GMSC: Gateway MSC; GMSK: Gaussian Minimum phase Shift Keying; GPRS: General Packet Radio Services; GSM: Global System for Mobile communications; HA: Home Agent; HeNB: Home evolved Node B; HLR: Home Location Register; HSS: Home Subscriber Server; HTTP: Hyper Text Transfer Protocol; IFDMA: Interlaced FDMA; IFFT: Inverse Fourier Transform; IFFT: Inverse Fast Fourier Transform; IoT: Internet of Things; IP: Internet Protocol; ISI: InterSymbol Interference; ISO: International Standards Organization; ITU: International Telecommunication Union; ITU: International Telecommunication Union; ITU-R: International Telecommunications Union-Radio; LAN: Local Area Network; LFDMA: Localized FDMA; LTE: Long-Term Evolution; MAC: Medium Access Control; ME: Mobile Equipment; MIMO: Multiple Input Multiple Output; MME: Mobility Management Entity; MMS: Multimedia Message Service; mmWave: millimeter Wave; MS: Mobile Station; MSC: Mobile services Switching Center or Moblle Switching Center; MSK:M inimum phase Shift Keying; MTSO: Mobile Telephone Switching Office; NA: North America; NFV: Network Function Visualization; NIC: Network Identification Card; NSS: Network Switching Subsystem; NSS: Network Switching Subsystem; NTT: Nippon Telegraph and Telephone; OFDM: Orthogonal Frequency-Division Multiplexing; OFDMA: Orthogonal Frequency Division Multiple Access; OSI: Open Systems Interconnection; OSS: Operation and Support Subsystems; P/S:Parallel-to-Serial; PAPR: Peak-to-Average Power Ratio; PCF: Packet Control Function; PCM: Pulse Code Modulation; PDSN: Packet Data Serving Node; PN: Pseudo-Noise random code; POTS: Plain Old Telephone System; PS: Packet Switched network; PSTN: Public Switched Telephone Network; QAM: Quadrature Amplitude Modulation; RACH: Random Access Channel; RAN: Radio Access Network; RB: Resource Block; RF:" Radio Frequency; RFID:Radio Frequency Identification; RNC: Radio Network Controller; RRH: Remote Radio Head; S-GW: Serving GateWay; S/P: Serial-to-Parallel; SAT: Supervisory Audio Tone; SC-FDMA: Single Carrier – Frequency Division Multiple Access;

Table 8. US 5 G smartphone frequencies.

5 G Sub-6 GHz BAND	Samsung	LG	Sony
	Galaxy S20 Ultra	V60 ThinO 5 G UW	Xperia Pro
n2 – 1900 MHz	P	P	P
n5 – 850 MHz	P	P	P
n41 – 2500 MHz	P	P	P
n66 – 1700 MHz	P	P	P
n71 – 600 MHz	P	P	P
5 G mmWave BAND			
n260 – 39 GHz	P	P	P
n261 – 28 GHz	P	P	P

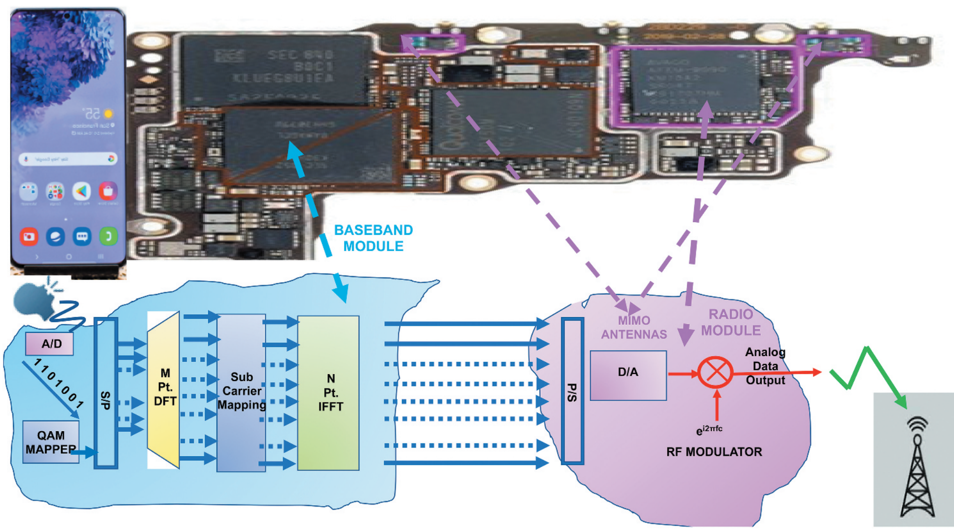


Figure 42. Block diagram of Samsung Galaxy S20 ultra transmitter and related hardware.

Table 9. 1 G to 5 G comparative performance.

GENERATIONS / PARAMETERS and FEATURES	1G	2G	3G	4G	5G
Peak Frequency	900 MHz	1900 MHz	2200 MHz	2300 MHz	6 GHz mm-Wave 25 -50 GHz
Bandwidth Downlink Data Rate	10 KHz Analog	300 Kb/s	3.0 Mb/s	1.0 Gb/s	20 Gb/s
Features	Analog Voice only Circuit Switched	Digitized Voice Circuit Switched + Text Messaging	Circuit Switched Digitized Voice + Packet Switched Data	Voice/Data Packet switched	Voice/Data Packet switched MIMO Antennas 10 ms >Latency≥ 1ms
Access Protocol	FDMA	TDMA/FDMA	CDMA	OFDMA Downlink SC-FDMA Uplink	OFDMA Downlink SC-FDMA Uplink

SDN: Software Defined Network; SDU: Selector and Distribution Unit; SF: Spreading Factor; SGSN: Serving GPRS Support Node; SIM: Subscriber Identification Module; SIMO: Single Input, Single Output; SISO: Single Input Single Output; SMS: Short Message Service; SMTP: Simple Mail Protocol; TCP: Transport Control Protocol; TDM: Time Division Multiplex; TDMA: Time Division Multiple Access; TO: Tandem Office; UDP: User Datagram Protocol; UE: User Equipment, i. e. a smartphone user; UMTS: Universal Mobile Telecommunications Service; USIM: Universal Mobile Telecommunications Service; SIM UTRAN: UMTS Terrestrial Radio Network Controllers; VLR: Visitor Location Register ; VM: Virtual Machine; VoIP: Voice over IP (Voice over Internet Protocol); WCDMA: Wide band Code Division Multiple Access; WIMAX: Worldwide Interoperability for Microwave Access

ORCID

Edvin Skaljo http://orcid.org/0000-0003-1710-3260

References

[1] P. Österlund. "Homenet Howto." Your guide to computer networks. Retrieved January 24, 2021. www.homenethowto.com

[2] V. H. Mac Donald, "The cellular concept," *Bell Syst. Tech. J.*, vol. 58(1):15-41. January 1979.

[3] Agrawal *et al.*, "Inside 3G wireless systems: the 1xEV-DV technology." Technology Review #2003-01, Tata Consulting Services, March 2003.

[4] E. Esteves. "The high data rate evolution of the cdma2000 cellular system," Multiaccess, Mobility and Teletraffic in Wireless Communications. Qualcomm Document. 5. pp. 61-72. Springer, Boston, MA. doi: 10.1007/978-1-4757-5916-7-6. Jan 2000.

[5] Chan *et al.*, "performance analysis of cellular CDMA high-speed data services," *Wirel. Commun. Mob. Comput*, vol. 6, pp. 505–521, 2006. DOI: 10.1002/wcm.293.

[6] M. Rumney, *LTE and the Evolution 70 4G Wireless*, Agilent Technologies 2000. Reprinted by London: John Wiley & sons, Ltd. 2009.

[7] H. C. Mayung, *et al. Peak-to-Average power ratio of single carrier FDMA signals with pulse shaping*, 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'06), Helsinki, Finland, 11-14 September, 2006

[8] C. Chlochina and H. Sari., *A review of OFDMA and single-carrier*, European Wireless Conference, Lucca, Italy, IEEE. 12-15 April, 2010 . pp. 706-710. I

[9] Habibi MA at al., A comprehensive survey of RAN architectures toward 5G mobile communication system. IEEE Access. 2019 May 28; 7:pp. 70371-421.

[10] Sikder *et al.*, *IoT enabled smart lighting systems for smart cities*, 2018 IEEE 8th Annual Computing and Communication Workshop and Conference, CCWC)

[11] W. Lam (n.d)., Criticality of 5G modem to RF integration; A look inside Samsung Galaxy S20Omdia Tech Informa. Retrieved January 24, 2021, from https://omdia.tech.informa.com/OM006104/Criticality-of-5G-Modem-to-RF-Integration-A-look-inside-Samsung-Galaxy-S20-Ultra

Appendix A. How does CDMA work?

There are two technologies involved in the use of **CDMA, C**ode **Di**vision **M**ultiple **A**ccess:

- Spread Spectrum, referred to in the literature as **DSSS, D**irect-**S**equence **S**pread-**S**pectrum, and
- Signal encoding with pseudo-random sequences referred to as **PN, P**seudo-**N**oise coding.

