

5G and Beyond: Past, Present and Future of the Mobile Communications

Carlos A. Gutierrez, *Senior Member, IEEE*, Oscar Caicedo, *Senior Member, IEEE*,
and Daniel U. Campos-Delgado*, *Senior Member, IEEE*

Abstract—The fifth-generation (5G) of mobile communications networks is emerging as a revolutionary technology that will accelerate the development of smart cities and the realization of the information society. This paper aims to provide an introduction to 5G for non-specialists, and a survey of this new technology for those already familiar with mobile communications, covering the conceptualization and the core technologies underpinning 5G networks. The paper also discusses the status of the commercial roll-out of 5G until 2020 from a worldwide perspective and gives a future view of mobile communications beyond 5G.

Index Terms—5G communications, Massive MIMO, Beamforming, mmWave communications, Mobile edge computing, Small cell stations, NOMA, network slicing.

I. INTRODUCTION

Mobile telephony is constantly evolving to meet the society's demands for more and better communication services. This process of continuous evolution is marked by the appearance of new standards that incorporate the latest technological advances and harmonize the coexistence of previous radio communication systems [1]. The great innovations that are triggered by the rise of a new standard (or a new set of standards) define new stages in the history of mobile telephony, which are classified as generations [2], [3]. We are currently moving towards the fifth generation (5G) of mobile telephone networks [4], [5]. This new generation is emerging with the promise of enabling services that will improve user experiences through access to ultra high definition and 360° video, work in the cloud, and immerse communications based on augmented and virtual reality. Unlike fourth generation (4G) networks, their 5G counterparts will offer native support for applications oriented to communications among machines. This new paradigm will be essential to provide all connectivity aspects related to smart cities, such as those required for Industry 4.0, intelligent transportation systems, e-health, and similar technologies.

The 5G revolution is built under the International Telecommunications Union (ITU) vision for the future of the international mobile telecommunications (IMT) for the year 2020 and beyond [6]. The use cases and minimum requirements related to the 5G technical performance were defined in a report published in 2017 by the ITU radio communications sector [7]. According to that report, 5G networks will reach

Carlos A. Gutierrez and Daniel U. Campos-Delgado are with Faculty of Sciences, and Daniel U. Campos-Delgado also with Instituto de Investigación en Comunicación Óptica, both in Universidad Autónoma de San Luis Potosí, S.L.P., Mexico, e-mails: cagutierrez@ieee.org, ducd@fciencias.uaslp.mx.

Oscar M. Caicedo-Rendon is with University of Cauca, Popayan, Colombia, email: omcaicedo@unicauca.edu.co.

*Corresponding author

a connectivity density ten times higher than 4G networks, a peak transmission rate twenty times greater, and an area traffic capacity hundred times greater. Apart from this, 5G latency times will be up to ten times shorter than previous 4G networks [8], [9]. Following the path traced by the ITU, the 3rd Generation Partnership Project (3GPP) has developed standards for 5G technologies. The first one was published in 2017, and the technology associated with that standard is known as 5G-New Radio (5G-NR) [10].

The commercial deployment of 5G-NR started at the end of 2018 [5]. However, the 5G ITU vision is still far from complete. In addition to the technological aspect, basic and applied research activities are still required to shape these new communication technologies to solve the social, economic, and environmental problems that plague the world. In the context of Latin America, the implications of regulatory policies and the release of new electromagnetic spectrum (EMS) bands should be thoroughly investigated, so that the 5G transformation will contribute effectively to reduce the digital gap in the region.

In this context, this paper presents a review of the background, the current status, and the future of 5G networks with the purpose of identifying the opportunity areas offered by this new technology, and to draw attention to technical and regulatory issues that are still open. The origins of the mobile telephony, along with the main innovations brought by each generation prior to 5G, are reviewed in Section II. The use cases, the technical requirements, and the network architecture of 5G networks are discussed in Section III. Sections IV and V address the main transmission and network management technologies for 5G. An overview of the current state of the 5G commercial deployment is detailed in Section VI. Section VII describes briefly a future perspective of the mobile telephony beyond 5G, and some conclusions and ending remarks are introduced in Section VIII. In fact, 5G networks have been the subject of numerous reviews, such as those presented in [8], [9], [11]–[13]. This paper complements the current 5G literature providing a integral review of both the radio access network (RAN) and the core network. In addition, this paper offers an overview of the current 5G worldwide deployment, and also of the ongoing initiatives that seek to extend the 5G capabilities with the focus towards the sixth generation (6G) of mobile telephony networks.

II. BACKGROUND OF 5G NETWORKS

The early stages of mobile telephony date to the invention of the wireless telegraph in the late nineteenth century, and its

applications in the naval industry, as a communication device between ships and shore stations [14]. Wireless telegraphy demonstrated the feasibility of long-distance communications based on the transmission of electromagnetic waves that propagate freely in the Earth's atmosphere. Furthermore, the technological advances in electronics at the beginning of the twentieth century, triggered by the invention of the vacuum tube, allowed to apply the principles of wireless telegraphy in commercial broadcasting and mobile telephony for military and police applications [15].

The first mobile system for civil use was the Mobile Telephone System (MTS), which was introduced in 25 cities of the United States in 1946 [15]. The MTS worked based on push-to-talk technology. The communication service was provided by a central station that linked to all the mobile users (MUs) of the city. This centralization of the service, along with the bandwidth restrictions that are imposed on every wireless system to regulate the EMS, significantly limits the capacity of the MTS. The concept of cellular telephone networks, developed at Bell Labs during the forties [16], allowed to overcome the problems of network saturation and fostered the commercial deployment of mobile telephony. The idea behind this concept is to increase the network capacity by dividing the coverage area into cells of smaller extension. Service is provided within each cell by a base station (BS), connected to the public telephone network. This concept allows different RF channels to be assigned to adjacent cells to reduce interference between them, and reuse RF channels in cells that are far enough to generate negligible interference. The cellular network architecture is the cornerstone of the mobile telephony for commercial use, where these communication networks are therefore referred as mobile cellular networks. Each generation of mobile cellular networks has been developed in a ten-year time frame, and the transition to a new generation is driven by the need to improve coverage, achieve higher transmission rates, and offer new mobile services [3], [17], as shown in Fig. 1.

The First Generation (1G) of mobile cellular networks entered the commercial market at the end of the seventies. Hence Japan was the first country to deploy a 1G network. In the Nordic countries (Norway, Sweden, Denmark, and Finland), 1G networks started to be implemented in 1981. By the mid-80's, there were 1G networks in UK, USA, Canada, and Mexico. This first generation was designed for voice services by using analog technology. It had a low spectral efficiency, and since there were no international agreements for the use and regulation of such networks, there were a wide variety of incompatible standards in operation. As a result, the service was restricted to a national—and even a regional—coverage. Despite their incompatibility, all the 1G networks were based on Frequency Division Multiple Access (FDMA) technology, and the calls were connected by circuit switching. In Japan and United States, sections of the 800 MHz band were reserved for 1G. Meanwhile, the 900 MHz band was reserved in Sweden and United Kingdom, whereas the 450 and 200 MHz bands were chosen in Germany and France, respectively [2].

The second Generation (2G) of mobile cellular networks

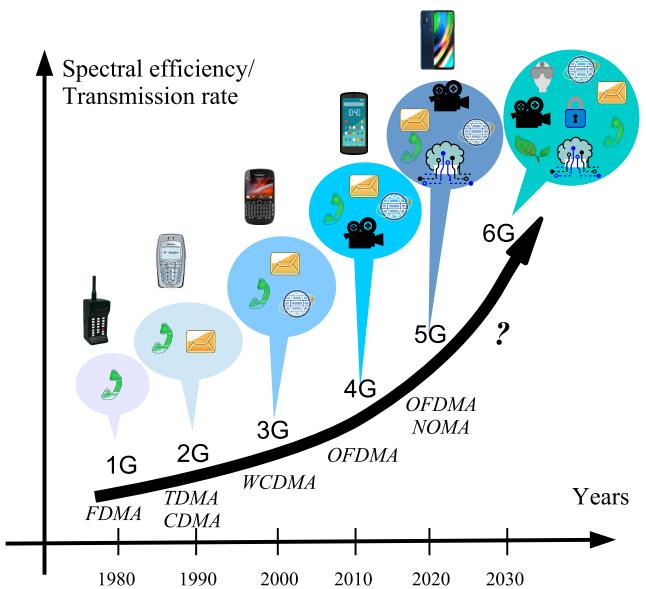


Fig. 1. Evolution of cellular technology.

started operations in the early 90's and was characterized by the switch to digital technologies. Spectral efficiency improved thanks to the use of time-division multiple access (TDMA) and code division multiple access (CDMA) techniques, as well as the introduction of channel coding techniques. An important innovation was the support of data transmission for short message service (SMS). 2G networks were a commercial success thanks to the development of international standards, especially, the Global System for Mobile Communications (GSM) and the Interim Standard 95 (IS-95) [18]. Both standards transmit on the 810 and 960 MHz bands over a circuit switched architecture. The theoretical data transmission speed of 2G was 115.2 kbps (with the high-speed circuit-switched data enhancement of GSM). The digital corner stone of the 2G service allowed the introduction of compact hand-held telephones with more efficient use of batteries. However, 2G networks transmitted data inefficiently, because the communications were based on circuit switching, which is suitable for voice traffic, but not for data traffic.

The transition to the third generation (3G) of cellular networks began in the early 2000s. One of the great innovations of 3G was the ability to support both circuit and packet switched links; the latter are required for services based on the Internet Protocol (IP). This generation allowed the incorporation of new mobile services, such as access to the Internet and video calls. The RAN of 3G is based on CDMA technology, and the main 3G standards are wideband CDMA (WCDMA) and CDMA 2000 [19]. Both standards are global in scope, and they achieve theoretical transmission speeds of up to 14.4 Mbps in the downlink (DL) [2]. 3G networks set the scene for the smart phones, which were equipped with a tactile and more visually attractive interface, as well as, a processing motor that supported a variety of mobile applications.

The fourth generation (4G) deployment began in early 2010, offering a peak transmission rate of up to 300 Mbps on the DL, and 75 Mbps on the uplink (UL); although with the most recent

updates, these networks achieve speeds of up to 1 Gbps on the DL [20]. The RAN of 4G is based on orthogonal frequency division multiple access (OFDMA), single carrier FDMA (SC-FDMA), scalable OFDMA (SOFDMA), and the multiple-input multiple-output (MIMO) transmission techniques. The main standards of 4G are Long Term Evolution (LTE) and its advanced version, the LTE Advanced (LTE-A) [21]. 4G networks enable voice, data, and multimedia services ubiquitously through subscription media, and are capable of supporting applications for high-definition video streaming and on-line gaming. 4G technology also supports mobile applications based on IP that require latency times of up to 20 ms.

III. CONCEPTUALIZATION OF 5G COMMUNICATIONS

A. Use Cases and Performance Requirements

4G networks will still continue to expand and to operate for several years supporting people-oriented communications services [20]–[22]. According to the latest version of the Ericsson Mobility Report, a co-existence of various technologies is expected by 2026, where the predominant ones are 5G, LTE, WCDMA / HSPA, and GSM / EDGE [23]. However, the current situation also demands services oriented to communication between machines, in order to support Industry 4.0 applications, intelligent transport systems, and home automation. Such applications pose new challenges, such as ensuring services in coverage areas with a high density of interconnected devices; low latency transmission for delay sensitive applications; the support of broadband links in high mobility conditions, similar to those found in high-speed trains; and the reconfigurability of the network for a flexible and efficient management of its resources. Hence, 5G networks seek to solve these challenges [11], [12].

The ITU defined three use cases for these new networks:

- Enhanced Mobile Broad-Band (EMBB): End-user centric applications characterized by requiring access to multimedia content, and demanding high data transfer rates with peak speeds of up to 20 Gbps and speeds perceived by the user of up to 100 Mbps, in addition to mobility support up to 500 km/h and an spectral efficiency three times higher than the ones provided by 4G. Examples of EMBB applications are voice and ultra high definition mobile video, as well as, virtual and augmented reality for immersive communications.
- Massive Machine-Type Communications (MMTC): Applications involving a large number of connected devices, which transmit a low volume of data with little sensitivity to delay and whose energy consumption it is typically low. Applications of this type include connected vehicles, pet / person care, and wireless sensor and actuator networks. The objective for this use case is to have a traffic capacity of 10 Mbps/m^2 , a connection density of $10^6 \text{ devices/km}^2$, and an energy efficiency of the network 100 times higher than 4G.
- Ultra Reliable Low Latency Communications (URLLC): Applications characterized by operating in real time, such as those related to transport security, remote surgeries,

TABLE I
PERFORMANCE INDICATORS FOR 4G AND 5G NETWORKS (INFORMATION FROM [7]).

Indicator	4G	5G
UL transmission peak rate	500 Mbps	10 Gbps
DL transmission peak rate	1 Gbps	20 Gps
UL peak spectral efficiency	6.75 b/s/Hz	15 b/s/Hz
DL peak spectral efficiency	15 b/s/Hz	30 b/s/Hz
Connectivity density	10^5 disp./km^2	10^6 disp./km^2
Mobility	350 km/h	500 km/h
Traffic capacity per area	0.1 M b/s/m ²	10 M b/s/m ²
Latency	20 ms	1 ms

and the automation of industrial processes, which require a low latency of up to 1 ms.

Table I summarizes the main performance requirements of these three use cases [7], and compares them with their counterpart of 4G. The key technologies identified by the ITU for meeting these requirements are reviewed in [24].

B. 5G Network Architecture

The requirements of high transmission data rates and high density connectivity can be provided at the physical layer and medium access control level by using efficient transmission techniques. However, the requirements of low latency, high mobility, and traffic capacity per area involve strategies at the core of the network. Thus, one of the main innovations of 5G is in the network architecture, whose design is much more flexible than previous generations thanks to the implementation of logical segmentation, *softwarization*, and network virtualization and its services.

Figure 2 shows a diagram of the general architecture of 5G networks, which consists of RAN, core network and user equipment. These elements are represented by Network Functions (NF) and their respective interfaces. Table II describes the NF of a 5G architecture. In [13] and [25], different architectural options to deploy the RAN are presented, and the architectures for core network deployment are studied in [26] and [27].

The 5G networks follow a service-based architecture (SBA), which provides important benefits to the network structure, for example

- (I) Easy update: each service can be modified with minimal impact on other services;
- (II) Extensibility: adding a new functionality involves adding a new service and its corresponding interfaces;
- (III) Modularity: a 5G system is made up of services of fine granularity (they implement a single NF) and low coupling (they use standardized interfaces) known as microservices [28], thereby promoting the logical segmentation of the network; and
- (IV) Opening and reuse: services can be exposed and called out through lightweight and standardized interfaces, promoting thereby the reuse of functionalities.

These benefits allow the 5G architecture to be flexible, programmable, and easy to automate. They also promote a significant cost reduction thanks to the deployment of NFs as virtual network functions (VNF) in small, medium or large data centers.

TABLE II
5G NETWORK FUNCTIONS.

Network Function	Description
Application function	Interact with the 5G core, especially PCF and NEF, to help in traffic routing.
Access and mobility management function	End point of the no-access layer which includes registration, connection, and access and mobility management.
Authentication server function	Supervise the authentication process of UEs.
Data network	Connectivity function to any data network.
Network data analytics function	Data analytics of the NFs.
Network exposure function (NEF)	Expose the functionalities and events of the NF by allowing the communication to external applications.
Network repository function	Register the NFs and allows their identification to facilitate composite services.
Network slice selection function	Select the logic network instances that serve the UE.
Policy control function (PCF)	Support the selected policies which are used by CP to control the network.
Radio access network	Allow the radio access to the NF by UP, and the CP by the core of the 5G network.
Session management function	Request of the session beginning, modification and liberation of the data, and IP assignment of the UEs.
Unified data management (UDM)	Support the generation of authentication credentials, user identification and subscriptions management.
Unified data repository	Allow to UDM and PCF to obtain subscriptions and policies information, respectively.
User equipment (UE)	Allow the final users to use the 5G network services.
User plane function	Manage mobility and QoS among radio access technologies, routes packages, and generate traffic records.

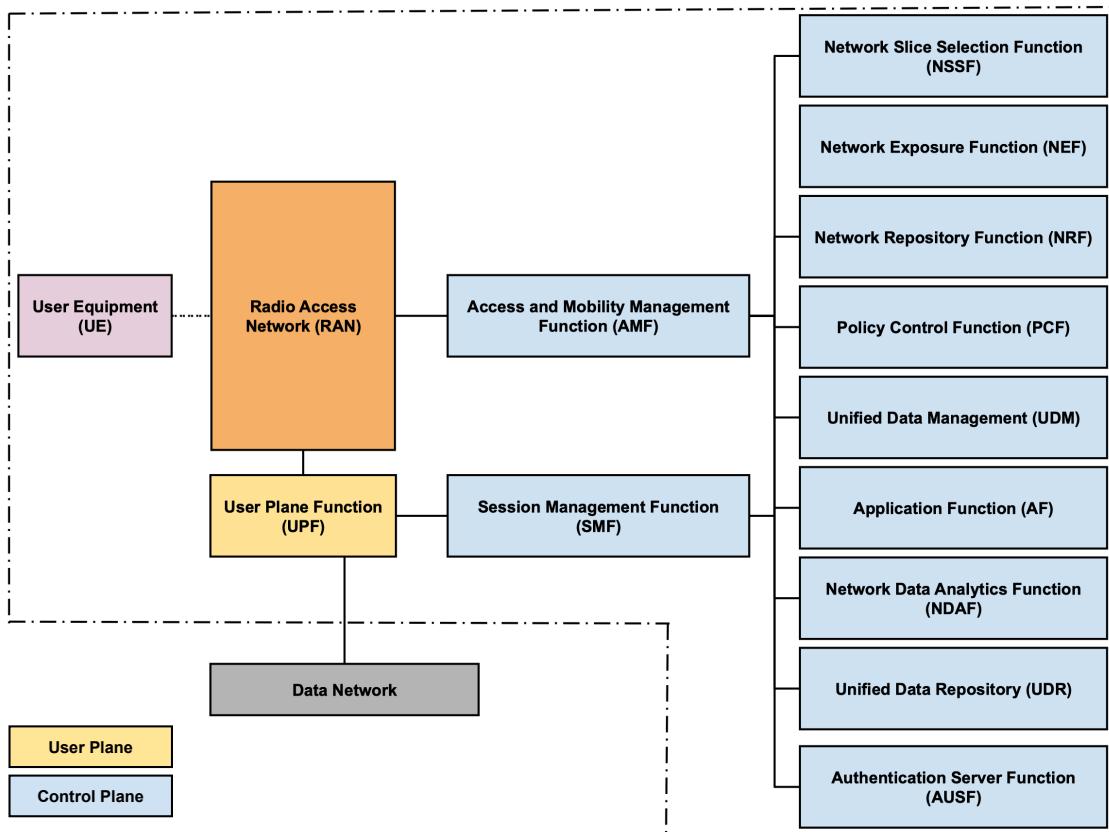


Fig. 2. 5G Network Architecture.

The control and user plane separation (CUPS) is another pillar of the 5G network architecture. While CUPS is not a new concept, it is fundamental to 5G, because it allows the separation of the evolved packets core (EPC) of 4G in the network functions of the user plane (UP), and of the control plane (CP) [29]. The NF of the UP can be deployed close to the application due to latency and data transfer rate requirements, which are critical to URLLC and EMBB use cases, respectively. On the other hand, the NF of the CP can be operated centrally to satisfy reliability requirements. The UP carries only data traffic, while the CP carries the network signaling. The NF of the UP and CP interact with each other

through standardized interfaces. A 5G service is implemented by NF chains (virtualized or not) belonging to the CP and UP. Mobile edge computing (MEC), fog computing (FC), cloud computing (CC), software defined networks, and network functions virtualization (NFV) are fundamental elements of the 5G network architecture. These technologies are reviewed in Section V.

IV. KEY TECHNOLOGIES FOR THE 5G RAN

The technology revolution prompted by 5G networks is best understood by identifying the dimensions of innovation

that determine the evolutionary course of wireless communication systems. These dimensions are [30]: 1) EMS availability, 2) spectral efficiency, 3) spatial efficiency, and 4) system efficiency. The first three dimensions are described in this section, while the fourth is discussed in Section V.

A. Millimeter Wave Communications

The first dimension of innovation refers to the EMS band (or bands) that the network has available to establish the wireless links between BSs and MUs. The relevant elements of this dimension are the central frequency and bandwidth of the available RF channels, which determine the range and capacity (in terms of maximum data transmission rate) of the link, respectively. One of the main innovations of 5G networks is in the inclusion of frequency links between 20 and 80 GHz, that is, links in the millimeter wave (mmW) band [31]. This concept will support services requiring peak transmission rates in the order of Gbps, as required for the EMBB use case. These transmission rates are difficult to achieve in EMS bands at lower frequencies, such as those reserved for 4G, where the bandwidth is limited to a few tens of MHz. In contrast, the bandwidth available in the mmW band is in the order of several GHz.

Increasing the system's operating frequency opens the access to wider bandwidths, but it also reduces the system's range, as high frequency electromagnetic waves are more susceptible to absorption and attenuation. Therefore, the mmW band is not suitable for indirect communications, i.e. in the absence of a line of sight (LOS) between transmitter and receiver. These limitations can be removed by employing antenna arrays to reduce the impact of attenuation. Moreover, the high atmospheric attenuation of the mmW band is convenient to reduce interference when a large number of devices need to be connected, as in the MMTC use case. However, further research is still needed for a good understanding of the propagation of RF signals at the mmW band. For example, further measurements campaigns are needed to increase the empirical information about propagation losses, time and frequency dispersion, angular spread and probability of LOS in the RF links at mmW frequencies [32].

On the other hand, it was recently shown in [33] that the mobile radio channel behaves as a non-stationary system as the bandwidth of the transmitted signal increases. The effects of these non-stationarities on the air interface of 5G networks have not been studied in detail. In addition, 5G technology will be deployed at different rhythms in big and small cities, and rural areas, as pointed out in [34]. This phenomenon can increase the digital gap between cities and rural areas. Therefore, it is important to look for alternatives that will allow a seamless migration to the mmW band.

B. Large Scale Antenna Arrays

The second dimension of innovation refers to the efficient use of the available EMS, measured with respect to the number of information bits that can be transmitted per each available Hz. 5G networks seek to boost the spectral efficiency

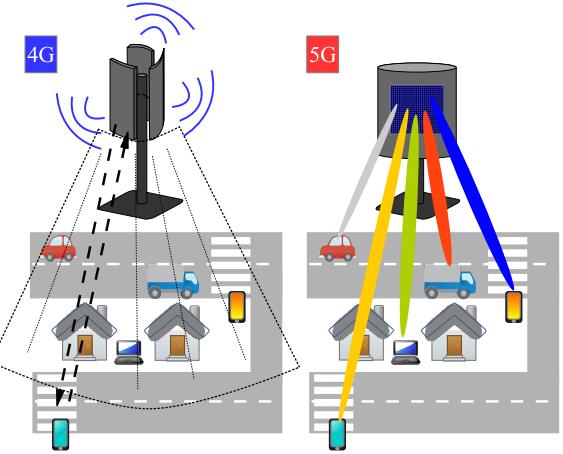


Fig. 3. Transmission schemes in 4G and 5G networks with and without beamforming, respectively.

by leveraging on the use of large-scale antenna arrays (LS-AA) operating in the mmW band. In this frequency band, it is possible to build a compact-profile LS-AA composed of hundreds or even thousands of antenna elements.

The LS-AA transmission techniques that will be implemented for 5G rely on beamforming [35]–[38] and massive MIMO (mMIMO) [39]–[41]. In beamforming, a LS-AA is employed to obtain a combined radiation pattern having a high directivity. This property permits to extend the communications range without increasing the signal-to-noise ratio (SNR) [42] or interference towards other users. Beamforming is also used as a medium access technique, known as space division multiple access (SDMA). By employing SDMA, the BS can maintain a simultaneous communication with multiple MUs on the same RF channel but using different radiation beams, each with a high signal to interference ratio (SIR) [43], as shown in Fig. 3. This figure also shows a 4G scenario in which the BS communicates with MUs using sector antennas with moderately directive radiation patterns. Beamforming is also used for data multiplexing in the spatial domain [44]. Some of the relevant research problems focused on LS-AA technologies are in the development of energy-efficient hybrid (analog and digital) precoding techniques of low complexity, as well as, in the simplification of the RF chains required for the implementation of a LS-AA [36], [44].

On the other hand, mMIMO transceivers employ LS-AAs on both sides of the link for spatial multiplexing of large data volumes. The main difference between mMIMO and MIMO is the capability of the former to perform multi-user detection (MUD) to increase the information flow [40]. The performance of MUD depends on the receiver's ability to eliminate the interference generated by other users. In this sense, precoding techniques are used for interference cancellation. These techniques require an accurate estimate of the channel state information (CSI) of each active MU, where this information can be obtained through pilot signals, or by feedback from MUs [45].

