The Path Towards 6G and Digital Reality

This eBook has been produced as part of the Digital Reality Initiative of the IEEE Future Directions Committee. However, the content is of sole responsibility of the author.

Table of Contents

I had, recently, an interesting conversation with some analysts looking at the implication of 6G (yes, 6 not 5). That in itself was surprising since most of the time analysts are looking at the next quarter. Yet, they were interested on what kind of impact 6G might have on telecom operators, telecom manufacturers, and on the semiconductor industry. Of course, looking that far down the line, they were also interested in understanding what type of services may require 6G connectivity.

I started the conversation saying that 6G does not exist (and I guess most of you would agree), but then I revealed that it was already here, in terms of "prodrome". Using this word may suggest that I see 6G as a disease, but that is not completely the case. In other words, looking at the past evolution and at the present situation, it may be possible to detect a few signs that can be used to make some predictions about 6G.

I touched on the following aspects:

- Lessons from the "G" evolution
- Spectrum efficiency
- Spectrum availability
- Processing capacity in devices
- Power requirement
- Network architecture
- Services that may require/benefit from 6G

The area of services is intertwined with the shift towards Digital Reality that has started, but is still far from being completed. Activities, publications, and events organized by the IEEE Digital Reality Initiative, brought to you by the IEEE Future Directions Committee, are centered around this transition (related to the execution of the Digital Transformation and to the rise of ambient intelligence). I'll discuss the various aspects in this eBook, and at the end, provide some thoughts on their implication pertaining to the path towards Digital Reality.

1. Lessons from “G” Evolution

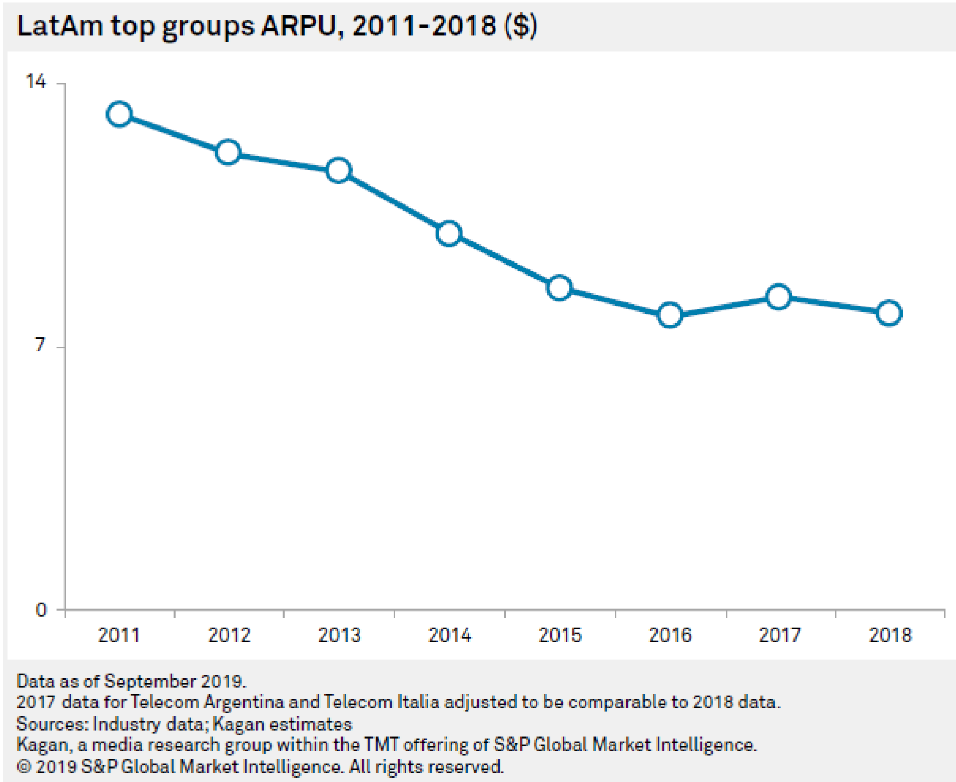

Figure 1. The Average Revenue per User (ARPU) in LATAM, as well as in many other areas, has been declining in spite of an increase in network use, data transfer, and an increase in network performances. Image credit: Global Market Intelligence, 2019

If you look back starting from 1G, each subsequent “G” (up to the 4th G) was often the result of technology evolution and/or the need of Wireless Telecom Operators in order to meet a growing demand. The market was expanding (more users/cellphones), and more network equipment was needed. Having a new technology that could decrease the per-element cost (with respect to capacity) was a great incentive to move from one “G” to the next. Additionally, the expansion of the market resulted in an increase of revenues.

The capital expenditures (CapEx) used to pay for expanding the network (for example, base stations and antennas) can be recovered, in a relatively short time, thanks to an expanding market even though the Average Revenue per User, or ARPU, was decreasing). Additionally, the operating expenses (OPEX) were also decreasing (again, measured against capacity).

The expanding market meant more handsets were sold, and the increase in production volumes resulted in a decreased price. More than that, the expanding market fueled innovation in handsets, with new models stimulating the top buyers to upgrade, and attracted new buyers with lower-cost older models. All in all, a virtual spiral has increased sales and increased the attractiveness of the wireless services (the “me too” effect).

It is in this “ensemble” that we can find the reason for the 10-year generation cycle. After ten years, a new G arrives on the market. New technology supports this, and equipment manufacturers (network and device) and telecom operators ride (and push) the wave for economic reasons.

How is it that an exponential technology evolution does not result in an exponential deceleration of a previous G in favor of the new one? Why is the10-year generation cycle so stable?

There are a few reasons why:

- The exponential technology evolution does not result in an exponential market adoption
- The market perception of “novelty” is logarithmic (you need something that is 10 times more performant in order for consumers to perceive it even just 2 times better), hence the logarithmic perception combined with an exponential evolution leads to a linear adoption
- New technology flanks existing ones (we still have 2G around even as 5G is starting to be deployed)

With the advent of 4G, the landscape has changed. In many Countries the market has saturated, the space for expansion has dwindled, and there is only replacement. Also, the coverage provided by the network has reached, in most places, 100% (or at least 100% of the area that is of interest to users). A new generation will typically cover a smaller area and expand over time. Hence the market (us as consumers) will stick to the previous generation since it is already available everywhere. This has a negative implication on the operators as the new generation is rarely appealing enough to sustain and justify a premium price.

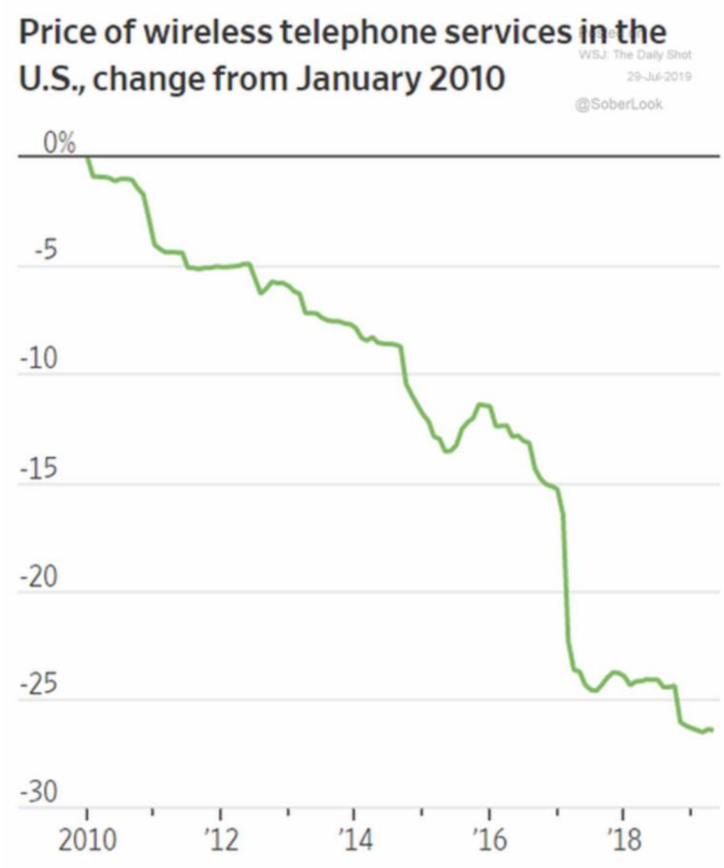

Figure 2. The price of wireless services has declined everywhere in the last twenty years. The graphic shows the decline in the US over the last ten years. Image credit: Bureau of Labor Statistics

An Operator will need to invest money to deploy the new "G," but revenues will not increase. If that is the case, why would an Operator do that? Well, because they have no choice. The new generation provides better performances and lower OPEX. If an Operator does not deploy the new "G", someone else will, which will attract customers and enable the network to run at a lower cost. This will allow industry competitors to offer lower prices that will undercut other offers.

5G is a clear example of this situation, and there is no reason to believe that 6G will be any different. As a matter of fact, the more availability and accessibility of a given G, the less the market is likely to pay a premium for the new G (for example, 4G is sufficient for users in most situations). By 2030, 5G will be fully deployed and people will get capacity and performance capabilities that will exceed their (wildest) needs.

Having a 6G network to provide 100 Gbp, versus 1 Gbps provided by 5G networks, is unlikely to entice customers to pay a premium. What is more likely to happen is that the "cost" of the new network will have to be "paid" by services, not by connectivity. This opens up quite a different scenario.

2. Spectrum Efficiency

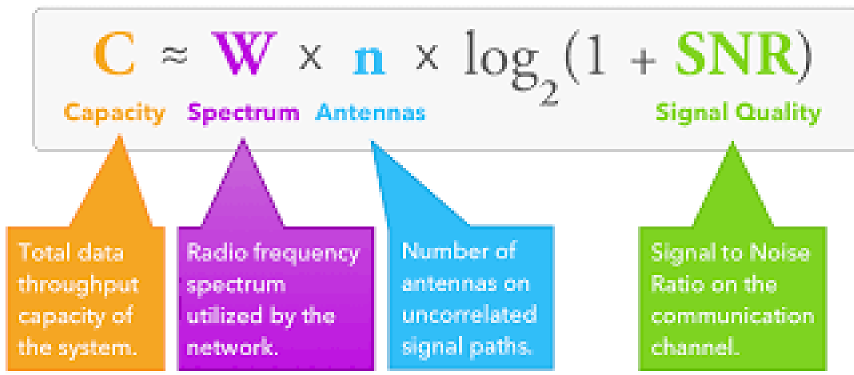

Figure 3. The Shannon theorem, expanded, to take into account the use of several antennas. In the graphic W stands for the spectrum band, B in the original Shannon theorem, and SNR for the Signal Noise ratio. Image credit: Waveform

Over the last 40 years, since the very first analogue wireless systems, researchers have managed to increase the spectrum efficiency, which means they can pack more and more information in the radio waves. Actually, with 4G, they have reached the Shannon limit. Claude Shannon, an American mathematician and electrical engineer, found a relation between signal power and the noise on a channel that limited the capacity. Shannon's findings stemmed from research and ideas that originated from Harry Nyquist and Ralph Hartly. Exceeding the Shannon limit will produce errors. For example, the signal will no longer be useful as you'll no longer be able to distinguish the signal from the noise):

$$C=B\log(1+S/N)$$

Where C is the theoretically available channel capacity (in bit/s), B is the spectrum band in Hz, S is the Signal power in W, and N is the Noise power in W).

Since the spectral efficiency is a function of the signal power, you cannot give an absolute number to it. By increasing the signal power you can overcome noise, hence packing more bits per Hz. In practice, you have some limit to the power dictated by the regulation (max V per meter allowed), the kind of noise in the transmission channel (extremely low for optical fiber, and much higher from wireless users in an urban area, and even higher in a factory), as well as the use and reliability of battery power.

Today, in normal usage conditions and with the best wireless system, the Shannon limit for a modern wireless system is around 4 bits per Hz (in other words, for every available Hz in the spectrum range allocated to a wireless transmission, you can squeeze in 4 bits. Due to the complexity of environmental conditions, you can find numbers from 0.5 to 13 in spectral efficiency).What I am indicating is a “compromise” just to give an idea of where we are. A standard 3G system may have 1 bit per Hz in spectral efficiency, whereas standard 4G reaches 2.5, and with QAM 64, spectrum efficiency can reach 4.

This limit has already been overcome using “tricks”, such as higher order modulation (like QAM 256 reaching 6.3 bit per Hz), and most importantly using Multiple Input Multiple Output (MIMO).

The latter is a really nice way to circumvent the Shannon limit. This limit is pertaining to the use of a single channel. Of course if you use more channels, you can increase the number of bits per Hz, as long as these channels do not interfere with one another. This is actually the key point! By using several antennas, in theory, I could create multiple channels—one for each antenna couple (transmitting and receiving). However, these parallel transmissions, using the same frequency and spectrum band, will be interfering with one another.