DSSS, Direct-**S**equence **S**pread-**S**pectrum is a technology dating back to **WW II** (**W**orld **W**ar **II**). It spreads the spectrum of the signal, before transmission, by a large amount, such as 50 times the original bandwidth B. The signal is then buried below the level of any interference such as noise or jamming introduced during transmission. Upon reception, the detected signal being phase locked to the transmitted signal is detected coherently, while the interference, having no fixed-phase relationship among its components undergoes incoherent detection.

The situation is analogous to optics: if both, interference and signal are broken up into a large number of frequency bands, say N, each represented by a sine wave in that band, power from the interference is on the order of N times the energy, e_f in each band times the incremental bandwidth, Δf, i. e. Ne_f; on the other hand, the detected power in the signal is the coherent sum of N amplitude (voltage), V, i. e. $(NV\}^2 = N^2V^2 = N^2E_f$ yielding an improvement in signal-to-interference ratio on the order of $N^2E_{f/}/Ne_f \sim N$. This ratio is called the *Processing Gain* or *Spreading Factor*. **DSSS, D**irect-**S**equence **S**pread-**S**pectrum is illustrated in Figure A1.

Note from the figure that expressing Spreading Factor, N = (#Chips/Bits) per unit time is the same as N = (#Chips/s)/(#Bits/s) = R_c (chip rate)/R_b (Bit rate).

Encoding has been around since classical times, but **PN, P**seudo-**N**oise didn't come into being until the 50s. The idea in **CDMA, C**ode **Di**vision **M**ultiple **A**ccess is to assign each user, a distinct orthogonal code that is applied to encode their respective signals. Orthogonal codes will be explained below. Just bear in mind that the transmitted signal is the sum of all the users'

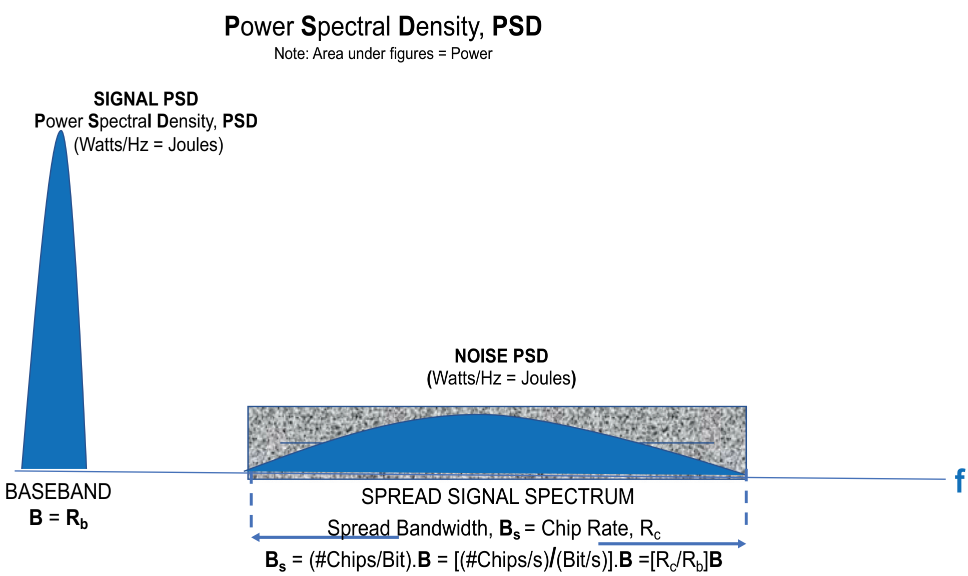


Figure A1. Spectrum spreading.

encoded signals. *Upon reception, applying a user's assigned specific code to the received composite signal only extracts that user's signal.*

To understand orthogonal codes, we first need to explain how they are generated, then applied. We follow here the elegant treatment of Carl Oliver

(https://mathaliveprojectcom.wordpress.com/) *A key advantage of such codes is to allow all users to share the same spread spectrum*

The **PN, P**seudo-**N**oise orthogonal code shown in Table A1 below is called a Walsh code and generated as indicated in the table.

The upper left quadrant is filled as indicated, 0 0 in the first row, and 0 1 in the second row. This pattern is repeated in the next quadrant on the right of the first one, and once again below that first one. The three quadrant are assembled in a square pattern by adding a similar quadrant except that its content is the complementary of the other quadrants, i. e. "0" replaced by "1" and "1" replaced by "0." The resultant pattern could be repeated several times, leading to larger matrices with longer codes. To illustrate the principle, we follow Oliver's demonstration and restrict ourselves to a "four digit" code. Each digit is called a **chip** generated at a speed N times the bit rate, with N being the code length or spreading factor equal to 4. In our example, N =4.

Before discussing the coding procedure, let us explain in simple terms the concept of orthogonality, in particular when it comes to functions that are discrete sequences of bit.

To understand orthogonality, we need to define first correlation, both cross-correlation between two different discrete sequences, such as two different Walsh codes and autocorrelation between two replicas of the same codes.

The correlation function is defined in calculus as the Integral of the product of two functions over a given interval as follows. Let one function be held fixed along the time axis, an slide the other in discrete steps. At each step multiply the two functions and add them, this last operation being what the Integration does.

With discrete sequence, the procedure is similar: multiply and add. This is illustrated in Table A2. Note that in order to carry out these two operations in the time domain, both sequence functions must be expressed in analog form by assigning the value +1 (for example +1V) to binary "0", and the value −1 (−1 V) to binary "1".

Take the cross-correlation case, which is a measure of dissimilarity between the two functions. Leave the first sequence (1-1-11) in place, and slide below the second sequence (0000) one bit-step at a time. For instance, the 1st and 5th row in the figure show:

1 −1 −1 1

1 1 1

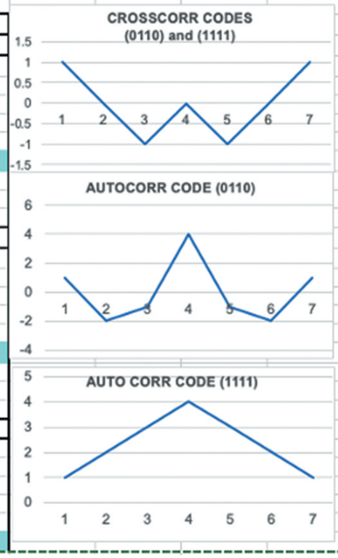
Multiply the corresponding columns and add, which gives −1. Continue sliding the second code sequence until the sum adds up to zero, yielding the top curve on the upper right side of the table. Note that the cross-correlation is negative, a measure of dissimilarity. Though the chosen codes are much too short for practical purposes, as the length increases by a factor of 40 or 400, the cross correlation "oscillations" between 0 and −2 will be reduced, flatten out and tend to zero. Zero cross correlation indicates orthogonality.

Table A1. 4-chip Walsh code generation.

0			0		0	0
0			1		0	1
0	0	0	0	Code # 1	User #2 Binary Message (11)	
0	1	0	1	Code # 2		
0	0	1	1	Code # 3		
0	1	1	0	Code # 4	User #1 Binary Message (01)	

Table A2. Cross- and auto-correlation codes.

CROSSCORRELATION and AUTOCORRELATION OF CODES						
CROSSCORRELATION BETWEEN BINARY CODES (0110) and (0000)						
1	-1	-1	1	ANALOG VERSION of CODE (0110)		
1	1	1	1	ANALOG VERSION of CODE (0000)		
1						
1	1					
1	1	1				
1	1	1	1			
1	0	-1	0	-1	0	1
AUTOCORRELATION CODE (01)						
1	-1	-1	1	ANALOG VERSION of CODE (0110)		
1						
-1	1					
-1	-1	1				
1	-1	-1	1			
1	-2	-1	4	-1	-2	1
AUTOCORRELATION CODE (11)						
1	1	1	1	ANALOG VERSION of CODE (1111)		
1						
1	1					
1	1	1				
	1	1	1			
1	2	3	4	3	2	1



Repeat the same procedure for the autocorrelation, except in these two cases the code sequences are multiplied by themselves. Note that in both cases, the autocorrelation is mostly positive, symmetric and tends to a peak, which for a 4-chip sequential code is equal to 4. As the number of chips increase, from 4 to N = 40 or 400 the negative “oscillations” will be greatly reduced and the peak will sharpen, increasing to N= 40 or 400. A long code will exhibit a narrow peak, and as the code length tends to infinity, the peak will narrow to a spike of zero width with infinite amplitude.

Thus, correlation is a measure of orthogonality, which in “real life” is defined as the “dot product” of two vectors. Zero cross correlation means the two vectors are orthogonal, that is, 90° apart; on the other hand, an autocorrelation peak of value N, signifies the two vectors are 0° apart and the value N, the dot product of both coincident vectors yields the magnitude of those vectors.

In order to reinforce the notions of orthogonality in the reader’s mind, we have added Table A3. It calculates the product between Walsh codes, either by binary XOR, or in decimal form, i.e. ANALOG. The result is converted to ANALOG, and the Averaged Sum of the Products yields 0 in all cases, thus proving orthogonality.

We now conclude this appendix by demonstrating the recovery of a specific user’s message embedded in the composite message generated by multiple users.