An open research problem relevant to 5G technologies focuses on the performance of mMIMO operating in combi-

nation with frequency division duplexing (FDD) [40]. Most of the work on mMIMO assumes a system operating under a time division duplexing scheme (Time-Division Duplexing, TDD) to avoid discrepancies with the channel estimation on the DL [39]. However, the tandem of mMIMO-FDD can accelerate the adoption of the mMIMO technology in 5G networks, since FDD schemes have been widely used in mobile telephony. Other relevant problems include the development of low complexity detection and precoding algorithms [40].

C. Non-Orthogonal Multiple-Access

The medium access techniques are also instrumental to boost spectral efficiency. The access to the RF channel prior to 5G was based on a principle of orthogonality, either by TDMA, FDMA, or CDMA, to avoid or limit interference between users. These techniques are referred to as orthogonal multiple access techniques (OMA). In OMA, the capacity of active users depend on the orthogonal resources available. On the other hand, the transmission based on non-orthogonal multiple access (NOMA) allows multiple users to transmit simultaneously on the same communication channel, and it employs MUD strategies to mitigate the multiple access interference [46]–[49]. NOMA helps to improve the system's spectral efficiency and the transmission rate of all active users, since it allows the massification of connectivity, and reduces the latency time [48]. These benefits are obtained in exchange for increasing the complexity of the receivers. There are several approaches to NOMA, with implementations at bit and symbol levels, but all seek to assign particular signatures to the MUs based on transmission power, code spreading, coding or interleaving [47], [49]. NOMA can be further enhanced by a MIMO transmission and reception strategy. By combining both technologies, it is possible to improve the channel capacity as compared to MIMO-OMA [50], [51].

In NOMA by power allocation [46], the transmission power is the diversity factor in the time domain that let identifying the information of each user. Thereby, the users located far away from the BS will transmit with a higher power than the closest users. The receivers employ successive interference cancellation (SIC) to iteratively separate the information from the interference, according to the transmission power of each user. This strategy is illustrated in Fig. 4. The main research problems at NOMA are related to the design of decoding techniques and the efficient management of transmission resources (*e.g.*, power and bandwidth). The information feedback (such as the CSI) between BSs and MUs, and the complexity reduction of the receivers are also relevant research problems [52].

In addition to the LS-AA technologies discussed above, the spectral efficiency also relies on new waveform designs and advanced modulation schemes for the transmission of information. The technology trends in these fields for 5G networks are not reviewed in this article due to space limitations, but the reader can refer to [24] for further information.

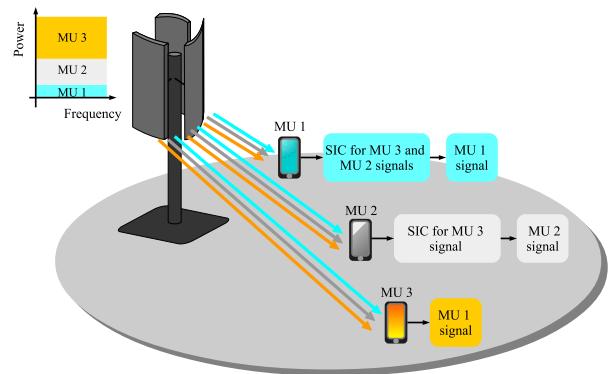


Fig. 4. NOMA Principle by Power Assignment.

D. Small Cell Stations

The third dimension of innovation relates to spatial efficiency, and it considers the improvements aimed at supporting a greater number of users connected per unit of area. The simplest strategy—yet effective—to improve spatial efficiency is the segmentation of the coverage area, as defined by the cellular network concept. As the density of connected users increases, it is convenient to reduce the size of the cells. Thus, another innovation of 5G is precisely in the network densification by the deployment of a large number of small cell stations, where each one has a short range, *i.e.* lower than 100 meters.

The benefits of deploying a great number of small cells are discussed in [30]. One of them is the ability to provide coverage in high-traffic areas and inside buildings. Small cells increase the network throughput for communication links in frequency bands below 6 GHz. However, a great number of cells (picocells and femtocells) leads to a significant energy consumption [53], [54], opening the need for novel strategies that promote energy efficiency [55] [56]. The deployment of small cells also faces regulatory and operational challenges, such as high fees due to the use of public infrastructure (such as lighting poles) to install these cells, and the different criteria to measure the health risk by exposure to high power electromagnetic fields [34]. Others challenges are related to the interconnection and management of a great number of small cells, where some of such problems and their solutions are discussed in the next section.

V. KEY TECHNOLOGIES FOR THE 5G CORE

The focus of the fourth dimension of innovation is on system efficiency, and it addresses primarily the 5G network core. In this dimension, the efficiency aims, for example, at the reduction of latency time, higher transmission rates, lower packet-losses, and reduction of energy consumption.

A. Logic Network Segmentation

One of the objectives of 5G networks relies on offering a wide range of services, each one with different requirements for quality of service (QoS), opening a range of opportunities for applications from various industry segments (also known

as verticals) [57], [27]. To achieve this goal, 5G systems must be able to provide network segments by applying the concept of logical segmentation [58], [59], whose principles are the automation of network operation, high reliability, scalability and isolation of segments belonging to different clients, support for *softwarization*, programmability and virtualization, hierarchical abstraction and segment customization, and elasticity of network resources [60]. The key technologies to realize these principles are artificial intelligence (AI), SDN, NFV, MEC, FC, and CC.

The logical network segmentation in 5G must intelligently overcome different challenges, such as the creation of admission and planning algorithms aimed at allocating resources to the segments during their creation and operation, the location of the network functions in the infrastructure, the control of intra-segment traffic, the end-to-end management of segments, the mobility management and the orchestration of the NFs that make up the segments, as well as, their security and privacy.

B. Artificial Intelligence

Several areas of AI, among which machine learning (ML) stands out, are useful for developing cognitive logical segments that are fundamental to achieve self-directed 5G networks. A self-directed network must autonomously deploy the perceive-plan-decide-act cycle, which must be accompanied by learning and feedback, both from the network as a whole, and from its constituent segments [61], [62]. A study on the use of supervised, unsupervised, and hybrid ML techniques in communication networks is presented in [63]. ML techniques have recently generated interest in the network domain to automate decision making, where reinforcement learning (RL), and deep reinforcement learning (DRL) are two examples [64].

In RL, an agent makes decisions periodically considering states (*e.g.*, the set of routers available to calculate a route) and actions (*e.g.*, send traffic along the calculated route). The agent observes the result (expressed as a reward, for example, optimized traffic in terms of delay) of its interaction with the environment (*e.g.*, a 5G network) and then automatically adjusts its strategy to achieve an optimal policy (*e.g.*, optimize delay on a logical segment of URLLC) [65], [66], [67]. RL approaches, such as Q-learning [68], and state-action-reward-state-action (SARSA) [69], slowly converge to the optimal policy when they need to explore and acquire knowledge about a large state-action set, making it difficult to use in large-scale networks such as 5G.

Deep learning has recently been used as a tool to overcome the limitations of RL, giving rise to another type of learning called DRL [70]. Deep learning techniques learn by discovering complex structures in data and build their computational models by using multiple layers of processing. Hence, a deep neural network (DNN) can create multiple levels of abstraction to represent data [71], [72]. Typical structures of this type of learning are feed-forward neural networks (FNN), and recurrent neural network (RNN). In FNN, there are no cycles or loops in the neural network. Therefore, the data moves in only one direction: from the input nodes, through the hidden nodes, and towards the output nodes [73]. Convolutional neural

networks (CNN) are the best known model of FNN with a wide variety of applications, including dynamic routing [74]. Unlike FNN, a RNN is recursive in nature. The connections between neurons produce specific cycles, meaning that the output depends not only on their immediate inputs, but also on the neuronal state of the previous step [75]. A RNN is designed to use sequential data, *i.e.* the current state is related to the previous one. Long short-term memory (LSTM) is one of the most representative RNNs [76] and is useful, for example, for the prediction of network traffic [77].

DRL uses deep neural networks to improve the speed and performance of RL algorithms [78]. From a high-level perspective, DRL implements a DNN to derive the optimal policy instead of the Q table used by *Q-learning*. The most relevant DRL algorithms are [64], [79]: DQL, DDQL, DQL with *prioritized experience replay*, DDQN, *distributional deep Q-Learning*, *deep Q-Learning with noisy nets*, and *rainbow deep Q-Learning*. Recently, DRL has been used in the context of IoT, HetNets, UAV, SDN and NFV, all of them involved in the conception of 5G. The integration of DRL in 5G networks could revolutionize the management mechanisms of logical segments, especially those related to planning and resource sharing. This integration would allow to evolve from techniques based on models, to those without a model, which learn deeply by interacting with the environment to satisfy both experience level agreements (XLA) and service level agreements (Service Level Agreements, SLA). These agreements are related to indicators of quality of experience (QoE) and QoS, respectively.

Due to the complexity of 5G communication systems, the following points will be fundamental for logical network segmentation based on DRL: (*i*) the definition of the sets of states and actions should cover the dynamics of the segments network; (*ii*) the modeling of the sets of states and actions should reduce the cost of learning; (*iii*) the reward should be as simple as possible to facilitate the learning process; (*iv*) the quantification of the reward should incorporate the time of response (the interaction with the environment is not instantaneous) and computation; and (*v*) the results provided by DRL agents must be compatible with the time scale in which events occur and are managed in a 5G network.

C. Software Defined Networks

A fundamental concept for network *softwarization* is SDN [80], [81], and this concept is essential for managing multiple virtual logical networks in the context of 5G [82]. The SDN architecture offers the following benefits [83], [84]: (*i*) clear separation of the data plane and CP, (*ii*) logical centralization of the control function, and (*iii*) implementation of the control function in *software* [85], [86].

A high-level view of the SDN architecture and its blueprints is presented in [87], [88], [89]. The data plane is made up of network elements specialized in packet handling (forwarding), and communicates with the CP through one or more interfaces in the south zone (South Bound Interfaces, SBI). OpenFlow [90] is the best known SBI due to its widespread use by network equipment vendors and research groups. The application

plane implements and organizes the logic of network functions (e.g., load balancing and routing). The application plane communicates its network requirements to the CP through one or more north bound interfaces (NBI). The CP translates the application requirements and applies them to the network elements through SBI. The CP also defines one or more east west bound interfaces (EWBI) that enable the implementation of physically distributed and logically centralized control to cope with large-scale and wide-area networks through their logical segmentation. The management plane is orthogonal to the other SDN planes to carry out its operation, administration and maintenance via management interfaces such as OF-Config [91], and the open vSwitch database management protocol, (OVSDB) [92]. The knowledge plane is transversal to the aforementioned planes, and its objective is to make SDN networks more intelligent by using ML [62], [61].

D. Virtualization of Network Functions

Virtual logical networks are essential to deploy 5G services. NFV, defined by the European Telecommunications Standards Institute (ETSI), is a technology that facilitates the deployment of multiple logical network segments over a shared infrastructure [93]. NFV decouples the NFs from the hardware on which they run to extract the physical resources of the infrastructure, divide logically and assign them to the VNFs that make up the virtual logical networks [94]–[96].

The main components of NFV are: the NFV infrastructure (NFVI), the VNFs, and the NFV management and network orchestration (MANO) [97]. NFVI is the combination of physical and virtual resources (hardware and software) that make up the deployment environment of the VNF. A VNF can implement one or more network functions and run on virtual resources, such as virtual machines running services, or containers running microservices. As a VNF can be composed of other VNFs, it can be deployed in several virtual machines or containers. MANO includes [98]: (i) an NFV orchestrating to perform the composition of NFVI resources, (ii) a responsible VNF manager (VNFM) for the life cycle of VNFs; and (iii) a virtual infrastructure manager (VIM) in charge of virtualizing, monitoring, configuring, and controlling network resources (physical and virtual) [99].

E. Edge Computing in 5G Networks

Users of 5G networks require real-time contextual information, as well as, low-latency communications in a given geographical area. The integration of the MEC concept to 5G allows meeting these requirements. MEC provides high bandwidth, low latency, real-time RAN information, and location awareness as well as CC capabilities at the edge of the 5G network [100], [101], [57]. These capabilities are offered through small data centers deployed to meet the QoE required by end users of 5G networks [102].

A collaboration platform built through MEC-5G integration must be generic in functionality and hosting, and provide diverse subsystems of services designed for different verticals. Secure access must be available to these subsystems to facilitate application deployment [103]. MEC offers caching and

virtualization services to reduce the volume of data transmitted in the core of the 5G network, and thus improve the use of resources [104].

FC is an alternative to MEC to provide real-time contextual information and low latency communications in a specific geographic area [105]. One of the fundamentals of FC is to support IoT by providing secure connections and services between and for things, data, people and processes [106], [107]; FC is essential for MMTC. IoT implies a profound change in today's Internet towards a network of interconnected things (objects). Things/objects help integrate intelligence into our environment by supporting the collection of information when interacting with the world (physical or virtual) [108]. FC improves CC by bringing cloud applications, computing power, storage and communication capabilities closer to end users [109], [110]. CC involves both data center hardware and software used to provide infrastructure, platform and applications as a service [111]. Bringing CC closer to end users and things (*i.e.* at the edge of the 5G network) enables high-performance and low-latency services along with an efficient use of resources.

FC comprises three layers: IoT, CC, and fog [112], [25]. Each layer is responsible for providing different functionalities [113], [114], [115]: The CC layer consists of one or more data centers that offer infrastructure, platform, and *software* as a service. The fog layer includes one or more nodes, which can reuse wireless interfaces and, furthermore, coexist with network elements, such as SBs or femto cell routers. A fog node is an entity (physical/virtual) that includes its computing, storage, and communication capabilities. This layer can contain multiple sublayers depending on the application requirements. The IoT layer is responsible for sending and receiving data to and from the fog layer which usually performs a first stage of analysis or data processing and, if necessary, sends it for further analysis to the CC layer. The IoT layer includes devices such as sensors that can interact with each other forming *ad hoc* networks.

VI. GLOBAL DEPLOYMENT OF 5G

The commercial deployment of 5G technology started in late 2018 in Asia. A brief historical description of this deployment will follow next (see Fig. 5), by emphasizing the leading companies and operators, as detailed in [5]. South Korea conducted the first successful deployment of a 5G trial network in February 2018 during the Winter Olympic Games in PyeongChang. Following this initial effort, the commercial launch of 5G in this country was carried out at the end of the same year by the companies SK Telecom, Korea Telecom, and LG Uplus. In fact, South Korea is so far the country with the widest 5G coverage and the highest transmission rates [116].

China represents the world's largest market for 5G technology. The deployment of 5G in China began in October 2019 through China Mobile, Unicom and China Telecom. Huawei is one of the big promoters of 5G technologies in that country [117]. In the Asia-Pacific region, Australia launched its first commercial 5G network in May 2019 through Telstra. New Zealand did the same in December through Spark and



Fig. 5. Commercial deployment of 5G mobile technologies by country and operators (first ones to provide the services) between 2018 and 2020.

Vodafone, and the Philippines in June with Globe Telecom. The Asia-Pacific region is the geographic area with the greatest interest and development in 5G, and it is expected to have two-thirds of the world's subscribers by 2024 [118].

In Europe, the commercial deployment of 5G technology started in UK in May 2019, by the telecommunication companies EE, Vodafone, O2 and Three. In Switzerland, the deployment was carried out by Swisscom and Sunrise in May 2019, while in Italy, it was leaded by Vodafone and Telecom Italia in June of the same year. Germany was another of the first countries to operate commercial 5G networks through Deutsche Telekom and Vodafone, starting operation in July 2019. Romania, Austria, Ireland, Finland, and Hungary joined this commercial deployment in the course of that same year. Other European countries, such as Spain, Netherlands, Sweden, Denmark and Belgium launched their first commercial 5G networks in 2020.

North America is another large market for 5G technology. In this region, Verizon carried out the first commercial 5G launch in May 2019, while AT&T and T-Mobile stepped forward in the second half of the same year. In Canada, the leadership is held by Rogers Communications and Bell Mobility, whose 5G commercial deployments began in 2020. Another important market is the Middle East, where Kuwait, Saudi Arabia and Qatar began operating their first commercial 5G networks in the period from April to June 2019. In Africa, the leading country in this technology is South Africa, where Rain and Vodacom started their 5G networks in September 2019 [5].

The current state of 5G technology in Latin America is incipient. The first commercial 5G network was launched in Uruguay in April 2019, and it is operated by Antel with the support of Nokia. According to the report in [4], aside from Uruguay, only Brazil has commercial 5G mobile phone networks, which are operated by Claro. In Colombia, there are 5G fixed Internet networks through Directv, while other

countries, such as Argentina, Chile, Peru, and Mexico are in the early stages of investment and deployment. In the Caribbean area, Puerto Rico and the Virgin Islands have 5G networks since December 2019, which are operated by T-Mobile. As described in [119], one of the great challenges in Latin America is to improve the cellular infrastructure to achieve QoS and coverage levels similar to those of leading countries, such as Japan, United Kingdom and United States. In fact, according to [120], in the next ten years, the development of mobile telephony in Latin America will still focus on 4G technology, which is expected to reach 67 % of the region's population by 2025, which will lead 5G integration in the near future.

The global status of the 5G deployment can be visualized based on the report in [4], where until September 2020, there were 397 wireless technology operators from 129 countries or independent territories, which have invested in tests, acquired licenses, planned network development, or launched 5G networks. Furthermore, 101 operators from 44 countries or territories have generated one or more services compatible with 3GPP standards for 5G. A great tool to visualize the global expansion of this technology are the interactive 5G coverage maps produced by OOKLA [116].

The potential health and environmental consequences of 5G technologies must be analyzed in detail during the global deployments. These implications are associated, among others, with the following factors. First, the 5G backbone technologies in the physical layer (mmW communications, LS-AA, and small cell stations) use much higher transmission frequencies than previous generations [121]. Thus, the penetration levels into biological tissue will be lower, but other organs such as the skin and eyes could be affected [122]. Second, an increase in the carbon footprint can be generated by the raise in the number of connected devices, especially to accommodate industrial IoT, and by the growth in data centers to support

virtualization, as the ones used in FC or MEC to enable RAN virtualization [123]. To address these challenges, the scientific community has requested that governments must invest in environmental impact studies to measure the possible health risks related to the 5G deployment, and then generate regulatory policies that protect the population health [124]. In addition, researchers must continue their development towards cost-effective and sustainable 5G networks from the point of view of carbon footprint and energy efficiency [125], [126].

VII. MOBILE TELEPHONY BEYOND 5G

5G networks will transform the mobile communications landscape by enabling ultra-reliable and ultra-low-latency links, capable of connecting people and machines alike. These networks will be essential to develop and deploy the full potential of Industry 4.0, intelligent transportation systems, IoT, and smart cities [127]. However, there are applications whose requirements could exceed the projected 5G capacities. For example, while 5G networks are the preferred technologies for connected vehicles, there is no guarantee that such networks will be able to meet the low latency required by autonomous vehicles. Furthermore, 5G networks focus mainly on two-dimensional communications, that is, land mobile communications. However, the growth of the aeronautical and aerospace industry, evidenced by the proliferation of UAVs, nano-satellites, and high-altitude platforms, poses a three-dimensional communications scenario comprising land, air, and water environments. In addition, the development of augmented reality technologies and haptic interfaces is leading to truly interactive communications, which are not considered in the 5G requirements. Two examples of this type of technology are the haptic or tactile Internet [128], and holographic communications [129].

While 5G networks are in an early stage of development, the work towards the definition and visualization of the 6G of mobile telephone networks is already underway. In July 2018, the ITU established the Network 2030 working group, whose task was to analyze the challenges and opportunities in the evolution of communications networks towards the year 2030 and beyond [130]. This group identified seven use cases for 2030 networks, and among them stand out: holographic communications, tactile Internet for remote operations, integrated terrestrial-space network, and industrial IoT with *cloudification*. Although the vision of what 6G networks will be like is not clear yet, the expectation is that these networks will enable ultra-massive communications between machines, with extremely low latencies (between 10 and 100 μ s), extremely reliable links (frame error rates $\sim 1\text{-}10^{-9}$), extremely low power consumption (1 pJ/b energy efficiency), and broadband link support under very high mobility conditions (users mobility over 1,000 km/h) [131], [132].

To achieve these goals, a change in the paradigm of mobile network design is needed, in which communications are no longer directed to people or machines, but rather to data [133], [134]. Furthermore, the efficient handling of large volumes of information in 6G applications will require the network to adopt AI as one of its pillars [135]. Finally, transmission

systems must operate in frequency bands even higher than those of 5G technologies to reach peak transmission rates in the order of Tbps. Transmission technologies being considered for 6G include optical laser and visible light communications (VLC), radio frequency links in the THz band, supermassive MIMO systems, holographic beamforming, smart surfaces, and multiplexing techniques based on the angular orbital momentum of electromagnetic waves [131]–[134].

6G networks will impose additional challenges related to the management and orchestration of three-dimensional cellular networks that will include both BSs and MUs based on UAVs. These networks will also require the evolution of admission and planning protocols of logical network segments generated by new use cases requiring, for example, extremely high data speed and reliability, *i.e.* a combination of EMBB and URLLC in 5G. In summary, 6G networks will allow us to imagine a scenario in which smart cities proliferate. While this scenario could be perceived as distant, the path towards a future of truly smart communication networks is being traced by the commercial roll-out of 5G networks.

VIII. CONCLUSIONS

The deployment of 5G networks will allow to provide mobile services for unprecedented use cases, namely, EMBB, MMTC and URLLC. However, this technological revolution will be gradual, and in the forthcoming years, a co-existence of different communication services will be maintained, mainly with 4G networks. In general, the primary change brought by the 5G technology focuses on the network architecture through concepts such as logical segmentation, *softwarization*, and network virtualization. In Latin America, the penetration of 5G networks is still sparse, and a gradual investment in 5G infrastructure is expected to follow after the consolidation of 4G services. Despite this regional scenario, worldwide activities aiming at the conceptualization of 6G networks have already begun.

ACKNOWLEDGMENTS

The authors are grateful for the funding from CONACYT through a Basic Science grant No. 254637.

5G and Beyond: Past, Present and Future of the Mobile Communications

Carlos A. Gutierrez, *Senior Member, IEEE*, Oscar Caicedo, *Senior Member, IEEE*,
and Daniel U. Campos-Delgado*, *Senior Member, IEEE*

Abstract—The fifth-generation (5G) of mobile communications networks is emerging as a revolutionary technology that will accelerate the development of smart cities and the realization of the information society. This paper aims to provide an introduction to 5G for non-specialists, and a survey of this new technology for those already familiar with mobile communications, covering the conceptualization and the core technologies underpinning 5G networks. The paper also discusses the status of the commercial roll-out of 5G until 2020 from a worldwide perspective and gives a future view of mobile communications beyond 5G.

Index Terms—5G communications, Massive MIMO, Beamforming, mmWave communications, Mobile edge computing, Small cell stations, NOMA, network slicing.

I. INTRODUÇÃO

A telefonia móvel está em constante evolução à procura de satisfazer as demandas da sociedade por mais e melhores serviços de informação e comunicação. O curso deste processo de evolução é marcado pelo aparecimento de padrões que refletem os últimos avanços tecnológicos e a coexistência dos sistemas de radiocomunicações [1]. As grandes inovações que se desencadeiam com o surgimento de um novo padrão (ou um novo conjunto de padrões) marcam sucessivamente novos estágios na história da telefonia móvel conhecidos como gerações [2], [3]. No momento estamos em trânsito à *Fifth Generation* (5G) de redes da telefonia móvel [4], [5]. Esta nova geração promete serviços que irão melhorar as experiências do usuário por meio do acesso a vídeo de ultra alta definição, vídeo 360°, trabalho e computação suportada na nuvem e comunicações imersivas baseadas em realidade aumentada e virtual. A diferença das redes de *Fourth Generation* (4G), as Redes 5G oferecem suporte nativo para aplicativos direcionados às comunicações entre máquinas. Isso será fundamental para o desenvolvimento da Indústria 4.0, os sistemas de transporte inteligente, a *e-health*, e em geral, para todos os elementos envolvidos na conectividade das cidades inteligentes.

O caminho para 5G está sendo construído com base na visão que estabeleceu a *International Telecommunications Union* (ITU), sobre o futuro das *International Mobile Telecommunications* (IMT) para o ano de 2020 e além [6]. Os casos de uso e requisitos de desempenho de 5G foram definidos num relatório publicado em 2017 pelo setor das radiocomunicações

Carlos A. Gutierrez and Daniel U. Campos-Delgado are with Faculty of Sciences, and Daniel U. Campos-Delgado also with Instituto de Investigación en Comunicación Óptica, both in Universidad Autónoma de San Luis Potosí, S.L.P., Mexico, e-mails: cagutierrez@ieee.org, ducd@fcienicias.uaslp.mx.