Here’s the good news: “interference” does not exist! Interference is not a property of waves. Waves do not interfere, and therefore, if a wave meets another wave, each one will continue undisturbed and unaffected. What actually happens is that an observer will no longer be able to distinguish one wave from the other at the point where they intersect. Therefore, the interference is a problem for the detector, not the waves. You can easily visualize this if you look at a calm sea where you may notice small waves in one area, and flat patches in another. These are areas where waves overlap and annihilate one another (a crest of one adds to the trough of the other, resulting in a flat area). If you have “n” transmitting antennas, “n+1” receiving antennas, and each antenna is separated from the others by at least half a wavelength, then you can distinguish the interference and obtain the signal. This is basically the principle of MIMO. To exploit it you would need sufficient processing power in order to manage all signals received in parallel by the antennas, and this is something that will be addressed in a future publication. For now, it is good to know that there is a way to circumvent the Shannon limit and expand the capacity of a wireless system.

6G will not just exploit massive MIMO, it will be able to do something amazing: spread signal processing across many devices, each one acting as an array of antennas. Rather than having a single access point in 6G, in theory at least, you can have an unlimited number of access points, thus multiplying the overall capacity. It would be like sending mail to many receivers. You may have a bottleneck at one point, but the messages will get to other points that, in turn, will be able to relay them to the intended receiver once available.

3. Spectrum Availability

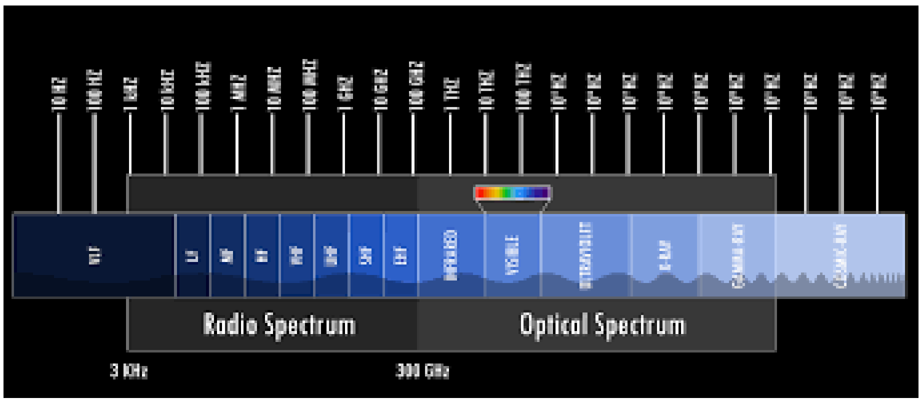

Figure 4. An overall view of the spectrum. In the picture the radio spectrum is going up to the 300 of Gbps, above that, in the THz starts the optical spectrum. Image credit: NASA

Electromagnetic fields are pervasive, and it can take a long time to identify them. In 1819, Ørsted (a Danish power company) was most likely the first to detect that something was going on. They noticed that placing a wire within the vicinity of a compass had an effect. A few years later, Faraday started to grasp what was going on, but it was Maxwell (in 1861/62) who laid the theoretical foundation of electromagnetic fields (by the way, implicitly determining that all electromagnetic fields propagate at a constant speed independently of the reference frame, and that light is an electromagnetic field!). For a crash course on Maxwell's equations, take a look at this.

An electromagnetic field is characterized by a frequency (f – the number of oscillations per second). An equivalent way of characterizing an electromagnetic field is by its wavelength (λ – the spatial distance between two crests). Multiplying the two, you get the speed of light (c):

$$\lambda f = c$$

Since the speed of light is constant, when you talk about the frequency of an electromagnetic field you are also talking about its wavelength. Physicists dealing with optics prefer to talk about wavelengths, and those dealing with electricity prefer to talk about frequency, however, it is exactly the same thing. Engineers are usually talking about frequency. This is what they do when talking about wireless systems and the spectrum used (range of frequencies).

It is important to understand that there is a continuum in the electromagnetic field frequencies. We use, as an example, 50 (or 60) Hz in powering our home (if the frequency is 0 you have a static electromagnetic field, also known as Direct Current, or DC).

We can generate an electromagnetic field of a given frequency using an oscillator. For example, an antenna is an oscillator that propagates a field in the air. An antenna doubles back as a detector of an electromagnetic field, converting it into a (tiny) electrical current oscillating at the same frequency. An electronic circuit can amplify this “signal” and process it.

This is the first important point when using electromagnetic fields in communications: you need an electronic circuit that can generate a field at the desired frequency while simultaneously “processing” the frequency. The higher the frequency, the trickier the electronic circuit. That is the reason why the evolution of electronics (Moore’s law) has made it possible to deal with rapidly growing frequencies. The ones that will be used for 5G (28/75 GHz, also called mm waves) require a more sophisticated use of electronics that was unavailable (at an affordable cost) just 10 years ago. However, now 5G deployments are using lower frequencies, comparable to the ones used by 4G. With the expected evolution in the next ten years, researchers are confident that it will be feasible to handle higher frequencies, over 100 GHz and up to a THz (1,000 GHz). Hence the expectation that 6G will be able to leverage higher frequencies. Since they have not been used before, they are available for new applications and a new spectrum for free!

Higher frequencies allow you to pack more bits per second, as explained in the previous sections. However, the higher the frequency, the bigger the propagation issues. Intuitively, the higher the frequency, the shorter the wavelength, and this means that smaller obstacles will stop the propagation. You can observe this in the aforementioned sea wave’s example. If a wave hits a small rock, it will go around it and continue undisturbed. On the other hand, if the rock is big (like an island), the waves will be blocked, resulting in a calm sea on the other side of the obstacle. In radio-communications there is a soft spot balancing propagation with bit carrying capacity that is between the 900 MHz and 3 GHz. If you go below this, you cannot carry much data, but if you go above that, propagation constraints will require the use of smaller and smaller cells (and that implies higher deployment cost).

Radio communications are based on electromagnetic fields, as are optical communications, and we are actually using a range of nearby frequencies, frequency bands, or "spectrums". As an example, in Italy, 5G was given the 3.6-3.8 GHz band, split in two slots of 80MHz and 2 slots of 20MHz (80+80+20+20= 200 =3,800-3600). Now, if you take the big slot, 80MHz, and you squeeze 6 bits per Hz, you get a capacity of 480 Mbps, which is much lower than the Gbps claimed in marketing campaigns. To reach that sort of capacity you would need to use mm waves, i.e. use the 26-28GHz, or higher frequencies, where more spectrum can be allocated. The problem, as previously mentioned, is that at those frequencies propagation will have negative effects, and you need to use very small cells (in the order of a hundred meters, versus km in a 4G network).

Propagation and cell size are the second crucial aspects as we increase frequency. For example, in 6G, talking about THz means a large capacity, but incredibly small cells that are measured in meters! A classic network based on cells would not work from an economic affordability standpoint. What would be required is a paradigm change with communication taking place among "users" most of the time, and only once in a while some of these users' devices will connect to the "big network." You no longer have fixed cells, but a mesh of moving cells continuously interacting with one another and creating a communication fabric that will connect when needed to the communication infrastructure. This will be addressed in a later section.

What is also clear is that the management of higher frequencies, and multiple bands, places very demanding requirements on the transmission points and, in particular, on the terminals.

4. Processing Capacity in the Devices

Figure 5. A graphic from the past. I discovered it browsing on slides describing the evolution of wireless systems. It goes back to 2013, at a time when we spoke about 4G as the best system ever. The point that struck me was the power of the terminal we keep in our hand. And, of course, it got much more powerful in the past 7 years, really opening the door to 6G. Image credit: Qualcomm

Addressing spectrum efficiency and availability in the previous posts, I continuously mentioned the need for processing power to both exploit the spectrum (efficiency), as well as to take advantage of available ones. For the former, the transmitting/receiving device needs to use multiple antennas and combine/process the signals received (or the ones that need to be sent). For the latter, the available spectrum lies in higher frequencies (the “good ones” are already taken), and you need more processing speed to tackle them. Additionally, there is an increased need to manage multiple bands in parallel, and this multiplies the need of processing capacity.

There are two parties involved in the communication process: the base station and the device (such as a cell phone, tablet, hotspot, etc.). The base stations are controlled by large corporations, the Operators, that in principle can afford to invest money for advanced processing capabilities (although they are also very frugal). Additionally, the base station and related antenna are industrial grade equipment not constrained by size, and therefore are able to use bulky heat dissipation systems. The user device is quite a different story as it must be affordable to the average user, it has to be small in size, and it cannot depreciate too quickly. For example, consumers will be highly unsatisfied if they will need to recharge the battery every half hour! Due to higher constraints for the device, this is one feature we have to analyze in order to understand the (possible) evolution.

Figure 6. An early “mobile phone”. This device was more like a suitcase than the small, slick phone we are accustomed to fitting in our pocket/purse. The big box contained both the electronics and the battery, both pretty bulky! Image credit: eBuyer

This crucial role of the “terminal” is not something we discovered recently. It has been this way since the very beginning.

Today’s 4G phones would have required a truck to carry them and their power supply if they were produced using the same electronics that were available 30 years ago. For example, the first 3G “phone” that I saw was actually a pile of racks carried by a van.

It is the evolution of electronics (Moore’s law) both in terms of performance and reducing the power consumption required for current wireless systems.

Power consumption and performance do not go hand in hand, even though we have seen the former decrease and the latter increase in almost perfect “sync”. It is a matter of compromise; Apple decided to deliver the iPhone 11 without 5G due to the lack of 5G coverage and because the 5G chips accessible last year depleted battery life much faster. In just one year, the situation has improved, and now you can obtain 5G chips that require a comparative consumption of power to a 4G chip.

Managing the lower spectrum bands (up to 3.8 GHz) and the higher -mm waves- spectrum bands (25-75 GHz) is not possible at this time due to cost and power constraints. The 5G smartphones that are on the market operate in the same spectrum bands of the 4G. In other words, from an engineering point of view, they cannot deliver the hyped Gbps speed. However, they can from the marketing point of view!

If we cannot produce a fully compliant 5G phone at this point in time, image the challenges in making a 6G phone with 10 times higher spectrum bands and massive MIMO, plus supporting all the node functionalities, in addition to the terminal functionalities. However, in ten years’ time, we can expect that electronics will be able to support (at least part of these) requirements. For a full analysis and understanding of how researchers hypothesize 6G, it may likely take an additional 15 years.

The reason why researchers can be confident about their predictions is due to the evolution records of terminals. Some of the smartphones currently on the market can control the NASA Apollo program — imagine controlling the LEM, a large Saturn rocket, with your phone! It is more nimble than the Cray-2 supercomputer of the 1980's, and it is even faster than the computer used to control the Orion spaceship, which was used by NASA for the Mars mission. Amazingly, all of this performance rests in the palm of your hand and costs a few hundred bucks.

6G is likely to require a hundred times more processing capacity than smartphones available today, and even more critical, it will require significantly enhanced batteries as it will be using them for a highly accelerated amount of time. In order to establish a local processing fabric, 6G will require the capability to communicate with other devices, even when not in use.

5. Power Requirement

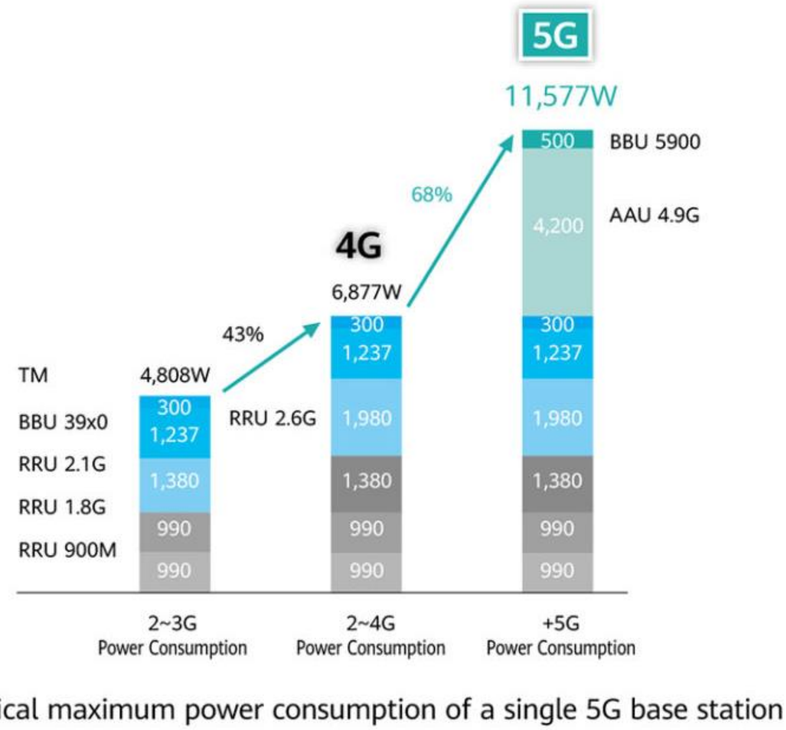

Figure 7. Maximum power consumption of a base station in a 3G (left) 4G (center) and 5G (right) network. Notice the sharp increase as more and more "intelligence" is integrated into the base station. The sharp increase in 5G is due to the intelligent antennas (AAU). Image credit: Huawei

All of these enhanced processing abilities on the device side go hand in hand with the increase of power demand, yet such an increase, as pointed out in a previous section, is not manageable because increased power means:

- increased dissipation in the form of heat (second law of thermodynamics);
- increased draining of the battery

A couple of potential approaches to resolve this issue include the manufacturing of more efficient chips, and decreasing the size of transistors (and thus the possibility of using lower voltage). Another possible solution is to program computer chips so that only specific parts of the chip are active when needed in order for the device to work. Keeping the performance of silicon at a lower level (not increasing the clock rate), and using more silicon (multi core), are additional enhancements that have made it possible to balance increased performances with lower (stable) power consumption.