Figure A2 shows how a 4-chip code (0110) is applied to a message consisting of two bits:(01) that are converted to analog as 1 and −1. Note the third waveform of the figure, labeled “Analog Coded Message”: each data bit is spread across the code by multiplying the analog version of data and code, We shall return to this observation after we develop Table A4 spread sheet (no relation to the spread spectrum!) that shows how the calculations are carried out. Figure A3 shows a complete system illustrating how the original message is recovered. Refer to the figure: On the left, the message, in analog form, modulates the 4-chip code as in Figure A2. Further to the right of Figure A3, the transmitted message goes through a correlator, in essence a multiplier followed by a low pass filter. The function of the multiplier, as its name implies, is to decode the coded signal by multiplying it by its 4-chip code, thus yielding the original message as shown in the far right.

In real life, the situation is more complex as the received message is a composite of all the users' individual messages. This is illustrated in the spread sheet of Table A4 which lists messages of two users, User #1 and User #2, with their respective assigned PN codes as per Table A1. Note how the data is spread (or repeated) 4 times across the 4-chip code. The two messages, in analog format, are added together. The bottom half of the table shows the composite message being multiplied by each of the users' code, then averaged, (*remember multiplication and averaging is what's correlation is about*), thus returning the respective original messages in analog format. The row below the returned analog messages shows their binary counterpart, identical to the original transmitted before coding.

Appendix B. How to make carriers orthogonal in OFDM?

In order to understand the orthogonality principle that makes OFDM transmission and reception possible in wireless mobile communications, refer to Figure A4, a simplified version of Figure 20: OFDM receiver, Hardware Implementation.

Figure A4. illustrates the principle of OFDM reception. The incoming data stream, made of rectangular pulses (bits) of duration Δt is represented by the functions $m_n(t)$ with n = 0,1,2 . . . . N-1, that modulate the respective subcarriers of frequency, f_n .

Suppose that upon reception, one wants to recover the baseband that modulates the subcarrier of f_0 frequency. It is necessary to resort to coherence modulation, by feeding the corresponding multiplier with a signal of identical frequency that is phase-locked to the incoming signal. The output of the multiplier is simply (after using elementary trigonometry):

$$s_0(t) = m_0(t)\sin^2(2\pi f_0 t) = m_0(t)/2 - m_0(t)\cos^2 2(2\pi f_0 t)/2 \tag{B1}$$

The first term on the right side of the equation is the recovered baseband, except for a normalization factor. It is shown on the right side of the figure, together with low-pass filter that eliminates the undesired second term at twice the subcarrier frequency.

Having recovered the desired modulation, an additional step is necessary because the received signal, a composite of N modulated subcarriers, is impressed simultaneously on the N multipliers, and specifically on the first modulator, n = 0 that carries the desired signal, $\neq m_0$ (t). How does one block off the reception of the other N-1 subcarriers? This is how orthogonality comes into play,

Since N modulated subcarriers are impressed simultaneously on the N multipliers, the composite signal is made of N signals, each the product of the local oscillator with the N subcarrier frequencies:

$$s_{np}(t) = m_0(t)\sin(2\pi f_n t)\sin(2\pi f_p t); \tag{B2}$$

$n \neq p$ are integers, multiple of the lowest frequency, running from 0 to (N-1).

The area under the curve defined by the product of the two sign waves is zero as long as n is different from p. This is known as the orthogonality condition, and is expressed as (ignoring the normalizing factors):

Table A3. Proof of code orthogonality.

0	0	0	0	Digital Code # 1
0	1	0	1	Digital Code # 2
0	1	0	1	XORed
1	-1	1	-1	ANALOG
0				AVERAGED SUM of PRODUCTS
1	1	1	1	Analog Code #1
1	-1	1	-1	Analog Code #2
1	-1	1	-1	Product
0				AVERAGED SUM of PRODUCTS
0	0	0	0	Digital Code # 1
0	0	1	1	Digital Code # 3
0	0	1	1	XORed
1	1	-1	-1	ANALOG
0				AVERAGED SUM of PRODUCTS
1	1	1	1	Analog Code # 1
1	1	-1	-1	Analog Code # 3
1	1	-1	-1	Product
0				AVERAGED SUM of PRODUCTS

$$\int_0^{2\pi} \mathit{sin}(\mathrm{n}2\pi \mathrm{f_n t}) \times \sin(\mathrm{p}2\pi \mathrm{f_p t})\, dt = 0 \text{ for n different from p} \tag{B3a}$$

$$\int_0^{2\pi} \mathit{sin}(\mathrm{n}2\pi \mathrm{f_n t}) \times \sin(\mathrm{p}2\pi \mathrm{f_p t})\, dt = 1 \text{ for n = p} \tag{B3b}$$

The proof of Eq. (B3) is elementary and omitted here as it can be found in any calculus textbook, and on the web as well. Furthermore, note from (B3b) that if n = p = 0, it reverts to Eq. (B1). Also observe that the integration over the period of the lowest frequency represents the function of a low pass filter.

Although the above expressions are easily derived, they become obvious by looking at the graph Figure A5: Example of two orthogonal sinewave subcarriers of two sinewave subcarriers. One sine wave is twice the frequency of the other, both extending over the period of the lowest frequency.

Look at the four quadrants of the figures, first quadrant from 0 to π/2, second from π/2 to π, third from π to 3π/2 and fourth quadrant from 3π/2 to 2π. The areas under the product of both waves in the first and fourth quadrant are identical except that the latter is the negative of the former, adding up to zero. Similarly, the area of the products in the third quadrant is the negative of the one in the second quadrant, adding up to zero. The area under the product of these two sinewave subcarriers is always zero, independently of the values of n and p as long as they are integers and multiples of the lowest frequency, i. e. f_o, $2f_o$, $3f_o$, ...$(N-1)f_o$.

Having established that subcarrier sine waves of different frequencies are orthogonal, we need to look at the Fourier transforms of the resultant waveforms. The transmission and reception processes carried out by the hardware multipliers and local oscillators have been replaced the IFFT, as shown in Figure 21 and described in the text. Since the IFFT operates in the frequency domain, one must ensure that the orthogonality condition is preserved through the Fourier Transformation. *It turns out that the Fourier transform, being a linear transformation does preserve the orthogonality*. Figure A6. provides a verification. The upper

Table A4. Messages of two users.

TABLE A1-4								
Message Recovery - 4(Chips/message Bit)								
"0"				"1"				USER #1 BINARY DATA: (01)
0	1	1	0	0	1	1	0	USER #1 BINARY CODE: (0110)
0	0	0	0	1	1	1	1	USER #1 SPREAD BINARY DATA
0	1	1	0	1	0	0	1	X0Red
1	-1	-1	1	-1	1	1	-1	ANALOG Encoded message
"1"				"1"				USER #2 BINARY DATA: (11)
0	0	0	0	0	0	0	0	USER #2 BINARY CODE: (0000)
1	1	1	1	1	1	1	1	USER #2 SPREAD BINARY DATA
1	1	1	1	1	1	1	1	X0Red
-1	-1	-1	-1	-1	-1	-1	-1	ANALOG Encoded message
								USER #1
0	-2	-2	0	-2	0	0	-2	ADDED ANALOG Encoded messages
1	-1	-1	1	1	-1	-1	1	USER #1 ANALOG CODE
0	2	2	0	-2	0	0	-2	ANALOG MULTIPLICATION
1				-1				AVERAGE OVER 4-CHIP CODE
"0"				"1"				BINARY DATA
								USER #2
0	-2	-2	0	-2	0	0	-2	ADDED ANALOG Encoded messages
1	1	1	1	1	1	1	1	ANALOG CODE USER_(11)
0	-2	-2	0	-2	0	0	-2	ANALOG MULTIPLICATION
-1				-1				AVERAGE OVER 4-CHIP CODE
"1"				"1"				BINARY DATA

left figure shows the Fourier transform of one data pulse; it is the classical sync function where the zeros of the amplitude spectrum are **1/Δt** apart in frequency. The lower left figure shows two sinewave of different frequencies modulated by a square pulse but separated in frequency by a multiple integer of 1/Δ**t** such that the nth frequency is $f_n = f_o + (n\text{-}1)/\Delta\mathbf{t}$, assuming there are N subcarriers, of frequencies labeled from 0 to (N-1). ***Thus, a corollary of orthogonality is equivalent to demanding that the main lobes of each of the interfering frequencies be down to zero when the frequency of interest is maximum***,

How does the IFFT then carry out the function of multiplication illustrated in Figure A4 in the frequency domain? Bear in mind that the incoming signal is a composite of all the modulated subcarriers. So, two conditions must be met:

(1) If the incoming signal, a sinusoidal frequency, say f_0 modulated by a square pulse (a bit) is multiplied by a local oscillator whose frequency matches the incoming signal frequency, the baseband signal, $m_0(t)$ is recovered, as per Eq. A2.1
(2) If the incoming signal is of a frequency f_n different from the local oscillator, say f_0, it must reject these extraneous signals.

Since the IFFT operates in the frequency domain, it must deal with Fourier Transforms, FT of these two quantities: Incoming signal and local oscillator. In order to obtain the FT of the multiplied time domain signals, according to Fourier Transform theory, the two Fourier Transforms must be convolved with each other to yield the Fourier Transform of the multiplied signal. Thus, a multiplication of time signals corresponds to a convolution of their respective Fourier Transforms.

We can prove the conditions (1) and (2) with some mathematics. It is standard Fourier Transform fare, but the details though straightforward are tedious, and we will give only the final results to prove these two assertions.

