5G – Wireless Communications for 2020

Barreto, André; Faria, Bruno; Portela Lopes de Almeida, Erika; Rodriguez, Ignacio; Lauridsen, Mads; Amorim, Rafael Medeiros de; Vieira, Robson

Published in: Journal of Communications and Information Systems

DOI (link to publication from Publisher):
10.14209/jcis.2016.14

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

? Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
? You may not further distribute the material or use it for any profit-making activity or commercial gain.
? You may freely distribute the URL identifying the publication in the public portal.

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.
Abstract—Existing cellular technologies are rapidly coming to their performance limits. This is due not only to the growth in data traffic and in the number of connected terminals, but also because we are on the verge of new era, where everyone and everything will be connected, with more demanding and varied requirements that cannot be satisfied by current networks. On account of this, efforts are being made all over the world to design new wireless technologies that will support the expected demands for the next decade. These technologies, embraced under the commercial name of 5th generation, are currently being studied, and in this tutorial paper we will give an overview of the main trends that are likely to make their way in the next-generation standards.

I. INTRODUCTION

The demand for mobile data is increasing annually. According to a Cisco report [1], IP traffic will surpass one zetabyte (one trillion gigabytes) in 2016, and two zetabytes in 2019. It is estimated that in 2019, for the first time in history, the traffic in mobile networks will surpass the traffic on fixed networks. Besides the mobile data increase, the expansion and the creation of services and applications that will impose new traffic requirements to the network are also expected. One example is the internet of things (IoT), that will allow the connection and control of objects, such as smart watches, wearable devices and sensors.

In parallel, an increase on the number of connected devices is also expected. In 2014, for example, the number of smartphones increased by 439 million, and together with mobile phones and tablets, they account for 7.4 billion devices connected to mobile networks [1]. With IoT, the number of connected devices is expected to reach 50 billion by 2019 [1]. Connecting such a large number of mobile devices, while providing the needed data rates and latency requirements, is certainly a challenge to any mobile technology. An important question emerges in this scenario: will the current systems be able to meet the diverse requirements of all these services?

In order to answer this question, it is important to first understand the state-of-the-art in terms of cellular wireless technologies. In 2008, the 3rd Generation Partnership Project (3GPP) long-term evolution (LTE) Release 8 was frozen. The theoretical peak data rate provided by LTE in the downlink is 300 Mbps and, in the uplink, 75 Mbps, depending on the user equipment category [2]. The minimum one-way latency in the U-plane is in the order of 5 ms [3]. In some regions, the evolution of LTE networks, LTE-Advanced (LTE-A), is also being deployed. LTE-A is capable of providing up to 3 Gbps in the downlink and 1.5 Gbps in the uplink, using techniques such as carrier-aggregation and $8 \times 8$ multiple-input-multiple-output (MIMO) [4]. Currently, LTE is not able to provide device-to-device connectivity, but this type of communication will be included in future releases, in order to support IoT applications. In fact, LTE is being redesigned to provide the requirements of machine-to-machine communications in LTE machine-to-machine (LTE-M) and a new radio technology capable of addressing IoT requirements is under standardization within 3GPP: the narrowband IoT (NB-IoT) [5, 6]. Despite being in constant evolution, there are some limitations on the LTE technology, including: the requirement for backward compatibility, insufficient support for different types of connected devices (from smart meters to intelligent cars, for example), a relatively high latency, that impacts real-time applications, and the lack of support for a large number of simultaneous accesses, due to the limits of the random access channel procedures (RACH) in LTE [7].

In order to understand whether the current systems will be able to support emerging mobile applications, it is also important to know their requirements. Since 2012, different initiatives in multiple consortia and organizations worldwide have been taken to discuss the future mobile and wireless communications applications and scenarios. In November 2012, the METIS (Mobile and Wireless Communications Enablers for the Twenty-twenty Information Society) consortium was formed in Europe with the aim of developing a concept for the future mobile and wireless communications [8]. In 2013, this consortium identified challenging scenarios to be addressed by the 5th generation (5G) of wireless networks. The scenarios describe different applications, with distinct communication requirements in terms of latency, throughput, reliability, number of devices to be supported by the network and energy consumption. In 2015, the International Telecommunications Union (ITU) has also set requirements for International Mobile Telecommunications (IMT 2020), such as: peak data rates of 10 Gbps, experienced data rates in the order of 100 Mbps to 1 Gbps and latency in the order of 1 ms [9].