Clearly, these are not specific issues for mobile devices, they apply to all chips. However, a major problem pertaining to mobile devices is that they rely on battery, and therefore, device manufacturers need to ensure at least 18 hours between recharges, keep size and weight within reasonable limits (actually this is both an ergonomic and a marketing requirement since bulky and heavy phones would likely be difficult to sell), and they cannot use active dissipation systems (like a fan).

A lot of research has gone into designing more efficient chips for smartphones, and today a Snapdragon chip, or the Apple A14 Bionic (used in smartphones and tablets), delivers the same processing capacity with half the power consumption of equivalent chips used in laptops, although at a higher cost. The signal processing function of the chips have adopted very sophisticated architectures and power saving tactics that have somehow balanced the increasing demand of advanced signal processing required by new wireless systems (MIMO, multi frequency bands).

6G will require further sophistication in processing, something that today's technology is unable to support (at price and performance point needed in the consumer market). Interestingly, the new chips are now embedding specific AI circuitry to support native AI processing. This will be crucial to support semantic-based communications at the core of 6G. In addition, the shift from silicon to graphene, or other single atom layer materials, like molybdenum disulfide, will also help support the increased capacity requirements.

Another thing to analyze is the device-side power consumption, as well as how the power is needed and used by a device like a smartphone. If 100, the following are power requirements to download data (figures refer to HSDPA, high-speed 3G):

- 85 is the amount needed for WiFi data transmission
- 50 for managing SMS
- 48 for making a voice call
- 32 for playing an MP3 file (music)
- 18 to illuminate the smart phone screen

So, in other words, data communications are by far the most power-demanding activity on a smartphone. However, in terms total power consumption, the situation changes significantly since data transmission is just a tiny fraction of the use of a smartphone: most time is spent on running apps, and some apps, such as video games, are very power intensive.

In a 6G scenario, we can expect that in relative terms, the data transmission part will remain the most expensive one. However, we will have to separate the low-range data communications that are using the THz band from the one using the GHz. The latter will compare with power requirements of today's 5G, although the signal processing involving multi-frequency bands will be more power-hungry than today. The former, using THz, will be using significantly less power in terms of transmission, due to the ultra-small cells, but a bit more for signal processing. Overall, THz transmission will be cheaper in terms of power used than today's 5G.

However, as is now true for 4G and even more so with 5G, the variety of services supported will include a few that will require quite a bit of in-device processing, and that will increase the use of power. I will address this later when discussing 6G services.

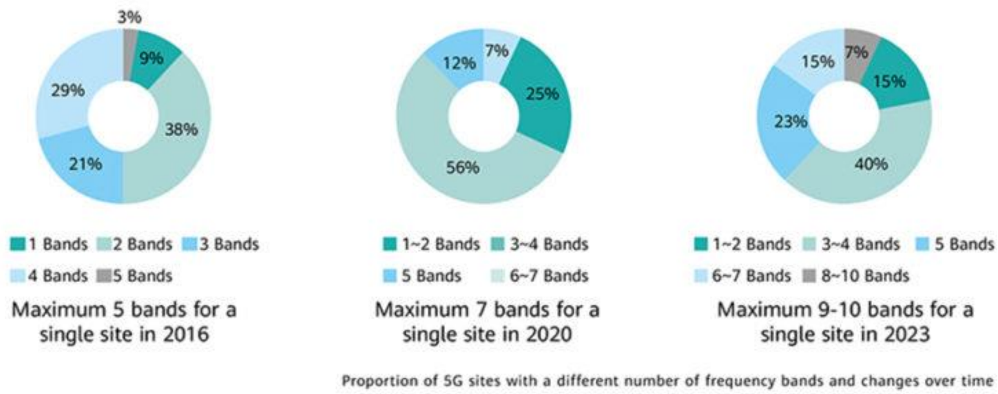

Figure 8. The adoption of additional frequency bands in each area covered by a base station will increase in the next few years, leading to an increase of power consumption. Image credit: Huawei

Power issues are not limited to devices, they also affect wireless network operational costs. As explained in the previous section, we have seen an increasing demand of power in base stations as we transitioned to more advanced cellular systems. Increased bandwidth, MIMO, and smart antennas all require more power. The decrease in power resulting from the use of optical fiber to replace copper in connecting the towers, base stations, and the core network is not enough to offset the increased demand of the other components). This power-demand increase is expected to continue as the deployment of more sophisticated edge networks (see the next section in this series) will take place in the coming years.

One additional reason the base station requires an increase in power is the need to manage more frequency bands.

As shown in Figure 8, the expectation is that over the next 4 years (by 2023), base stations managing 5 frequency bands or more will grow from 19% to 37% (they were only 3% in 2016). These figures are even more significant if you consider the total expected increase in the number of base stations within the next 3 years. The small-cell market growth in the upcoming 4 years has an estimated CAGR of 23% in the US, and this is similar to the growth in networks that will be deploying 5G).

6G will further expand the number of frequency bands, and this will further increase power consumption. Additionally, as I will discuss in the next section, we are going to see a growing intelligence at the edges, with massively distributed data bases/data centers, and computing at the edges. Add to this, actually, “multiply this” with the increased number of “base stations, antennas, nodes” that will be required by a 6G communications fabric, and you start to get an idea of the power consumption problem.

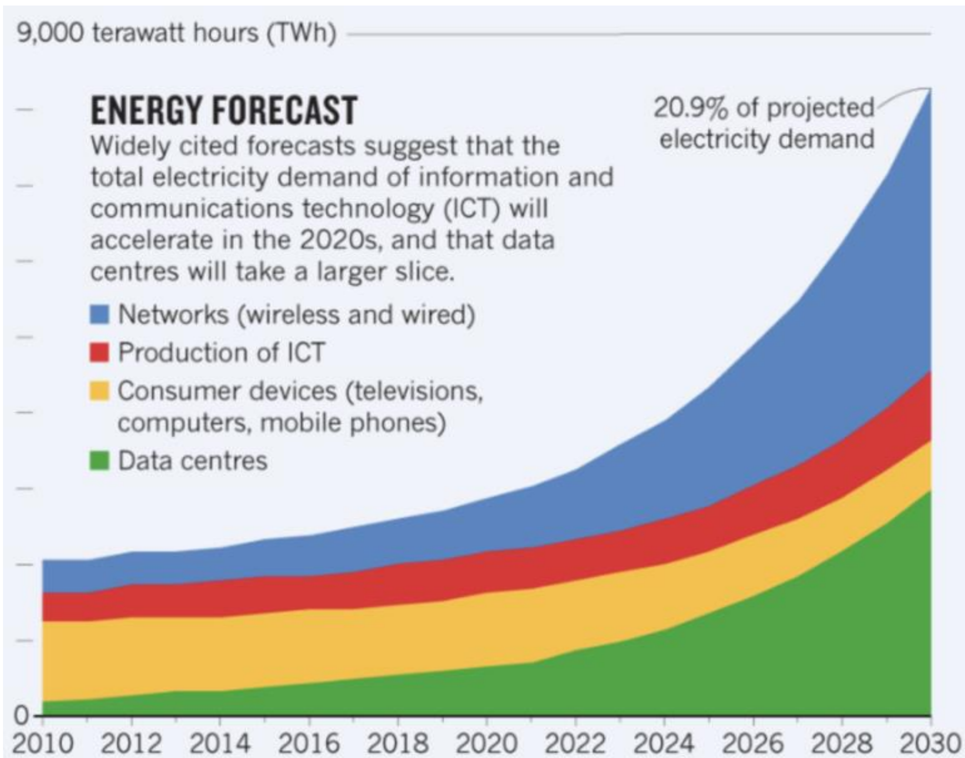

Figure 9. The forecasted increase of electrical power demands in the ICT areas. As shown in the graphics, the main contributors to the growth are the communications networks and the data centers. Image credit: IEA

The power demand for ICT is expected to reach one-fifth of the overall electrical power-consumption worldwide by 2030 – 3 times more than its level in 2010. In absolute terms, the growth is even larger considering that the power-demand is (slowly) shifting from fossil fuel to electricity (mostly due to a shift in electrical transport).

Looking at the graphics in Figure 9 above, one can see the impressive share of power being used by the communication networks (both wireline and wireless), and by data centers. The latter are the ones that are seeing the highest increase.

We can expect this trend to continue beyond 2030, with 6G giving further impulse to electrical power use at the edges (processing, data storage and transmission).

I am not actually confident about the data on consumer devices showing a decreased demand. As I observed in Italy, there has been an increased demand in this sector from almost zero in 1990, to some 3 TWh in 2020. During this time, the network power-consumption, for one of the largest telecom operators in Italy, has remained stable at 2TWh (that Operator has halved due to competition and the number of clients, but it has increased the network traffic by an order of magnitude).

For an in depth analysis and forecast of future power-demand in communication networks, and the analysis of its various components, take a look at this article. Independently of the various sources, all seem to agree that in the future data centers will be the main user of power in this area.

6. Network Architecture

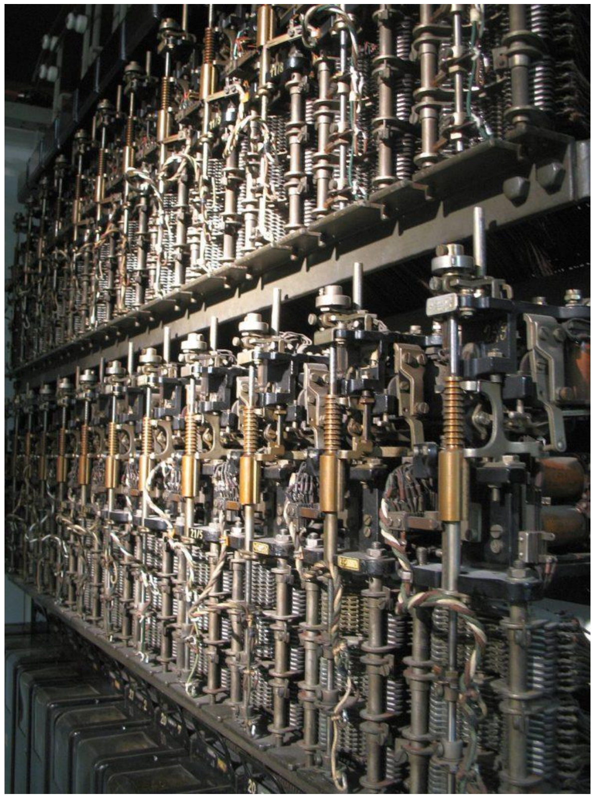

Figure 10. An electromechanical exchange of the past, using Strowger switches for automatic routing of calls. Image credit: 99percentinvisible

The telecommunication network, seen from a mile above, has kept evolving towards a flatter and flatter architecture. It was very hierarchical in the beginning; a call from Siena to Philadelphia would have to be routed via Florence, Rome, London, and then New York in order to reach its destination in Philadelphia. All routing was pre-cabled, and mechanical routers (see the photo) with no intelligence had to route the call through subsequent steps, which is why the calls were unable to go directly to the intended contact.

This hierarchy started to flatten with the advent of computers, the so called "Intelligent Network". Now, a computer looks at what resources (wires) are available and instructs the network to allocate a path to that specific contact. At that point, a call from Siena to Philadelphia could follow one route, and the next could follow a completely different one.

Wireless networks (networks supporting wireless terminations) appeared at a later point in time when computers were already widely used and, therefore, were born with much flatter and flexible architectures.

Actually, the subsequent "G" saw an ever increasing adoption of software and computers. The 2G, also known in Europe as GSM (Group Special Mobile), was dubbed by technicians as the "Great Software Monster" to emphasize the big role software played, which we can put in perspective if we compare the 4/5 G use of software to the one in the GSM!