We list below the expressions for the **F**ourier **T**ransforms, **FT** of the incoming signal of arbitrary frequency f_n and the sinusoidal **L**ocal **O**scillator, **LO** of frequency f_0. Then, we follow by showing up the expressions that result from the convolution of the two transforms and plot the results in a graph. (If the reader is interested in the details of the calculations, they will be found in the "cloud" in a not too distant future.

The **FT, F**ourier **T**ransforms expressions below are indicated between brackets. To simplify the presentation, we have dropped the normalization factors, which do not affect the results.

$$\text{Input FT } \{\{m_n(t)\ \sin 2\pi f_n t\} = \text{sinc}\pi(f + f_n)\Delta t - \text{sinc}\pi(f - f_n)\Delta t \tag{B4}$$

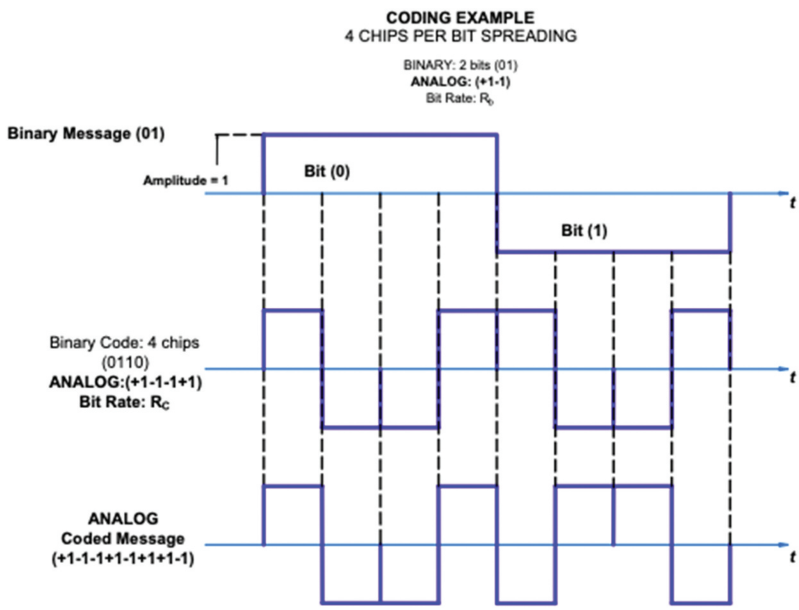


Figure A2. Coding example.

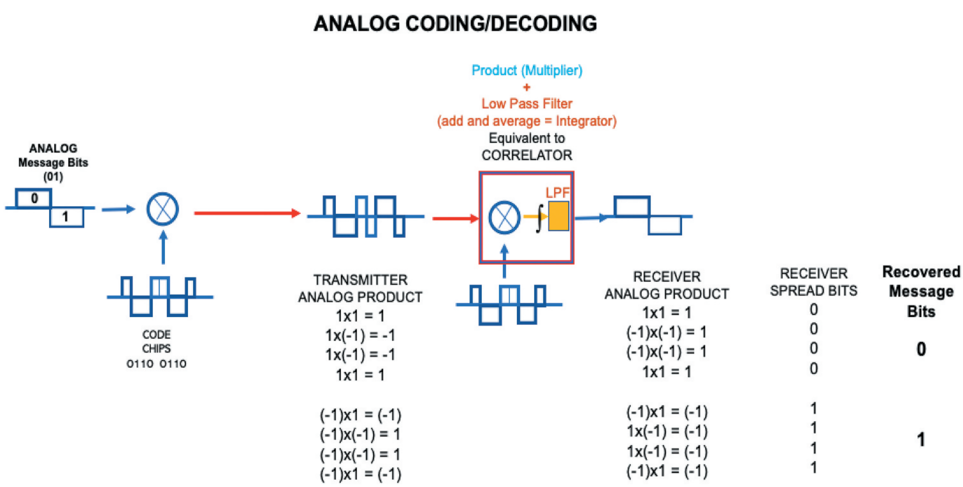


Figure A3. Analog coding/decoding – message recovery

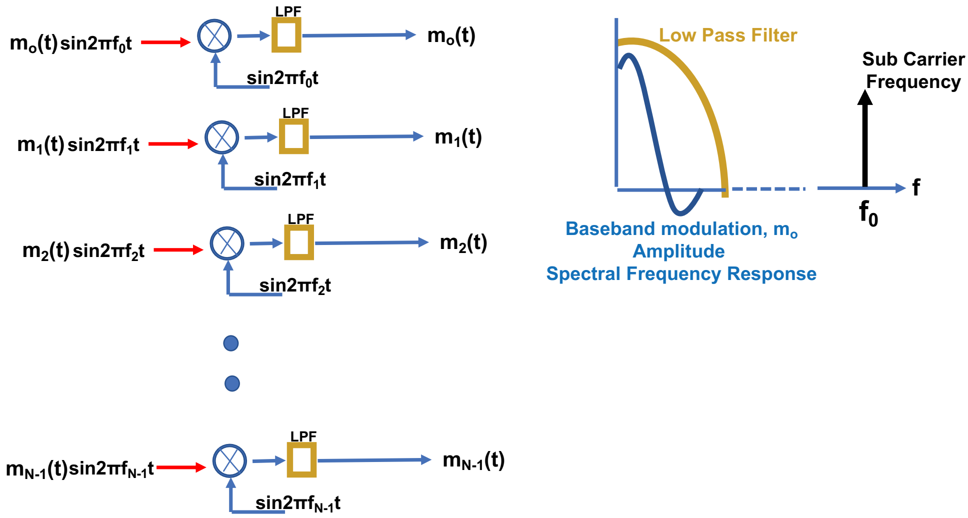


Figure A4. Simplified version of ofdm receiver.

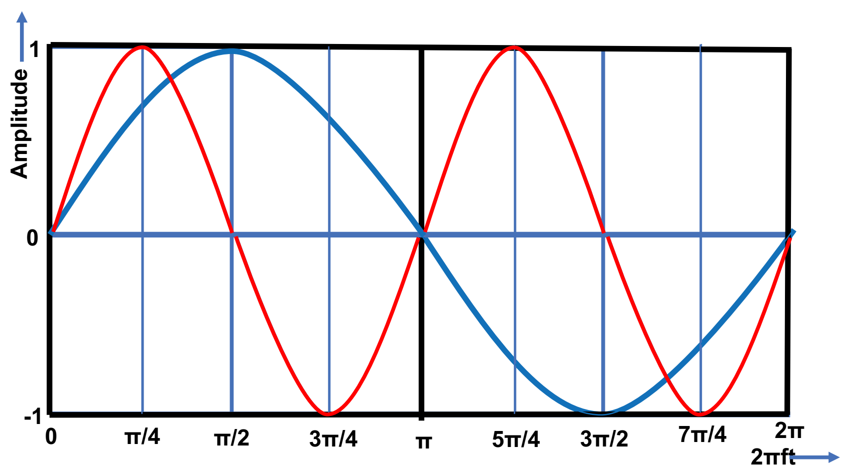


Figure A5. Example of two orthogonal sinewave subcarriers.

$$\text{LO FT } \{\sin 2\pi f_0 t\} = (f + f_0) - (f - f_0) \tag{B5}$$

In the above expressions, the function sinc(x) is the well-known **[sin(x)]/x** function and **δ(x)**, the impulse or Dirac delta function.

The convolution, between the two FT, expressed as

$$\operatorname{sinc}\pi(f + f_n)\Delta t - \operatorname{sinc}\pi(f - f_n)\Delta t\} \ast \{\boldsymbol{\delta}(f + f_0) - \boldsymbol{\delta}(f - f_0)\} \tag{B6}$$

yields the following four terms,plotted in Figure A7:

$$\operatorname{sinc}\pi[f + (f_n - f_0)\Delta t], \operatorname{sinc}\pi[f - (f_n - f_0)\Delta t], \tag{B7a}$$

$$\operatorname{sinc}\pi[f + (f_n + f_0)\Delta t], \operatorname{sinc}\pi[f - (f_n + f_0)\Delta t], \tag{B7b}$$

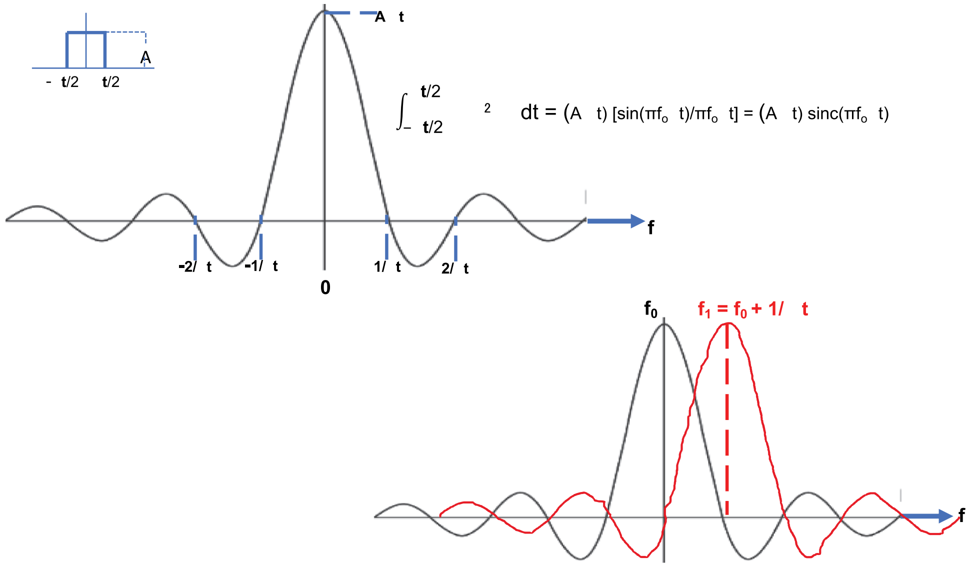


Figure A6. Preservation of orthogonality in the frequency domain.