Oscar M. Caicedo-Rendón is with University of Cauca, Popayán, Colombia, email: omcaicedo@unicauca.edu.co.

*Corresponding author

da ITU [7]. De acordo com esse relatório, as redes 5G atingirão densidades de conectividade dez vezes superior as atingidas pelas redes 4G, taxas de transmissão de pico vinte vezes maiores e o tráfego por área cem vezes maior. Além disso, os tempos de latência de 5G serão até dez vezes inferiores aos de 4G [8], [9]. Segundo o caminho pela ITU, o *3rd Generation Partnership Project* (3GPP) desenvolveu padrões para tecnologias 5G. O primeiro deles foi publicado em 2017, a tecnologia associada com esse padrão é conhecida como *5G-New Radio* (5G-NR) [10].

O lançamento comercial de 5G-NR começou no final de 2018 [5]. No entanto, a visão da ITU de 5G está longe de ser atingida. Além do aspecto tecnológico, é fundamental a realização de atividades de pesquisa básica e aplicada que permitam o uso dessas novas tecnologias para resolver problemas sociais, econômicos e ambientais no mundo. Na América Latina, implicações da política regulatória e a liberação de novas faixas do *Electromagnetic Spectrum* (EMS) devem ser investigadas minuciosamente para que a transformação promovida por 5G reduza a exclusão digital na região.

Este artigo apresenta uma revisão dos antecedentes, a situação atual, e o futuro das redes 5G visando divulgar as áreas de oportunidade oferecidas por esta tecnologia e chamar a atenção para questões técnicas e regulatórias que ainda estão por ser resolvidas. As origens da telefonia móvel e as principais inovações trazidas por gerações antes de 5G são revisados na Seção II. Os casos de uso, os requisitos técnicos e a arquitetura de 5G são discutidos na Seção III. As Seções IV e V revisam as principais tecnologias de transmissão e gerenciamento de rede para 5G. O estado atual da implantação comercial 5G é abordado na Seção VI. A Seção VII conclui o artigo com uma perspectiva do telefone móvel além de 5G. Redes 5G foram objeto de vários artigos de revisão, sobressaindo os disponíveis em [8], [9], [11]–[13]. Este artigo complementa a literatura 5G atual fornecendo uma revisão abrangente de ambas as tecnologias para a *Radio Access Network* (RAN), como as que lidam com a *Core Network* (CN). Além disso, este artigo oferece uma visão global da implantação atual de 5G e iniciativas para estender as capacidades 5G com foco em 6G. Já que a maioria das siglas usadas nas comunicações de telefones celulares são regularmente identificados em inglês, essa convenção será usada em todo o artigo.

II. ANTECEDENTES DE 5G

As origens da telefonia móvel remontam à invenção do telégrafo sem fio no final do século XIX, e suas aplicações na indústria naval como instrumento de comunicação entre navios

e estações costeiras [14]. A telegrafia sem fio demonstrou a viabilidade das comunicações a longa distância com base na transmissão de ondas eletromagnéticas que se propagam livremente na atmosfera terrestre. O progresso feito no início do século vinte na área de eletrônica – desencadeada pela invenção do tubo de vácuo – permitiu aplicar os princípios da telegrafia sem fio na transmissão comercial e telefonia móvel de natureza militar e policial [15].

O primeiro sistema de telefonia móvel para uso civil foi o *Mobile Telephone System* (MTS), introduzido em 25 cidades nos Estados Unidos em 1946 [15]. O MTS funcionava a partir de técnicas *push-to-talk*. O serviço do MTS era fornecido por uma estação central que atendia a todos os *Mobile Users* (MU) da cidade. Esta centralização do serviço, junto com restrições estritas na largura de banda, impõe a qualquer sistema sem fio para regular o compartilhamento de EMS, limitaram a capacidade desta rede consideravelmente. O conceito de uma rede de telefonia celular móvel desenvolvido na Bell Labs durante a década de quarenta [16] permitiu a superação de problemas de saturação da MTS e foi o catalisador para a telefonia móvel comercial. A ideia por trás desse conceito é aumentar a capacidade do sistema, dividindo a área de cobertura em células de extensão menor, cada uma servida por uma *Base Station* (BS) conectada à rede telefônica pública. Isso permite que diferentes canais de frequência de rádio sejam atribuídos a células adjacentes para reduzir a interferência entre elas, e reutilizar canais em células que estão suficientemente distantes uns das outras, de modo que a interferência entre eles seja insignificante graças à atenuação experimentada pelas ondas eletromagnéticas se propagando na atmosfera. A arquitetura de rede celular é a base da telefonia celular para uso civil, por isso também é conhecido como telefonia celular. Cada uma das gerações de redes de telefonia celular se desenvolveu ao longo de um período de dez anos, e a transição para uma nova geração sempre foi motivado pela necessidade de melhorar a cobertura, alcançar taxas de transmissão maiores e oferecer novos serviços móveis [3], [17], como mostrado na Fig. 1.

A *First Generation* (1G) de redes celulares entrou no mercado no final da década de setenta, sendo o Japão o primeiro país a implantar uma rede 1G. A partir de 1981 as redes 1G foram implementadas nos países nórdicos: Noruega, Suécia, Dinamarca e Finlândia. Em meados dos anos 80 já havia redes 1G no Reino Unido, Estados Unidos, Canadá e México. Esta primeira geração foi concebida para serviços de voz usando tecnologia analógica, por isso apresentou uma baixa eficiência espectral. Além disso, pela falta de acordos internacionais para a operação de redes 1G houve uma ampla variedade de padrões operacionais incompatíveis. Como resultado, o serviço era restrito à cobertura nacional e até regional. Apesar da incompatibilidade entre as diferentes redes 1G, todas usaram a tecnologia *Frequency Division Multiple Access* (FDMA) e as ligações foram veiculadas por comutação de circuitos. No Japão e nos Estados Unidos, a faixa operacional reservada para 1G era 800 MHz. Enquanto isso, na Suécia e no Reino Unido reservaram a faixa de 900 MHz, e na Alemanha e na França as faixas de 450 e 200 MHz, respectivamente [2].

A *Second Generation* (2G) de redes celulares surgiu no início dos anos 90 e foi caracterizada pela mudança para a

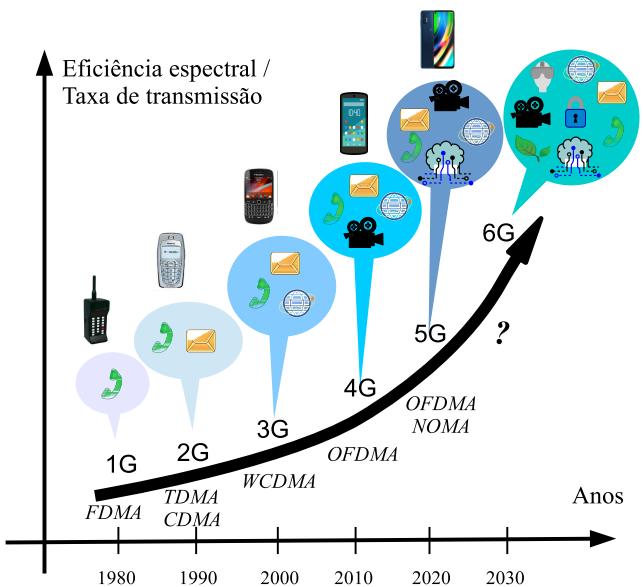


Fig. 1. Evolução da telefonia celular.

tecnologia digital. A eficiência espectral melhorou graças à adoção das técnicas *Time-Division Multiple Access* (TDMA) e *Code Division Multiple Access* (CDMA), bem como a introdução de técnicas de codificação de canal. Uma inovação importante foi a inclusão do *Short Message Service* (SMS). As redes 2G tiveram um grande sucesso comercial graças ao desenvolvimento de escopo internacional, por exemplo, o *Global System for Mobile Communications* (GSM) e o *Interim Standard 95* (IS-95) [18]. Ambos os padrões transmitem nas faixas de 810 e 960 MHz em uma arquitetura de circuito comutado. A velocidade teórica de transmissão de dados de 2G é 115,2 kbps (com atualização *High-Speed Circuit-Switched Data* de GSM), e a digitalização do serviço possibilitou a introdução de dispositivos compactos com uso eficiente de baterias. As redes 2G transmitem dados de maneira ineficiente porque usam comutação de circuitos. Este tipo de comutação é adequado para tráfego de voz, mas não para tráfego de dados.

A transição para a *Third Generation* (3G) de redes celulares começou no início do ano 2000. Uma das grandes inovações do 3G foi a capacidade de desempenho dos links por comutação de circuitos e pacotes; estes são os últimos necessários para serviços baseados no *Internet Protocol* (IP). Esta geração permitiu a incorporação de novos serviços móveis, como o acesso para a Internet e chamadas de vídeo. A tecnologia CDMA é fundamental para os sistemas de 3G e seus principais padrões são a *Wideband CDMA* (WCDMA) e CDMA 2000 [19]. Ambos os padrões são globais em escopo e teoricamente fornecem velocidades de transmissão de até 14,4 Mbps no *downlink* (DL) [2]. Com as redes de 3G veio o conceito de Telefones “inteligentes”, com uma interface visual mais atraente e a incorporação de aplicativos móveis.

O lançamento das redes de 4G começou no início de 2010, oferecendo taxas de máximas de transmissão de 300 Mbps no DL e 75 Mbps no *uplink* (UL), embora com as atualizações mais recentes, essas redes alcancem velocidades de até 1 Gbps no DL [20]. As tecnologias base de 4G são *Orthogonal*

gonal Frequency-Division Multiple Access (OFDMA), Single Carrier FDMA (SC-FDMA), Scalable OFDMA (SOFDMA), e Multiple-Input Multiple-Output (MIMO). Os principais padrões 4G são Long Term Evolution (LTE) e sua versão avançada, LTE Advanced (LTE-A) [21]. As inovações das redes 4G tornaram possível o fornecimento de serviços de voz, dados e multimídia onipresente por meio de assinatura, bem como gerar aplicativos de vídeo de alta definição e plataformas para jogos online interativos. A tecnologia 4G também permitiu gerar soluções totalmente móveis com base em IP, atingindo tempos de latência de até 20 ms.

III. CONCEITUAÇÃO DE COMUNICAÇÕES 5G

A. Casos de Uso e Requisitos de Desempenho

As redes 4G expandirão e continuarão em operação por vários anos apoiando os serviços de comunicações orientados para as pessoas [20]–[22]. De acordo com a última versão do Ericsson Mobility Report, espera-se em 2026 uma coexistência de diversas tecnologias para atender usuários móveis, sendo as predominantes: 5G, LTE, WCDMA / HSPA e GSM / EDGE [23]. No entanto, a situação atual também exige serviços orientados para comunicação entre máquinas, a fim de apoiar aplicativos para a Indústria 4.0, os sistemas de transporte inteligente, e a automação residencial. Tais aplicativos representam novos desafios, como garantir o serviço em áreas de cobertura com alta densidade de dispositivos interconectados; transmissão de baixa latência para aplicativos sensível ao atraso; suporte de enlaces de banda larga em condições altamente móveis, semelhantes às encontradas em trens de alta velocidade; e a reconfigurabilidade da rede para uma gestão flexível e eficiente dos seus recursos. As redes 5G buscam resolver esses desafios [11], [12]. A ITU definiu três casos de uso para essas novas redes:

- *Enhanced Mobile Broad-band (EMBB)*: aplicativos centrados no usuário final caracterizados por exigir acesso a conteúdo multimídia que precisa de altas taxas de transferência de dados com velocidades de pico de até 20 Gbps e velocidades percebida pelo usuário de até 100 Mbps, além de suporte de mobilidade de até 500 km/h e eficiência espectral três vezes maior do que o fornecido por 4G. Exemplos de aplicações EMBB são voz e vídeo móvel de ultra alta definição, bem como realidade virtual e aumentada para experiências de comunicação envolvente.
- *Massive Machine-Type Communications (MMTC)*: aplicativos envolvendo um grande número de dispositivos conectados, transmitindo um pequeno volume de dados com baixa sensibilidade ao atraso e pouco consumo de energia. Os aplicativos deste tipo incluem veículos conectados, cuidado de animais/pessoas, e redes de atuadores e sensores sem fio. O objetivo para este caso de uso, é ter uma capacidade de tráfego de 10 Mbps/m², uma densidade de conexão de 10⁶ dispositivos/km² e uma eficiência energética da rede 100 vezes maior que 4G.
- *Ultra Reliable Low Latency Communications (URLLC)*: aplicativos caracterizados por operar em tempo real, como aqueles relacionados à segurança do transporte,

TABELA I
INDICADORES DE DESEMPENHO PARA REDES 4G E 5G (DADOS EXTRAÍDOS DE [7]).

Indicador	4G	5G
Taxa máxima de transmissão UL	500 Mbps	10 Gbps
Taxa máxima de transmissão DL	1 Gbps	20 Gps
Eficiência espectral máxima UL	6.75 b/s/Hz	15 b/s/Hz
Eficiência espectral máxima DL	15 b/s/Hz	30 b/s/Hz
Densidade de conectividade	10 ⁵ disp./km ²	10 ⁶ disp./km ²
Mobilidade	350 km/h	500 km/h
Capacidade de tráfego por área	0.1 M b/s/m ²	10 M b/s/m ²
Latência	20 ms	1 ms

cirurgias remotas e automação de processos industriais, que requerem latências baixas de até 1 ms.

A Tabela I resume os principais requisitos de desempenho desses três casos de uso [7], e os compara com seus correspondentes 4G. As principais tecnologias identificadas pela ITU para atender a esses requisitos são revisados em [24].

B. Arquitetura da Rede 5G

Os requisitos de altas taxas de transmissão e suporte com uma alta densidade de conexões, podem ser fornecidos para nível de camada física e controle de acesso ao meio usando técnicas de transmissão eficientes. No entanto, os requisitos de baixa latência, alta mobilidade e capacidade de tráfego por área envolvem estratégias na CN. Assim, uma das principais inovações do 5G está na arquitetura da rede, cujo desenho é muito mais flexível do que o de gerações precedentes graças à implementação de conceitos de segmentação lógica, *softwarization*, e virtualização da rede e seus serviços.

A Fig. 2 mostra um diagrama da arquitetura das redes 5G, que consiste na RAN, a CN, e terminais de usuário. Esses elementos são representados como *Network Functions* (NF) e seus respectivas interfaces. A Tabela II descreve resumidamente cada NF da arquitetura 5G. Diferentes opções arquitetônicas para implantar o RAN são apresentados em [13] e [25], e as arquiteturas para a implantação da CN são estudadas em [26] e [27].

As redes 5G seguem um padrão arquitetônico baseado em *Service-Based Architecture* (SBA). A SBA contribui importantes benefícios para a estrutura da rede, por exemplo:

- (i) *Atualização fácil*: cada serviço pode ser modificado com impacto mínimo no resto.
- (ii) *Extensibilidade*: a adição de uma nova funcionalidade envolve a adição de um novo serviço e seu correspondente interfaces.
- (iii) *Modularidade*: um sistema 5G é composto por serviços de granularidade fina (eles implementam uma única NF) e baixo acoplamento (usam interfaces padronizadas) conhecidos como microsserviços [28], promovendo assim a segmentação lógica de rede.
- (iv) *Abertura e reutilização*: os serviços podem ser expostos e invocados por meio de interfaces leves e padronizadas, incentivando assim o reaproveitamento de funcionalidades.

Esses benefícios permitem que a arquitetura 5G seja flexível, programável e fácil de automatizar, bem como promove uma redução significativa de custos graças a implantação

TABELA II
FUNÇÕES DE REDE DO 5G.

Função de rede	Descrição
Função de aplicativo	Interage com o núcleo da rede 5G, principalmente PCF e NEF, para auxiliar no gerenciamento de tráfego.
Função de gerenciamento de acesso e mobilidade	Ponto final do estrato de não-acesso, inclui funções de registro, conexão, gerenciamento de mobilidade e acessibilidade.
Função do servidor de autenticação	Suporta o processo de autenticação dos UE.
Rede de dados	Função que representa a conexão a qualquer rede de dados.
Função de análise de dados de rede	Analisa dados das NF.
Função de exposição de rede (NEF)	Expõe as funcionalidades e eventos das NF, permitindo sua comunicação com aplicações externas.
Função de repositório de funções de rede	Registra as NF e permite seu descobrimento para facilitar a composição de serviços.
Função de seleção de segmento lógico de rede	Seleciona o conjunto de instâncias de rede lógica que atendem um UE.
Função de controle de política (PCF)	Superta a definição de políticas que são utilizadas pelo CP para governar o comportamento da rede.
Rede de acesso rádio	Fornecce acesso rádio que conecta com as NF de UP, e o CP do núcleo da rede 5G.
Função de gerenciamento de sessão	Administra o estabelecimento, modificação e liberação de sessões de dados e aloca endereços IP aos UE.
Gerenciamento unificado de dados (UDM)	Inclui suporte para a geração de credenciais de autenticação, identificação de usuário e gerenciamento de assinaturas.
Repositório unificado de dados	Permite que UDM e PCF obtenham informação de assinatura e políticas, respectivamente.
Equipo de usuário (UE)	Permite aos usuários finais usarem os serviços fornecidos pela rede 5G.
Função do plano do usuário	Administra mobilidade e QoS entre tecnologias de acesso rádio, direciona pacotes e gera relatórios de tráfego.

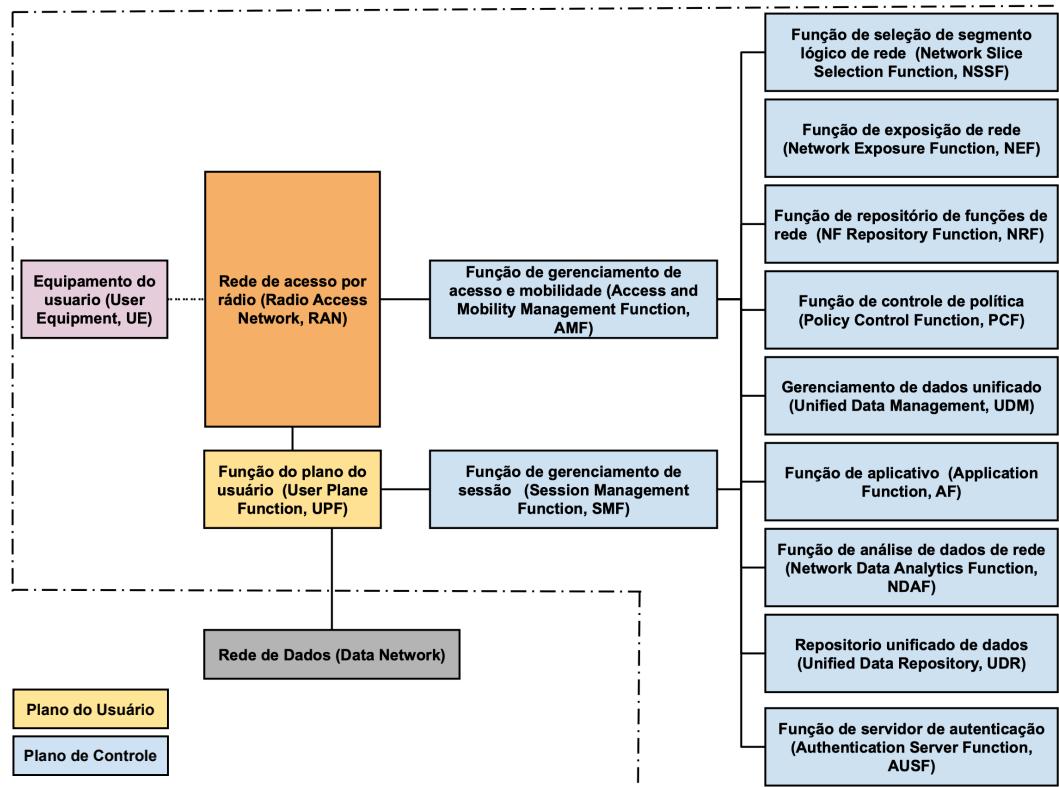


Fig. 2. Arquitetura da rede 5G.

de *Virtual Network Functions* (VNF) em pequenos, médios ou grandes centros de dados.

O conceito *Control and User Plane Separation* (CUPS) é outro dos pilares da arquitetura de redes 5G. CUPS, embora não seja um novo conceito, é fundamental no 5G porque permite a separação do *Evolved Packet Core* (EPC) de 4G em funções de rede do *User Plane* (UP) e funções de rede do *Control Plane* (CP) [29]. As NF do UP podem ser implantadas perto do aplicativo que eles suportam para satisfazer requisitos de latência e taxa de transferência, essenciais para os casos de uso URLLC e EMBB, respectivamente. Por outro lado, as NF do CP podem operar centralmente para satisfazer os requisitos de confiabilidade. O UP carrega apenas tráfego de dados,

enquanto o CP carrega a sinalização da rede. As NF dos UP e CP interagem entre si por meio de interfaces padronizadas. Um serviço 5G é implementado através da composição de cadeias de NFs (virtualizadas ou não) pertencentes aos CP e UP.

Tecnologias fundamentais para a realização da 5G CN são *Mobile Edge Computing* (MEC), *Fog Computing* (FC), *Cloud Computing* (CC), *Software-Defined Networking* (SDN), e *Network Function Virtualization* (NFV). Essas tecnologias são apresentadas na Seção V.

IV. TECNOLOGIAS CHAVE PARA 5G RAN

A transformação tecnológica proporcionada pelas redes 5G é melhor compreendida ao identificar as dimensões da

inovação naquilo que determina o curso evolutivo dos sistemas de comunicações sem fio. Essas dimensões são [30]: 1) disponibilidade de EMS, 2) eficiência espectral, 3) eficiência espacial, e 4) eficiência do sistema. As primeiras três dimensões são descritas nesta seção, enquanto a quarta é discutida na seção V.

A. Comunicações por Ondas Milimétricas

A primeira dimensão da inovação refere-se à faixa (ou faixas) do EMS que a rede tem à sua disposição para estabelecer os links sem fio entre os BS e os MU. Os elementos relevantes desta dimensão são a frequência de operação e largura de banda, que determinam o alcance e capacidade (em termos da máxima taxa de transmissão) do enlace, respectivamente. Uma das principais inovações das redes 5G está na inclusão de enlaces em frequências entre 20 e 80 GHz, ou seja, enlaces na faixa das ondas *millimeter Wave* (mmW) [31]. Isto visa apoiar serviços exigindo taxas de transmissão de pico na ordem de Gbps, conforme esperado para o caso de uso EMBB. Essas velocidades de transmissão são difíceis de alcançar em faixas de frequência menores, como as utilizadas por 4G, em que a largura da banda é limitada a algumas dezenas de MHz, enquanto que a faixa mmW oferece larguras de banda de vários GHz.

Embora o aumento da frequência de operação permite larguras de banda maiores, também reduz o alcance do sistema, porque as ondas eletromagnéticas se tornam mais suscetíveis à atenuação de absorção como sua frequência aumenta. A faixa MmW também não é adequada para comunicações indiretas, ou seja, comunicações em que não há *Line of Sight* (LOS), entre transmissor e receptor, devido à atenuação excessiva que sofrem as ondas mmW ao interagir com objetos com os que enlaces indiretos podem ser criados. Essas limitações podem ser resolvidas através do emprego de arranjos de antenas para reduzir o impacto da atenuação que são convenientes quando um grande número de dispositivos precisa ser conectado como no caso de uso MMTC. Além disso, ainda existem inúmeras questões relacionadas à disseminação de sinais nesta faixa de frequência. Por exemplo, [32] sinaliza a necessidade de realizar medições de campo para aumentar a informação empírica sobre as perdas de propagação, a dispersão temporal, a dispersão em frequência, a propagação angular, e a probabilidade de LOS dos enlaces na faixa de ondas mmW.

Por outro lado, em [33] é mostrado que o canal móvel de propagação se comporta como um sistema não estacionário conforme a largura de banda do sinal transmitido aumenta. Os efeitos dessa não estacionariedade sobre os benefícios de interface aérea das redes 5G não foram estudados em detalhe. Em [34] observa-se que a implantação da tecnologia na faixa mmW seguirá ritmos diferentes nas grandes e pequenas cidades. Isso pode aumentar a lacuna entre grandes cidades e áreas rurais, de modo que é importante buscar alternativas para reduzir o impacto desta implantação arritmica.