The Internet and IP (Internet Protocol) were co-opted in 3G, and of course this is an integral part of 4/5G and, no question about it, will be part of 6G.

It was not just the network that became computerized and leveraged software. Devices, and in particular cellphones (now smartphones), have also become computers that run software. Interestingly, smartphones have been progressively taking the upper-hand in supporting services. As smartphones become more involved in the support and delivery of services, this progressively leaves the (Operators') network migrating into the Cloud. This is providing flexibility and is decoupling the network from services. Additionally, it is also shifting ownership of services (and related revenues) from telecom Operators to third parties.

This evolution is particularly evident in the wireless network (systems) in research pertaining to a service network and an infrastructure network (the one comprising wires, base stations, antennas…).

Actually, the "service network" is completely separate from the infrastructure network, and it is completely flat. Services are created independently of the infrastructure, and if sufficient performance is available, it can work on any infrastructure (any "G" can support any service).

This is great for the users, but a nightmare for the Telcom industry because:

- Users can access services wherever they are, independently of the network available in a particular location, and will therefore only perceive the service and not the network. As long as the performances are sufficient, a user will be happy.
- Service creators and providers design the services in such a way so that they will work on a broad range of networks, and so that they can operate a wide range of performances. Since they are making money from the use of the services, the more people that can use them, the more the revenues.
- A new G, such as 5G, can deliver enhanced performances, however, service providers are unlikely to create services that require higher performances since this would restrict the number of users that could leverage the service. It is better for them to create services that can run on Gs and smartphones that are widely available.
- Telcos are facing the problem of deploying a new G as it requires a premium price, and there are basically no services that would make an evident difference, which can hinder the return on investment. For example, 5G: telling a user that they would be able to

download a movie in a few seconds while on the move, versus several minutes with 4G, would not make sense being that streaming services are designed to work perfectly well within 4G performances.

The separation of the service network from the physical (classical) infrastructure has reached a point of no return since the service network is actually enabled by an intermediate layer—the service enabler layer (cloud, servers, and much more) which will be addressed later.

In the past, an Operator could design services to leverage new network features and thus provide an incentive to customers to buy onto their new network, even if it was in the earlier stage of deployment. Now, the service providers fueling the service network do not have any interest in investing to promote a new network (a new "G") until that becomes widespread and can sustain a large market.

In spite of all this progress, we are still far from a really flat network infrastructure and from a completely flexible one, so there is still plenty of space for further evolution.

All things considered, it may be unsurprising that devices such as smartphones are driving service evolution. What is surprising is that they are also driving the "physical network" evolution!

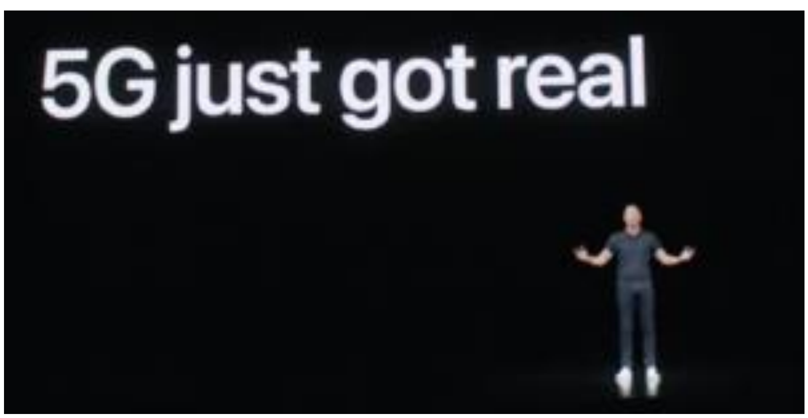

Figure 11. The thought provoking statement during a presentation from Apple, of the new iPhone, by the Chairman of Verizon: iPhone 12 makes 5G "real"! Image taken from the Apple iPhone 12 event

Watching the iPhone 12 presentation, I noticed the statement of Hans Vestberg, Chairman and CEO of Verizon, stating that "5G just got real" because of the advent of the iPhone 12. It is the device that makes 5G a reality, not the other way round—the 5G network that allows the development of new devices!

This is not really new. For those that have been working for decades in telecommunications, the shift took place at least 10 years ago. In the last century, device producers looked at the Network Operators (and manufacturers) to understand what the future network could provide, and based on that, they design their products. Now, it is the other way round. Network Operators look at device manufacturers to understand what kind of devices will be on the market (and their penetration), and based on those details they plan for new network deployments. Seen in this light, it may not be surprising that the Chair of Verizon considers the launch of the iPhone 12 (and the expected uptake from the market) as the sign of a real advent of 5G. It is no longer the dog wagging the tail, it is the tail wagging the dog.

It is not just a market issue, it is a technical issue. In the following example, I refer to smartphones. However, it is more general and applies to all communication devices such as vehicles, robots, etc. Smartphones are convenient since there are so many of them. They create "industrial volume" which allows their chips to power other devices. Smartphones are embedding network functionalities such as coding, access network selection, and more will be embedded in the future.

As networks are becoming increasingly "softwarized", both in terms of functionalities and control, it is obvious that a smartphone can pick up more and more functionalities that used to be confined to the network.

Telecom Operators have always been striving for optimization of resources (to decrease CAPEX), and software/computers have been providing tremendous support. Additionally, the possibility of automating network operations and management through software (thus decreasing OPEX) has resulted in self-configuration and self-management. Network equipment has become more intelligent, and the network architecture has evolved (and is evolving) to take advantage of this.

Of course, telecommunication networks are complex systems comprising hundreds of thousands of parts, and we can find new, fully computerized parts along old electromechanical ones. In the last decade, researchers and telecom engineers have found a way to cope with this diversity (up to a certain extent) by increasingly virtualizing functions, thus being able to operate on a network composed of software. Specific interfaces (actuators) will take care of execution at the physical level, taking into account the variety of equipment. This approach is known as Network Function Virtualization (NFV).

As network resources are virtualized, they are basically computers with software implementing the desired functionality (as is the case with the latest network equipment). It then becomes possible to operate the entire network at a virtual level, in the cyberspace (this is the DX applied to network infrastructure). We can determine what resources are available and how to use them in the most effective way, taking into account the demand for transport (and the need of services). In other words, we can define the network via software: SDN - Software Defined Network.

Clearly, operating in the cyberspace provides many advantages, like a global visibility of resources and all the flexibility in the world (the cyberspace is "flat"), thus decreasing both CAPEX (better use of resources means that you need less of them) and OPEX (many activities can be performed automatically and autonomously by the virtual network). Obviously, the intelligence of the resulting network is further enhanced by applying machine learning and, in general AI. Machine learning can "learn" how the resources are being requested, analyze and record traffic and service patterns, and act accordingly in their reservation and assignment.

Network equipment used to be “smarter” in the core and "dumber" at the edges. The reason is simple: smarter equipment used to cost a lot of money, dumber ones were cheaper. Hence, since you have fewer equipment in the core and more at the edges, it made sense to make the core equipment smarter and have it to take care of the "dumber" edge.

However, the situation is now different. The decrease in costs of electronics and computers make smarter edge affordable, so now it is no longer true that the edge is necessarily dumb—it can participate in the global intelligence of the network. SDN can therefore be read in a different way, rather than Software Defined Network (with the Telco using it to allocate resources). It can become Software Defined Networking, i.e. the edge of the network defines how to use the resources. The core, in this sense, becomes a type of warehouse for resources —which one to use and how to connect them can be decided by the edge. We are taking a step in the 6G direction.

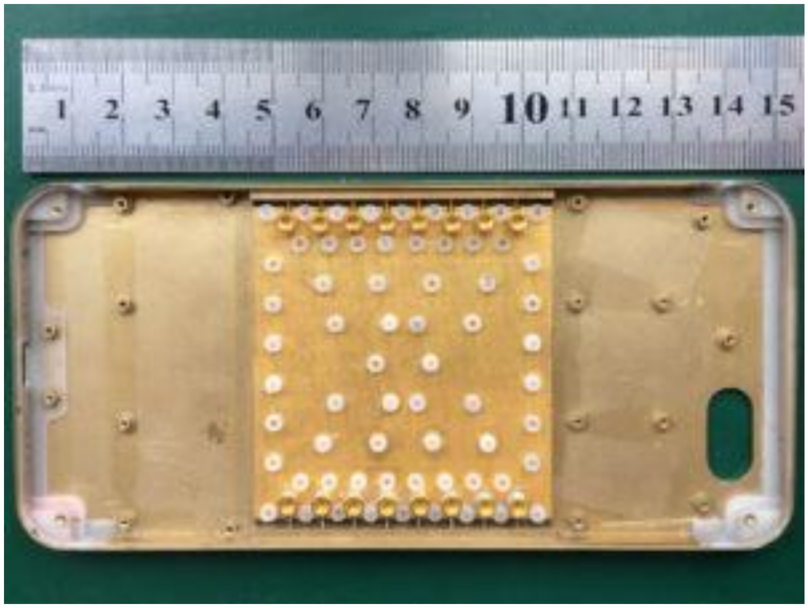

Figure 12. A prototype of phased-array antennas for mm wave communications fitting inside a smartphone. Each of the dots is an antenna, Electronic circuits steer the electrical signal to one or more of these elemental antennas, forming multiple beams that sustain higher bandwidth communications. Image credit: Shanghai University and IEEE

Wireless networks are everything but wireless. The "core" part is basically undistinguishable from the wireline network. However, their edges are quite different, of course, composed of base stations, poles, antennas … but still very "fixed" indeed. For many years, these edges have been "passive", dumb. Their role was to convert the radio signal into an electrical current that could flow on a wire (or into photons that can flow in a fiber), and the other way round.

In the last few years, antennas have become progressively smart—able to manage several radio communications in parallel. Phased-array antennas, see the photo above, enable lower power emission (since the radio beam is focused on the receiving antenna, rather than being spread around), and higher bandwidth (since several communication beams can be used in parallel - MIMO). This is part of the 4G network, and an essential component of the 5G network. You can bet that 6G will make use of even more sophisticated antennas.

In addition, the edge is now slowly being transformed into an active part of the network with the ultimate goal of becoming a network by itself, connected to other networks, one of those being the core network. The base station is becoming a power horse, encompassing a growing amount of storage and processing power, so big in fact that it can be mistaken for a small cloud.

Indeed, the edge evolution is moving towards the support of cloud at the edge (starting to morph into a fog), and computing at the edge. This is in the 5G architecture and it will be clearly part of the 6G.

However, with 6G, the edge will extend to include the devices be it smartphones, robots, vehicles, or … "people". We have the potential of seeing this starting in 5G in the second part of this decade.

The need for an intelligent edge in 5G has been proven by several trials[1]. Very low latency can only be achieved if the communication is managed by the edge, as you cannot rely on the core network. This is the case when vehicle to vehicle communications are needed for collision avoidance at an intersection (incoming vehicles have to coordinate the order in which they cross the intersection).

With edge cloud and edge computing, a possible evolution of the 5G in the second part of this decade, and a sure thing for 6G, we start to lose the concept of a core terminated by an edge. Rather, we are shifting towards a network architecture where edges are just other networks. If this reminds you of the Internet architecture, that's good because this is a similar case. Networks can be added and they interplay with one another, mostly trying to manage all traffic internally and diverting the traffic to other "contiguous" networks when the traffic is not intended to remain local. This requires intelligence, and as a matter of fact, it also requires awareness of the context. It is no longer a predefined structure where it is known from the start what resources will be involved and how to seize them, rather it becomes a dynamic allocation that changes depending on the originator and on the context at that particular time. Sometimes, this is also known as a "semantic" network, because the network needs to understand the meaning of that communication "session" and work out a way to accommodate it.

OK, it is now time to consider a "real" wireless network!

Smartphones and complex IoTs (or aggregates of IoTs such as vehicles) have been addressed as "terminals". In fact, a few years have passed, and in several occasions, this name no longer applies. Think of the tethering supported by your smartphone, think of the WiFi network in more modern cars, or an Alexa, AppleTV—all of them are devices that double up as a Network Node. You'll find the basic functions of routing, channel selection, data storage, and processing that we are accustomed to seeing in network nodes. Your smartphones act as a network node and access gateway for your computer. Your vehicle WiFi is a (local) network supporting devices like your phone, which is seen as a server, enabling it to provide access to the data it stores, as

[1] https://enterpriseiotinsights.com/20190322/channels/news/edge-is-the-only-way-to-make-5g-v2x-safe

well as to some of the applications that were downloaded on it. Alexa connects and orchestrates a variety of IoT in your home (Alexa, switch on the light…).

This transition from being a "terminal" to becoming a network node is a crucial and disrupting evolution paving the way to 6G, and it is already here!