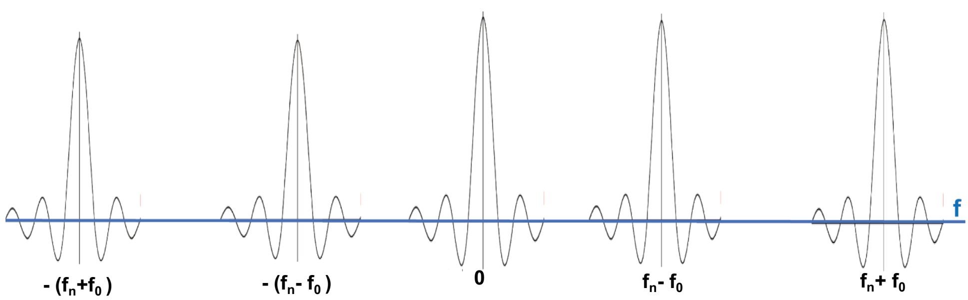


Figure A7. Convoluted FT between input signal, f_n and LO, f_0. Note that when $f_n = f_0$, the two curves f_n-f_0 collapse onto the curve centered at f = 0, and the two $\pm(f_n + f_0)$ spread further apart to $2f_0$

All four are replica of the **FT** of one of the composite modulating signals, $m_0(t) + m_1(t) + m_2(t) \ldots$ .up to $m_{N-1}(t)$, in this case $m_n(t)$ plotted in Figure A7. They are located at frequencies $\pm(f_n - f_0)$ and $\pm(f_n + f_0)$, thus do not interfere with the baseband $m_0(t)$ input recovered after multiplication shown in the right-hand of Figure A4.; they are filtered out by the Low-Pass filter. Also shown in the figure, is the FT of the recovered baseband modulation, $m_0(t)$ whose frequency f_n matches the oscillator frequency, $f_n = f_0$. This curve is shown in the center of the figure. (For simplicity, we have made a change of coordinates that shows the lowest frequency $f_0 = 0$ though this translation along the frequency axis does not invalidate any of the results.) Note that in this case, the plot reduces to three identical responses located, respectively, at the three frequencies, f_0, $2f_0$ and $-2f_0$. The results are in complete agreement with those obtained with "hardware" multiplication in the time domain, which yields a Baseband signal and a term at the $2f_0$ which is blocked off by the Low Pass Filter

Appendix C. Short tutorial on QAM

An optical carrier of frequency $f_c = \omega_c/2\pi$ modulated by a digital signal of bandwidth $B << f_c$ undergoes both, amplitude modulation A(t) and phase modulation Φ(t), and can be represented by the equation

$$S(t) = A(t)\cos[\omega_c t - \Phi(t)] \tag{C1}$$

A more useful representation is to isolate the baseband signal from the rapidly changing carrier by expanding the above expression to

$$S(t) = A(t)\cos\Phi(t)\cos(\omega_c t) + A(t)\sin\Phi(t)\sin(\omega_c t) \tag{C2}$$

that lends itself to the simple two-dimensional cartesian representation shown in Figure A8

$$I(t) = A(t)\cos\Phi(t) \tag{C3}$$

plotted on the horizontal axis and called the In-phase component, and

$$Q(t) = A(t)\sin\Phi(t) \tag{C4}$$

plotted on the vertical axis and called the Quadrature component.
Thus, the signal represented by these two orthogonal components

$$S(t) = I(t)\cos(\omega_c t) + Q(t)\sin(\omega_c t) \tag{C5}$$

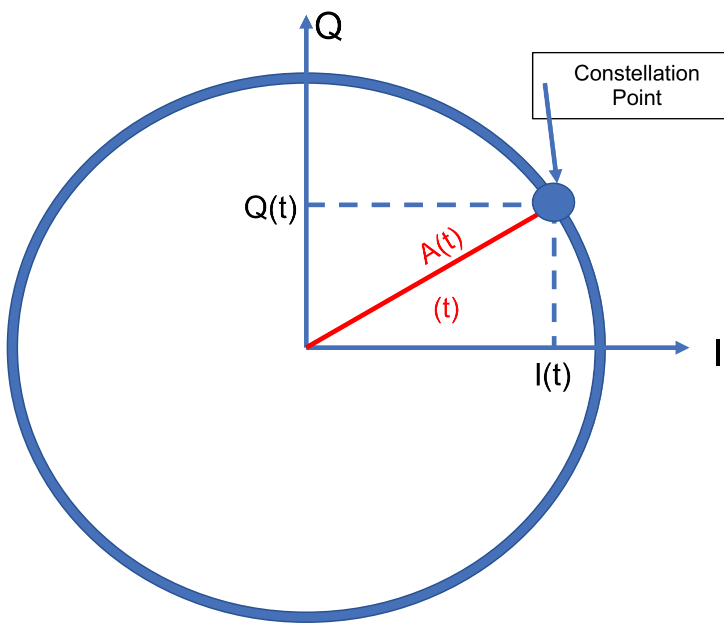


Figure A8. Example of signal constellation point for a specific value of Amplitude, A and phase.

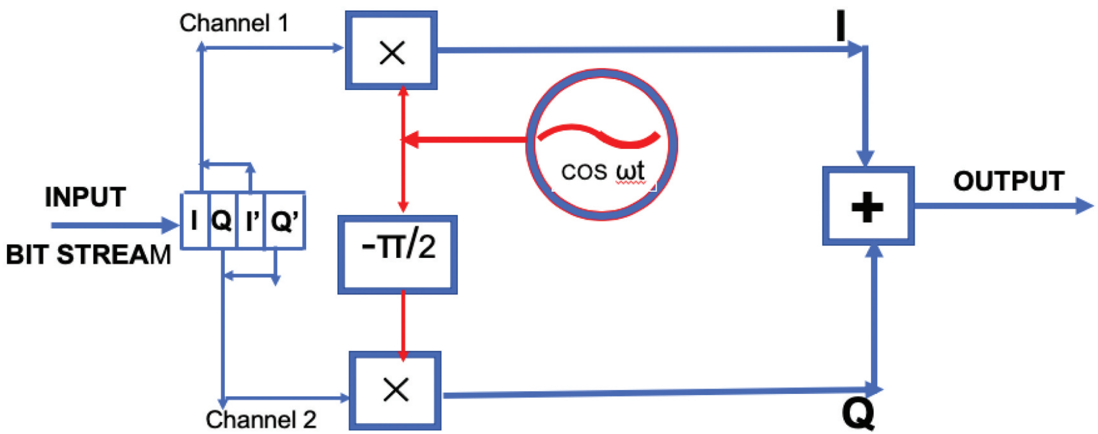


Figure A9. Typical QAM modulator showing interlacing of bits.

Table A5. Example 1: 16 QAM calculations.

A	B	C	D	E	F	G	H
	16-QAM: 8 PHASES with 2 AMPLITUDES = 16						
		PHI=45, 90, 135, 180, 225, 270, 315, 360				A= 3, 1	
PHI	COS PHI	SIN PHI	A	I = A COS PHI	Q = SIN PHI	SQRT(E*E + F*F) = A	SYMBOLS
45	0.707106781	0.707106781	3	2.121320344	2.121320344	3	0000
45	0.707106781	0.707106781	1	0.707106781	0.707106781	1	0001
90	6.12574E-17	1	3	1.83772E-16	3	3	0010
90	6.12574E-17	1	1	6.12574E-17	1	1	0011
135	-0.707106781	0.707106781	3	-2.121320344	2.121320344	3	0100
135	-0.707106781	0.707106781	1	-0.707106781	0.707106781	1	0101
180	-1	1.22515E-16	3	-3	3.67545E-16	3	0110
180	-1	1.22515E-16	1	-1	1.22515E-16	1	0111
225	-0.707106781	-0.707106781	3	-2.121320344	-2.121320344	3	1000
225	-0.707106781	-0.707106781	1	-0.707106781	-0.707106781	1	1001
270	-1.83772E-16	-1	3	-5.51317E-16	-3	3	1010
270	-1.83772E-16	-1	1	-1.83772E-16	-1	1	1011
315	0.707106781	-0.707106781	3	2.121320344	-2.121320344	3	1100
315	0.707106781	-0.707106781	1	0.707106781	-0.707106781	1	1101
360	1	-2.4503E-16	3	3	-7.35089E-16	3	1110
360	1	-2.4503E-16	1	1	-2.4503E-16	1	1111

Table A6. Example 2: 16 QAM calculations.