B. Matrizes de Antenas em Grande Escala

A segunda dimensão de inovação refere-se ao uso eficiente do EMS disponível, medido considerando a quantidade bits

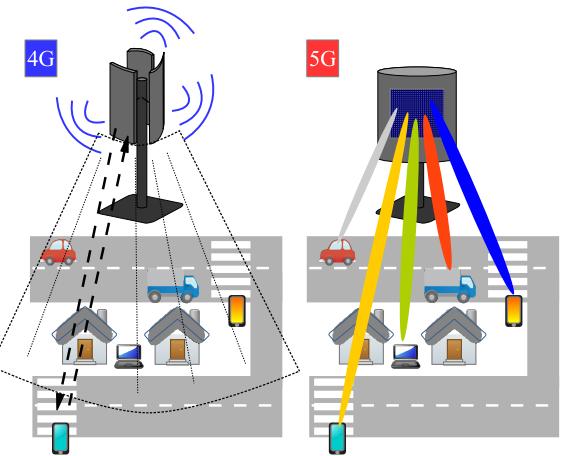


Fig. 3. Transmissão em uma rede 4G sem conformação de feixe e em uma rede 5G com conformação de feixe.

de informação que podem ser transmitidos para cada Hz disponível. Redes 5G buscam aumentar a eficiência espectral de uma forma nunca antes visto usando *Large-Scale Antenna Arrays* (LSAA). O uso dessas técnicas de transmissão é viável e oportuno considerando que 5G incluirá enlaces entre o BS e os MU na faixa mmW. Para esta faixa de frequência é possível construir antenas de um perfil compacto que podem ser integradas a arranjos compostos de centenas ou mesmo milhares delas.

As técnicas LS-AA que figuram no contexto de 5G são as de conformação de feixe [35]–[38] e *massive MIMO* (mMIMO) [39]–[41]. A conformação do feixe, usa LS-AA para obter um padrão de radiação combinado de grande diretividade, o que permite estender o alcance de conectar e obter uma melhor *Signal to Noise Ratio* (SNR) [42] sem aumentar a interferência para outros usuários. A conformação de feixe também é usada no *Space Division Multiple Access* (SDMA). Usando SDMA, o BS pode se comunicar simultaneamente com vários MU no mesmo canal de radiofrequência separando os usuários por meio de enlaces focados, cada um com uma alta proporção no *Signal to Interference Ratio* (SIR) [43]. A Fig. 3 mostra, em primeiro lugar, um cenário 4G em que o BS comunica-se com vários MU usando antenas setoriais com padrões de radiação moderadamente diretivos. Segundo, um Cenário 5G em que o BS se comunica simultaneamente com vários MU usando enlaces focados. A conformação de feixe também é usada para a multiplexação espacial de dados [44]. Alguns problemas de pesquisa relacionados com essas tecnologias de arranjos de antenas estão no desenvolvimento de técnicas de pré-codificação híbrida (análogo e digital) de baixa complexidade e alta eficiência energética, bem como a simplificação das cadeias de RF [36], [44].

Os sistemas mMIMO empregam LS-AA em ambos os lados do link para realizar multiplexação espacial de grandes volumes de dados. A principal diferença entre mMIMO e MIMO consiste em que o primeiro incorpora *Multi-User Detection* (MUD) para aumentar o fluxo de informações [40]. O sucesso de MUD depende da capacidade do receptor para cancelar a interferência gerada por outros usuários. Para isso,

são utilizadas técnicas de pré-codificação, que exigem uma estimativa precisa da *Channel State Information* (CSI) de cada MU ativo. Essas informações podem ser obtidas através de sinais piloto, ou por *feedback* dos MU [45].

Um problema por resolver no contexto de mMIMO está em seu funcionamento num esquema de *Frequency-Division Duplexing* (FDD) [40]. A maioria dos trabalhos em mMIMO assumem um sistema operando sob um esquema de *Time-Division Duplexing* (TDD) com a intenção de evitar os problemas de estimativa de canal no DL [39]. No entanto, FDD pode acelerar a implantação da tecnologia mMIMO em redes 5G, uma vez que os esquemas FDD foram amplamente utilizados na telefonia móvel. O desenvolvimento de algoritmos de detecção e pré-codificação de baixa complexidade também é um problema por resolver para a realização de mMIMO [40].

C. Acesso Múltiplo não Ortogonal

As técnicas de acesso ao meio também são fundamentais para aumentar a eficiência espectral. As estratégias de acesso ao médio anteriores ao 5G usam o princípio de ortogonalidade, seja em TDMA, FDMA ou CDMA para evitar ou limitar a interferência entre os usuários. Essas estratégias são conhecidas como *Orthogonal Multiple Access* (OMA). Em OMA, a capacidade de usuários ativos depende dos recursos ortogonais disponíveis. Por outro lado, o *Non-Orthogonal Multiple Access* (NOMA) permite que vários usuários usem simultaneamente o mesmo canal de comunicação, independentemente de ocorrer interferência de acesso múltiplo [46]–[49]. NOMA atenua essa interferência usando estratégias de MUD. Em geral, NOMA ajuda a melhorar a eficiência espectral da comunicação sem fio e a taxa de transmissão dos usuários ativos, permite a massificação da conectividade, e reduz o tempo de latência [48]. Eses benefícios são obtidos em troca de aumentar a complexidade dos receptores para dar-lhes a capacidade de realizar o MUD. Dentro da filosofia NOMA, existem várias implementações no nível de bits ou símbolos, mas todas procuram atribuir assinaturas para MUs em função da potência de transmissão, códigos de espalhamento, codificação ou intercalação [47], [49]. NOMA pode ser aprimorado considerando uma estratégia de transmissão e recepção MIMO. Combinando ambas as estratégias é possível melhorar a capacidade do canal comparando com MIMO-OMA [50], [51].

No NOMA por atribuição de potência [46], o poder de transmissão é o fator de diversidade no domínio do tempo que permite identificar as informações de cada usuário; assim, os usuários mais distantes da BS irão transmitir com uma potência maior do que um telefone celular próximo. Os receptores empregam *Successive Interference Cancellation* (SIC) para separar iterativamente as informações na interferência de acordo com a potência de transmissão de cada usuário. Esta estratégia pode ser visualizada na Fig. 4. Os principais problemas de pesquisa no NOMA são relacionadas ao desenvolvimento de técnicas de decodificação e gestão eficiente de recursos de transmissão (e.g., a potência e largura de banda), bem como com as estratégias com *feedback* de informação (como CSI) entre BS e MU, que permitam o cancelamento da interferência, e uma redução na complexidade do receptor [52].

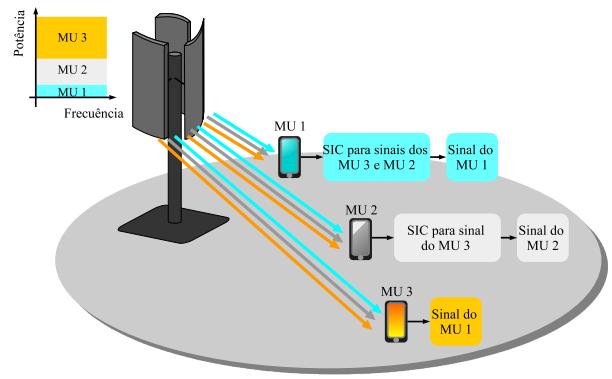


Fig. 4. Princípio NOMA para alocação de potência.

Além de tecnologias anteriores, eficiência espectral também depende de esquemas de modulação avançados e novas formas de onda para a transmissão de informações. As tendências do 5G neste campo não são revisadas neste artigo devido a limitações de espaço, mas o leitor pode encontrar informações sobre isso em [24].

D. Células de Curto Alcance

A terceira dimensão da inovação, relacionada com a eficiência espacial, inclui melhorias destinadas a apoiar um maior número de usuários conectados por unidade de área. A estratégia mais simples – embora eficaz, para melhorar a eficiência espacial é segmentar a área de cobertura, como é sugerido pelo conceito de rede celular. De fato, sob medida que aumenta a densidade de resultados de usuários conectados é conveniente reduzir o tamanho das células. Outra inovação importante de 5G reside precisamente na densificação exibindo um grande número de células pequenas, cujo alcance é inferior a 100 metros.

Vários benefícios associados às células de curto alcance são apresentados em [30], entre os quais se destaca a capacidade de fornecer cobertura em áreas de alto tráfego, bem como por dentro de edifícios, e aumentar o fluxo efetivo da rede em faixas de radiofrequência abaixo de 6 GHz. No entanto, um grande número de células (pico e femto) leva a um crescimento significativo no consumo de energia [53] [54], abrindo a necessidade de novos esquemas de eficiência de energia [55] [56]. A implantação de células de curto alcance também enfrenta obstáculos relacionados principalmente com questões regulatórias e operacionais, como altas taxas devido ao uso de infraestrutura pública (como postes de serviços de iluminação) para instalar essas células e diferentes critérios para medir o risco para a saúde da exposição aos campos eletromagnéticos de alta potência [34]. Em adição a estes problemas, existem outros relacionados à interconexão e administração de um grande número de células de curto alcance; alguns deles e suas soluções são apontados.

V. TECNOLOGIAS CHAVE PARA O NÚCLEO DO 5G

A quarta dimensão da inovação, com foco na eficiência do sistema, envolve principalmente o núcleo da rede 5G. A eficiência buscada está relacionada, por exemplo, à redução

dos tempos de latência, aumento de taxas de transmissão, a redução da perda de pacotes e a diminuição do consumo de energia.

A. Segmentação Lógica da Rede

Um dos objetivos do 5G é oferecer uma ampla gama de serviços, cada um com diferentes requisitos de *Quality of Service* (QoS) abrindo uma gama de oportunidades para aplicações de vários segmentos da indústria (também conhecidos como verticais) [57] [27]. Para atingir esse objetivo, os sistemas 5G devem ser capazes de fornecer pedaços de rede aplicando o conceito de segmentação lógica [58], [59] que tem como princípios a automação da operação de rede, a alta confiabilidade, a escalabilidade e isolamento dos segmentos pertencentes a diferentes clientes, o suporte para *softwarization*, programabilidade e virtualização, a abstração hierárquica e personalização de segmento, e a elasticidade dos recursos da rede [60]. As tecnologias chave para atingir esses princípios são *Artificial Intelligence* (AI), SDN, NFV, MEC, FC, e CC.

A segmentação lógica nos sistemas de 5G deve superar de forma inteligente vários desafios, como a criação de novos algoritmos de admissão e *scheduling* destinados a atribuir recursos para os segmentos durante sua criação e operação, a alocação das (V)NF na infraestrutura (física ou virtual), o controle de tráfego intra-segmento, o gerenciamento fim a fim dos segmentos, o isolamento entre segmentos, a gestão da mobilidade e orquestração das (V)NF que constituem os segmentos, bem como sua segurança e privacidade.

B. Inteligência Artificial

Diversas áreas da IA, entre as quais se destaca *Machine Learning* (ML), são úteis para desenvolver segmentos lógicos cognitivos que são fundamentais para alcançar redes 5G autodirigidas. Uma rede autodirigida deve implantar de forma autônoma o ciclo Perceber-Planejar-Decidir-Agir, que deve ser acompanhado de aprendizagem e *feedback*, tanto da rede como um todo como dos seus segmentos constitutivos [61] [62]. Um estudo sobre o uso de técnicas de ML supervisionadas, não supervisionadas, e híbridas em redes de comunicação está disponível em [63]. As técnicas de ML que recentemente geraram maior interesse no domínio de redes para automatizar a tomada de decisões são as de *Reinforcement Learning* (RL) e *Deep Reinforcement Learning* (DRL) [64].

Em RL, um agente toma decisões periodicamente considerando estados (e.g., conjunto de roteadores disponíveis para calcular uma rota) e ações (e.g., enviar tráfego ao longo da rota calculada). O agente observa o resultado (expresso como recompensa, por exemplo, tráfego otimizado em termos de atraso) de sua interação com o ambiente (e.g., uma rede 5G) e, em seguida, ajusta automaticamente sua estratégia para alcançar uma política ideal (e.g., otimizar o atraso num segmento lógico do tipo URLLC) [65] [66] [67]. As abordagens de RL, como Q-learning [68] e *State–Action–Reward–State–Action* (SARSA) [69], convergem lentamente para a política ideal quando precisam explorar e adquirir conhecimento sobre conjuntos de estados e ações grandes, dificultando o uso em redes 5G de grande escala.

Deep Learning (DL) foi usado recentemente como uma ferramenta para superar as limitações de RL, dando origem a outro tipo de aprendizagem denominado DRL [70]. Técnicas de aprendizado profundo aprendem descobrindo estruturas complexas nos dados e constroem seus modelos computacionais usando várias camadas de processamento; uma *Deep Neural Network* (DNN) pode criar vários níveis de abstração para representar os dados [71] [72]. As estruturas típicas focadas neste tipo de aprendizagem são as *Feedforward Neural Network* (FNN) e *Recurrent Neural Network* (RNN). Em FNN, não há ciclos ou loops na rede neural. Portanto, os dados são movidos numa direção: desde os nódulos de entrada, através dos nódulos escondidos, e para os nódulos de saída [73]. As *Convolutional Neural Networks* (CNN) são o modelo FNN mais conhecido com uma grande variedade de aplicações, incluindo roteamento dinâmico [74]. Ao contrário das FNN, uma RNN é recursiva. As conexões entre os neurônios produzem ciclos específicos, o que significa que uma saída depende não apenas de suas entradas imediatas, mas também do estado neuronal da etapa anterior [75]. Uma RNN é projetada para usar dados sequenciais, quando o estágio atual está relacionado às etapas anteriores. *Long Short-Term Memory* (LSTM) é uma das RNN mais representativas [76] e é útil, por exemplo, para previsão de tráfego de rede [77].

DRL usa redes neurais profundas para melhorar velocidade e desempenho dos algoritmos RL [78]. De uma perspectiva de alto nível, DRL implementa uma DNN para derivar a política ideal em vez da tabela Q usada pelo *Q-learning*. Os algoritmos DRL mais relevantes são [64] [79]: DQL, DDQL, DQL com *Experiência Priorizada Replay*, DDQN, *Distributional Deep Q-Learning*, *Deep Q-Learning com Noisy Nets*, e *Rainbow Deep Q-Learning*. Recentemente, o DRL tem sido usado para resolver problemas no contexto de *Internet of Things* (IoT), HetNets, UAV, SDN e NFV, todos eles envolvidos na concepção de 5G. A integração do DRL em 5G poderia revolucionar os mecanismos de gerenciamento de segmento lógico, especialmente aqueles relacionados ao planejamento e compartilhamento de recursos. Esta integração permitiria evoluir as técnicas baseadas em modelos para aquelas sem um modelo, que aprendem profundamente interagindo com o meio ambiente para satisfazer tanto os *eXperience Level Agreements* (XLA) quanto os *Service Level Agreements* (SLA); esses acordos estão relacionados a indicadores de *Quality of Experience* (QoE) e QoS, respectivamente.

Devido à complexidade dos sistemas de comunicação de 5G, os seguintes pontos serão fundamentais para a segmentação de rede lógica com base em DRL: (i) a definição dos conjuntos de estados e ações deve abranger a dinâmica dos segmentos de rede; (ii) o modelo dos conjuntos de estados e ações deve permitir a redução do custo de aprendizagem; (iii) a recompensa deve ser o mais simples possível para facilitar o processo de aprendizagem; (iv) a quantificação da recompensa deve incluir tanto o tempo de cálculo quanto o tempo de obtenção (a interação com o ambiente não é instantânea); e (v) os resultados fornecidos pelos agentes DRL devem ser compatíveis com a escala de tempo em que eventos ocorrem e são gerenciados em uma rede 5G.

C. Redes Definidas por Software

SDN [80] [81], um conceito essencial para a *softwarization* de redes, é essencial para gerenciar várias redes lógicas virtuais no contexto de 5G [82]. A arquitetura SDN oferece os seguintes benefícios [83] [84]: (*i*) separação dos planos de dados e controle, (*ii*) centralização lógica da função de controle, e (*iii*) implementação da função de controle em *software* [85] [86].

Em [87] [88] [89] se apresenta uma visão de alto nível da arquitetura SDN e seus planos. O plano de dados é composto por elementos de rede especializados no tratamento de pacotes (encaminhamento) e se comunica com o plano de controle através de uma ou mais *South Bound Interfaces* (SBI). OpenFlow [90] é o SBI mais conhecido devido ao seu uso difundido por vendedores de equipamentos de rede e grupos de pesquisa. O plano de aplicativos implementa e organiza a lógica das funções rede (e.g., balanceamento de carga e roteamento). O plano de aplicativo comunica seus requisitos de rede para o plano de controle através de uma ou mais *North Bound Interfaces* (NBI). O plano de controle traduz os requisitos das aplicações e os aplica aos elementos da rede por meio de SBI. O plano de controle também define uma ou mais *East West Bound Interface* (EWBI) que permitem implementar um controle logicamente centralizado e fisicamente distribuído para lidar com redes de área larga e grande escala através de sua segmentação lógica. O plano de gerenciamento é ortogonal aos outros planos de SDN para realizar sua operação, administração e manutenção via interfaces de gerenciamento, como OF-Config [91], e *Open vSwitch Database Management Protocol* (OVSDB) [92]. O plano de conhecimento é transversal aos planos acima mencionados, e visa tornar as redes SDN mais inteligentes com o uso de técnicas de ML [62] [61].

D. Virtualização de Funções de Rede

As redes lógicas virtuais são essenciais para implantar os serviços de 5G. NFV, definido pelo *European Telecommunications Standards Institute* (ETSI), é uma tecnologia que facilita a implantação de vários segmentos lógicos de rede em uma infraestrutura compartilhada [93]. NFV desacopla as NF do hardware em que são executadas para abstrair recursos físicos da infraestrutura, dividi-los logicamente e atribuí-los aos VNF que compõem as redes lógicas virtuais [94]–[96].

Os principais componentes de NFV são: a *NFV Infrastructure* (NFVI), as VNF, e o *Management and Network Orchestration* (MANO) [97]. NFVI é a combinação de recursos físicos e virtuais (hardware e software) que compõem o ambiente de implementação de VNF. Uma VNF pode implementar uma ou mais funções de rede e ser executada em recursos virtuais, como máquinas virtuais rodando serviços ou contêineres rodando microserviços. Como uma VNF pode ser composta por outras VNF, pode ser implantada em diversas máquinas virtuais ou contêineres. MANO inclui [98]: (*i*) um orquestrador de NFV para realizar a composição dos recursos de NFVI, (*ii*) um *Virtual Network Functions Manager* (VNFM) responsável pelo gerenciamento do ciclo de vida das VNF; e (*iii*) um *Virtual Infrastructure Manager* (VIM)

responsável pela virtualização, monitoramento, configuração e controle dos recursos (físicos e virtuais) de rede [99].

E. Computação na Borda da Rede 5G

Os usuários da rede 5G exigem informações contextuais em tempo real, bem como baixa latência de comunicação numa determinada área geográfica. A integração do conceito MEC ao 5G permite atender a esses requisitos. MEC fornece alta largura de banda, baixa latência, conhecimento da localização e informações da RAN em tempo real, bem como recursos de CC na borda da rede 5G [100] [101] [57]. Esses recursos são oferecidos por meio de pequenos centros de dados implantados com a finalidade de atender a QoE exigida pelos usuários das redes 5G [102].

Uma plataforma de colaboração construída através da integração MEC-5G deve ter funcionalidade genérica bem como hospedar e fornecer diferentes subsistemas de serviços concebidos para diferentes verticais. Acesso seguro deve estar disponível nesses subsistemas para facilitar o desenvolvimento de aplicações [103]. MEC oferece serviços de memória cache e virtualização para reduzir o volume de dados transmitidos no núcleo da rede 5G e, portanto, melhorar o uso de recursos [104].

FC é uma alternativa à MEC para fornecer informações contextuais em tempo real, bem como baixa latência na comunicação numa determinada área geográfica [105]. Um dos fundamentos de FC é oferecer suporte à IoT, fornecendo conexões e serviços seguros entre e para coisas, dados, pessoas e processos [106] [107]; FC é essencial para MMTC. IoT implica uma mudança profunda na Internet atual em direção a atingir uma rede de coisas interconectadas (objetos). Coisas / objetos ajudam a integrar inteligência em nosso ambiente, apoiando a coleta de informações quando eles interagem com o mundo (físico ou virtual) [108]. A FC melhora a CC trazendo aplicações, poder computacional, armazenamento e os recursos da comunicação na nuvem para os usuários finais [109] [110]. CC envolve *hardware* e *software* de centros de dados usados para fornecer infraestrutura, plataforma e aplicativos como serviço [111]. Trazer CC para mais perto dos usuários finais e as coisas (ou seja, para borda da rede 5G) permite aprovisionar serviços de alto desempenho e baixa latência com uso eficiente dos recursos da infraestrutura.

FC inclui três camadas: IoT, CC, e névoa [112] [25]. Cada camada é responsável por fornecer diferentes funcionalidades [113] [114] [115]. A camada CC consiste em um ou mais centros de dados que oferecem infraestrutura, plataforma e *software* como um serviço. A camada de névoa inclui um ou mais nós que podem reutilizar interfaces sem fio e, além disso, coexistem com elementos de rede, como BS ou roteadores de femtocélulas. Um nó de névoa é uma entidade (física/virtual) que inclui seus recursos de computação, armazenamento e comunicações. Esta camada pode conter várias subcamadas de acordo com os requisitos da aplicação. A camada de IoT é responsável por enviar e receber dados de e para a camada de névoa que geralmente realiza um primeiro estágio de análise ou processamento de dados e, se necessário, envia dados para uma análise mais detalhada na camada CC. A camada de IoT



Fig. 5. Implementação comercial de tecnologia móvel 5G por países e operadoras (primeiros a fornecer o serviço) entre 2018 e 2020.

incli dispositivos como sensores que podem interagir uns com os outros formando redes ad-hoc.

VI. ESTADO DA IMPLEMENTAÇÃO GLOBAL DO 5G

A implantação comercial da tecnologia 5G começou a final de 2018 na Ásia. A seguir teremos um relato histórico desta implantação (ver Fig. 5), enfatizando nas principais empresas e operadoras, conforme explicado em [5]. A Coreia do Sul apresentou a primeira implementação bem-sucedida de uma rede 5G de teste em fevereiro de 2018 durante os Jogos de Inverno em PyeonChang. Após esse impulso inicial, o lançamento comercial 5G foi realizado nesse mesmo país em dezembro pelas empresas SK Telecom, da Coreia Telecom e LG Uplus. Deve-se notar que a Coreia do Sul é o país na atualidade, que tem a cobertura 5G mais ampla e com as taxas de transmissão mais altas [116].

A China representa o maior mercado mundial para a tecnologia 5G. A implantação do 5G começou naquele país em outubro de 2019 por meio da China Mobile, Unicom e China Telecom, a empresa Huawei [117] sendo uma de suas grandes impulsionadores. Na região da Ásia-Pacífico, a Austrália começou operar suas primeiras redes 5G comerciais em maio de 2019 por meio da Telstra, a Nova Zelândia fez isso em dezembro por meio da Spark e da Vodafone, e as Filipinas em junho com a Globe Telecom. Assim, a Ásia-Pacífico é a região com maior interesse em 5G e deverá ter dois terços dos assinantes globais para 2024 [118].

Na Europa, o desenvolvimento comercial da tecnologia 5G começou no Reino Unido em maio de 2019, patrocinado pela EE, Vodafone, O2 e Three. Na Suíça, a implantação comercial foi realizada pela Swisscom e Sunrise a partir de maio de 2019, enquanto na Itália esteve no comando a Vodafone e Telecom Itália desde junho do mesmo ano. A Alemanha foi outro dos primeiros países a operar redes 5G comerciais

através da Deutsche Telekom e Vodafone, iniciando operações em julho de 2019. Romênia, Áustria, Irlanda, Finlândia e Hungria juntaram-se a esta implantação comercial no decorrer desse mesmo ano. Outros países europeus, como Espanha, Holanda, Suécia, Dinamarca e Bélgica ativaram suas redes 5G comerciais em 2020.

A América do Norte é outro grande mercado para a tecnologia 5G, onde a Verizon realizou o primeiro lançamento comercial 5G em maio de 2019, enquanto AT&T e T-Mobile o fizeram no segundo semestre daquele mesmo ano. No Canadá, a implantação é liderada por Rogers Communications e Bell Mobility, onde o lançamento comercial de 5G começou em 2020. Outro principal mercado é o Oriente Médio, onde Kuwait, Emirados Árabes Unidos e Catar começaram a operar suas primeiras redes comerciais 5G no período de abril a junho de 2019. Na África, o país líder nessa tecnologia é a África do Sul, onde Rain e Vodacom ativaram suas redes 5G em setembro de 2019 [5].