Actually, in the 5G architecture (not yet implemented, although we start to see the first signs in the "private 5G"), there is the possibility of handing off the session control to the device. At this point, it is no longer a "terminal" but a network node). This means that the device, for example your smartphone, can request and negotiate network resources to set up a communication session, and can change these resources and their aggregation through the session as new needs arise. This is not possible in the 4G architecture, but it becomes possible in the 5G. The examples I gave before are possible using WiFi, but when using the Operator's network this will not be possible, which is the reason why my car requires me to switch on the iPhone WiFi in order to manage the iPhone as a server, despite it being equipped with AppleCar.

The shift from "terminal" to "network node" kills the last hierarchy remaining in the telecommunications network. A smartphone that is a network node can use SDN (Software Defined Networking) to set up the chain/pool of resources required. While this is great from the application point of view, it is the final nightmare for the Telcos that have already seen their networks become "dumb pipes" and now might find out they will no longer provide connectivity services (the only ones left), rather becoming a commodity warehouse of a bunch of network equipment on shelves that can be picked up on demand by (application) users, and that of course will be in competition with other third-party network equipment (included those provided via NFV). Network functions such as authentication, end to end security, store and forward… can (and are starting to) drift away from the network (soft SIM, biometric identification...) and be provided by third parties.

Of course once we shift the focus from the network edge (base stations, poles, antennas, etc.) to the "terminals", we are stepping into a real wireless network—one that moves around on wheels (and legs)!

Moving around requires a completely different approach to the management of communications resources. This is the area of mesh networks, well explored in the military sector where the goal is to create a reliable, self-configurable network out of thousands of devices (that may be parachuted from above, as an example). This is an area where Small World theory, autonomous systems, and the underlying Artificial Intelligence are leading the evolution (and there is plenty of research needed). These three areas also require plenty of processing power and low power communications to sustain massively distributed processing.

Smartphones, vehicles, and IoTs will become network nodes that can create a network bottom-up (eventually, when needed, connecting to the big pipes and reserving the resources required). Incidentally, this is the way Nature creates communications infrastructures among a pride of lions, a flock of sheep, and a swarm of bees!

This will become a technical possibility (as mentioned on a limited scale it is already here), and even more important, it will be a feasible approach to 6G from an economic standpoint, shifting the investment from Telcos to the edges (and to consumer electronics, i.e. to each of us).

An interesting twist brought by a network stemming from the edges, and with terminals that double up into network nodes, is the change of the traffic capacity paradigm. In today's networks, the more terminals that are connected, the lower the capacity available to each of them (the capacity depends on the base station/frequency, and spectrum available, and this capacity has to be divided by the number of users at a given time). In a 6G network, it is the other way around. Every additional device that can act as a network node increases the overall capacity, so the more users, the more capacity!

7. Services that may Require/Benefit from 6G

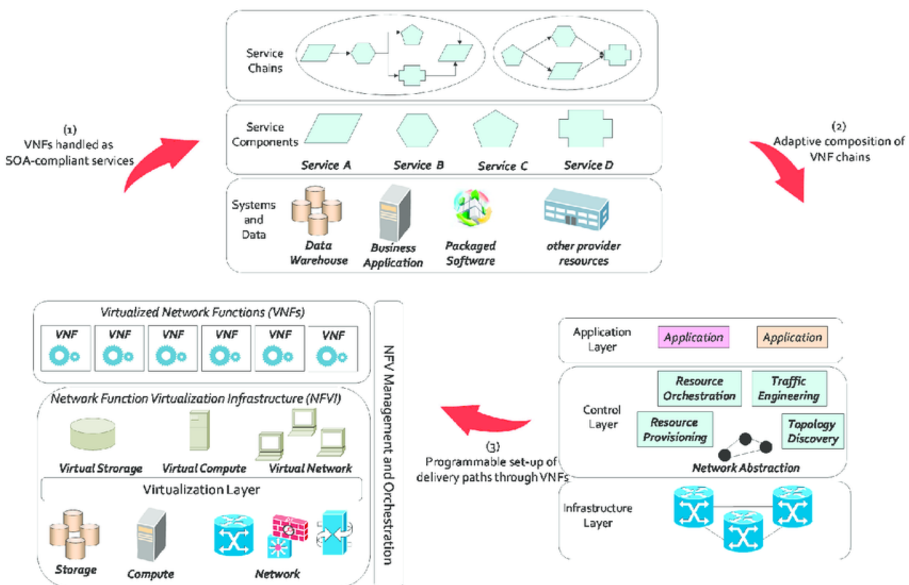

Figure 13. Service Oriented Architecture -SOA- may be seen as a pre-cursor of the future. Network Service-based evolution. Image credit: Federica Paganelli, Barbara Martini

Every time a new wireless system is envisioned, the possibilities on what that new system can mean begin to evolve. Amazing new services are foreseen, a good portion of them just inherited from the hype on the previous system, and a few that are new and unlikely to materialize once the new system, eventually, is deployed.

As previously mentioned, one reason is that the service providers, those that are creating and deploying services, are targeting the market that can access the available infrastructures, not the niches that a new system can support (initially). Another reason is that if a service is indeed “amazing” (which means the service has a market and customers willing to pay), you can bet that companies will try their best to deploy it on the existing infrastructures, using tricks and shortcuts.

As seen in the last fifteen years, a leap in service quality and variety is not related to a new wireless system, rather to the appearance and availability of new devices.

Hence, rather than reveling on the huge bandwidth that 6G might provide and its extremely low latency (in the part where local peer-to-peer communications are involved), we should better take a look at the expected evolution of devices such as smartphones, wearables, vehicles, and IoTs—a kind of exercise that requires a crystal ball, and goes along with the same accuracy you can expect from a crystal ball.

There is also another reason to take this device-centric approach, and that is the topsy-turvy creation, deployment, and growth of 6G as I see it from today’s standpoint.

We have been living in a service network-based evolution, meaning that services have been evolving in sync with the network availability (and evolution). In ten years’ time, in addition to the 6G paradigm and architecture, we are likely to shift in a network service-based evolution where the network evolution will be driven (or “created”) by the new services.

Once we accept the idea that a terminal becomes a network node, and that its behavior/functionality will be software-based, it becomes obvious that new “terminals” will create “new” networks, and the overall network creation and evolution will be a consequence of their aggregation.

It will no longer be a quest to find, given some (increased) network performances, a service that can become the silver bullet to exploit them and stimulate market adoption, rather a variety of services, locally taking advantage of technology evolution, will be creating local, expanding networks, which eventually will reshape the overall network.

Services will require new devices to support and make them desirable (user friendly, seamlessly pervading daily processes in life and in business) and affordable (the device cost/price will have to match market possibility).

As a result, it makes more sense to look at device roadmaps than network roadmaps. It also means that the big players will come from the industry of “manufacturing tools” and from consumer electronics, not from network equipment manufacturers or Telecom Operators.

Is this totally new, an evolution coming out of the blue? Not at all! We have been talking about Software Defined Network, and even Service Defined Network for quite a while. The idea was to use the softwarization of current networks (NFV, SDN) to carve a network slice that would fit at best a given service requirement.

The difference is that here we were using an existing network, in the future the very creation of a service (along with the deployment of devices supporting it) will create the network (lead to the evolution of the network).

Smartphones

Figure 14. A variety of ideas have been floating around on the possible evolution of smartphones, from becoming (literally) transparent, thin and flexible to the point of folding or rolling up, to being incorporated into a wearable that can be embedded into the body. Image credit: NewSoftwaresNet

If past experience is of any indication, 6G will reach the peak of hype by 2030 and will become a widespread reality by 2035. By 2030, we should have the iPhone 22 and Galaxy S30, shouldn’t we?

Well, had you asked me approximately 10 years ago if by 2030 we will still have a cellphone, I would have said “no, that thin brick will no longer be part of our daily life. It will have been replaced by wearable and embedded connectivity.”

I was wrong. The cellphone has already disappeared, possibly starting to fade away 5 years ago, and now it is completely gone.

Since the iPhone 5 (and related Galaxy), the launch events have been focusing on several characteristics, but never mention that the new phone was indeed a “phone”—able to make calls. That part has simply faded away. This year’s launch was an exception since there was the

5G to use as a marketing lever (according to Shelley Palmer, at the moment, “5G is 95% marketing hype, and 5% marketing hype!”[2]). Even then, the “call” feature was completely neglected, and instead focused on the possibility of downloading a movie on the go in just a couple of seconds (why you may want to do that remains a mystery…).

We can expect that, as the 5G hype fades away, future product launches will again forget about the “communications” aspects of the phone.

There have been, in the past decade, many suggestions from designers and futurists on the evolution of the smartphone. Designers tend to imagine the use of smart materials that make different form factors possible (for example, transparent objects—crystalline batteries may indeed support transparency —or very thin, foldable, rollable devices…). The latter tend to think of the future phone as embedded in wearable devices and, eventually, in the body with no need for a physical interface. A voice on one side and smart glasses on the other would provide all the I/O functionality required.

While the former designs may become reality, I doubt the latter. Not because of technology shortcoming, but because of the need to “sell” some physical object to the market (to make money).

Therefore, I would bet that in the next decade we will still be buying some sort of smartphone that will operate as a gateway to a variety of interfaces, and as a platform to software and data (both personal and global data). It will play the role of an authentication device, a gateway to the body network, and a network node/gateway to local and global networks. However, it will remain a physical object with a design that will “sell” it over and over.

The role of orchestrator/network node/gateway to the body network and to the local network will be based on the 6G and on the intelligence that goes along with it.

Most likely, the smartphone will keep increasing its “capability” in term of:

- Processing power: it is reasonable to expect at least 13 times more power by 2030, and 50 times more by 2050;
- Storage capacity: low-end models may have 1 TB, and high-end 10 TB capacity;
- Autonomy: most likely, the issue of autonomy will be less crucial as wireless charging will be the norm, and there will be plenty of wireless chargers in the ambient space so that smartphones will be charging all day long;
- Display: although a screen will likely remain an integral part of the phone for marketing purposes, the smartphone will make use and coordinate the display on a variety of

[2] https://www.shellypalmer.com/2020/10/get-iphone-12/

ambient surfaces (car windshields, walls, table tops, etc.) and “body” screens (for example smart glasses, wearables, and much further down the line, embedded displays);

- Intelligence: Artificial Intelligence (AI) will be embedded in the smartphone and will cooperate with Intelligence providers (we are starting to see the first example with Google and Facebook).
    - Personal Digital Twins will be hosted on the smartphone as well as in cyberspace;
    - Augmented Reality will be managed in the smartphone (intelligence), also mediated by the Personal Digital Twin;
    - Virtual Reality will also be orchestrated by the smartphone, and delivered through specific devices.
- Orchestration: the smartphone will create communication environments connecting the cyberspace to the physical space, under the mediation of the Personal Digital Twins and Ambient Digital Twins.

Whether you want to consider this a smartphone or not, it is probably a matter of name. As a matter of fact, the name doesn’t fit well with the present smartphone if you think about the way we are using it. Most of the time, it is not for making or receiving calls, it is a personal computer that can, among many other things, make and receive calls. We can make calls from our car and from our computer, yet we don't call them "phones". This will be an even more inaccurate name in the future. The key property of the future smartphone will be its personal relation with me (you), much more of a personal extension of myself and an augmenter of my capabilities. In this sense, 6G will support services that extend my capabilities from talking and understanding different languages, to navigating new and complex problem spaces thanks to artificial intelligence and seamless communications with my senses (and brain).

Wearables

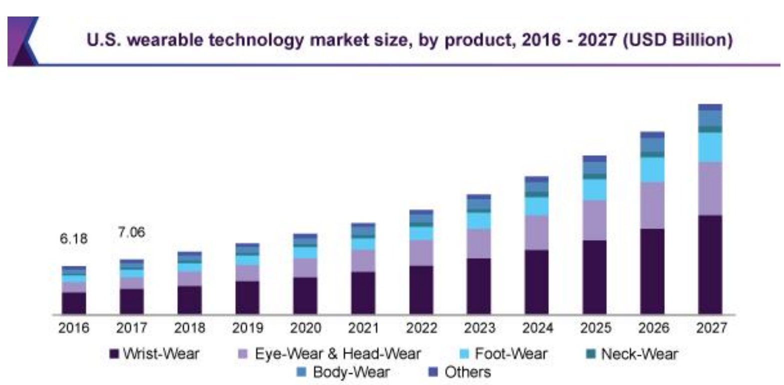

Figure 15. The US expected the growth of the wearable market from approximately 10 B$ in 2020 to double by 2027. Image credit: GrandViewResearch

Another interesting area of evolution that can influence the design, deployment, and uptake of 6G is the one of wearables.

Wearables have been around for over a decade in the consumer market as smart watches, fitness sensors, and in industrial settings as tools for specific needs in the production/operation process.