Achieving such challenging performance targets will not be a straightforward task and will require a joint research and standardization effort between industry and academia. In this paper, the focus is on presenting an overview of the ongoing standardization activities, technical challenges and proposed solutions to achieve 5G requirements. In Section II, the scenarios defined by METIS and ITU are presented,

A. Barreto, B. Faria, and R. Vieira are with Ektrum, Brasilia, Brazil, (e-mail: {andre.barreto, bruno.faria, robson.vieira}@ektrum.org.br), and were with INDT (Institute for Technology Development).
E. Almeida is with INDT, Brasilia, Brazil (e-mail: erika.almeida@indt.org.br).
R. Amorim, I. Rodriguez, E. Almeida and M. Lauridsen are with Aalborg University, Denmark, (e-mail: {rma, irl, eplda, ml}@es.aau.dk).

The Associate Editor coordinating the review of this manuscript and approving it for publication was Prof. Lisandro Lovisolo.

Digital Object Identifier (DOI): 10.14209/jcis.2016.14
as well as the ongoing 5G standardization activities. In Section III, we present new paradigms for radio network architecture; while in Section IV we discuss possible changes at the air interface. Section V discusses radio propagation and modeling considering the envisioned network architecture and current spectrum utilization and limits. Section VI presents the hardware implementation challenges of the proposed solutions and, finally, in Section VII we present the final remarks.

II. 5G SCENARIOS AND STANDARDIZATION

The first step in defining a new communication standard is to identify the performance targets and requirements to be achieved by the new technology, which are driven by the new use cases. Since 2012, different organizations around the globe have been discussing the 5G scenarios that will serve as guidance in the definition of a new cellular wireless communication standard. In Europe, in 2013, the 5G Infrastructure Public Private Partnership (5GPPP) was proposed by a group of 31 industry companies and major research and development centers, with the aim of delivering solutions, architectures, technologies and standards for the ubiquitous 5G communication infrastructures of the next decade [10]. In 2013, METIS presented the definition of the challenging 5G scenarios [8]. In 2014, 4G Americas started a work group project that provided a summary of the 5G initiatives landscape [10]. In 2015, ITU also delivered a document with its view [9]. After the definition of those scenarios, study activities were initialized in 3GPP, also in 2015 [11]. In this section, we present the scenarios defined by METIS and ITU, and the 3GPP ongoing activities.

A. METIS and ITU Requirements

METIS identified five challenging scenarios to be addressed by 5G wireless networks, regarding the requirements for new applications in industrial environment and smart cities, high-speed and high-reliability services, and a massive number of connected devices:

1) Amazingly fast: provision of very high data rates for mobile users.
2) Great service in a crowd: reliable access even in very crowded conditions.
3) Ubiquitous things communicating: efficiently handling of a multitude of connected gadgets. For example, the network should be capable of complying with the requirements of very-low rate sensors with transmission power constraints, while, at the same time, supporting smartphones running TV-On-Demand applications. In other words, the network architecture must be scalable and flexible.
4) Best experience follows you: focus on mitigating the negative effects that mobility may impose to the user experience, such as high number of handovers (increasing the chance of a failure) and a propagation channel that varies rapidly in time. The capability of handling this scenario is primal to services like smart cities and smart cars.
5) Super real-time and reliable connections: complying with applications with strict latency and reliability requirements. Such scenario may take place, for instance, in industrial and logistical services, which require real-time communication, high reliability and availability, for instance: remote control of vehicles (cars, drones, and others) and wireless automation of large industrial process plants.

After the closure of METIS project in 2015, the METIS-II project was started. The main objectives of METIS-II are developing the overall 5G radio access architecture; collaborating with the 5GPPP for a common evaluation of 5G radio access network; and preparing concerted actions towards regulatory and standardization bodies.

In September 2015, ITU has also released its vision for the future development of IMT 2020 and beyond [9]. One of the roles of ITU Radio communication Sector (ITU-R) is to ensure rational, efficient and economical use of the radio-frequency spectrum by radio communication services. To achieve this objective, ITU-R continuously observes society’s demands for mobile communications systems, so that the requirements and definitions for new communication services are aligned with these demands. Considering IMT-2020, the trends identified by ITU and that will drive the technology evolution are:

1) Very-low-latency and high-reliability human-centric communication: instantaneous connectivity experience, fundamental to cloud services and virtual reality.
2) Very-low-latency and high-reliability machine-centric communication: focus on real-time constraints of machine to machine communications, such as driverless cars, emergency and disaster response, smart grids, etc.
3) High user density: satisfactory end-user experience with a high traffic density per unit area, such as shopping malls, stadiums, open air festivals or other public events.
4) High quality at high mobility: enabling successful deployment of applications on moving user equipments, such as those located within cars or high-speed trains.
5) Enhanced multimedia services: providing high-definition multimedia in many areas beyond entertainment, such as medical treatment, safety, and security.
6) Internet of things: assuming that in the future a lot of devices will be connected to the network, with varying levels of energy consumption, transmission power, latency requirements, cost, and many other critical indices for stable connection.
7) Convergence of applications: focus on the requirements of new applications that are increasingly being delivered over IMT 2020, for example: e-Government (digital interaction between government agencies and their citizens, employees, businesses and other government agencies), public protection and disaster relief communication.
8) Ultra-accurate positioning applications: to provide the requirements of location-based service applications, such as navigation services for unmanned vehicles and drones.

Although there are differences between the scenarios
presented by ITU and METIS, there are common trends. The new communication standard will need to be able to handle a plethora of devices with different traffic requirements and hardware specifications: from sensors with low data rates and long battery lives, to smartphones with high data rate requirements. Furthermore, both organizations envision support of seamless mobility and ultra-reliability as an essential part of 5G. In this context, considering the challenging scenarios described by ITU and METIS, a technological breakthrough in existing network technologies is necessary in order to handle all described requirements. That necessity is driving major industry players in discussions for the design of new technological standards. The ongoing standardization activities in 3GPP are presented in the next subsection.

B. 3GPP and ITU Ongoing Standardization Activities

Considering IMT-2020, the first step towards a complete specification was taken in 2015, with the publication of ITU’s view, detailed in Section II-A. Also in 2015, ITU discussed 5G in its World Radiocommunication Conference 2015 (WRC-15). This event, that takes place every three to four years, debates the international rules on the usage of radio-frequency spectrum. In 2015 event, the adoption of new spectrum bands for 5G wireless networks was an important part of the discussion, and WRC-15 agenda item 10, discussed spectrum allocations below 6 GHz, that might be suitable for 5G. However, studies on frequencies above 6 GHz shall be presented on the next WRC, scheduled for 2019.

The complete ITU schedule for 5G is shown in Figure 1, where the numbering from #18 to #36 stands for the meeting number of Working Party 5D (WP 5D), responsible for the overall radio system aspects of International Mobile Telecommunications (IMT) [13]. The next steps are planned to take place in 2016 and 2017, when ITU is expected to release the technical and performance IMT-2020 requirements, and the evaluation criteria and methods. From 2017 to 2019, ITU will receive technical proposals for IMT-2020, releasing its specifications at the end of 2020.

The communication protocols themselves are, however, not defined by ITU, which only validates whether the requirements are satisfied. The organization that is leading the development of 5G wireless networks is 3GPP. 3GPP announced in March 2015 a preliminary time line, aligned with the one presented in Figure 1, in which they plan to release specifications for a new technology that satisfies the requirements of IMT-2020 by June 2019.

Before approval of a new technical specification group (TSG) on 5G RAN (radio access network) at 3GPP, a workshop took place in September 2015, with industrial and academy communities presenting their views on the next generation radio technology and on the work required toward a new RAN. More than seventy presentations were made, covering the full range of requirements that will be important to TSG RAN work items for the next five years. A workshop summary can be found in a document prepared by the RAN workshop chairman [14]. Basically, three use cases were discussed:

- Enhanced mobile broadband.
- Massive machine type communications (MTC).
- Ultra-reliable and low latency communications.

The need for new radio access technology (RAT) was also discussed. It is likely that there will be a new RAT as part of 5G and it will not be backward compatible with LTE-A. Two phases were identified for the specification work. The first phase shall address a subset of commercial needs, yet to be agreed on, to be standardized as part of 3GPP Release 15, due in 2018. The proposals will be submitted during the proposals...
for “IMT-2020” phase, shown in 1. The full 5G specifications to be submitted to IMT-2020 shall be ready in the second phase as part of Release 16, due in December 2019.

In parallel, before 5G protocol standardization, work on channel modeling for frequencies above 6 GHz has already started on RAN #69 meeting [15]. Channel modeling and 5G candidate spectrum will be discussed in Section V-B.