O estado atual da tecnologia 5G na América Latina é extremamente incipiente. A primeira rede 5G comercial surgiu em Uruguai, operada pela empresa Antel com a ajuda de Nokia desde abril de 2019. De acordo com o relatório disponível em [4], além do Uruguai, somente o Brasil possui redes comerciais de Telefonia móvel 5G, operadas pela empresa Claro. Na Colômbia existem redes de internet fixa 5G por meio da Directv, enquanto outros países, como a Argentina, o Chile, o Peru e o México estão nos estágios iniciais de investimento e implantação. Na área do Caribe, Porto Rico e as Ilhas Virgens têm redes 5G desde dezembro 2019, operadas pela T-Mobile. Conforme descrito em [119], um dos grandes desafios para a implantação de 5G na América Latina consiste na melhoria da infraestrutura celular para atingir níveis de QoS e cobertura próximos aos de países líderes como o Japão, Reino Unido e Estados Unidos. De acordo com [120], nos próximos dez anos, o desenvolvimento em telefonia móvel para a América

Latina ainda terá como foco a tecnologia 4G, que espera-se consiga 67% da população da região até 2025, que irá cimentar a integração do 5G num futuro próximo.

Em resumo, o estado da implantação da tecnologia 5G no mundo pode ser dimensionado conforme o relatório disponível em [4], onde até setembro do ano 2020, existem 397 operadoras de tecnologia sem fio de 129 países ou territórios independentes, que investiram em testes, adquiriram licenças, planejaram o desenvolvimento da rede, ou implantaram redes 5G. Além disso, 101 operadoras de 44 países ou territórios têm gerado um ou mais serviços compatíveis com os serviços 3GPP para 5G. Uma ótima ferramenta para visualizar o progresso mundial desta tecnologia são os mapas interativos de cobertura 5G desenvolvidos pela OOKLA [116].

Durante a implantação global do 5G também é de extrema importância enfrentar os desafios relacionados com possíveis consequências na saúde e no ambiente. Essas implicações estão associadas, entre outros, aos seguintes fatores. Primeiro, as tecnologias chave de 5G na camada física (comunicações por ondas milimétricas, arranjos de antenas a grande escala, e células curto alcance) empregam frequências de transmissão muito maiores do que as empregadas pelas gerações anteriores [121]. Assim, os níveis de penetração no tecido biológico será menor, mas outros órgãos, como a pele e os olhos podem ser afetados [122]. Segundo, um aumento da pegada de carbono gerada pelo aumento no número de dispositivos conectados, especialmente para acomodar a IoT industrial, e o crescimento em tamanho tanto dos centros de dados da CC usados para suportar à virtualização da CN quanto dos utilizados na FC o MEC para habilitar a virtualização da RAN [123]. Para enfrentar esses desafios, a comunidade científica solicitou que os governos investirem em estudos de impacto ambiental para dimensionar os possíveis riscos com a implantação de 5G, e imediatamente gerar políticas regulatórias que protejam a saúde da população [124]. Além disso, os pesquisadores devem continuar a trabalhar arduamente para obter redes 5G eficientes e sustentáveis do ponto de vista da pegada de carbono e eficiência energética [125] [126].

VII. TELEFONIA MÓVEL ALÉM DO 5G

As redes 5G transformarão o cenário das comunicações celulares permitindo enlaces ultra confiáveis e de ultra baixa latência, capazes de conectar pessoas e máquinas. Essas redes serão essenciais para o desenvolvimento e para implantar todo o potencial da Indústria 4.0, os sistemas de transporte inteligente, a IoT, a gestão eficiente de energia e as cidades inteligentes [127]. No entanto, existem aplicativos cujos requisitos podem exceder as capacidades projetadas para 5G. Por exemplo, embora 5G esteja se moldando para ser a tecnologia certa para veículos conectados pode ser não suficiente para garantir a latência extremamente baixa exigida por veículos autônomos. Além disso, as redes 5G concentram-se principalmente nas comunicações em dois dimensões, isto é, nas comunicações móveis terrestres. No entanto, o crescimento da indústria aeronáutica e aeroespacial, evidenciado pela proliferação dos UAV, os nano-satélites e as plataformas de alta altitude, levantam um cenário de comunicações tridimensionais, ou seja, ambientes terra, ar e água. Além disso,

a maturidade que têm alcançado as tecnologias de realidade aumentada e interfaces hapticas está abrindo o caminho para as comunicações interativas, que não são contemplados no 5G. Dois exemplos desse tipo de comunicação são a Internet tático (háptico) [128] e as comunicações holográficas [129].

Embora o 5G esteja em um estágio inicial, diversos trabalhos para definir as características da sexta geração de redes de telefonia móvel já estão em andamento. Em julho de 2018, a ITU estabeleceu o grupo temático Rede 2030 cuja tarefa foi analisar os desafios e oportunidades na evolução das redes de comunicação em direção ao ano 2030 e além [130]. Este grupo identificou sete casos de uso para as redes 2030, incluindo comunicações holográficas, o Internet tático para operações remotas, a rede espaço-terra integrada e o IoT industrial com *cloudification*. Embora a visão de como serão as redes 6G ainda não consolida-se, a expectativa é que essas redes permitam comunicações ultra-massivas entre máquinas, com latências extremamente baixas (entre 10 y 100 μ s), extrema confiabilidade (taxas de erro de quadro de 1-10⁻⁹), consumo de energia extremamente baixo (eficiência energética de 1 pJ/b), e suporte de enlaces de banda larga em condições de mobilidade muito alta (usuários mudando de lugar a más de 1,000 km/h) [131], [132].

Para atingir essas características, é necessária uma mudança no paradigma no desenho da rede de telefonia móvel, em que as comunicações não são mais direcionadas às pessoas ou máquinas, mas para os dados [133], [134]. Além disso, a gestão eficiente dos grandes volumes de informação que é exigida por aplicativos 6G exige que essas redes adotem a AI como um dos seus pilares e não apenas como ferramenta de suporte [135]. Finalmente, os sistemas de transmissão devem operar em faixas de frequência ainda maiores do que 5G para ser capazes de atingir taxas de transmissão de pico na ordem dos Tbps. As Tecnologias de transmissão sendo consideradas para 6G incluem *Visible Light Communications* (VLC), enlaces de radiofrequência na faixa dos THz, sistemas MIMO supermassivos, conformação de feixe holográfico, superfícies inteligentes, e multiplexação com base no momento orbital angular das ondas eletromagnéticas [131]–[134].

6G vai impor desafios adicionais relacionados a, primeiro, o gerenciamento e orquestração da tridimensionalidade das redes celulares que incluirão aos BS e UM baseados em UAV. Em segundo lugar, a evolução dos algoritmos de admissão e planejamento de segmentos lógicos de rede gerados devido ao aparecimento de novos casos de uso exigindo, para exemplo, velocidade de dados e confiabilidade extremamente alta; ou seja, uma combinação de 5G EMBB e URLLC. As redes 6G nos permitem imaginar um cenário onde abundam as cidades inteligentes. Embora este cenário seja percebido de longe, a estrada começou a ser construída com a ascensão das redes 5G.

VIII. CONCLUSÕES

A implantação comercial das redes 5G permitirá estender os aplicativos móveis para os casos de uso EMBB, MMTC e URLLC. No entanto, esta mudança tecnológica será gradual, e nos próximos anos permanecerá uma coexistência de serviços,

principalmente com redes 4G. Em geral, a mudança central da tecnologia 5G está focada na arquitetura de rede por meio de conceitos como segmentação lógica, *softwarization*, *programmability* e virtualização de rede. Em particular, na América Latina, a penetração das redes 5G ainda é baixa e espera-se consolidação nos próximos anos dos serviços 4G e um início de investimento em infraestrutura 5G. Apesar disso, em todo o mundo os esforços de pesquisa e visualização das necessidades para comunicações 6G têm começado.

AGRADECIMENTOS

Os autores agradecem o financiamento do CONACYT por meio do projeto de Ciências Básicas nº 254637.

5G and Beyond: Past, Present and Future of the Mobile Communications

Carlos A. Gutierrez, *Senior Member, IEEE*, Oscar Caicedo, *Senior Member, IEEE*,
and Daniel U. Campos-Delgado*, *Senior Member, IEEE*

Abstract—The fifth-generation (5G) of mobile communications networks is emerging as a revolutionary technology that will accelerate the development of smart cities and the realization of the information society. This paper aims to provide an introduction to 5G for non-specialists, and a survey of this new technology for those already familiar with mobile communications, covering the conceptualization and the core technologies underpinning 5G networks. The paper also discusses the status of the commercial roll-out of 5G until 2020 from a worldwide perspective and gives a future view of mobile communications beyond 5G.

Index Terms—5G communications, Massive MIMO, Beamforming, mmWave communications, Mobile edge computing, Small cell stations, NOMA, network slicing.

I. INTRODUCCIÓN

La telefonía móvil evoluciona constantemente buscando satisfacer las demandas de la sociedad por más y mejores servicios de información y comunicaciones. El rumbo de este proceso de evolución está marcado por la aparición de estándares que recogen los últimos avances tecnológicos y armonizan la coexistencia de los sistemas de radiocomunicaciones [1]. Las grandes innovaciones que se desencadenan con el surgimiento de un nuevo estándar (o de un nuevo conjunto de estándares) marcan a su vez nuevas etapas en la historia de la telefonía móvil, a las que se les clasifica en generaciones [2], [3]. Actualmente nos encontramos transitando hacia la quinta generación (Fifth Generation, 5G) de redes de telefonía móvil [4], [5]. Esta nueva generación promete servicios que mejorarán las experiencias de uso a través del acceso a vídeo de ultra alta definición, vídeo en 360°, trabajo y cómputo en la nube, y comunicaciones inmersivas basadas en la realidad aumentada y la realidad virtual. A diferencia de las redes de cuarta generación (Fourth Generation, 4G), las redes 5G ofrecen soporte nativo para aplicaciones orientadas a las comunicaciones entre máquinas. Esto será fundamental para el desarrollo de la Industria 4.0, el transporte inteligente, la *e-health*, y en general para todo aspecto de conectividad relacionado con las ciudades inteligentes.

El camino hacia 5G se está construyendo sobre la visión que estableció la Unión Internacional de Telecomunicaciones (International Telecommunications Union, ITU), acerca del futuro de las telecomunicaciones móviles internacionales (International Mobile Telecommunications, IMT) para el año 2020 y más

Carlos A. Gutierrez and Daniel U. Campos-Delgado are with Faculty of Sciences, and Daniel U. Campos-Delgado also with Instituto de Investigación en Comunicación Óptica, both in Universidad Autónoma de San Luis Potosí, S.L.P., Mexico, e-mails: cagutierrez@ieee.org, ducd@fciencias.uaslp.mx.

Oscar M. Caicedo-Rendon is with University of Cauca, Popayan, Colombia, email: omcaicedo@unicauca.edu.co.

*Corresponding author

allá [6]. Los casos de uso y requisitos de rendimiento de 5G están definidos en un reporte publicado en el 2017 por el sector de radiocomunicaciones de la ITU [7]. Según este reporte, las redes 5G alcanzarán densidades de conectividad diez veces mayores que las redes 4G, tasas pico de transmisión veinte veces mayor, y tráfico por área cien veces mayor. Aparte de esto, los tiempos de latencia de 5G serán hasta diez veces menores que los de 4G [8], [9]. Siguiendo el camino trazado por la ITU, el Proyecto Asociación de 3ra Generación (3rd Generation Partnership Project, 3GPP) ha elaborado estándares para tecnologías 5G. El primero de ellos fue publicado en 2017, la tecnología asociada con ese estándar es conocida como 5G-New Radio (5G-NR) [10].

El despliegue comercial de 5G-NR comenzó a finales del 2018 [5]. Sin embargo, la visión de la ITU sobre 5G está lejos de completarse. Además del aspecto tecnológico, es fundamental realizar actividades de investigación básica y aplicada que permitan emplear estas nuevas tecnologías para resolver los problemas sociales, económicos y ambientales que aquejan al mundo. En el contexto de América Latina, las implicaciones de la política regulatoria y de la liberación de nuevas bandas del espectro electromagnético (Electromagnetic Spectrum, EMS) deben investigarse a fondo para que la transformación de 5G reduzca la brecha digital en la región.

Este artículo presenta una revisión de los antecedentes, la situación actual y el futuro de las redes 5G con la intención de dar a conocer las áreas de oportunidad que ofrece esta tecnología, y atraer atención a los problemas técnicos y regulatorios que continúan abiertos. Los orígenes de la telefonía móvil y las principales innovaciones aportadas por las generaciones anteriores a 5G se revisan en la Sección II. Los casos de uso, los requisitos técnicos y la arquitectura de las redes 5G se discuten en la Sección III. Las secciones IV y V revisan las principales tecnologías de transmisión y gestión de red para 5G. El estado actual del despliegue comercial de 5G se aborda en la Sección VI. La Sección VII describe brevemente una perspectiva de la telefonía móvil más allá de 5G, y finalmente, la Sección VIII presenta comentarios finales y conclusiones. Cabe mencionar que las redes 5G ya han sido objeto de numerosos artículos de revisión, entre los que destacan [8], [9], [11]–[13]. Este artículo complementa la literatura actual sobre 5G proporcionando una revisión integral tanto de las tecnologías para la red de acceso por radio (Radio Access Network, RAN), como de las que se ocupan para la red de núcleo. Además, este artículo ofrece una visión global del despliegue actual de 5G y de las iniciativas para extender las capacidades de 5G con miras hacia 6G. Dado que la mayoría de los acrónimos utilizados dentro de las comunicaciones

móviles se identifican regularmente en inglés, se empleará esta convención a lo largo del artículo.

II. ANTECEDENTES DE LAS REDES 5G

Los orígenes de la telefonía móvil se remontan a la invención del telégrafo inalámbrico a finales del siglo diecinueve, y a sus aplicaciones en la industria naval como instrumento de comunicación entre barcos y estaciones de tierra [14]. La telegrafía inalámbrica demostró la factibilidad de las comunicaciones a larga distancia basadas en la transmisión de ondas electromagnéticas que se propagan libremente en la atmósfera terrestre. Los avances alcanzados a principios del siglo veinte en materia de electrónica—desencadenados por la invención del tubo de vacío—permitieron aplicar los principios de la telegrafía inalámbrica en la radiodifusión comercial y la telefonía móvil de carácter militar y policial [15].

El primer sistema de telefonía móvil de uso civil fue el *Mobile Telephone System* (MTS), el cual se introdujo en 25 ciudades de Estados Unidos en 1946 [15]. El MTS funcionaba a partir de técnicas de pulsar para hablar. El servicio en MTS era proporcionado por una estación central que atendía a todos los usuarios móviles (Mobile Users, MUs) de la ciudad. Esta centralización del servicio, junto con las estrictas restricciones en ancho de banda impuestas a todo sistema inalámbrico para regular el uso compartido del EMS, limitó considerablemente la capacidad de esta red. El concepto de red celular de telefonía móvil desarrollado en Bell Labs durante la década de los cuarenta [16] permitió superar los problemas de saturación de MTS y fue el catalizador de la telefonía móvil comercial. La idea detrás de este concepto es incrementar la capacidad del sistema dividiendo el área de cobertura en celdas de menor extensión, cada una atendida por una estación base (Base Station, BS) conectada a la red pública de telefonía. Esto permite asignar diferentes canales de radiofrecuencia a celdas adyacentes para reducir la interferencia entre ellas, y reutilizar canales en celdas que están suficientemente alejadas unas de otras, de manera que la interferencia entre ellas es despreciable gracias a la atenuación que experimentan las ondas electromagnéticas al propagarse en la atmósfera. La arquitectura de red celular es la piedra angular de la telefonía móvil de uso civil, por ello es también conocida como telefonía celular. Cada una de las generaciones de redes de telefonía celular se ha desarrollado en un marco de diez años, y la transición hacia una nueva generación siempre ha estado motivada por la necesidad de mejorar la cobertura, alcanzar tasas de transmisión más altas y ofrecer nuevos servicios móviles [3], [17], como se muestra en la Fig. 1.

La primera generación (First Generation, 1G) de redes celulares entró al mercado a finales de los años setenta, siendo Japón el primer país en desplegar una red 1G. A partir de 1981 se implementaron redes 1G en los países nórdicos: Noruega, Suecia, Dinamarca y Finlandia. Para mediados de los 80's ya se tenían redes 1G en Reino Unido, EUA, Canadá y México. Esta primera generación fue concebida para servicios de voz utilizando tecnología analógica, por lo que presentaba una baja eficiencia espectral. Además, debido a que no existían acuerdos internacionales para la operación de las redes 1G, había

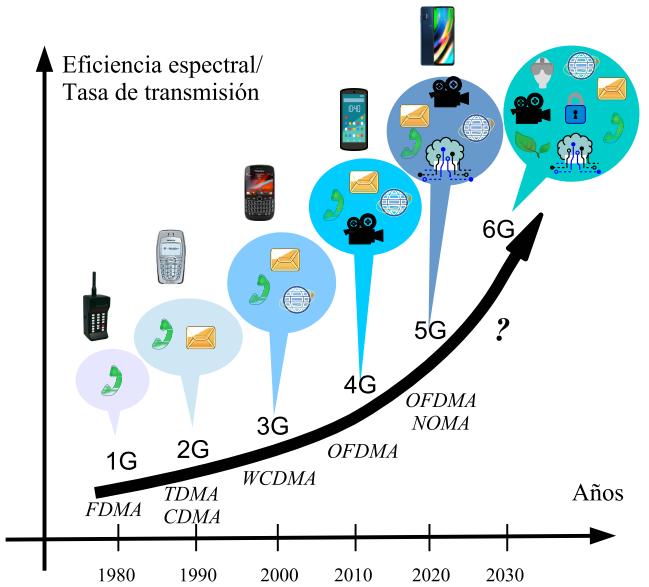


Fig. 1. Evolución de la telefonía celular.

una gran variedad de estándares en operación. Como resultado el servicio estaba restringido a coberturas nacionales e incluso regionales. A pesar de la incompatibilidad entre las diferentes redes 1G, todas empleaban tecnología de acceso múltiple por división en frecuencia (Frequency Division Multiple Access, FDMA) y las llamadas se enlazaban por conmutación de circuitos. En Japón y Estados Unidos, la banda de operación reservada para 1G fue la de 800 MHz. Mientras tanto en Suecia y Reino Unido se reservó la banda de 900 MHz, y en Alemania y Francia se utilizaron las bandas de 450 y 200 MHz, respectivamente [2].

La segunda generación (Second Generation, 2G) de redes celulares surgió a principios de los 90's y se caracterizó por el cambio a la tecnología digital. La eficiencia espectral mejoró gracias a la adopción de técnicas de acceso múltiple por división en tiempo (Time-Division Multiple Access, TDMA) y de acceso múltiple por división de códigos (Code Division Multiple Access, CDMA), así como a la introducción de técnicas de codificación de canal. Una innovación importante fue la inclusión del servicio de datos para mensajes cortos (Short Message Service, SMS). Las redes 2G fueron un gran éxito comercial gracias al desarrollo de estándares de ámbito internacional, por ejemplo, el *Global System for Mobile Communications* (GSM) y el *Interim Standard 95* (IS-95) [18]. Ambos estándares transmiten en las bandas de 810 y 960 MHz sobre una arquitectura de conmutación de circuitos. La velocidad de transmisión de datos teórica de 2G era de 115.2 kbps (con la mejora *High-Speed Circuit-Switched Data* de GSM), y la digitalización del servicio permitió introducir dispositivos compactos con uso eficiente de las baterías. Las redes 2G transmiten datos de manera ineficiente porque utilizan conmutación de circuitos. Este tipo de conmutación es adecuado para tráfico de voz, pero no para el tráfico de datos.

La transición hacia la tercera generación (Third Generation, 3G) de redes celulares comenzó a principios del 2000. Una de las grandes innovaciones de 3G fue la capacidad de realizar

enlaces tanto por conmutación de circuitos como de paquetes; estos últimos son necesarios para servicios basados en el protocolo de Internet (Internet Protocol, IP). Esta generación permitió incorporar nuevos servicios móviles, como el acceso a Internet y las video-llamadas. La tecnología CDMA es la base de 3G y sus principales estándares son CDMA de banda ancha (Wideband CDMA, WCDMA) y CDMA 2000 [19]. Ambos estándares son de ámbito global y en teoría proporcionan velocidades de transmisión de hasta 14.4 Mbps en el enlace descendente (Downlink, DL) [2]. Con 3G surgieron los teléfonos “inteligentes”, con un interfase visual más atractiva y la incorporación de aplicaciones móviles.

El despliegue de 4G comenzó a principios del 2010, ofreciendo tasas de transmisión pico de hasta 300 Mbps en el DL y de 75 Mbps en el enlace ascendente (Uplink, UL), aunque con las actualizaciones más recientes, estas redes alcanzan velocidades de hasta 1 Gbps en el DL [20]. Las tecnologías base de 4G son el acceso múltiple por división de frecuencias ortogonales (Orthogonal Frequency-Division Multiple Access, OFDMA), el FDMA de portadora sencilla (Single Carrier FDMA, SC-FDMA), el OFDMA de portadora escalable (Scalable OFDMA, SOFDMA), y las técnicas de transmisión de múltiples entradas y múltiples salidas (Multiple-Input Multiple Output, MIMO). Los principales estándares de 4G son el *Long Term Evolution* (LTE) y su versión avanzada, el *LTE Advanced* (LTE-A) [21]. Las innovaciones de las redes 4G han permitido proveer servicios de voz, datos y multimedia de forma ubicua a través de medios de suscripción, así como generar aplicaciones de vídeo de alta definición, y plataformas interactivas de juego en línea. La tecnología 4G también ha permitido generar soluciones completamente móviles basadas en IP, alcanzando tiempos de latencia de hasta 20 ms.

III. CONCEPTUALIZACIÓN DE LAS COMUNICACIONES 5G

A. Casos de Uso y Requisitos de Rendimiento

Las redes de 4G seguirán en expansión y continuarán en operación durante varios años más dando soporte a servicios de comunicaciones orientados a personas [20]–[22]. Según la última versión de Ericsson Mobility Report, se espera para 2026 una co-existencia de diversas tecnologías para atender a los usuarios móviles, siendo las preponderantes: 5G, LTE, WCDMA/HSPA, y GSM/EDGE [23]. Sin embargo, la coyuntura actual también demanda servicios orientados a la comunicación entre máquinas, con la finalidad de soportar aplicaciones de la Industria 4.0, los sistemas de transporte inteligente, y la automatización del hogar. Tales aplicaciones plantean retos nuevos, como garantizar el servicio en zonas de cobertura con una gran densidad de dispositivos interconectados; la transmisión con baja latencia para aplicaciones sensibles al retraso; el soporte de enlaces de banda ancha en condiciones de alta movilidad, similares a las que se encuentran en trenes de alta velocidad; y la reconfigurabilidad de la red para un manejo flexible y eficiente de sus recursos. Las redes 5G buscan solventar estos retos [11], [12].

La ITU definió tres casos de uso para estas nuevas redes:

- **Banda Ancha Móvil Mejorada (Enhanced Mobile Broadband, EMBB):** Aplicaciones centradas en el usuario final

TABLA I
INDICADORES DE RENDIMIENTO DE LAS REDES 4G Y 5G (DATOS EXTRAÍDOS DE [7]).

Indicador	4G	5G
Tasa pico de transmisión UL	500 Mbps	10 Gbps
Tasa pico de transmisión DL	1 Gbps	20 Gbps
Eficiencia espectral pico UL	6.75 b/s/Hz	15 b/s/Hz
Eficiencia espectral pico DL	15 b/s/Hz	30 b/s/Hz
Densidad de conectividad	10^5 disp./km ²	10^6 disp./km ²
Movilidad	350 km/h	500 km/h
Capacidad de tráfico por área	0.1 M b/s/m ²	10 M b/s/m ²
Latencia	20 ms	1 ms

caracterizadas por requerir acceso a contenido multimedia que necesita tasas altas de transferencia de datos con velocidades pico de hasta 20 Gbps y velocidades percibidas por el usuario de hasta 100 Mbps, además de soporte de movilidad de hasta 500 km/h y una eficiencia espectral tres veces superior a la proporcionada por 4G. Ejemplos de aplicaciones EMBB son la voz y el vídeo móvil de ultra alta definición, así como la realidad virtual y la realidad aumentada para experiencias de comunicación inmersivas.