Due to the need to keep consumption down (particularly in the consumer market), wearables are seen as add-ons to other devices (smartphones in the consumer market), acting as sensors. Communications take place (usually) via Bluetooth at a frequency of 2.4GHz. The use of low energy Bluetooth, BLE, is usually the domain of IoT, although it has started to be used in wearables as well.

The wearable market is expected to grow significantly, doubling in the US between 2020 and 2027 (see figure 15), and tripling worldwide from 32.63 B$ in 2019, to over 100 B$ in 2027 at a CAGR of 15.6%. This growth is likely to fuel the adoption of new technology and increase the functionality.

5G is limited to the mm waves, as they are in the range between 30GHz and 300GHz. However, current deployment of 5G uses ranges very similar to that of 4G frequencies (in the 3GH range), so public 5G is not using mm waves. This is now starting to occur in the private 5G deployed in industrial environments, and 6G will use (also) sub-mm waves (between 300GHz - 3THz). At this frequency, propagation is very limited, but this is not an issue for wearables requiring a connection with a hub (such as a smartphone) just a meter away, or with another wearable nearby in a Body Area Network—BAN. Currently, BAN are used in hospital environments to connect sensors on patients, and are based on the IEEE 802.15.6 standard (a variation of WiFi suitable for low power transmission) operating in the 2.4 GHz band.

The evolution of electronics will make the use of sub-mm waves possible and convenient, and thus personalized healthcare is an area where 6G can provide services. This is in sync with the evolution towards proactive and personalized healthcare (actually, accelerated by the pandemic by increasing the need to use remote healthcare to off-load the burden of hospitals, and for almost real-time monitoring of patients at home).

Indeed, healthcare services may be expected to use (and steer the evolution of) wearables and create a fertile market for 6G.

Notice that the 6G will be created at local levels by the wearable (or by the hub to which wearables connect), it will not be deployed necessarily by third-parties. In other words, the wearables will be creating a local 6G network, not the other way around.

In addition to healthcare, we can expect wearable screens—first in the shape of glasses, and in the distant future, in the form of electronic contact lenses (I'll consider implants in the IoT discussion). This will also stimulate new services, particularly in the area of augmented reality to morph into digital reality.

In the industrial environment we can expect increased use of wearables as a way to create a context for the interplay of human and autonomous systems (autonomous robots). One way of seeing them is as a massively distributed intelligent sensors network. This is likely another important area for 6G growth.

Vehicles

The automotive industry, including truck manufacturing and operation, is undergoing two major, converging changes that will completely modify the road/transportation landscape in the next decade, impacting people's perception of value and the way they'll look at private transportation.

On one hand, we are seeing a shift from internal combustion engines to electric engines, and on the other, we are seeing an evolution towards self-driving vehicles:

- Electric vehicles require a completely different "refueling" infrastructure, on-board intelligence to use it, and the ability to coordinate with other vehicles the access to refueling points. The grid also requires evolution to match the increasing demand and its distribution. In addition, of course, they change the manufacturing process, use fewer components, and require a different supply chain involving a different set of providers.
- Self-driving vehicles require full awareness of the context and predictive intelligence. Part of this derives from real-time, low latency communication with the environment and with other vehicles. Both on-board intelligence and distributed intelligence are required for safe operation, and for decreasing manufacturing and operating costs. In addition, self-driving vehicles change the perception of value for the end-users that will shift from ownership of a product to the ownership of a service (a transportation means).

As you can infer from the previous points, there are common characteristics in both evolution, intelligence, and communications (both involving autonomy).

5G is potentially serving part of the need, but it would need to be ubiquitous and ensure low latency. The former will require time, at least until the middle of this decade, and the latter requires an edge architecture that can handle local communications, thus insuring single digit latency.

6G will be designed with a bottom-up architecture based on autonomous edges and autonomous nodes (terminals), both with local intelligence and ready to form a distributed intelligence. Additionally, peer-to-peer communications based on mm and sub-mm waves will be managed, thus supporting direct vehicle-to-vehicle communications at close range.

The possibility to create (mesh) networks at the edges is an important plus. Vehicles will become network nodes and will be able, in urban environments, to create mesh networks.

Notice that vehicles today have embraced both the possibility to use passengers' (and the drivers) smartphones as a gateway, and the embedding of a "phone" with its own SIM card to support vehicle to public-network communications. Normally, the vehicle is not supporting real-time, on-board sensor connectivity. Most brands prefer to accumulate data locally and relay them asynchronously, and sometimes a record is stored in the electronic "key" and used during maintenance procedures. Additionally, there is a clear separation between communications supporting entertainment, and communications for the vehicle operation. This is most likely to remain a requirement in the future to ensure safe operation and encapsulate critical transactions. For example, you might be annoyed by someone hacking into your stereo, but you definitely do not want a hacker taking control of your engine.

In the long term, the availability of a communications platform at the edges, formed by vehicles, will support the delivery of stored, forwarded, and community services by several third-parties. 6G will likely be an ideal fabric for this type of service offer.

Internet of Things (IoT)

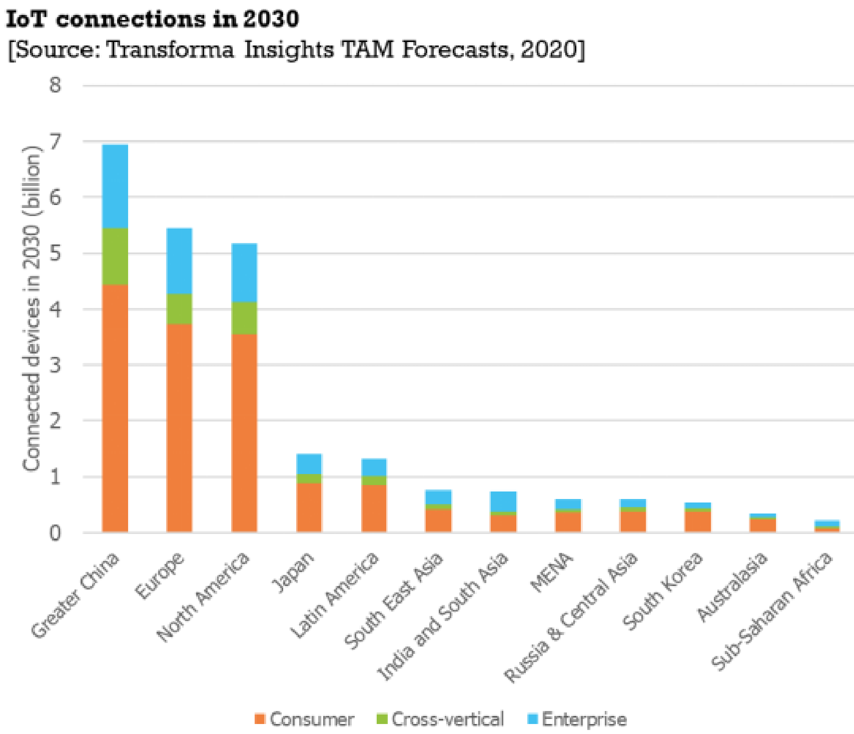

Figure 16. An estimate of IoT-connected devices, by 2030, segmented by region and sector. Notice that there are several other estimates showing significantly different numbers. Image credit: Transforma Insights 2020

Internet of Things, IoT, has an ubiquitous presence in today's world. Google it and you will get numbers ranging from 20 billion to over 50 billion as of 2020. A 30 billion difference in the estimate is no peanut, and this reflects both the difficulty in counting them, as well as the fuzziness in their definition.

When you say that a "Thing" is a part of the IoT, it implies that it is connected to the Internet. However, how do you define "connected"? Does it need to have a direct link and a unique address? Or will a unique identity be sufficient? Is a device such as a smartphone an IoT? It has connectivity, of course, and it has a unique address. What about the accelerometer in the smartphone? Or the phones microphone, temperature sensor, or the compass? They all are connected to the internet, although via the smartphone, and they can all be uniquely identified. We have over 3 billion active smartphones in the world as of 2020, each of them with at least 10 sensors. Should these be counted as IoT? That will make for over 30 billion IoT just by themselves…

As you see the landscape is fuzzy, but the overall message is clear: IoTs are playing a significant role in todays (and even more so, tomorrow's) network and service areas. What is notable in IoT is the huge variety in terms of communication requirements. You have billions of them that periodically send a few bytes, other billions that are streaming continuously in high-def video, billions that can wait for quite a bit of time to send their data, and others that require an immediate response.

There are IoTs that connect directly to the network (and to a data processing center), others that process data locally and only send polished up information, and plenty in between that form local clusters where raw data is processed.

There are IoTs engaged in one-way communications (such as sensors), and others that are entertaining two-way communications, partly asynchronously and partly via hand-shaking.

5G, with its mm waves, can use very small antennas (the size of an antenna is related to the wavelength it needs to manage), hence it can be ideal for small size IoTs. 6G, with a sub-mm wavelength, can be even better! Additionally, the smaller the cell, the less power required for data transmission—an important bonus in many IoT applications. Notice that 4G, and of course later evolutions, support native IP, thus reducing the need to process a more complex stack, hence decreasing power demand on the IoT side.

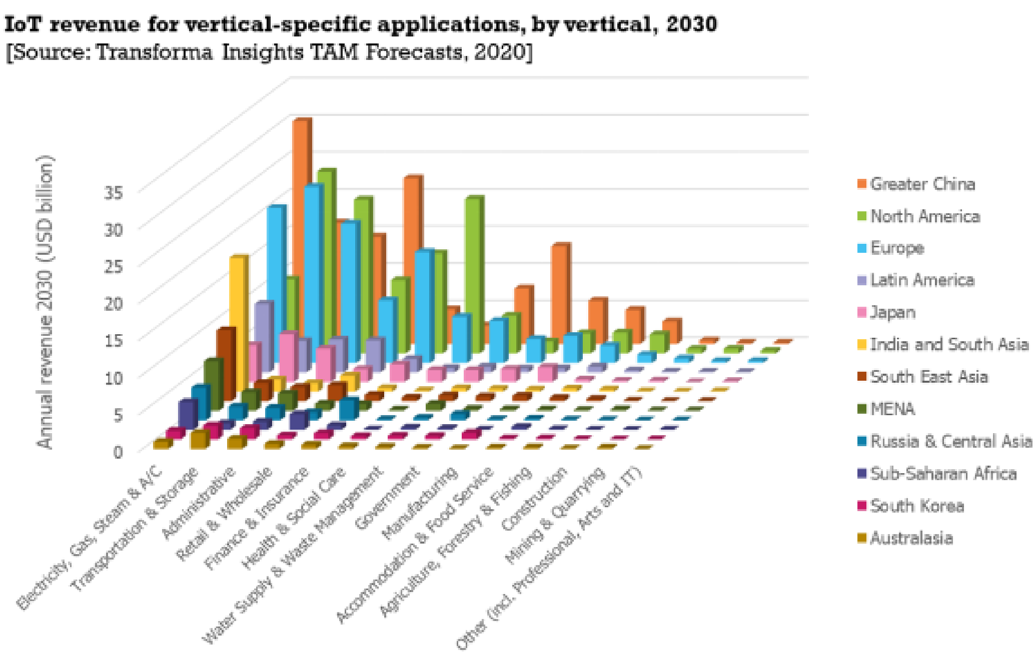

Figure 17. Expected revenues by 2030 of IoT devices segmented by region, and sector. Image credit: Transforma Insights 2020

IoT will keep growing, becoming a "dense" presence in any ambient. Some of them (a significant portion) will be embedded into products (as they are today in smartphones and in vehicles), and the overall trend is towards open-data framework and public API, allowing third-party services to access the data "created/harvested" by IoT. With 6G, one can assume that part of these IoTs will become points (sources) forming a massively distributed data center at the edges (ambient data center). Processing-power and local artificial intelligence are likely to create ambient awareness and support fully autonomous systems (see next).