		16-QAM: 4-I values x 4-Q Values = 16		
I	Q	PHI = ArcTan(Q/I)	SQRT(E*E + F*F) = A	SYMBOLS
3	3	45	4.242640687	0000
1	3	71.56505118	3.16227766	0001
-1	3	108.4	3.16227766	0010
-3	3	135	4.242640687	0011
3	1	18.43494882	3.16227766	0100
1	1	45	1.414213562	0101
-1	1	135	1.414213562	0110
-3	1	161.6	3.16227766	0111
3	-1	341.6	3.16227766	1000
1	-1	315	1.414213562	1001
-1	-1	225	1.414213562	1010
-3	-1	198.4	3.16227766	1011
3	-3	315	4.242640687	1100
1	-3	288.4	3.16227766	1101
-1	-3	251.6	3.16227766	1110
-3	-3	225	4.242640687	1111

defines the signal constellation point for a specific value of A and Φ. Each pair of value is related to one of the 16 code words from 0000 to 1111 via a lookup table.

The advantage of breaking up the signal into two orthogonal components, is that being in quadrature they do not interfere with each other and as a result carry twice as much data in the same bandwidth. This orthogonal property is the same property used in WDM with different

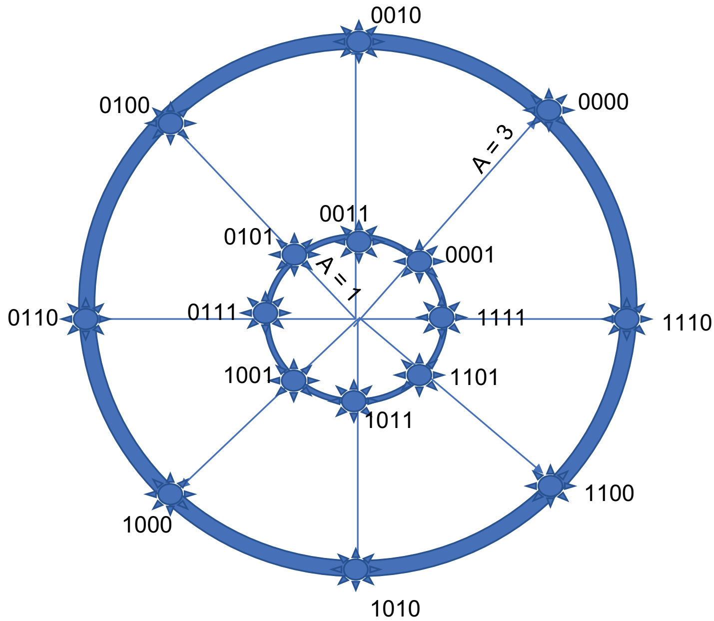


Figure A10. Example 1 16-QAM: constellation map.

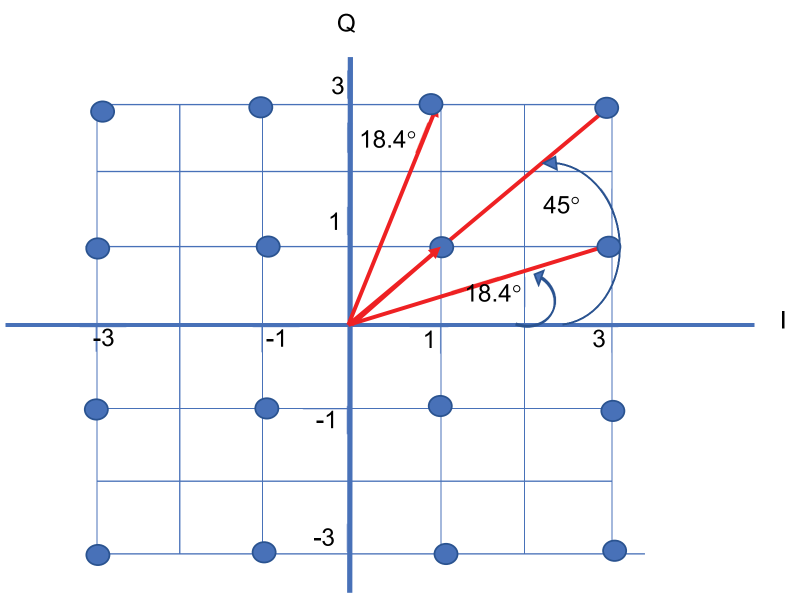


Figure A11. Example 2, 16-QAM constellation map.

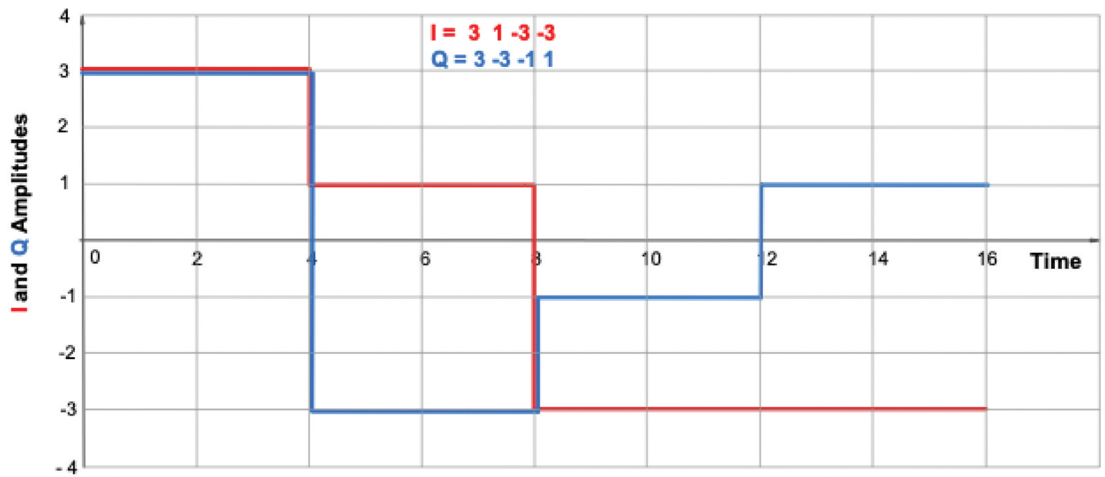


Figure A12. I and Q values for the first four symbols of the bit stream.

wavelengths or in Polarization Mode transmission with two polarized fields at right angle to each other.

Let us proceed now to describe a 16-QAM modulation and demodulation process. Consider the following bit stream (Eq. 6) describing the baseband signal, each bit occurring at a rate C bit/sec. Thus, the duration of a bit is 1/C sec. If we assign bits per symbol, we will 2^4 =16 different levels, a mix of amplitudes and phases. Each symbol will last four times as long, or 4/C and will toggle at ¼ of the bit rate C:

0000 0101 1111 0010 0110 1111 0011 1011 (C6)

Now, interlace the bits between two channels, Channel 1 and Channel 2 so as to produce symbols of 4 bits with every other bit going to one channel or the other.

Channel 1: 0000 1101 0111 0111 (C7)
Channel 2: 0011 1100 1011 0111 (C8)

The local oscillator generates the optical carrier. The In-Phase component modulates the carrier as A(t) cosΦ(t) cos(ω_ct), and the Quadrature component, which is multiplied by the cosine function shifted by π/2 modulates the carrier as A(t) sinΦ(t) sin(ω_ct).

For a 16-QAM modulator, each symbol is 4 bits long. Hence the bit stream shown in Figure A9 goes into a 4-bit shift register, and the bits are dumped alternately between Channel 1 and Channel 2 so the first symbol in Channel 1 is I I' I' I''' or 0000, and the first symbol in Channel 2 is Q Q'Q''Q'''or 0011.

The demodulation is the inverse process. The received signal given by Eq. (2) is sent through two different channels. The first channel undergoes multiplication by cos(ω_ct) followed by a low-pass filter which gets rid of all the optical frequency components: cos(ω_ct), cos(2ω_ct), sin(2ω_ct), thus yielding the baseband I(t) = A(t) cosΦ(t). The second channel performs a multiplication by sin(ω_ct) followed by a low pass filter and returns the baseband Q(t) = A(t) sinΦ(t). Both I and Q are dumped into a four-bit register that restores the original sequence of the bit stream, I, Q, I', Q', I'', Q'', I''', Q'''.

Two examples of 16-QAM constellations are givenin Figures A10 and A11. The assignment of the 16 4-bit code numbers to each of the 16 combinations of A and Φ is not optimized for minimum error probability via the so-called Gray code. The calculations and the lookup table assignments shown below serve merely to illustrate the process.

Finally, we show below (Figure A12) a plot of I and Q values for the first four sequences of symbols shown in the binary streams, Eq. (C7) and (C8) derived from the Lookup Table A6 of Example 2

Appendix D. Discrete Fourier transforms in OFDM

Discrete **F**ourier **T**ransforms (**DTF**) play a very important role in wireless transmission of signals, particularly between base stations (i.e. Cell Towers) and mobile users (i.e. smart phones). The use of **DTF** simplifies greatly transmitter and receiver implementation. With N parallel transmitted signals in parallel, as is the case, each signal modulates an **RF** (**R**adio Frequency around 2 GHz for example) subcarrier at one of the N frequencies. At the receiving end, N local oscillators at N corresponding frequencies are required for coherent detection. N can be a large number, anywhere from 48 to 1024 or higher, thus requiring such a large amount of "hardware" that renders the scheme impractical.

DTF, Discrete **F**ourier **T**ransforms, comes to the rescue by replacing this unwieldy mass of hardware with an algorithmic software. The adoption of OFDMA (Optical Frequency-Division Multiplexing Access) in 4 G LTE-Advanced wireless communications owes it primarily to **DTF**, a simple and cost-effective Access Technology.