- **Comunicaciones Masivas de Tipo Máquina (Massive Machine-Type Communications, MMTC):** Aplicaciones que involucran una gran cantidad de dispositivos conectados, los cuales transmiten un volumen bajo de datos con poca sensibilidad al retraso y cuyo consumo de energía es típicamente bajo. Aplicaciones de este tipo incluyen vehículos conectados, cuidado de mascotas/personas, y redes de sensores y actuadores inalámbricos. El objetivo para este caso de uso es tener una capacidad de tráfico de 10 Mbps/m², una densidad de conexión de 10^6 dispositivos/km², y una eficiencia energética de la red 100 veces superior a la de 4G.
- **Comunicaciones de Gran fiabilidad y Baja Latencia (Ultra Reliable Low Latency Communications, URLLC):** Aplicaciones caracterizadas por operar en tiempo real, como las relacionadas con la seguridad del transporte, cirugías remotas y la automatización de procesos industriales, las cuales requieren latencias bajas de hasta 1 ms.

La Tabla I resume los principales requisitos de rendimiento de estos tres casos de uso [7], y los compara con sus contrapartes de 4G. Las tecnologías clave que identificó la ITU para cubrir estos requisitos se revisan en [24].

B. Arquitectura de la Red 5G

Los requerimientos de tasas de transmisión altas y soporte de una gran densidad de conexiones, se pueden proveer a nivel de capa física y de control de acceso al medio utilizando técnicas de transmisión eficientes. Sin embargo, los requisitos de baja latencia, alta movilidad y capacidad de tráfico por área involucran estrategias en el núcleo de la red. Así, una de las principales innovaciones de 5G está en la arquitectura de la red, cuyo diseño es mucho más flexible que el de las generaciones precedentes gracias a la implementación de conceptos de segmentación lógica, *softwarización*, y virtualización de la red y sus servicios.

TABLA II
FUNCIONES DE LA RED 5G.

Función de Red	Descripción
Función de aplicación	Interactúa con el núcleo de la red 5G, especialmente PCF y NEF, para ayudar en la gestión de tráfico.
Función de gestión de acceso y movilidad	Punto final del estrato de no-acceso, incluye funciones de registro, conexión, gestión de movilidad y accesibilidad.
Función del servidor de autenticación	Soporta el proceso de autenticación de los UE.
Red de datos	Función que representa la conexión a cualquier red de datos.
Función de análisis de datos de red	Analiza los datos de las NF.
Función de exposición de red (NEF)	Expone las funcionalidades y eventos de las NF, habilitando su comunicación con aplicaciones externas.
Repositorio de funciones de red	Registra las NF y permite su identificación para facilitar la composición de servicios.
Función de selección del segmento lógico de red	Selecciona el conjunto de instancias de red lógica que atienden un UE.
Función de control de política (PCF)	Soporta la definición de políticas que son utilizadas por el CP para administrar el comportamiento de la red.
Red de acceso radio	Proporciona el acceso radio que conecta con las NF del UP, y el CP del núcleo de la red 5G.
Función de gestión de sesión	Gestiona el establecimiento, modificación y liberación de sesiones de datos y asigna IP a los UE.
Gestión unificada de datos (UDM)	Incluye soporte para la generación de credenciales de autenticación, identificación de usuario y gestión de suscripciones.
Repositorio unificado de datos	Permite que UDM y PCF obtengan información de suscripción y políticas, respectivamente.
Equipo de usuario (UE)	Permite a los usuarios finales usar los servicios proporcionados por la red 5G.
Función de plano de usuario	Gestiona movilidad y QoS entre tecnologías de acceso radio, dirige paquetes y genera informes de tráfico

La Fig. 2 muestra un diagrama de la arquitectura general de las redes 5G, la cual consta de la RAN, la red de núcleo y las terminales de usuario. Estos elementos son representados como funciones de red (Network Functions, NF) y sus respectivas interfaces. La Tabla II describe brevemente las NF de la arquitectura de 5G. Diferentes opciones arquitectónicas para desplegar la RAN son presentadas en [13] y [25], y las arquitecturas para el despliegue de la red de núcleo se estudian en [26] y [27].

Las redes 5G siguen un patrón arquitectónico basado en servicios (Service-Based Architecture, SBA). SBA aporta importantes beneficios a la estructura de la red, por ejemplo:

- (I) *Fácil actualización*: cada servicio puede ser modificado con un impacto mínimo en el resto.
- (II) *Extensibilidad*: la adición de una nueva funcionalidad implica agregar un nuevo servicio y sus correspondientes interfaces.
- (III) *Modularidad*: un sistema 5G se compone de servicios de granularidad fina (implementan una sola NF) y bajo acoplamiento (usan interfaces estandarizadas) conocidos como microservicios [28], fomentando así la segmentación lógica de la red.
- (IV) *Apertura y reutilización*: los servicios pueden ser expuestos e invocados a través de interfaces ligeras y estandarizadas, incentivando así la reutilización de funcionalidades.

Estos beneficios permiten que la arquitectura de 5G sea flexible, programable y de fácil automatización, así como también promueve una reducción significativa de costos gracias al despliegue de las NF como funciones virtuales de red (Virtual Network Functions, VNF) en pequeños, medianos o grandes centros de datos.

La separación de los planos de control y usuario (Control and User Plane Separation, CUPS) es otro de los pilares de la arquitectura de las redes de 5G. CUPS, aunque no es un concepto nuevo, es fundamental en 5G porque permite la separación del núcleo de paquetes evolucionado (Evolved Packet Core, EPC) de 4G en funciones de red del plano de usuario (User Plane, UP), y funciones de red del plano de control (Control Plane, CP) [29]. Las NF del UP pueden ser desplegadas cerca de la aplicación que soportan para satisfacer requisitos de latencia y tasa de transferencia, fundamental para

los casos de uso URLLC y EMBB, respectivamente. Por su parte, las NF del CP pueden operar de manera centralizada para satisfacer requisitos de confiabilidad. El UP transporta solo tráfico de datos, mientras que el CP se encarga de la señalización de red. Las NF del UP y CP interactúan entre sí a través de interfaces estandarizadas. Un servicio en 5G es implementado mediante la composición de cadenas de NF (virtualizadas o no) pertenecientes a CP y UP.

La computación móvil en el borde (Mobile Edge Computing, MEC), la computación en la niebla (Fog Computing, FC), la computación en la nube (Cloud Computing, CC), las redes definidas por *software* (Software-Defined Networking), y la virtualización de funciones de red (Network Functions Virtualization, NFV) son fundamentales para construir los pilares, y por tanto, la arquitectura de las redes de 5G. Estas tecnologías se presentan en la sección V.

IV. TECNOLOGÍAS CLAVE PARA 5G RAN

La transformación tecnológica originada por las redes 5G se comprende mejor identificando las dimensiones de innovación que determinan el rumbo evolutivo de los sistemas de comunicaciones inalámbricos. Estas dimensiones son [30]: 1) la disponibilidad de EMS, 2) la eficiencia espectral, 3) la eficiencia espacial, y 4) la eficiencia del sistema. Las primeras tres dimensiones se describen en esta sección, mientras que la cuarta se discute en la sección V.

A. Comunicaciones por Ondas Milimétricas

La primera dimensión de innovación se refiere a la banda (o bandas) del EMS que la red tiene a su disposición para establecer los enlaces inalámbricos entre las BS y los MU. Los elementos relevantes de esta dimensión son la frecuencia de operación y el ancho de banda, los cuales determinan el rango y la capacidad (en términos de la máxima velocidad de transmisión alcanzable) del enlace, respectivamente. Una de las principales innovaciones de las redes 5G está en la inclusión de enlaces a frecuencias entre 20 y 80 GHz, es decir, enlaces en la banda de las ondas milimétricas (millimeter Wave, mmW) [31]. Con esto se busca dar soporte a servicios que requieran tasas pico de transmisión en el orden de los Gbps, como los previstos para el caso de uso EMBB. Estas velocidades de transmisión son difíciles de alcanzar en bandas

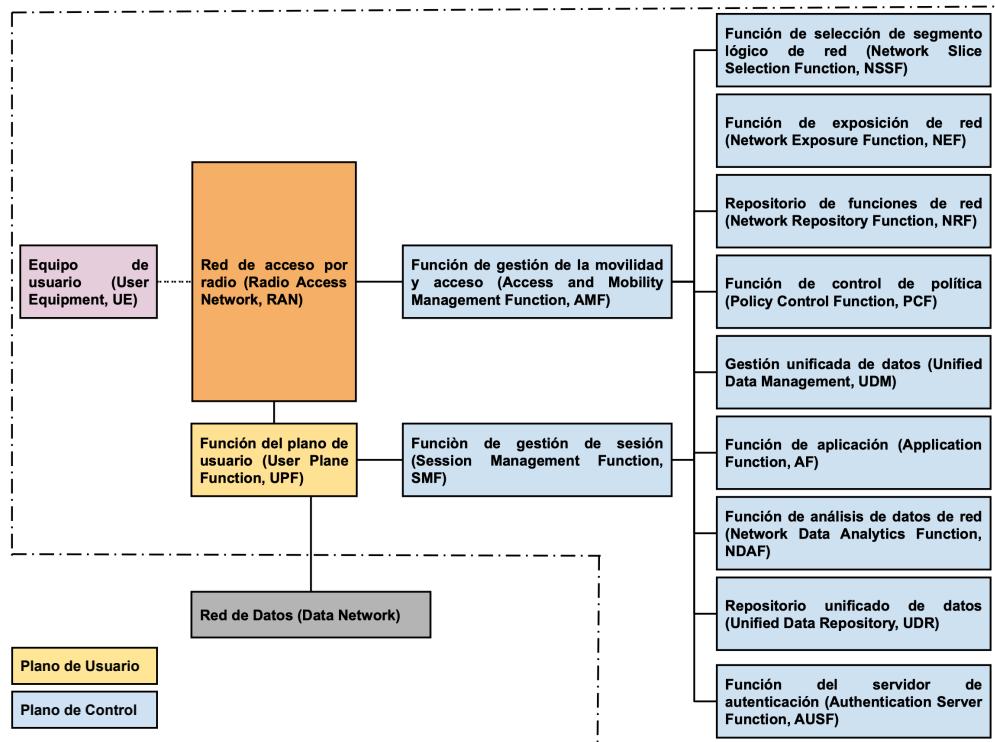


Fig. 2. Arquitectura de la red 5G.

de frecuencia más bajas, como las de 4G, en las que el ancho de banda está limitado a algunas decenas de MHz, mientras que la banda mmW ofrece anchos de banda de varios GHz.

Aunque incrementar la frecuencia de operación permite anchos de banda mayores, también reduce el rango del sistema, ya que las ondas electromagnéticas se vuelven más susceptibles a la atenuación por absorción a medida que su frecuencia aumenta. La banda mmW tampoco es adecuada para las comunicaciones indirectas, es decir, comunicaciones en las que no existe línea de vista (Line of Sight, LOS) entre transmisor y receptor, debido a la atenuación excesiva que sufren las ondas mmW al interactuar con los objetos con los que se podrían crear enlaces indirectos. Estas limitaciones se pueden solventar empleando arreglos de antenas para reducir el impacto de la atenuación, o incluso resultan convenientes cuando se requiere conectar una gran cantidad de dispositivos, como en el caso de uso MMTC. Al margen de esto, aún existen numerosas interrogantes relacionadas con la propagación de señales en esta banda de frecuencias. Por ejemplo, en [32] se identifica la necesidad de llevar a cabo mediciones de campo para incrementar la información empírica sobre las pérdidas por propagación, la dispersión temporal, la dispersión en frecuencia, la dispersión angular y la probabilidad de LOS de los enlaces en la banda mmW.

Por otro lado, en [33] se muestra que el canal móvil de propagación se comporta como un sistema no-estacionario a medida que aumenta el ancho de banda de la señal transmitida. Estos comportamientos no-estacionarios previstos para las redes 5G no se han estudiado a detalle en la literatura. En

[34] se observa que el despliegue de la tecnología en la banda mmW seguirá ritmos diferentes en ciudades grandes y poblaciones pequeñas. Esto puede incrementar la brecha digital entre las grandes urbes y las zonas rurales, por lo que es importante buscar alternativas para reducir su impacto.

B. Arreglos de Antenas de Gran Escala

La segunda dimensión de innovación se refiere al uso eficiente del EMS disponible, medido en relación a la cantidad de bits de información que se consigue transmitir por cada Hz disponible. Las redes 5G buscan incrementar la eficiencia espectral de manera nunca antes vista empleando arreglos de antenas de gran escala (Large-Scale Antenna Arrays, LS-AA). El uso de estas técnicas de transmisión resulta viable y oportuno considerando que 5G incluirá enlaces entre las BS y los MU en la banda mmW. Para esta banda de frecuencias es posible construir antenas de un perfil compacto integrables con arreglos compuestos por cientos o incluso miles de elementos.

Las técnicas LS-AA que figuran en el contexto de 5G son las de conformación de haz [35]–[38] y MIMO masivo (massive MIMO, mMIMO) [39]–[41]. En la conformación de haz se emplea LS-AA para obtener un patrón de radiación combinado de alta directividad, el cual permite extender el rango del enlace y obtener una mejor relación señal a ruido (Signal to Noise Ratio, SNR) [42] sin incrementar la interferencia hacia otros usuarios. La conformación de haz es también empleada en el acceso al medio por división espacial (Space Division Multiple Access, SDMA). Mediante SDMA, la BS puede comunicarse simultáneamente con varios MU sobre el

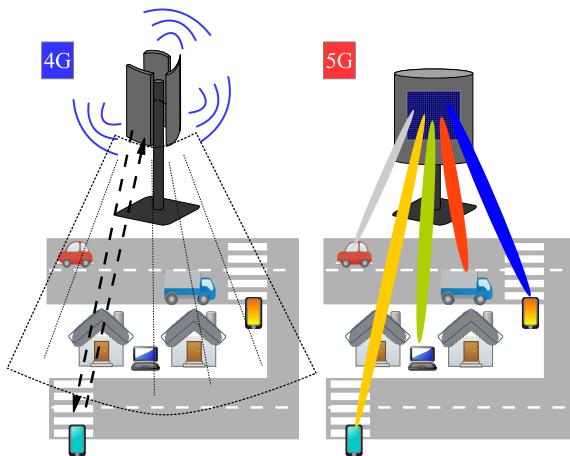


Fig. 3. Transmisión en una red 4G sin conformación de haz y en una red 5G con conformación de haz.

mismo canal de radiofrecuencia separando a los usuarios a través de enlaces focalizados, cada uno con una alta relación señal a interferencia (Signal to Interference Ratio, SIR) [43]. La Fig. 3 muestra, primero, un escenario 4G en el que la BS se comunica con los MU empleando antenas sectoriales con patrones de radiación medianamente directivos. Segundo, un escenario 5G en el que la BS se comunica simultáneamente con varios MU usando enlaces focalizados. La conformación de haz también se emplea para la multiplexación espacial de datos [44]. Algunos problemas de investigación relacionados con estas tecnologías de arreglos de antenas están en el desarrollo de técnicas de pre-codificación híbridas (analógicas y digitales) de baja complejidad y alta eficiencia energética, así como la simplificación de las cadenas de RF [36], [44].

Los sistemas mMIMO emplean LS-AA en ambos lados del enlace para realizar multiplexado espacial de grandes volúmenes de datos. La principal diferencia entre mMIMO y MIMO consiste en que el primero incorpora detección multi-usuario (Multi-User Detection, MUD) para incrementar el flujo de información [40]. El éxito de la MUD depende de la capacidad del receptor de cancelar la interferencia generada por otros usuarios. Para ello se emplean técnicas de precodificación, las cuales requieren de una estimación precisa de la información del estado del canal (Channel State Information, CSI) de cada MU activo. Esta información puede obtenerse mediante señales piloto, o por realimentación de los MU [45].

Una interrogante en torno a mMIMO está en su funcionamiento en un esquema de duplexado por división de frecuencia (Frequency-Division Duplexing, FDD) [40]. La mayoría de los trabajos sobre mMIMO asumen un sistema operando bajo un esquema de duplexado por división en tiempo (Time-Division Duplexing, TDD) con la intención de evitar los problemas de estimación del canal en el DL [39]. No obstante, el tandem mMIMO-FDD puede acelerar el despliegue de la tecnología mMIMO en redes 5G, ya que los esquemas FDD han sido empleados ampliamente en la telefonía móvil. Otros problemas abiertos incluyen el desarrollo de algoritmos de detección y pre-codificación de baja complejidad [40].

C. Acceso Múltiple No-Ortogonal

Las técnicas de acceso al medio son también fundamentales para incrementar la eficiencia espectral. En las estrategias de acceso al medio previas a 5G se utiliza un principio de ortogonalidad, ya sea en TDMA, FDMA, o CDMA para evitar o limitar la interferencia entre usuarios. Estas estrategias son referidas como técnicas de acceso múltiple ortogonal (Orthogonal Multiple Access, OMA). En OMA, la capacidad de usuarios activos depende de los recursos ortogonales disponibles. Por otro lado, la estrategia de transmisión de acceso múltiple no ortogonal (Non-Orthogonal Multiple Access, NOMA) permite que varios usuarios utilicen simultáneamente el mismo canal de comunicación, sin importar si se produce interferencia de acceso múltiple [46]–[49]. NOMA mitiga dicha interferencia empleando estrategias de MUD. En general, NOMA ayuda a mejorar la eficiencia espectral del sistema de comunicación inalámbrico y la tasa de transmisión de todos los usuarios activos, permite la masificación de la conectividad, y reduce el tiempo de latencia [48]. Estos beneficios se obtienen a cambio de incrementar la complejidad de los receptores para darles la capacidad de realizar la MUD. Dentro de la filosofía de NOMA existen diversas implementaciones a nivel de bits o símbolos, pero todas buscan asignar firmas a los MU en función de la potencia de transmisión, códigos de esparcimiento, codificación o intercalamiento [47], [49]. NOMA se puede potencializar aún más al considerar una estrategia de transmisión y recepción MIMO. Al combinar ambas estrategias es posible mejorar la capacidad del canal comparando con MIMO-OMA [50], [51].

En NOMA por asignación de potencia [46], la potencia de transmisión es el factor de diversidad en el dominio del tiempo que permite identificar la información de cada usuario; así los usuarios más alejados de la BS transmitirán con una potencia mayor que un móvil cercano. Los receptores emplean cancelación de interferencia sucesiva (Successive Interference Cancellation, SIC) para desacoplar de forma iterativa la información de la interferencia según la potencia de transmisión de cada usuario. Esta estrategia se puede visualizar en la Fig. 4. Los principales problemas de investigación en NOMA están relacionados con el desarrollo de técnicas de decodificación y manejo eficiente de los recursos de transmisión (e.g., la potencia y el ancho de banda), así como con las estrategias de realimentación de información (como la CSI) entre las BS y los MU, que permitan la cancelación de la interferencia, y una reducción de complejidad en los receptores [52].

Además de las tecnologías anteriores, la eficiencia espectral también depende de esquemas avanzados de modulación y de nuevas formas de onda para la transmisión de información. Las tendencias de la tecnología 5G en este campo no se revisan en este artículo por limitaciones de espacio, pero el lector puede encontrar información al respecto en [24].

D. Celdas de Corto Alcance

La tercera dimensión de innovación, relacionada con la eficiencia espacial, comprende las mejoras dirigidas a soportar una mayor cantidad de usuarios conectados por unidad de área. La estrategia más simple—aunque efectiva, para mejorar la

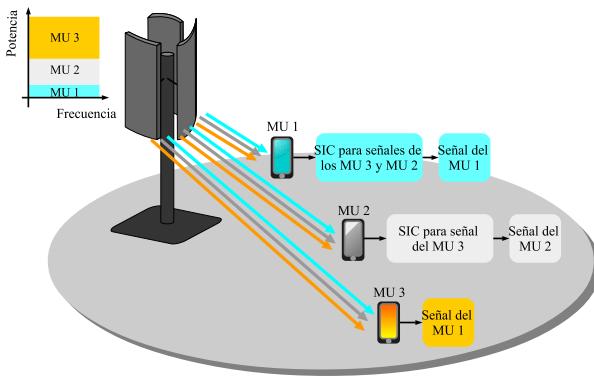


Fig. 4. Principio de NOMA por asignación de potencia.

eficiencia espacial es segmentar el área de cobertura, tal como lo sugiere el concepto de la red celular. De hecho, a medida que aumenta la densidad de usuarios conectados resulta conveniente reducir el tamaño de las celdas. Otra innovación importante de 5G radica precisamente en la densificación mediante el despliegue de un gran número de celdas pequeñas, cuyo alcance es menor a 100 metros.

Varios beneficios asociados con las celdas de corto alcance se presentan en [30], entre los que destacan la capacidad de proporcionar cobertura en zonas de alto tráfico así como al interior de edificaciones, y aumentar el caudal eficaz de la red en bandas de radiofrecuencia por debajo de los 6 GHz. No obstante, un alto número de celdas (pico y femto) conduce a un crecimiento significativo en el consumo de energía [53], [54], abriendo la necesidad por novedosos esquemas de eficiencia energética [55], [56]. El despliegue de las celdas de corto alcance también enfrenta obstáculos relacionados principalmente con cuestiones regulatorias y operativas, como altas cuotas por el uso de la infraestructura pública (como en postes de iluminación) para instalar estas celdas y diferentes criterios para medir el riesgo a la salud por la exposición a campos electromagnéticos de alta potencia [34]. Además de estos problemas, existen otros relacionados con la interconexión y administración de un gran número de celdas de corto alcance; algunos de ellos y sus soluciones se discuten a continuación.

V. TECNOLOGÍAS CLAVE PARA EL NÚCLEO DE 5G

La cuarta dimensión de innovación, enfocada en la eficiencia del sistema, involucra primordialmente al núcleo de la red 5G. La eficiencia buscada se relaciona, por ejemplo, con la reducción de los tiempos de latencia, el incremento de las tasas de transmisión, la reducción de la pérdida de paquetes y la disminución del consumo de energía.

A. Segmentación Lógica de la Red

Uno de los objetivos de 5G es ofrecer una amplia gama de servicios, cada uno con diferentes requisitos de calidad de servicio (Quality of Service, QoS) abriendo un abanico de oportunidades para aplicaciones provenientes de los diversos segmentos de la industria (también conocidos como verticales) [57], [27]. Para lograr este objetivo, los sistemas 5G deben ser capaces de proporcionar trozos de red mediante la aplicación

del concepto de segmentación lógica [58], [59] que tiene como principios a la automatización del funcionamiento de la red, la alta confiabilidad, la escalabilidad y el aislamiento de los segmentos pertenecientes a diferentes clientes, el soporte a *softwarización*, programabilidad y virtualización, la abstracción jerárquica y personalización de segmentos, y la elasticidad de los recursos de red [60]. Las tecnologías clave para realizar estos principios son la inteligencia artificial (Artificial Intelligence, AI), SDN, NFV, MEC, FC, y CC.

La segmentación lógica de red en 5G debe superar de manera inteligente diferentes desafíos, como la creación de algoritmos de admisión y planificación destinados a asignar recursos a los segmentos durante su creación y operación, la ubicación de las funciones de red en la infraestructura, el control del tráfico intra-segmento, la gestión extremo a extremo de los segmentos, el aislamiento entre segmentos, la gestión de la movilidad y orquestación de las NF que constituyen los segmentos, así como también de su seguridad y privacidad.

B. Inteligencia Artificial

Varias áreas de la AI, entre las que destaca el aprendizaje automático (Machine Learning, ML), son útiles para desarrollar segmentos lógicos cognitivos que son fundamentales para lograr redes 5G auto-dirigidas. Una red auto-dirigida debe desplegar de manera autónoma el ciclo Percibir-Planificar-Decidir-Actuar, el cuál debe estar acompañado de aprendizaje y realimentación, tanto de la red como un todo, como de sus segmentos constitutivos [61] [62]. Un estudio sobre el uso de técnicas de ML supervisadas, no supervisadas, e híbridas en redes de comunicaciones está disponible en [63]. Las técnicas de ML que recientemente han generado mayor interés en el dominio de las redes para automatizar la toma de decisiones son el aprendizaje por refuerzo (Reinforcement Learning, RL) y el aprendizaje por refuerzo profundo (Deep Reinforcement Learning, DRL) [64].

En RL, un agente toma decisiones periódicamente considerando estados (*e.g.*, conjunto de enrutadores disponibles para calcular una ruta) y acciones (*e.g.*, enviar tráfico a lo largo de la ruta calculada). El agente observa el resultado (expresado como una recompensa, por ejemplo, tráfico optimizado en términos de retraso) de su interacción con el entorno (*e.g.*, una red 5G) y luego ajusta automáticamente su estrategia para lograr una política óptima (*e.g.*, optimizar el retraso en un segmento lógico de URLLC) [65], [66], [67]. Los enfoques de RL, como Q-learning [68] y Estado-Acción-Recompensa-Estado-Acción (State–Action–Reward–State–Action, SARSA) [69], convergen lentamente a la política óptima cuando necesitan explorar y adquirir conocimiento sobre un conjunto estados-acciones grande, dificultando su uso en redes de gran escala como las de 5G.