Robots

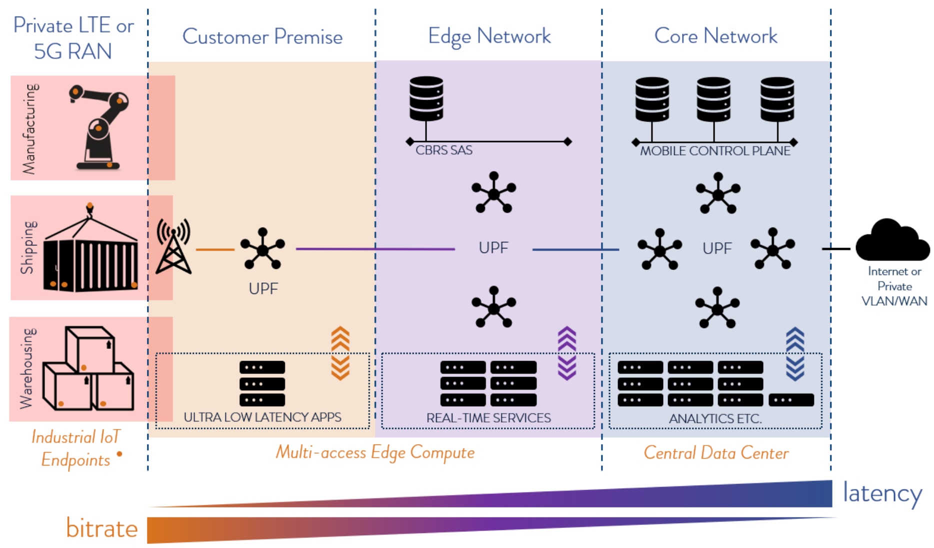

Figure 18. The already existing "private LTE" is evolving and expanding into private 5G. The low latency available at the edges, when using 5G, creates an effective communications fabric inside factories and small areas where robots/autonomous systems operate (such as a shipping yard). Communications takes place among robots on the customer premises, involving, when required, the edge network. Applications requiring ultra-low latency reside on the customer premises with the edge supporting real-time functionality, accepting some 10ms delay. Image credit: Metaswitch

A growing number of robots are turning into avid communications users. According to the latest statistics (World Robotics 2020[3]), there are approximately 2.7 million industrial robots at work around the world in factories. For example, in Singapore there is one robot for every 10 human workers in a factory. Industrial robots are just the tip of the iceberg once you consider robots in mining, healthcare, the supply and delivery chain (Amazon alone has 200,000 robots working in warehouses[4] as of January 2020[3]), and domestic robots, such as robo vacuums, which are expected to become part of 79 million homes in the next four years[5].

It is important to consider that robots are becoming more and more aware of their environment, acting as autonomous systems (this goes hand in hand with awareness), and cooperating with one another and with humans. All of these capabilities require communication.

The reason I started by making a reference to industrial robots is because this is the area where we are starting to see an evolution that will pave the way to 6G: private 5G[6].

5G has been designed to deliver very low latency. However, this is only possible in the single digits (less than 10 ms) only if one considers the radio link (the radio access network); meaning that once you enter the core network, it becomes a gamble to ensure low latency.

On the other hand, very low latency is something required when you are dealing with "local" communications. For example, if you were to communicate with a rover on Mars, your latency will be measured in minutes, not in ms! Autonomous systems, like the ones that are evolving in factories, may require low latency. If these autonomous systems are operating and collaborating through cyberspace, such as is the case in the use of Digital Twins at stage III and IV, then they may require low latency, which is something that can only be ensured if communications are managed locally. This is where "private 5G" becomes handy.

Siemens, one of the major providers of factory tools and robots, is integrating private 5G in its offer and Mindsphere. Their platform supporting operation through Digital Twins is evolving to include private 5G in local area communications.

Shifting the focus from the core network to the devices using the network, and providing a local

---

[3] https://industryeurope.com/sectors/automationandrobotics/global-population-of-industrial-robots-at-record-high-says-ifr/

[4] https://roboticsandautomationnews.com/2020/01/21/amazon-now-has-200000-robots-working-in-its-warehouses/28840/

[5] https://www.industryweek.com/technology-and-iiot/article/22028128/79-million-homes-globally-will-have-a-robot-in-residence-by-2024

[6] https://www.metaswitch.com/private5g

communications fabric that is completely in synch with service requests, is a step towards 6G. It is also a significant step in changing the perception from communications as a Telco dependent service to a local facility that can be provided by third-parties, starting from those that are providing other tools for the factory. Actually, these providers are starting to see the communications fabric as an integral part of their offer. The next step will be to use the local communication fabric (and services) to customize the core network, by taking over the network slicing (SDN and NFV from the edges).

In the next decade, we will likely see autonomous systems creating a communications bubble around them, and as these communication bubbles overlap, they create a communication fabric. Most importantly, these bubbles and the context they create are "service-based". It will be up to each specific service, and service instance, to create the communications gateways to access functionalities and share data.

8. Wrapping it up

OK, time to make sense of the title, and in particular, of the second part—6G … it is already here!

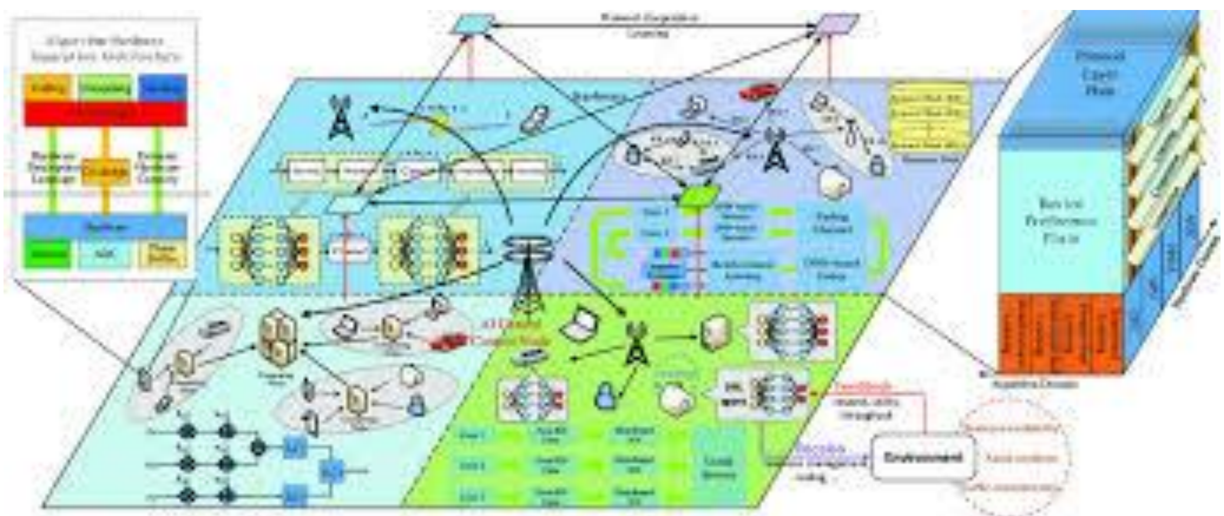

Figure 19. A possible architecture for 6G. Notice the emphasis on the various types of users, and the central role of artificial intelligence in orchestrating autonomous players. Image credit: Khaled Ben Letaief et al. Roadmap to 6G

The Quest for Bigger Capacity

Often we use capacity and speed interchangeably, yet they represent quite different things. The Amazon River has a huge "capacity" (209,000 cubic meters per second), but its "speed" is just about 2.5 km/h. A mountain river can reach a speed ten times that, yet its capacity is negligible if compared to the Amazon.

In cellular systems, the capacity is related to the number of Hz per bit multiplied by the available spectrum. As electronics become more performant, we use higher frequency and allocate more spectrum, hence we can increase the capacity of a cell. However, the capacity perceived by the user (usually measured in "speed", Gbps) depends on the number of users accessing that cell at the same time. If I have a cell providing a 1Gbps capacity, and there are 10 users accessing it, the maximum capacity per user (assuming we are dividing the cell capacity in equal parts) will be 100 Mbps (actually, the capacity per user will be less than that, but you get the point).

A straightforward way to increase the capacity of a wireless system is to deploy more cells so that there will be fewer users in each cell, thus reducing the need to divide the cell capacity to serve more users. As frequency increases, we have to deploy smaller cells, but this could be done at lower frequencies. You can deploy 100 meter cells in 4G!

With 5G we will be moving into mm waves. Verizon has already started the deployment in some areas of mm waves, and those higher frequencies (mm waves are 30 GHz up to 300 GHz) can support a broader spectrum, hence higher per cell capacity. Combine that with smaller cells and you get an exponential increase in capacity and perceived speed per user.

The use of mm waves is paving the way to 6G, which will make use of mm and sub-mm waves. It will also use longer waves, as the ones used in 4G and at this time by 5G. So, the quest for bigger capacity has not ended. As technology evolution will support more, we will deploy more and use more. The adoption of mm waves is a step in the 6G direction.

The Quest for more Flexible Architecture

Telecom Operators have always been looking to decrease CAPEX and OPEX (up-front investment and operating expenses), and the network architecture (from the very beginning, with electromechanical systems and human "switches") has been designed to minimize cost. The advent of software, and even before with automatic switches, has completely changed the landscape by introducing higher flexibility in the network.

The advent of cellular phones, basically computers you hold in your hand, has further increased the flexibility, making it possible to shift some network functions to the edges, and have customers pay to support those functions by “buying” the cellphone. As a matter of fact, if you were to deploy a fixed network from scratch, you would pay for a given coverage and capacity, let’s say 100 something, and that cost would have to be borne by the Telco Operator. If that network is a wireless network, the cost would still be around 100. However, that cost will be split into 30% for network equipment—including base stations—borne by the Telco, and 70% for the terminals, a cost borne by the end-user. A wireless network has the added benefit of smooth scaling in which a Telco will deploy more antennas and base stations as the traffic demand grows, while simultaneously increasing revenues. However, this is not the case for a wireline network as the Operator has to invest up–front and, in the last mile, be able to connect customers as demand increases.

The 5G architecture provides for increased flexibility by shifting (it is not required, it is a possibility) the session control to the terminals. This is actually a bit of 6G embedded in 5G. 6G terminals will become network nodes and will have the capability to create networks bottom-up (or outside-in).

The network slicing, NFV and SDN, which are part of the 5G architecture will be the stepping stones for the 6G architecture.

Artificial distributed intelligence will be the glue used to orchestrate all autonomous entities, as it happens in swarms of bees and flocks of starlings.

The Quest for New Players

Figure 20. Industry 4 is one of the first beneficiaries of private 5G. These new networks are regarded more as a service than as an infrastructure. Siemens is leading in this area, as one of the main global robot/machinery for industry. Image credit: Pixabay

The telecommunications biz was born as a vertically integrated sector. The separation between "Operators" and "Manufacturers" took place immediately, but the two worlds have remained closely interconnected. The International Telecommunication Union (ITU) sought the participation of both Operators and Manufacturers. The members of ITU, which is a branch of the United Nations (UN), are Countries, not private entities. This further underlines that telecommunications is a business under the wing of the State. In many Countries, the relation between the Operator and the Manufacturer was very close, to the point that the business plans and the evolution plans were defined in cooperation.

It is only in the last 20 years that the whole telecom sector underwent a radical change with newcomers entering the arena—the fragmentation of both the infrastructure and the service offering. The edges of the infrastructure in the past were tightly controlled by the Operator. You were not allowed to buy your own phone, nor connect a device to the network; you could only use equipment provided by your Operator. Now, it is up to the customer to select whatever terminal (phone, modem, IoT, media-set), and even the network edge has become fuzzy and not controlled by the Operator—think about your home network using power lines, WiFi… or the connectivity provided in shops, restaurants, malls, and even in many cities from the municipality...

Private networks are on the rise, and "private 5G" may become the forerunner of the future 6G bottom-up construction.

Private 5G is in between a network and a service offer. Siemens has rights to use 100MHz of spectrum in the 3.7-3.8GHz in Germany, and they are starting to deploy an integrated offer of industrial robots that communicate on the factory plan using a fully private 5G network.

This private spectrum is made available in Germany to any company requiring it for their own use at a competitive price. For example, a company intending to cover 500,000 square meters, already a sizeable area, would be required to pay 16,000€ (almost $20,000 USD) for a 10-year period, a very low price indeed for a company compared to what they would have to pay for a private network provided by a Telco.

Municipalities are considering massive WiFi 6 deployment, and a few LiFi. The latter makes sense as many municipalities are swapping older lighting systems with LEDs, saving power and money (in the long term). LED lighting can easily be transformed into a broadcasting infrastructure, each lamp pole becoming the antenna/cell for the surrounding area. Likewise, malls, department stores, movie theatres, and stand-alone stores are creating local networks that are no longer serving as access points, rather as a service network to deliver a specific set of services.

This is gradually changing the network and service landscape, this is happening now, and this will just be taking up steam in this decade. 6G, from a service standpoint, will be a smooth evolution of a trend that is already visible now.

Industry 4.0 will play an important role in this evolution by bringing the end-customer closer to production, increasing the “softwarization” of products, and transforming products into services (or flanking services to products).

Digital Twins are entering into new areas, most notable is the interest towards “personal digital twins” applied in healthcare, knowledge, and parallel collaboration (avatars). This will increase the demand on the network and, more importantly, increase the need for autonomous and local processing/intelligence. 6G will both leverage and support this massively distributed local intelligence, and be the fabric supporting the emergence of a global intelligence.

The Quest for Intelligence

Networks were born as mechanical systems, not really a surprise if one considers the time of their “invention”. Even the first “robots” were mechanical systems. Intelligence in a mechanical system is not necessarily “null”— the basic building stones for intelligence are present in (some) mechanical systems:

- the ability to remember
- the ability to learn

Cogs and levers can be organized in a way to have these properties, but of course they are quite limited in “quantity”, and therefore in “quality”.