The resultant algorithm developed to carry out the **DTF, D**iscrete **F**ourier **T**ransform is a very rapid one, computationally; it is called, **F**ast **F**ourier **T**ransform, or **FFT**. From now on, we shall use the terminology, **FFT** instead of DFT.

Refer now to Figure A13 (Figure 21 in the text), the transmitter block diagram where all the hardware has been replaced by an **FFT** processor. As shown in the diagram below, Figure A13 the signal from the N parallel outputs is band-limited, and as such can be reproduced by sampling it at the Nyquist rate equal to $2f_{fmax}$. Since f_{fmax} = B/2, that means the signal can be sampled at B rate, which corresponds to a time interval T_{OFDM}/N =1/B. Thus, N =BT_{OFDM} samples over the time symbol time interval T_{OFDM} can reproduce the original signal faithfully.

We can now express the signal in the time domain as the sum of N signals as follows:

(1) Make a change of coordinates to simplify the final expression, and labels the N signals, 0, k_1, k_2 . . . .$k_{(N-1)}$ where 0 corresponds to the RF carrier frequency f_c.
(2) Each of the N baseband signals of complex amplitude modulate a sinusoidal signal whose frequency f_k = f_c +kΔf. k is an integer ranging from 0 to (N-1), which ensures the orthogonality of the sinusoidal signals.
(3) The composite signal s(t) is written in complex form, with the understanding that the actual time signal is the real part of the complex expression below:

$$s(t) = \sum_{0}^{N-1} A_k e^{i2\Pi f_k t} \tag{D1}$$

with

$$f_k = f_c + k\Delta f \tag{D2}$$

(4) Equation (D1), with the help of (D2) can now be rewritten,

$$s(t) = \sum_{0}^{N-1} A_k e^{i2\Pi k\Delta f t} \tag{D3}$$

As stated above, the original signal is faithfully reproduced with N samples, T_{OFDM}/N apart, at succeeding times nT_{OFDM}/N, where n ranges from 0 to N-1. So, we can replace s(t) in (A3) by its n sampled values at times t = nT_{OFDM}/N,

$$S\left(nT_{OFDM}/N\right) = e^{i2\pi f_c t} \sum_{0}^{N-1} A_k e^{i2\pi k\Delta f nT}/N \tag{D4}$$

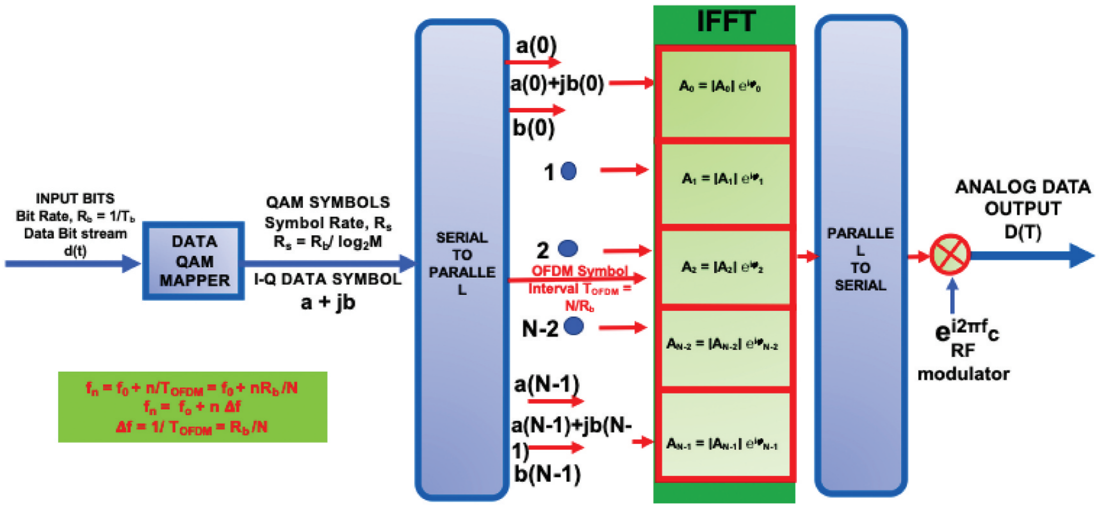


Figure A13. IFFT transmitter.

where we have dropped the subscript in T_{OFDM} for convenience. Recall from Figure 18 in the text that the subcarriers are orthogonal when they are spaced Δf apart. This follows from the fact that the unmodulated sinusoidal carriers expressed in exponential form in Equation (D1) above are orthogonal. The spectrum of modulated subcarriers of Figure 18 was obtained by having a rectangular pulse of duration T (same as T_{OFDM} since we dropped the subscript) modulating sinusoidal waves as described in Equation (D1). The resulting expression Is recognized as a Fourier Transform. Since the Fourier Transform is a linear transformation, the orthogonality of the sinusoidal carriers is preserved, if the frequency separation in Figure 18 is Δf = 1/T. If we substitute Δf = 1/T in Equation (D4) the sampled time signal becomes:

$$S\left({}^{nT_{OFDM}}/_{N}\right) = e^{i2\pi f_c t} \sum_{0}^{N-1} A_k e^{{}^{i2\pi kn}/_{N}} \tag{D5}$$

which is recognized as a discrete Fourier transform multiplied by the RF carrier f_c! The fast computational algorithm being used in the transmission is the **FFT, F**ast **F**ourier **T**ransform. Technically speaking, (D5) is an **I**nverse **F**ourier **T**ransform, or **IFFT** as referred to in the literature of wireless communications. ***Recall that a Fourier transform converts a signal from the time domain to the frequency domain, and an inverse transform, transforms*** *(as in (D5))* ***a signal from the frequency domain to the time doma*****in.**

Thus, the complex set of oscillators and multipliers has been replaced by a simple and elegant software operation, **IFFT**. It transmits and stores in N bins the value of the complex coefficients A_k from the time signal, Equation (D1) on the respective k-th subcarriers, Figure A13 (Figure 21 in text). Then it samples them making use of the orthogonality condition, Δf = 1/T and returns the inverse discrete Fourier inverse transform Equation (D5), which after parallel to serial conversion (i. e. multiplexing) delivers the time version of the modulated signals at each carrier frequency. The time signal then modulates an RF carrier for transmission through the air.

Recapping: Sampling of the band limited signal (D1) by the IFFT yields the time signal of (D5) as an inverse Fourier transform consisting of the sum of complex components in the frequency domain!

Appendix E. Peak-to-Average Power (PAPR) ratio in OFDM

For simplicity consider 3 subcarriers, like 2 harmonics of the same frequency:

$X(x) = \sin x + \sin 2x + \sin 3x = X_1 + X_2 + X_3$

Or in exponential notation

$X(x) = e^{ix} + ei^{2x} + e^{i3x}$

PAPR is defined in dB as:

$PAPR = 10Log_{10}Max[X(x).X^*(x)]/E[X(x).X^*(x)]$,

E = Expectation of X.X*, where the asterisk indicates conjugate value.

$Max[X.X^*] = 3 +$ the maximum of each of the 3 cross-products, $2 + 2 + 2 = 9$

For N subcarriers, $Max[X.X^*] = N^2$

$E[X.X^*] = E([X_1 + X_2 + X_3).(X_1^* + X_2^* + X_3^*)] = 3 +$ the expectation of the sum of cross-products is $0 = 3$

For N subcarriers, $E[X.X^*] = N$

For N subcarriers the ratio of $Max[X.X^*]/E[X.X^*] = N2/N = N$

Hence, $\mathbf{PAPR = 10Log_{10}N}$, for N subcarriers.

Appendix F. Frequency domain interpretation of SC-FDMA M-QAM symbols

The value of each symbol is broken down into its InPhase (I) and Quadrature (Q) components assigned to bins in the **FFT**. The FFT output returns both, the frequency of the I and Q components of each symbol, in terms of its amplitude and phase expressed in the complex form $Ae^{i\emptyset}$. Note that QPSK symbol amplitudes remain constant and of the same value for the duration of the SC-FDMA symbol interval, though their I and Q components do vary with time.

This is nicely illustrated in Figure A14 (Figure 24 in the text), with a simple QPSK modulation example of M = 4 Subcarriers, taken from Rumney [6]. It shows in the upper left corner a line tracing in time the succession of the 4 symbols being fed to the **FFT** after sampling the waveform at 4 times the **SC-FDMA** symbol interval

The resultant I and Q spectra as well as the spectrum of the spectral values of the amplitude and phase of the resultant complex vectors are shown in Figures A15-A17, respectively. For the reader, who can stomach a bit of math, some of the gory details are spelled out in these 3 figures! (In Figure A17, the QPSK Amplitude spectrum is shown in the middle of the figure at the bottom, and the phase spectrum at the right.)