El aprendizaje profundo ha sido utilizado recientemente como una herramienta para superar las limitaciones de RL, dando lugar a otro tipo de aprendizaje denominado DRL [70]. Las técnicas de aprendizaje profundo aprenden descubriendo estructuras complejas en los datos y construyen sus modelos computacionales utilizando múltiples capas de procesamiento;

una red neuronal profunda (Deep Neural Network, DNN) puede crear múltiples niveles de abstracción para representar los datos [71], [72]. Las estructuras típicas enfocadas en este tipo de aprendizaje son las redes neuronales prealimentadas (Feedforward Neural Network, FNN) y las recurrentes (Recurrent Neural Network, RNN). En FNN, no hay ciclos ni bucles en la red neuronal. Por tanto, los datos se mueven en una sola dirección: desde los nodos de entrada, a través de los nodos ocultos, y hacia los nodos de salida [73]. Las redes neuronales convolucionales (Convolutional Neural Network, CNN) son el modelo más conocido de FNN con una amplia variedad de aplicaciones, incluyendo enrutamiento dinámico [74]. A diferencia de FNN, una RNN es recursiva. Las conexiones entre neuronas producen ciclos específicos, significando que una salida depende no solo de sus entradas inmediatas, sino también del estado neuronal del paso anterior [75]. Una RNN está diseñada para usar datos secuenciales, cuando la etapa actual está relacionada con las etapas anteriores. La memoria a corto plazo (Long Short-Term Memory, LSTM) es una de las RNN más representativas [76] y es útil, por ejemplo, para la predicción del tráfico de red [77].

DRL utiliza redes neuronales profundas para mejorar la velocidad y el rendimiento de los algoritmos de RL [78]. Desde una perspectiva de alto nivel, DRL implementa una DNN para derivar la política óptima en lugar de la tabla Q usada por *Q-learning*. Los algoritmos DRL más relevantes son [64], [79]: DQL, DDQL, DQL con *Prioritized Experience Replay*, DDQN, *Distributional Deep Q-Learning*, *Deep Q-Learning With Noisy Nets*, y *Rainbow Deep Q-Learning*. Recientemente, DRL ha sido utilizado para abordar problemas en el contexto de la Internet de las cosas (Internet of Things, IoT), HetNets, UAV, SDN y NFV, todas ellas envueltas en la concepción de 5G. La integración de DRL en 5G podría revolucionar los mecanismos de gestión de los segmentos lógicos, especialmente los relacionados con la planificación y el uso compartido de recursos. Esta integración permitiría evolucionar de técnicas basadas en modelos, a aquellas sin un modelo, las cuales aprenden profundamente interactuando con el entorno para satisfacer tanto los acuerdos de nivel de experiencia (eXperience Level Agreements, XLA), como los de servicio (Service Level Agreements, SLA); estos acuerdos están relacionados con indicadores de calidad de experiencia (Quality of Experience, QoE) y QoS, respectivamente.

Debido a la complejidad de los sistemas de comunicación de 5G, los siguientes puntos serán fundamentales para la segmentación lógica de red basada en DRL: (i) la definición de los conjuntos de estados y acciones deberá abarcar la dinámica de los segmentos de red; (ii) el modelado de los conjuntos de estados y acciones deberá reducir el costo de aprendizaje; (iii) la recompensa deberá ser lo más simple posible para facilitar el proceso de aprendizaje; (iv) la cuantificación de la recompensa deberá incorporar el tiempo de su obtención (la interacción con el ambiente no es instantánea) y cómputo; y (v) los resultados proporcionados por los agentes de DRL deberán ser compatibles con la escala de tiempo en la que ocurren y son gestionados los eventos en una red 5G.

C. Redes Definidas por Software

SDN [80], [81], un concepto primordial para la *softwarización* de redes, es fundamental para administrar múltiples redes lógicas virtuales en el contexto de 5G [82]. La arquitectura SDN ofrece los siguientes beneficios [83], [84]: (i) clara separación del plano de datos y control, (ii) centralización lógica de la función de control, e (iii) implementación de la función de control en *software* [85], [86].

En [87], [88], [89] se presenta una vista de alto nivel de la arquitectura SDN y de sus planos. El plano de datos está compuesto por elementos de red especializados en el tratamiento de paquetes (re-envío), y se comunica con el plano de control a través de una o más interfaces de la zona sur (South Bound Interfaces, SBI). OpenFlow [90] es la SBI más conocida debido a su uso generalizado por parte de los vendedores de equipos de red y los grupos de investigación. El plano de aplicaciones implementa y organiza la lógica de las funciones de red (e.g., balanceo de carga y enrutamiento). El plano de aplicación comunica sus requisitos de red al plano de control mediante una o más interfaces de la zona norte (North Bound Interface, NBI). El plano de control traduce los requisitos de la aplicación y los aplica sobre los elementos de la red a través de SBI. El plano de control también define una o más interfaces de la zona este/oeste (East West Bound Interface, EWBI) que permiten implementar un control lógicamente centralizado y físicamente distribuido para hacer frente a redes de área amplia y gran escala mediante su segmentación lógica. El plano de gestión es ortogonal a los otros planos de SDN para realizar su operación administración y mantenimiento vía interfaces de gestión como OF-Config [91], y el protocolo de gestión de commutadores Open vSwitch (Open vSwitch Database Management Protocol, OVSDB) [92]. El plano de conocimiento es transversal a los planos mencionados, y tiene como objeto hacer redes SDN más inteligentes a partir del uso de ML [62], [61].

D. Virtualización de Funciones de Red

Las redes lógicas virtuales son primordiales para desplegar los servicios de 5G. NFV, definida por el instituto europeo de normas de telecomunicaciones (European Telecommunications Standards Institute, ETSI), es una tecnología que facilita el despliegue de múltiples segmentos lógicos de red sobre una infraestructura compartida [93]. NFV desacopla las NF del hardware en el cual se ejecutan para abstraer los recursos físicos de la infraestructura, dividirlos lógicamente y asignarlos a las VNF que componen las redes lógicas virtuales [94]–[96].

Los componentes principales de NFV son: la infraestructura de NFV (NFV Infrastructure, NFVI), las VNF, y el gestor/orquestador de NFV (MANagement and Network Orchestration, MANO) [97]. NFVI es la combinación de los recursos físicos y virtuales (hardware y software) que forman el entorno de implementación de las VNF. Una VNF puede implementar una o más funciones de red y ejecutarse en recursos virtuales, tal como máquinas virtuales ejecutando servicios o contenedores ejecutando microservicios. Como una VNF puede estar compuesta por otras VNF, ella puede ser desplegada en varias máquinas virtuales o contenedores. MANO incluye [98]: (i)

un orquestador de NFV para realizar la composición de los recursos de NFVI, (ii) un gestor de las VNF (Virtual Network Functions Manager, VNFM) responsable de administrar el ciclo de vida de las VNF; y (iii) un gestor de infraestructura (Virtual Infrastructure Manager, VIM) a cargo de virtualizar, monitorizar, configurar y controlar los recursos (físicos y virtuales) de red [99].

E. Computación en el Borde de la Red 5G

Los usuarios de las redes 5G requieren información contextual en tiempo real además de comunicaciones de baja latencia en una área geográfica determinada. La integración del concepto MEC a 5G permite cubrir estos requerimientos. MEC proporciona un alto ancho de banda, baja latencia, conocimiento de la ubicación e información de la RAN en tiempo real así como capacidades de CC en el borde de la red 5G [100], [101], [57]. Estas capacidades son ofrecidas a través de pequeños centros de datos desplegados para satisfacer la QoE requerida por los usuarios finales de las redes de 5G [102].

Una plataforma de colaboración construida mediante la integración MEC-5G debe ser genérica en funcionalidad así como también alojar y proporcionar diferentes subsistemas de servicios concebidos para diferentes verticales. El acceso seguro debe estar disponible en estos subsistemas para facilitar la implementación de aplicaciones [103]. MEC ofrece servicios de memoria caché y virtualización para reducir el volumen de datos transmitidos en el núcleo la red 5G y, de este modo, mejorar el uso de recursos [104].

FC es una alternativa a MEC para proporcionar información contextual en tiempo real además de comunicaciones de baja latencia en una área geográfica determinada [105]. Uno de los fundamentos de FC es respaldar IoT proporcionando conexiones seguras y servicios entre y para cosas, datos, personas y procesos [106], [107]; FC es esencial para MMTC. IoT implica un cambio profundo en la Internet actual hacia una red de cosas (objetos) interconectadas. Las cosas/objetos ayudan a integrar la inteligencia en nuestro entorno al soportar la recopilación de información cuando interactúan con el mundo (físico o virtual) [108]. FC mejora la CC al acercar las aplicaciones, la potencia computacional, el almacenamiento y las capacidades de comunicación de la nube a los usuarios finales [109], [110]. CC involucra tanto el hardware como el software de los centros de datos utilizados para proporcionar infraestructura, plataforma y aplicaciones como un servicio [111]. Acercar la CC a los usuarios finales y las cosas (*i.e.*, al borde de la red 5G) permite proveer de servicios de alto rendimiento y baja latencia con un uso eficiente de los recursos.

FC incluye tres capas: IoT, CC, y niebla [112], [25]. Cada capa es responsable de proporcionar diferentes funcionalidades [113], [114], [115]. La capa de CC consta de uno o más centros de datos que ofrecen infraestructura, plataforma y *software* como un servicio. La capa de niebla incluye uno o más nodos, los cuales pueden reutilizar interfaces inalámbricas y, además, coexistir con elementos de red, como SB o enrutadores de femtoceldas. Un nodo de niebla es una entidad (física/virtual)

que incluye sus capacidades de computación, almacenamiento y comunicaciones. Esta capa puede contener múltiples subcapas según los requisitos de las aplicaciones. La capa de IoT es responsable de enviar y recibir datos hacia y desde la capa de niebla que generalmente realiza una primera etapa de análisis o procesamiento de datos y, si es necesario, los envía para un análisis más detallado a la capa de CC. La capa de IoT incluye dispositivos como sensores que pueden interactuar entre sí formando redes *ad hoc*.

VI. ESTADO DEL DESPLIEGUE GLOBAL DE 5G

El despliegue comercial de la tecnología 5G comenzó a finales del 2018 en Asia. Enseguida se realizará un breve recuento histórico de este despliegue (ver Fig. 5), enfatizando en las compañías y operadores líderes, según lo expuesto en [5]. Corea del Sur presentó la primera implementación exitosa de una red de prueba 5G en febrero de 2018 durante los Juegos de Invierno en PyeonChang. Siguiendo este empuje inicial, el lanzamiento comercial de 5G se realizó en este mismo país en diciembre por parte de las compañías SK Telecom, Korea Telecom y LG Uplus. Cabe resaltar que Corea del Sur es hasta ahora el país que tiene la cobertura 5G más amplia y con las mayores tasas de transmisión [116].

China representa el mercado mundial más grande para la tecnología 5G. El despliegue de 5G comenzó en ese país en octubre de 2019 a través de China Mobile, Unicom y China Telecom, siendo la compañía Huawei [117] uno de sus grandes impulsores. En la región Asia-Pacífico, Australia comenzó a operar sus primeras redes comerciales 5G en mayo de 2019 a través de Telstra, Nueva Zelanda lo hizo en diciembre a través de Spark y Vodafone, y Filipinas en junio con Globe Telecom. De esta manera, la región Asia-Pacífico es el área geográfica con mayor interés en 5G, y se espera que cuente con dos terceras partes de los subscriptores mundiales para 2024 [118].

En Europa, el desarrollo comercial de la tecnología 5G comenzó en el Reino Unido a partir de mayo de 2019, esto auspiciado por EE, Vodafone, O2 y Three. En Suiza, el despliegue comercial fue realizado por Swisscom y Sunrise desde mayo de 2019, mientras que en Italia estuvo a cargo de Vodafone y Telecom Italia desde junio del mismo año. Alemania fue otro de los primeros países en operar redes 5G comerciales a través de Deutsche Telekom y Vodafone, iniciando operaciones en julio de 2019. Rumania, Austria, Irlanda, Finlandia y Hungría se unieron a éste despliegue comercial en el transcurso de ese mismo año. Otros países Europeos, como España, Holanda, Suecia, Dinamarca y Bélgica, activaron sus redes 5G comerciales en 2020.

Norteamérica es otro gran mercado para la tecnología 5G, donde Verizon llevó a cabo el primer lanzamiento comercial 5G en mayo de 2019, mientras que AT&T y T-Mobile lo hicieron en el segundo semestre de ese mismo año. En Canadá, el liderazgo lo llevan Rogers Communications y Bell Mobility, donde el despliegue comercial de 5G se inició en 2020. Otro mercado importante es el Medio Este, donde Kuwait, Emiratos Árabes Unidos y Catar comenzaron a operar sus primeras redes comerciales 5G en el periodo de abril a junio de 2019.



Fig. 5. Despliegue comercial de la tecnología móvil 5G por países y operadores (primeros en proveer el servicio) entre 2018 y 2020.

En África, el país líder en esta tecnología es Sudáfrica, en donde Rain y Vodacom activaron sus redes 5G en septiembre de 2019 [5].

El estado actual de la tecnología 5G en Latinoamérica es sumamente incipiente. La primera red comercial 5G surgió en Uruguay y es operada por la compañía Antel de la mano de Nokia desde abril de 2019. Según el reporte en [4], aparte de Uruguay, solo Brasil cuenta con redes comerciales de telefonía móvil 5G, las cuales son operadas por la compañía Claro. En Colombia se tienen redes 5G de internet fijo a través de Directv, mientras que otros países, como Argentina, Chile, Perú y México se encuentran en etapas tempranas de inversión y despliegue. En el área del Caribe, Puerto Rico y las Islas Vírgenes cuentan redes 5G desde diciembre de 2019, las cuales son operadas por T-Mobile. Como se describe en [119], uno de los grandes retos en Latinoamérica consiste en mejorar la infraestructura celular para alcanzar niveles de QoS y cobertura cercanos a los de países líderes como Japón, Reino Unido y Estados Unidos. De hecho, según [120], en los próximos diez años el desarrollo en telefonía móvil para Latinoamérica se centrará aún en la tecnología 4G, que se espera alcance al 67 % de la población de la región para 2025, lo cual cimiente la integración de 5G en un futuro cercano.

En resumen, el estado mundial del despliegue de la tecnología 5G se puede dimensionar según lo reportado en [4], donde hasta septiembre de este año, existen 397 operadores de tecnología inalámbrica de 129 países o territorios independientes, los cuales han invertido en pruebas, adquirido licencias, planificado en desarrollo de red, o puesto en marcha redes 5G. Más allá, 101 operadores de 44 países o territorios han generado uno o varios servicios compatibles con servicios 3GPP para 5G. Una gran herramienta para visualizar el avance mundial de esta tecnología son los mapas interactivos de cobertura 5G elaborados por OOKLA [116].

Durante el despliegue global de 5G es también de suma

importancia abordar desafíos relacionados con las posibles consecuencias en la salud y el medio ambiente. Estas implicaciones están asociadas entre otros a los siguientes factores. Primero, las tecnologías pilares de 5G en la capa física (comunicaciones por ondas milimétricas, arreglos de antenas de gran escala, y celdas de corto alcance) emplean frecuencias de transmisión mucho mayores que las de generaciones anteriores [121]. De este modo, los niveles de penetración en el tejido biológico serán menores, pero otros órganos como la piel y los ojos podrían verse afectados [122]. Segundo, es factible un incremento de la huella de carbono generado por el aumento en el número de dispositivos conectados, especialmente para dar cabida a la IoT industrial, y el crecimiento tanto del tamaño de los centros de datos de la CC usados para soportar la virtualización del núcleo de la red, como de los empleados en FC o MEC para habilitar la virtualización de la RAN [123]. Para abordar estos desafíos, la comunidad científica ha solicitado que los gobiernos inviertan en estudios de impacto ambiental para dimensionar los posibles riesgos con el despliegue de 5G, y enseguida generar las políticas regulatorias que protejan la salud de la población [124]. Además, los investigadores deben continuar trabajo arduamente para lograr redes 5G eficientes y sostenibles desde el punto de vista de la huella de carbono y la eficiencia energética [125], [126].

VII. TELEFONÍA MÓVIL MÁS ALLÁ DE 5G

Las redes 5G transformarán el panorama de las comunicaciones móviles habilitando enlaces ultra confiables y de ultra baja latencia, capaces de conectar a personas y máquinas por igual. Estas redes resultarán fundamentales para desarrollar y desplegar todo el potencial de la Industria 4.0, los sistemas de transporte inteligente, el IoT, el manejo eficiente de energía y las ciudades inteligentes [127]. No obstante, existen aplicaciones cuyos requerimientos podrían exceder las capacidades proyectadas para 5G. Por ejemplo, aunque 5G se perfila como la

tecnología idónea para vehículos conectados, no hay garantía de que sea capaz de satisfacer la latencia extremadamente baja requerida por vehículos autónomos. Además, las redes 5G se enfocan principalmente en las comunicaciones en dos dimensiones, es decir, en las comunicaciones móviles terrestres. Sin embargo, el crecimiento de la industria aeronáutica y aeroespacial, evidenciado por la proliferación de los UAV, los nano-satélites, y las plataformas de gran altitud, plantea un escenario de comunicaciones tridimensional, es decir entornos terrestre, aéreo y acuático. También, la madurez que han alcanzado las tecnologías de realidad aumentada e interfaces hapticas está abriendo el camino hacia las comunicaciones interactivas, las cuales no están contempladas en 5G. Dos ejemplos de este tipo de comunicaciones son el Internet táctil o haptico [128] y las comunicaciones holográficas [129].

Aunque 5G está en una etapa inicial, los trabajos para definir las características de la sexta generación de redes de telefonía móvil ya están en curso. En julio de 2018 la UIT estableció el grupo temático Network 2030 cuya tarea fue analizar los retos y oportunidades en la evolución de las redes de comunicaciones hacia el año 2030 y más allá [130]. Este grupo identificó siete casos de uso para las redes 2030, entre los que destacan las comunicaciones de tipo holográfico, el Internet táctil para operaciones remotas, la red integrada terrestre-espacial, y el IoT industrial con *cloudificación*. Aunque la visión de como serán las redes 6G aún no se consolida, la expectativa es que estas redes habiliten las comunicaciones ultra masivas entre máquinas, con latencias extremadamente bajas (de entre 10 y 100 μ s), confiabilidad extrema (tasas de error de trama de 1- 10^{-9}), consumo de potencia extremadamente bajo (eficiencia energética de 1 pJ/b), y soporte de enlaces de banda ancha en condiciones de muy alta movilidad (usuarios desplazándose a más de 1,000 km/h) [131], [132].

Para poder alcanzar estos objetivos se necesita un cambio en el paradigma del diseño de las redes de telefonía móvil, en el que las comunicaciones ya no se orienten a personas o máquinas, sino a los datos [133], [134]. Además, el manejo eficiente de los grandes volúmenes de información que demandarán las aplicaciones de 6G requiere que el diseño de la red adopte a la inteligencia artificial como uno de sus pilares y no solamente como una herramienta de soporte [135]. Finalmente, los sistemas de transmisión deberán operar en bandas de frecuencias aún más altas que las de 5G para poder alcanzar tasas pico de transmisión en el orden de los Tbps. Las tecnologías de transmisión que se están considerando para 6G incluyen comunicaciones ópticas por láser y luz visible (visible light communications, VLC), enlaces de radiofrecuencia en la banda de los THz, sistemas MIMO super masivos, conformación de haz holográfico, superficies inteligentes, técnicas de multiplexión basadas en el momento orbital angular de las ondas electromagnéticas [131]–[134].

6G impondrá desafíos adicionales relacionados con, primero, la gestión y orquestación de la tridimensionalidad de las redes celulares que incluirá tanto a las BS, como a los MU basados en UAV. Segundo, la evolución de los algoritmos de admisión y planificación de segmentos lógicos de red generada por la aparición de nuevos casos de uso requiriendo de, por ejemplo, velocidad de datos y confiabilidad extremamente

altas; es decir, una combinación de 5G EMBB y URLLC. Las redes 6G permiten imaginar un escenario en el que proliferan las ciudades inteligentes. Aunque este escenario se percibe lejano, el camino ha comenzado a construirse con la aparición de las redes 5G.

VIII. CONCLUSIONES

El despliegue comercial que se ha iniciado de las redes 5G permitirá extender las aplicaciones móviles en las vertientes: EMBB, MMTC y URLLC. Sin embargo, este cambio tecnológico será paulatino, y en los próximos años se mantendrá una co-existencia de servicios, principalmente con redes 4G. En general, el cambio modular de la tecnología 5G se enfoca en la arquitectura de red por medio de los conceptos como segmentación lógica, *softwarización*, programabilidad, y la virtualización de red. En particular, en Latinoamérica la penetración de las redes 5G es aún baja, y en los próximos años se prevé una consolidación de servicios 4G y un inicio de la inversión en infraestructura 5G. A pesar de esto, a nivel mundial ya se han iniciado los esfuerzos de investigación y visualización de las necesidades hacia las comunicaciones 6G.

AGRADECIMIENTOS

Los autores agradecen el financiamiento de CONACYT a través del proyecto de Ciencia Básica No. 254637.

REFERENCES

- [1] R. W. Jones, “The global framework for radiocommunications,” *ITU News*, no. 3, pp. 29–31, Apr. 2006.
- [2] M. K. Roberts, M. A. Temple, R. F. Mills, and R. A. Raines, “Evolution of the air interface of cellular communications systems toward 4G realization,” *IEEE Commun. Surveys Tuts.*, vol. 8, no. 1, pp. 2–23, 2006.
- [3] R. Baldemair *et al.*, “Evolving wireless communications: Addressing the challenges and expectations of the future,” *IEEE Veh. Technol. Mag.*, vol. 8, no. 1, pp. 24–30, 2013.
- [4] Global mobile Suppliers Association. 5G market: Snapshot september 2020. [Online]. Available: <https://gsacom.com>
- [5] GSMA Documents. 5G global launches & statistics. [Online]. Available: https://www.gsma.com/futurenetworks/ip_services/understanding-5g/5g-innovation/
- [6] International Telecommunications Union, *IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond*. International Telecommunications Union-Radiocommunications Sector, Recommendation ITU-R M.2083-0, Sep. 2015.
- [7] ———, *Minimum requirements related to technical performance for IMT-2020 radio interface(s)*. International Telecommunications Union-Radiocommunications Sector, Report ITU-R M.2410-0, Nov. 2017.
- [8] M. Shafi *et al.*, “5G: A tutorial overview of standards, trials, challenges, deployment, and practice,” *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1221, 2017.
- [9] A. Ghosh, A. Maeder, M. Baker, and D. Chandramouli, “5G evolution: A view on 5G cellular technology beyond 3gpp release 15,” *IEEE Access*, vol. 7, pp. 127 639–127 651, 2019.
- [10] 3rd Generation Partnership Project, *3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Release 15 Description; Summary of Rel-15 Work Items (Release 15)*. 3GPP-3GPP TR 21.915 V15.0.0, Sep. 2019.
- [11] A. Gupta and R. K. Jha, “A survey of 5G network: Architecture and emerging technologies,” *IEEE Access*, vol. 3, pp. 1206–1232, 2015.
- [12] J. G. Andrews *et al.*, “What will 5G be?” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.
- [13] M. A. Habibi, M. Nasimi, B. Han, and H. D. Schotten, “A comprehensive survey of RAN architectures toward 5G mobile communication system,” *IEEE Access*, vol. 7, pp. 70 371–70 421, 2019.
- [14] G. Marconi, “Wireless telegraphic communication – Nobel lecture,” *NobelPrize.org. Nobel Media AB*, pp. 1–27, Dec. 1909.