The very first telecommunications networks had to rely on external (human) intelligence, with operators understanding what connection was required and assigning the required resources. In 1891, Almon Brown Strowger patented the rotary dial, which was the stepping stone for automatic switching, and the very first spark of intelligence.

This was a local intelligence—we had to wait for computers and electronic signaling systems (SS#7) leveraging this local intelligence to create a global network intelligence.

The pervasive presence of computers, first in control and then in the core of the network (becoming network equipment), has grown ever since, and today’s networks are computer networks with all the flexibility and “intelligence” that can be mustered by computers (software).

In these last few years, the network has become more and more softwarized, decoupling its functionality from the hardware on one side, and spreading them from the core to the terminals (such as smartphones).

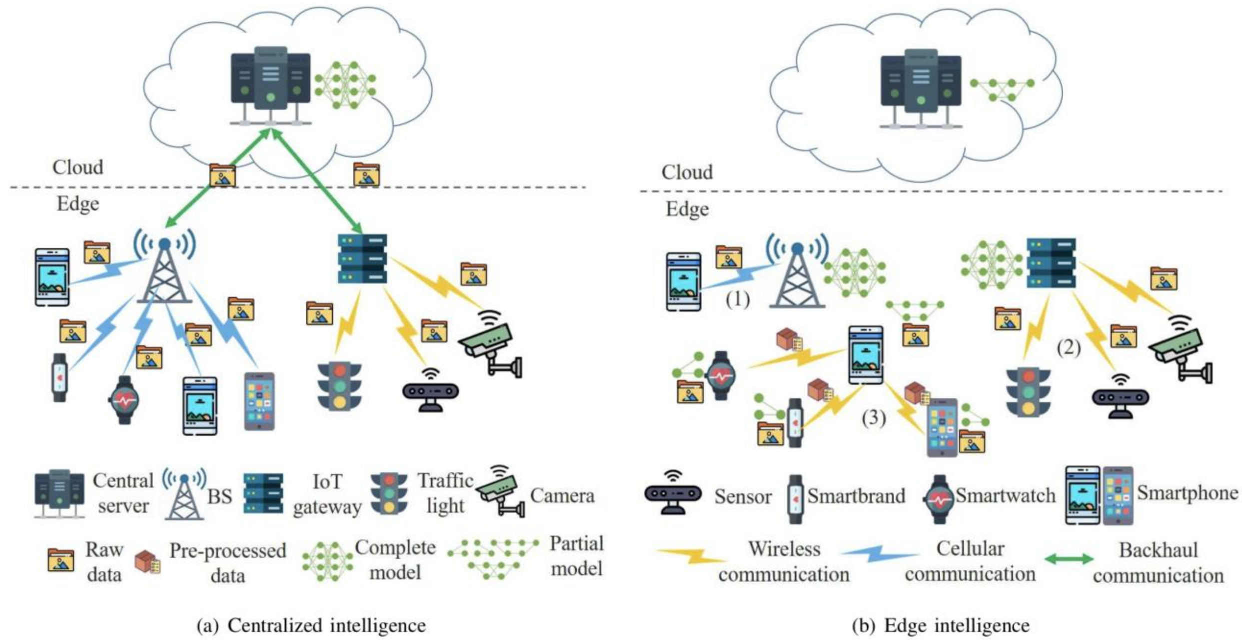

Figure 21. A nice comparison of traditional intelligence to edge intelligence from the perspective of implementation. In traditional intelligence, all data must be uploaded to a central cloud server, whereas in edge intelligence, intelligent application tasks are done at the edge with locally-generated data in a distributed manner. Caption and image credit: Dianlei Xu, Tong Li, Yong Li, Xiang Su, Sasu Tarkoma, Tao Jiang, Jon Crowcroft. Edge Intelligence: Architectures, Challenges and Applications

This process is not concluded, and, in a way, it has just begun. The shift as previously noted is towards data fabric supporting communications, rather than a communication fabric supporting access to data.

The data centers, which have been connected in the last two decades to offer services, have started to be dematerialized in a way, becoming a cloud. A service is in the cloud if it is no longer tied to a specific data center, increasing reliability and effectiveness.

Now, the cloud is taking advantage of the edges in what some call the “fog”. Actually, the shift is towards the inclusion of terminals in the fog and their seamless participation to the overall network intelligence. This process has already started with home media centers that can participate in the delivery of content. Software and intelligence at the edges evaluate usage and make predictions that, in turn, shape the behavior of the network locally and globally.

6G will be a seamless evolution of this trend that is leveraging intelligence at the edges.

Figure 22. A flock of starlings drawing the shape of a bunny in the sky. The shape is emerging from un-coordinated, but mutually influenced, behavior of each starling. Image credit: Pinterest

6G is going to be a massively distributed system in terms of intelligence. As a matter of fact, the underpinning of intelligence is a distributed system as far as we see in Nature. Our brain is a massively distributed system processing different streams of data arriving from our senses, and most importantly, multiplying these streams and the information derived from them by thousands, possibly millions, of parallel processes. All these processes are reunited in an emerging awareness, partially and in a new processing state. The paradigm of the brain as a state machine, which is not completely correct, can be used as a (partial) model. The main issue with a finite state machine model is that it ends up in mechanical processing however complex it could be, and there are debates on the fact that intelligence could be assimilated to a mechanical model paradigm.

Nevertheless, it is widely accepted that to have an intelligent behavior emerging, you need a complex system with loosely connected processes. Intelligence is not the mechanical sum of all processing (i.e. it cannot be pinpointed in any single part), rather it is an emerging property of the system.

The usual example is the behavior of a swarm of bees, the choreography of starling flocks (murmurations). There is no bee, or no starling orchestrating the behavior—no place where we can pinpoint intelligence. Intelligence is an emerging property of massive parallel behavior of individuals that, taken as single entities, have no awareness of what is going on, although the behavior of each individual influences all the others.

In telecommunications, the quest for intelligence (directed to make the most with the available resources, hence decreasing OPEX, and to decrease the investment for accommodating new requests, hence decreasing CAPEX), has focused on a centralized approach since creating intelligence is a costly business. Getting a centralized view of the network resources is cheaper than creating distributed intelligence and coordinating them. Reaching out to the edges of the network and to all resources/equipment is crucial, and a specific communications layer was designed (Signaling System #7).

There is an alternative approach to coordination, and that is having an emerging system-wide intelligence. This is the approach that has been followed by the Internet, and the reason is that the Internet is a collection of autonomous, independent networks where no one has master privilege. It would have been extremely difficult, impossible actually, to come to an agreement on a centralized control. Each network had to become aware of its surrounding and make the most of it.

Today, we have plenty of intelligence at the edges. Chips in smartphones are delivering more processing power and can host more data than the whole processing power of a telecommunications network back in the 90s, when the Intelligent Network was created.

Differently from the Intelligent Network Architecture (INA), the intelligence of devices like smartphones has not been designed to provide a global intelligence, rather to serve as local intelligence supporting services locally, and, only marginally, local networks (connectivity with other devices/sensors).

In addition to the “organizational” problem faced by the Internet, today’s variety of devices are facing the additional problem of their number. There are now billions of smart devices (compare this to the thousands of networks, a 4-6 order of magnitude difference), and it is simply impossible to achieve a centralized control and coordination (both cost and effectiveness to manage atomicity of transactions and their sheer number).

Additionally, the edges of the telecommunications networks have an evolution of their own, an evolution that is not coordinated by any central entity, but by market forces (leveraging on available technology). As an example, there is not a single plan to develop a seamless augmented reality device, but many companies around the world are pursuing this goal, using different approaches and with different business models. Yet, each of them is trying to make the most of whatever is, and will be, available.

This creates a chaotic environment that is conducive to the emerging of intelligence. Notice that chaotic does not mean that everything is independent of everything else. Actually, this kind of chaos is highly influenced by the context. Even in Brownian, motion molecules influence one another.

What is fascinating is that this emerging intelligence is not restricted to the devices. It is actually a mélange of devices and our intelligence, the very first sign of an emerging Digital Reality.

9. Digital Reality

Cyberspace has become so pervasive in terms of usage, and its access is so seamless in every moment of our life that we are starting to lose the perception of a separation between the reality of atoms and the ones of bits. They are starting to merge into a single perceived reality.

Part of this is the effect of pervasive connectivity—a smartphone has become an extension of our senses and, partially, of our brain. It is not just the smartphone, it is Alexa and its siblings. It is the multiple screens that open windows in the cyberspace in our homes. Just around the corner are even more sophisticated devices such as smart glasses, smart windshields, smart lenses, and more that will overlay, seamlessly, the world of bits onto the world of atoms.

Our "intelligence" depends on our senses, both in maturing (getting smarter) and in analyzing the context to find a reaction fitting the situation. In turn, our intelligence is a social intelligence—it depends on what we learn from others and by the way we interact with others.

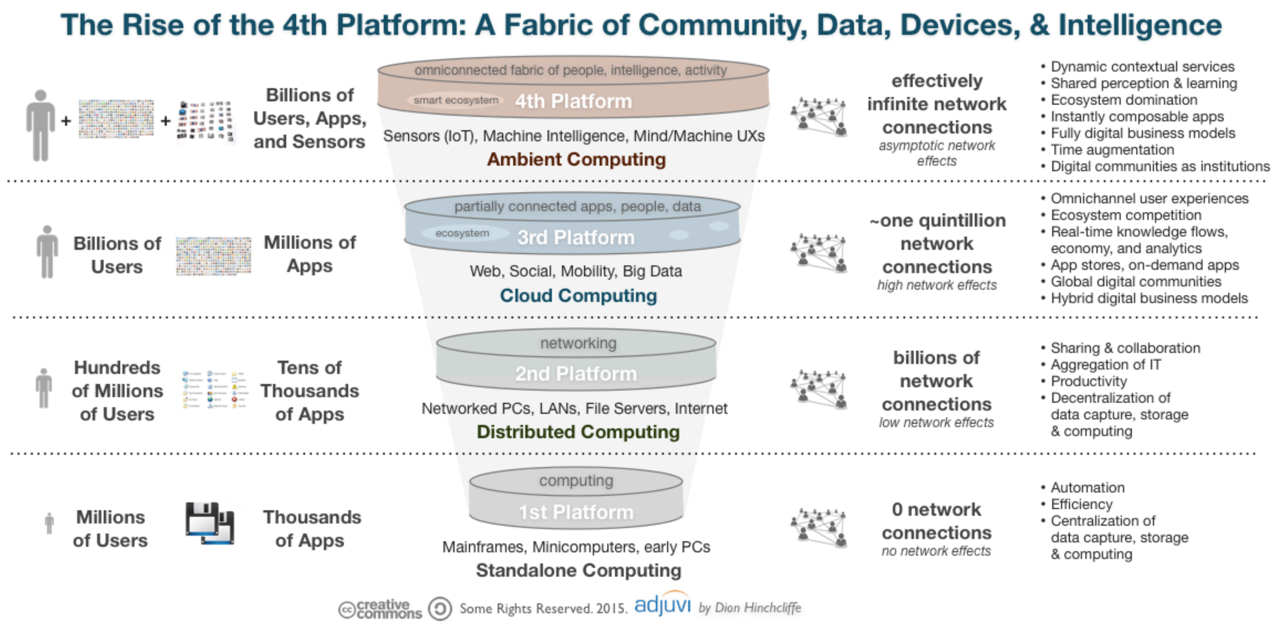

Figure 23. People, data, and devices are the substrata of an intelligence fabric that will become more and more pervasive in this decade and will be taken for granted until the next one. Image credit: Dion Hinchcliffe

Now, and more so in the future, a significant portion of these "others" will be machines, software, artificial intelligence, etc.

Likewise for the ambient intelligence. It is evolving from single points of intelligence to a global ambient intelligence, arising from the interplay of machines, software, and, soon enough, ourselves. Artificial intelligence is now learning from the context, and we are part of the context.

Clearly this requires the support for continuous interactions, a multi-sided interaction that does not necessarily require a central network. Devices and software can interact with one another using their own communications capability. This is what 5G is potentially supporting, and this is what 6G will be leveraging in creating communications fabric.

Artificial Intelligence is becoming the big player in this decade, not because of quantum leaps, rather because of its pervasiveness in business and in everyday life.

The ongoing Digital Transformation fuels the adoption of Artificial Intelligence since it provides the data and the operating space (the cyberspace).

This pervasiveness is well represented in figure 23 (actually an "old" figure going back to 2015, but still right to the point). Today, we are basically at the stage of transitioning from Cloud

Computing to Ambient Computing. This transition will most likely take all of this decade to complete, but once completed it will reshape service architectures and be the underpinning of Digital Reality.

6G will stem from this pervasive ambient intelligence.