The four Amplitude spectral lines, each in a bin of Δf = 15 kHz make up the frequency representation of the QPSK symbols Figure A18. More importantly, the four symbols, unlike OFDMA are not sent in parallel, but sequentially in time, thus avoiding the beating and resulting constructive and destructive phase interferences between the parallel transmission of the OFDMA subcarriers, each carrying a QPSK symbol. Such a phase interference is what

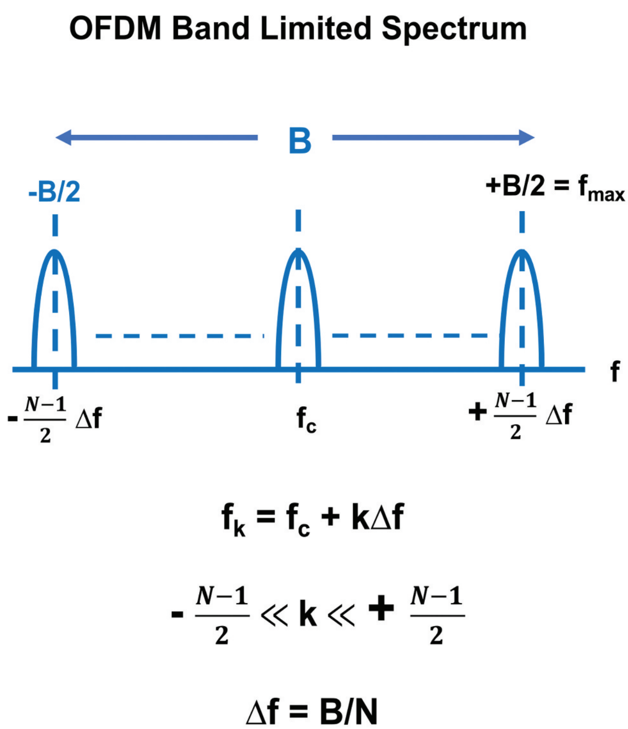


Figure A14. OFDM band limited signal

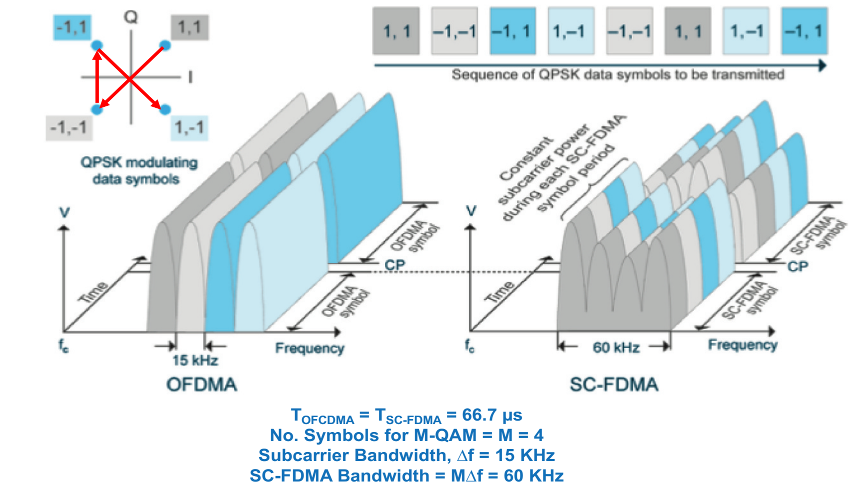
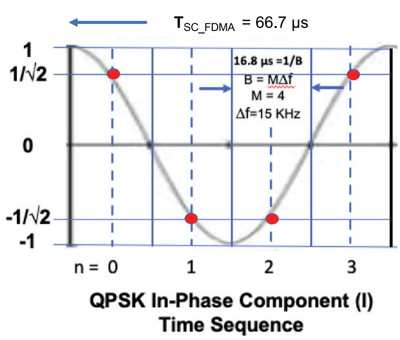
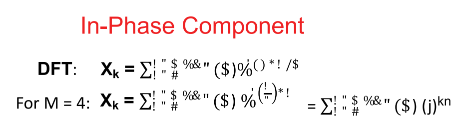


Figure A15. Symbol transmission with OFDMA AND SC-FDMA.





n / X_k	0	1	2	3	X_k	$\lvert X_k \rvert$
X_0	(1/√2)	(-1/√2)	(-1/√2)	(1/√2)	0	0
X_1	(1/√2)	(-1/√2) (+i)	(-1/√2)(-1)	(1/√2)(-i)	√2(1-i)	2
X_2	(1/√2)	(-1/√2) (-1)	(-1/√2)(+1)	(1/√2)(-1)	0	0
X_3	(1/√2)	(-1/√2) (-i)	(-1/√2)(-1)	(1/√2)(+i)	√2(1+i)	2

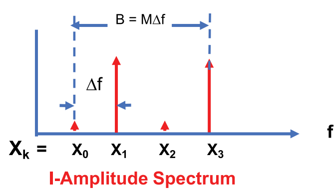
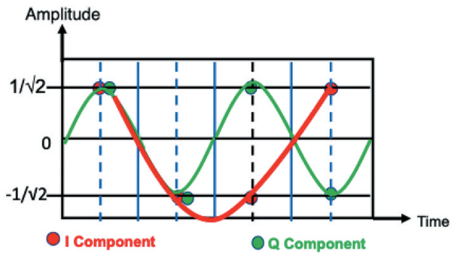


Figure A16. Amplitude frequency spectrum of QPSK I-component.

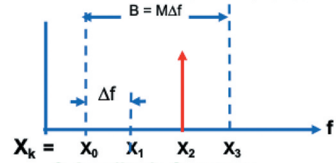
accounts for the highs and lows in signal strength that account for the resultant high **PAPR, P**eak-to-**A**verage **P**ower **R**atio of OFDMA, and is avoided in SC-FDMA



QPSK I and Q Symbol Components Time Sequence

$$\mathbf{X_k} = \sum_{n=0}^{n=M-1} x(n)e^{j2\pi kn/M}$$

$$M = 4: \quad \mathbf{X_k} = \sum_{n=0}^{n=M-1} x(n)\,(j)^{kn}$$



Q-Amplitude Spectrum

Quadrature Component

n / X_k	0	1	2	3	X_k	$\lvert X_k \rvert$
X_0	(1/√2)	(-1/√2) (1)	(1/√2)	(-1/√2) (1)	0	0
X_1	(1/√2)	(-1/√2) j	(1/√2) (-1)	(-1/√2) (-j)	0	0
X_2	(1/√2)	(-1/√2) (-1)	(1/√2) (1)	(-1/√2) (-1)	2√2	2√2
X_3	(1/√2)	(-1/√2) (-j)	(1/√2) (-1)	(-1/√2) (j)	0	0

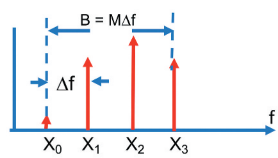
Figure A17. Amplitude frequency spectrum of QPSK Q-component.

$$\mathbf{X_k} = \sum_{n=0}^{n=M-1} x(n)e^{i2\pi kn/M} \qquad M = 4: \mathbf{X_k} = \sum_{n=0}^{n=M-1} x(n)\,e^{j\left(\frac{\pi}{2}\right)kn} = \sum_{n=0}^{n=M-1} x(n)\,(j)^{kn}$$

COMPLEX X_K SPECTRAL VALUES IN TERMS OF AMPLITUDE AND PHASE

n / X_k	0	1	2	3	X_k	$\lvert X_k \rvert$
X_0	$1\ e^{i\pi/4}$	$1\ e^{-i3\pi/4}$	$1\ e^{i3\pi/4}$	$1\ e^{-i\pi/4}$	0	0
X_1	$1\ e^{i\pi/4}$	$j\ e^{-i3\pi/4}$	$-1\ e^{i3\pi/4}$	$-j\ e^{-i\pi/4}$	√2-j√2	2
X_2	$1\ e^{i\pi/4}$	$-1\ e^{-i3\pi/4}$	$1\ e^{i3\pi/4}$	$-1\ e^{-i\pi/4}$	2√2	2√2
X_3	$1\ e^{i\pi/4}$	$-j\ e^{-i3\pi/4}$	$-1\ e^{i3\pi/4}$	$j\ e^{-i\pi/4}$	√2+j√2	2

$\phi_1 = -\pi/4$
$\phi_2 = 0$
$\phi_3 = +\pi/4$



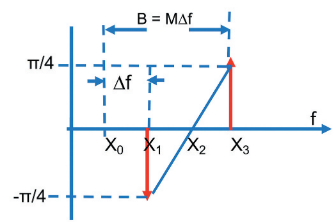


Figure A18. Spectral amplitude and phase components of QPSK complex wave.

Appendix G. How the cloud works

WHAT IS THE CLOUD?

THE CLOUD is a global network of digital resources hooked together and meant to operate as a single ecosystem

(See: https://azure.microsoft.com/en-us/overview/what-is-the-cloud/)

RESOURCES refer to digital hardware, software, and services found in data centers and computing architectures.

VIRTUALIZATION refers to combining different resources together and presenting them (a virtual instance) to the user as a representation of a physical device.

A virtual instance (functional simulation of a machine) may be configured and reconfigured dynamically without hardware intervention.

The cloud is conceptually a "pool" of resources and virtual instances providing the functions of traditional hardware and software, which can be configured dynamically without making physical changes.

Such resources and virtual instances, located anywhere in the world, are accessible *securely* through the Internet.

SUMMARIZING:

- The Cloud provides on-demand configuration and availability.
- The Cloud distributes services and functionality worldwide viat he internet.
- The Cloud provides secure remote access and increased fault tolerance while reducing costs compared to localized and/or fixed hardware platforms.