- [15] IEEE Communications Society, *A Brief History of Communications*. New Jersey: IEEE, 2002.
- [16] D. H. Ring, "Mobile telephony – Wide area coverage," Bell Telephone Laboratories Inc., Murray Hill, NJ, USA, Tech. Rep., Dec. 1947, Internal Tech. Memo.
- [17] D. Raychaudhuri and N. B. Mandayam, "Frontiers of wireless and mobile communications," *Proc. IEEE*, vol. 100, no. 4, pp. 824–840, 2012.
- [18] T. Halonen, J. Romero, and J. Melero, *GSM, GPRS and EDGE Performance: Evolution towards 3G/UMTS*, 2nd ed. John Wiley & Sons, Ltd, 2013.
- [19] P. Lescuyer, *UMTS: origins, architecture and the standard*. Springer Science & Business Media, 2004.
- [20] J. Lee *et al.*, "LTE-advanced in 3GPP Rel -13/14: an evolution toward 5G," *IEEE Commun. Mag.*, vol. 54, no. 3, pp. 36–42, Mar. 2016.
- [21] B. Furht and S. A. Ahson, *Long Term Evolution: 3GPP LTE radio and cellular technology*. CRC Press, 2009.
- [22] B. A. Bjerke, "LTE-advanced and the evolution of LTE deployments," *IEEE Wireless Commun. Mag.*, vol. 18, no. 5, pp. 4–5, Oct. 2011.
- [23] Ericsson. Ericsson mobility report november 2020. [Online]. Available: <https://www.ericsson.com/en/mobility-report>
- [24] International Telecommunications Union, *Future technology trends of terrestrial IMT systems*. International Telecommunications Union-Radiocommunications Sector, Report ITU-R M.2320-0, Nov. 2014.
- [25] K. Liang, L. Zhao, X. Chu, and H. Chen, "An integrated architecture for software defined and virtualized radio access networks with fog computing," *IEEE Network*, vol. 31, no. 1, pp. 80–87, 2017.
- [26] Y. Choi and N. Park, "Slice architecture for 5G core network," in *ICUFN*, 2017, pp. 571–575.
- [27] C. H. T. Arteaga, A. Ordoñez, and O. M. C. Rendon, "Scalability and performance analysis in 5G core network slicing," *IEEE Access*, vol. 8, pp. 142 086–142 100, 2020.
- [28] J. S. Orduz, G. D. Orozco, C. H. Tobar-Arteaga, and O. M. C. Rendon, "μvims: A finer-scalable architecture based on microservices," in *LCN*, 2019, pp. 141–148.
- [29] Z. Sun *et al.*, "Building dynamic mapping with cups for next generation automotive edge computing," in *CloudNet*, 2019, pp. 1–6.
- [30] H. Viswanathan and M. Weldon, "The past, present, and future of mobile communications," *Bell Labs Technical Journal*, pp. 8–21, 2014.
- [31] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.
- [32] M. Shafi *et al.*, "Microwave vs. millimeter-wave propagation channels: Key differences and impact on 5G cellular systems," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 14–20, Dec. 2018.
- [33] C. A. Gutiérrez *et al.*, "Doppler shift characterization of wideband mobile radio channels," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 12 375–12 380, Dec. 2019.
- [34] K. Husenovic, I. Bedi, S. Maddens, I. Bozsoki, D. Daryabwite, N. Sundberg, M. Maniewicz *et al.*, "Setting the scene for 5G: Opportunities & challenges," *International Telecommunications Union*, vol. 56, 2018.
- [35] I. A. Rumyancev and A. S. Korotkov, "Survey on beamforming techniques and integrated circuits for 5G systems," in *EExPolytech*, 2019, pp. 76–80.
- [36] I. Ahmed *et al.*, "A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3060–3097, 2018.
- [37] T. Okuyama *et al.*, "5G experimental trials of 4.5 GHz band digital beamforming in dense urban area," in *PIMRC*, 2018, pp. 1130–1131.
- [38] S. Parkvall *et al.*, "5G wireless access - trial concept and results," in *GLOBECOM*, 2015, pp. 1–6.
- [39] D. C. Araújo *et al.*, "Massive MIMO: survey and future research topics," *IET Communications*, vol. 10, no. 15, pp. 1938–1946, 2016.
- [40] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: Ten myths and one critical question," *IEEE Commun. Mag.*, vol. 54, no. 2, pp. 114–123, Feb. 2016.
- [41] M. A. Albreem, M. Juntti, and Shahabuddin, "Massive MIMO detection techniques: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3109–3132, 2019.
- [42] B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.
- [43] D. Hwang, S. S. Nam, and J. Yang, "Multi-antenna beamforming techniques in full-duplex and self-energy recycling systems: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 160–167, 2017.
- [44] O. El Ayach *et al.*, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wirel. Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
- [45] P. von Butovitsch *et al.*, "Advanced antenna systems for 5G networks," *Ericsson White Paper GFMC-18:000530*, pp. 1–15, 2018.
- [46] S. M. R. Islam, N. Avazov, O. A. Dobre, and K. Kwak, "Power-domain non-orthogonal multiple access (noma) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2017.
- [47] L. Dai *et al.*, "A survey of non-orthogonal multiple access for 5G," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2294–2323, 2018.
- [48] N. Mitton *et al.*, "A tutorial on nonorthogonal multiple access for 5G and beyond," *Wireless Communications and Mobile Computing*, vol. 2018, p. 9713450, 2018.
- [49] M. B. Shahab, R. Abbas, M. Shirvanimoghaddam, and S. J. Johnson, "Grant-free non-orthogonal multiple access for iot: A survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 1805–1838, 2020.
- [50] Y. Liu, G. Pan, H. Zhang, and M. Song, "On the capacity comparison between mimo-noma and mimo-oma," *IEEE Access*, vol. 4, pp. 2123–2129, 2016.
- [51] M. Zeng *et al.*, "On the sum rate of MIMO-NOMA and MIMO-OMA systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 4, pp. 534–537, 2017.
- [52] B. Makki, K. Chitti, A. Behravan, and M. Alouini, "A survey of noma: Current status and open research challenges," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 179–189, 2020.
- [53] A. Abrol and R. K. Jha, "Power optimization in 5g networks: A step towards green communication," *IEEE Access*, vol. 4, pp. 1355–1374, 2016.
- [54] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 4, pp. 524–540, 2011.
- [55] K. Chang, K. Chu, H. Wang, Y. Lin, and J. Pan, "Energy saving technology of 5g base station based on internet of things collaborative control," *IEEE Access*, vol. 8, pp. 32 935–32 946, 2020.
- [56] A. Dataesatu, P. Boonsrimuang, K. Mori, and P. Boonsrimuang, "Energy efficiency enhancement in 5g heterogeneous cellular networks using system throughput based sleep control scheme," in *ICACT*, 2020, pp. 549–553.
- [57] A. A. Barakabitze, A. Ahmad, R. Mijumbi, and A. Hines, "5G network slicing using sdn and nfv: A survey of taxonomy, architectures and future challenges," *Computer Networks*, vol. 167, p. 106984, 2020.
- [58] X. Zhou, R. Li, T. Chen, and H. Zhang, "Network slicing as a service: enabling enterprises' own software-defined cellular networks," *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 146–153, 2016.
- [59] C. Campolo, A. Molinaro, A. Iera, and F. Menichella, "5G network slicing for vehicle-to-everything services," *IEEE Wireless Communications*, vol. 24, no. 6, pp. 38–45, 2017.
- [60] I. Afolabi *et al.*, "Network slicing and softwarization: A survey on principles, enabling technologies, and solutions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2429–2453, 2018.
- [61] S. Ayoubi *et al.*, "Machine learning for cognitive network management," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 158–165, 2018.
- [62] A. Mestres *et al.*, "Knowledge-defined networking," *SIGCOMM Comput. Commun. Rev.*, vol. 47, no. 3, p. 2–10, sep 2017.
- [63] R. Boutaba *et al.*, "A comprehensive survey on machine learning for networking: evolution, applications and research opportunities," *J. Internet Serv. Appl.*, vol. 9, no. 1, pp. 16:1–16:99, 2018.
- [64] N. C. Luong *et al.*, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3133–3174, 2019.
- [65] D. M. Casas-Velasco, O. M. C. Rendon, and N. L. S. da Fonseca, "Intelligent routing based on reinforcement learning for software-defined networking," *IEEE TNSM*, pp. 1–1, 2020.
- [66] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.
- [67] Mausam and A. Kolobov, *Planning with Markov Decision Processes*. Morgan and Claypool, 2012.
- [68] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.
- [69] Y.-H. Wang, T.-H. S. Li, and C.-J. Lin, "Backward q-learning: The combination of sarsa algorithm and q-learning," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 9, pp. 2184–2193, 2013.
- [70] H. Li *et al.*, "Deep reinforcement learning: Framework, applications, and embedded implementations," in *ICCAD*, 2017, p. 847–854.
- [71] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1.

- [72] S. Amiri, S. Salimzadeh, and A. S. Z. Belloum, "A survey of scalable deep learning frameworks," in *eScience*, 2019, pp. 650–651.
- [73] T. L. Fine, *Feedforward neural network methodology*. Springer Science & Business Media, 2006.
- [74] S. K. Sahu, P. Kumar, and A. P. Singh, "Dynamic routing using inter capsule routing protocol between capsules," in *UKSim-AMSS*, 2018, pp. 1–5.
- [75] A. Sherstinsky, "Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network," *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, 2020.
- [76] F. A. Gers, J. A. Schmidhuber, and F. A. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, p. 2451–2471, Oct. 2000.
- [77] H. Lu and F. Yang, "A network traffic prediction model based on wavelet transformation and lstm network," in *IEEE ICSESS*, 2018, pp. 1–4.
- [78] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A proc. survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, 2017.
- [79] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.
- [80] N. Feamster, J. Rexford, and E. Zegura, "The road to sdn: An intellectual history of programmable networks," *SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 2, pp. 87–98, Apr. 2014.
- [81] P. Lin *et al.*, "A quick survey on selected approaches for preparing programmable networks," in *AINTEC*, 2011, pp. 160–163.
- [82] Z. Shu and T. Taleb, "A novel qos framework for network slicing in 5G and beyond networks based on sdn and nfv," *IEEE Network*, vol. 34, no. 3, pp. 256–263, 2020.
- [83] J. d. J. Gil Herrera and J. F. B. Vega, "Network functions virtualization: A survey," *IEEE LATAMT*, vol. 14, no. 2, pp. 983–997, Feb 2016.
- [84] I. Afolabi *et al.*, "Network slicing and softwarization: A survey on principles, enabling technologies, and solutions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 2429–2453, thirdquarter 2018.
- [85] D. Kreutz *et al.*, "Software-defined networking: A comprehensive survey," *Proc. IEEE*, vol. 103, no. 1, pp. 14–76, Jan 2015.
- [86] B. A. A. Nunes *et al.*, "A survey of software-defined networking: Past, present, and future of programmable networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1617–1634, Third 2014.
- [87] S. Bailey *et al.*, "SDN architecture overview," *Open Network Foundation*, 2013.
- [88] F. Estrada-Solano, A. Ordóñez, L. Z. Granville, and O. M. C. Rendon, "A framework for SDN integrated management based on a CIM model and a vertical management plane," *Elsevier Computer Communications*, vol. 102, pp. 150–164, 2017.
- [89] J. A. Wickboldt *et al.*, "Software-defined networking: management requirements and challenges," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 278–285, 2015.
- [90] N. McKeown *et al.*, "Openflow: enabling innovation in campus networks," *ACM SIGCOMM Comp. Commun. Rev.*, vol. 38, no. 2, pp. 69–74, 2008.
- [91] Tomás Cejka and Radek Krejčí, "Configuration of open vswitch using of-config," in *NOMS*, 2016, pp. 883–888.
- [92] C. Caba and J. Soler, "Apis for qos configuration in software defined networks," in *NetSoft*, 2015, pp. 1–5.
- [93] B. Chatras, U. S. Tsang Kwong, and N. Bihannic, "Nfv enabling network slicing for 5G," in *ICIN*, 2017, pp. 219–225.
- [94] K. Giotis, Y. Kryftis, and V. Maglaris, "Policy-based orchestration of nfv services in software-defined networks," in *NetSoft*, 2015, pp. 1–5.
- [95] L. Mamatas, S. Clayman, and A. Galis, "A flexible information service for management of virtualized software-defined infrastructures," *Netw.*, vol. 26, no. 5, p. 396–418, Sep. 2016.
- [96] F. Bari *et al.*, "Orchestrating virtualized network functions," *IEEE TNMS*, vol. 13, no. 4, pp. 725–739, 2016.
- [97] ETSI GS NFV 002, "Network functions virtualization (NFV); architectural framework v1.1.1," ETSI, Tech. Rep., October 2013.
- [98] S. R. Ali, *Network Function Virtualization*. Cham: Springer International Publishing, 2019, pp. 131–156.
- [99] B. Han, V. Gopalakrishnan, L. Ji, and S. Lee, "Network function virtualization: Challenges and opportunities for innovations," *IEEE Commun. Mag.*, vol. 53, no. 2, pp. 90–97, 2015.
- [100] T. X. Tran, A. Hajisami, P. Pandey, and D. Pompili, "Collaborative mobile edge computing in 5G networks: New paradigms, scenarios, and challenges," *IEEE Commun. Mag.*, vol. 55, no. 4, pp. 54–61, 2017.
- [101] T. Taleb *et al.*, "On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1657–1681, 2017.
- [102] C. Tselios and G. Tsolis, "On qoe-awareness through virtualized probes in 5G networks," in *CAMAD*. IEEE, 2016, pp. 159–164.
- [103] S. Nunna *et al.*, "Enabling real-time context-aware collaboration through 5G and mobile edge computing," in *ITNG*. IEEE, 2015, pp. 601–605.
- [104] Y. C. Hu *et al.*, "Mobile edge computing — a key technology towards 5G," *ETSI white paper*, vol. 11, no. 11, pp. 1–16, 2015.
- [105] E. Baccarelli *et al.*, "Fog of everything: Energy-efficient networked computing architectures, research challenges, and a case study," *IEEE Access*, vol. 5, pp. 9882–9910, 2017.
- [106] M. H. Miraz, M. Ali, P. S. Excell, and R. Picking, "A review on internet of things (iot), internet of everything (ioe) and internet of nano things (iont)," in *ITA*, 2015, pp. 219–224.
- [107] IEEE-SA Standards Board, "IEEE standard for adoption of openfog reference architecture for fog computing," *IEEE Std 1934-2018*, pp. 1–176, 2018.
- [108] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645 – 1660, 2013.
- [109] A. Brogi and S. Forti, "Qos-aware deployment of iot applications through the fog," *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1185–1192, 2017.
- [110] R. Bruschi, F. Davoli, P. Lago, and J. F. Pajo, "A scalable sdn slicing scheme for multi-domain fog/cloud services," in *NetSoft*, 2017, pp. 1–6.
- [111] M. Armbrust *et al.*, "A view of cloud computing," *Commun. ACM*, vol. 53, no. 4, p. 50–58, Apr. 2010.
- [112] X. He, Z. Ren, C. Shi, and J. Fang, "A novel load balancing strategy of software-defined cloud/fog networking in the internet of vehicles," *China Communications*, vol. 13, no. Supplement2, pp. 140–149, 2016.
- [113] R. Jain and S. Paul, "Network virtualization and software defined networking for cloud computing: a survey," *IEEE Commun. Mag.*, vol. 51, no. 11, pp. 24–31, 2013.
- [114] D. Bendouda, A. Rachedi, and H. Haffaf, "An hybrid and proactive architecture based on sdn for internet of things," in *IWCNC*, 2017, pp. 951–956.
- [115] Y. Xiao and M. Krantz, "Qoe and power efficiency tradeoff for fog computing networks with fog node cooperation," in *IEEE INFOCOM*, 2017, pp. 1–9.
- [116] SPEEDTEST TM . Ookla 5G map. [Online]. Available: <https://www.speedtest.net/ookla-5g-map>
- [117] Huawei Technologies Co., "Green 5G: Bulding a sustainable world," *White paper*, pp. 1–26, August 2020.
- [118] Informa Telecoms & Media Limited. 5G world map: A global breakdown of deployment. [Online]. Available: <https://mt.knect365.com/5gworldevent/5g-world-map-a-global-breakdown-of-deployment-blog/>
- [119] GSMA Latin America, "Infrastructure deployment in latin america," *GSMA Documents*, 2015. [Online]. Available: <https://www.gsma.com/latinamerica/infrastructure-deployment-in-latin-america/>
- [120] GSM Association, "The mobile economy latin america 2019," *GSMA Documents*, pp. 1–52, 2019. [Online]. Available: <http://www.gsma.com/>
- [121] D. H. Gultekin and P. H. Siegel, "Absorption of 5g radiation in brain tissue as a function of frequency, power and time," *IEEE Access*, vol. 8, pp. 115 593–115 612, 2020.
- [122] L. Hardell and M. Carlberg, "Health risks from radiofrequency radiation, including 5g, should be assessed by experts with no conflicts of interest," *Oncology Letters*, vol. 20, no. 4, pp. 1–11, 2020.
- [123] A. Fehske, G. Fettweis, J. Malmodin, and G. Biczok, "The global footprint of mobile communications: The ecological and economic perspective," *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 55–62, 2011.
- [124] M. Simkó and M. Mattsson, "5g wireless communication and health effects—a pragmatic review based on available studies regarding 6 to 100 ghz," *International Journal of Environmental Research and Public Health*, vol. 16, no. 18, 2019.
- [125] Q. Wu, G. Y. Li, W. Chen, D. W. K. Ng, and R. Schober, "An overview of sustainable green 5g networks," *IEEE Wirel. Commun.*, vol. 24, no. 4, pp. 72–80, 2017.
- [126] A. Bohli and R. Bouallegue, "How to meet increased capacities by future green 5g networks: A survey," *IEEE Access*, vol. 7, pp. 42 220–42 237, 2019.
- [127] C. Del-Valle-Soto, L. J. Valdivia, R. Velázquez, L. Rizo-Domínguez, and J.-C. López-Pimentel, "Smart campus: An experimental performance comparison of collaborative and cooperative schemes for wireless sensor network," *Energies*, vol. 12, no. 16, 2019.
- [128] G. Fettweis, "The tactile internet: Applications and challenges," *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 64–70, Mar. 2014.

- [129] Y. Huo, P. T. Kovacs, T. J. Naughton, and L. Hanzo, "Wireless holographic image communications relying on unequal error protected bitplanes," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 7136–7148, Aug. 2017.
- [130] International Telecommunications Union. (2019.) Focus group on technologies for network 2030. [Online]. Available: <https://www.itu.int/en/ITU-T/focusgroups/net2030/>
- [131] E. Calvanese Strinati *et al.*, "6G the next frontier: From holographic messaging to artificial intelligence using subterahertz and visible light communication," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 42–50, Sep. 2019.
- [132] Z. Zhang *et al.*, "6G wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 28–41, Sep. 2019.
- [133] H.-C. Yang and M.-S. Alouini, "Data-oriented transmission in future wireless systems: Toward trustworthy support of advanced Internet of things," *IEEE Veh. Technol. Mag.*, vol. 13, no. 3, pp. 78–83, Sep. 2019.
- [134] B. Zong *et al.*, "6G technologies: Key drivers, core requirements, system architectures, and enabling technologies," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 18–27, Sep. 2019.
- [135] M. Elsayed and M. Erol-Kantarci, "AI-enabled future wireless networks: Challenges, opportunities, and open issues," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 70–77, Sep. 2019.

ACRÓNIMOS

3GPP	Proyecto Asociación de Tercera Generación (3rd Generation Partnership Project)
5G-NR	5G New Radio
CDMA	Múltiple acceso por división por códigos (Code-Division Multiple Access)
CSI	Información del estado del canal (Channel State Information)
DL	Enlace descendente (Downlink)
EMBB	Banda Ancha Móvil Mejorada (Enhanced Mobile Broadband)
EMS	Especro electromagnético (Electromagnetic Spectrum)
FDMA	Múltiple acceso por división en frecuencia (Frequency Division Multiple Access)
GSM	Sistema global para comunicaciones móviles (Global System for Mobile Communications)
IP	Protocolo de Internet (Internet Protocol)
IMT	Telecomunicaciones móviles internacionales (International Mobile Telecommunications)
ITU	Unión Internacional de Telecomunicaciones (International Telecommunications Union)
OFDMA	Múltiple acceso por división en frecuencia ortogonal (Orthogonal Frequency Division Multiple Access)
OMA	Múltiple acceso ortogonal (Orthogonal Multiple Access)
OVSDB	Protocolo de gestión de comutadores Open vSwitch (Open vSwitch Database Management Protocol)
QoS	Calidad de servicio (Quality of Service)
LTE	Evolución de largo plazo (Long Term Evolution)
MAC	Control de acceso al medio (Medium Access Control)
MMTC	Comunicaciones Masivas de Tipo Máquina (Massive Machine-Type Communications)
NOMA	Múltiple acceso no-ortogonal (Non-Orthogonal Multiple Access)
PHY	Capa física (physical layer)
SARSA	Estado-Acción-Recompensa-Estado-Acción (State-Action-Reward-State-Action)
SC-FDMA	Múltiple acceso por división en frecuencia de portadora sencilla (Single-Carrier Frequency Division Multiple Access)
SOFDMA	Múltiple acceso por división en frecuencia ortogonal escalable (Scalable Orthogonal Frequency Division Multiple Access)
SIC	Cancelación de interferencia sucesiva (Successive Interference Cancellation)
SMS	Servicio de mensajes cortos (Short Message Service)
TDMA	Múltiple acceso por división en tiempo (Time-Division Multiple Access)
UL	Enlace ascendente (Uplink)
UM	Usuario móvil
UE	Equipo de usuario (User Equipment)
UMTS	Sistema de telecomunicaciones universales móviles (Universal Mobile Telecommunication System)
URLLC	Comunicaciones de Gran fiabilidad y Baja Latencia (Ultra Reliable Low Latency Communications)
WCDMA	Múltiple acceso por división por códigos anchos (Wide Code-Division Multiple Access)
WIMAX	Acceso por microondas para la interoperabilidad a nivel mundial (Worldwide Interoperability for Microwave Access)



Carlos A. Gutiérrez recibió el título de doctor en filosofía en sistemas de comunicaciones móviles por la Universidad de Agder (UiA), Noruega, en el 2009. De 2009 a 2011 se desempeñó como profesor-investigador de tiempo completo en el área de electrónica de la Escuela de Ingeniería de la Universidad Panamericana Campus Bonaterra. En 2012 se incorporó a la Facultad de Ciencias de la Universidad Autónoma de San Luis Potosí, ocupando un puesto de profesor-investigador de telecomunicaciones. Sus líneas de investigación incluyen temas relacionados con el modelado y simulación de canal para sistemas de comunicaciones móviles y el diseño de transceptores de radio frecuencia basados en esquemas de múltiples antenas y múltiples portadoras. El Dr. Gutiérrez ha servido como Editor Asociado de la revista IEEE Vehicular Technology Magazine, Editor Académico de la revista Mobile Information Systems y como Editor Invitado de las revistas Wireless Communications and Mobile Computing, Modelling and Simulation in Engineering, Procedia Technology, y Research in Computing Science, así como miembro del comité organizador de varias conferencias internacionales.



Oscar Caicedo es profesor titular de la Universidad del Cauca (UNICAUCA), Colombia. Posee un título de doctorado en Ciencias de la Computación de la Universidad Federal do Rio Grande do Sul (UFRGS), Brasil (2015). El profesor Caicedo actualmente participa como editor asociado en IEEE Communications Magazine, editor invitado en IEEE LATAMT e Wiley IJNM, y revisor en revistas de renombre internacional como IEEE TNSM, IEEE Access, Springer JONS, Springer JISA, Wiley IJCS, Elsevier Computer Networks, y Elsevier Computer Communications. Además ha sido TPC co-chair de IEEE Latincom, miembro del TPC de importantes eventos internacionales como CloudNet, EuCNC, Latincom, y NetSoft, y servirá como TPC co-chair en GLOBECOM 2022. Sus intereses de investigación incluyen la gestión de redes y servicios, el aprendizaje automático y su aplicabilidad en redes de comunicación, redes 5G y futuras, la virtualización de funciones de red, y las redes programables.



Daniel U. Campos-Delgado es doctor en ingeniería eléctrica por la Universidad Estatal de Luisiana (LSU), EUA, 2001. Desde 2001, se incorporó a la Facultad de Ciencias de la UASLP como profesor de tiempo completo. Sus líneas de investigación se enfocan en estimación y detección, algoritmos de optimización, y procesamiento digital de señales. En estas áreas cuenta con proyectos de colaboración internacional: University of California (Santa Barbara), Texas A&M University, e Instituto de Bioingeniería Fisiología Molecular (Milán); y en el periodo Agosto/2014 a Mayo/2015 fue profesor visitante en el Departamento de Ingeniería Biomédica de Texas A&M University. El Dr. Campos Delgado es miembro de la Academia Mexicana de Ciencias (AMC), y miembro "Senior" en la IEEE. En 2001, la Facultad de Ingeniería de LSU le otorgó el "Exemplary Dissertation Award", y en 2009 y 2013 obtuvo reconocimientos como Investigador Joven por la UASLP y la AMC. A partir de mayo/2019, fue asignado como Editor Asociado en la revista IEEE LATAMT, y desde octubre/2018, revisor de Mathematical Reviews/MathSciNet. El Dr. Campos-Delgado ha sido revisor de numerosas revistas de sociedades internacionales como IEEE, IET, OSA y American Mathematical Society, y recientemente fue seleccionado como miembro del banco de expertos de la Agencia Estatal de Investigación, Ministerio de Ciencia e Innovación (España). A partir de junio/2020, se unió al comité editorial de Teoría de las Comunicaciones dentro de la revista Frontiers in Communications and Networks.