III. RADIO NETWORK ARCHITECTURE AND TECHNIQUES

The architecture of current wireless networks is mostly centered on the base stations (BS) or access points (AP), which are fixed nodes that provide wireless access to user equipment (UE) located inside their coverage areas. Up to now, network evolution has focused mainly on BS deployment densification, which has been to satisfy the coverage and capacity necessities of mobile users. As a consequence of this strategy, radio networks systems are nowadays composed of different types of cells. Depending on the number of users connected to the BS and the extension of the covered area, they can be classified as macro-, micro-, pico-, or, more recently, femto-cells, as illustrated in Figure 2.a. The cells can also be classified according to propagation mechanisms, as shown in Figure 2.b, and detailed in Section V. Although this densification approach was sufficient in the past, the demanding 5G requirements, e.g., increased capacity, reliability, reduced latency or different types of IoT users, represents a challenge from the network architecture perspective.

Even though there is no consensus on how exactly a 5G network will be, some trends are already observed. Some of the architectural concepts envisaged for 5G are being investigated in terms of physical network structure and in terms of signalling protocols and network management. Some of these concepts are only incremental changes over existing mechanisms while others represent a total disruption with current solutions. In this section, an overview of the main architectural techniques envisioned for 5G are described in terms of potential gains and challenges that need to be overcome in the next few years.

A. Ultra Dense Networks and Heterogeneous Networks

Heterogeneous networks (HetNets) are a well-known concept in 3G and 4G networks. Since small cells are inevitable to satisfy even higher traffic demands of 5G networks, a further densification of cells in the network is expected [16, 17]. The so-called ultra-dense networks (UDN) will bring new challenges to the network, such as mobility management, synchronization of indoor cells without a GPS access, backhauling and network operation, management and maintenance, among others, but also other advantages, such as a reduced transmit power, and possibly lower delay spreads. Besides those aspects, UDN will also bring new propagation, interference and mobility scenarios.

B. Coordinated Multipoint

Coordinated multipoint (CoMP) [18] is one of the main techniques proposed for LTE-A, already standardized by 3GPP since Release 11. It consists in a cooperation among neighboring cells, that can simultaneously transmit/receive the signal to/from a given user using the same resource blocks (resource block are the allocation unity for time and frequency in LTE [19]). This technique allows an increased throughput, particularly for users at the cell edge, as in the downlink the signal received by these users is reinforced by the transmission of different BSs, and in the uplink the signal can be jointly processed by several BSs, similar to a system with multiple receive antennas.

In 5G, CoMP is likely to remain a relevant technique, and, with smaller cells, synchronization problems due to propagation delays, which are critical for this technique, tend to become less significant. Besides that, the use of interference cancellation techniques at the UE terminals and MIMO systems should improve the efficiency of CoMP.

CoMP is also used to obtain performance gains by means of interference coordination. There are 3 basic CoMP transmission modes:

- Joint scheduling: when different nodes transmit the same information to a UE, and this combines the information received by multiple sources, improving the SINR.
• Beamforming: which is the coordination between different nodes in the phase to be applied over the transmitted signal, in order to make it more directive, preventing neighbor cells to transmit in the same direction in a given channel.
• Uplink CoMP: which is used to combine the information received from a UE in different nodes in higher layers.

There are several studies showing the gains produced by techniques similar to CoMP and use of robust mechanisms of interference suppression in the received signal quality for LTE-A [20] and 5G HetNets [21]. However, all of them demand significant increase in processing complexity at the transmitter or at the receiver, which may impose a limitation in their use for low-latency services, expected to be a part of next generation of mobile networks [22].

C. Cloud Radio Access Networks, Distributed Antenna Systems and Virtual Cells

In current cellular networks, (de)coding, (de)modulation and other signal processing tasks are performed by modems located at each BS. With cloud RAN [23], these baseband processing units would be moved to a centralized server located in the cloud, and the BS would be just a ‘dumb’ equipment with RF down/up converters, that would forward the DL(UL) signal to(from) this server. By doing this, the modem resources could be pooled, so that they could run more efficiently and make network maintenance, upgrade and expansion easier. On the other hand, the transmission of raw data from the BSs to the servers may result in higher backhaul capacity requirements.

Besides that, as CoMP already shows, the system performance could be improved if the signal is processed jointly considering antennas at distinct locations. In a cloud RAN scenario, the centralized processing units could be employed also to jointly process the signal of several neighboring BS. For example, different BS could be seen as distinct antenna elements providing diversity gain inside a distributed antenna system (DAS).

Furthermore, the association of different BS to a central server does not need to be static. BS clusters can be dynamically formed so that they will cooperate and be processed jointly. In this manner, all resource management tasks can be regarded as a single virtual cell in each cluster [24].

D. RAN Cache

Caching is a concept that is widely employed in computer networks, in which regularly accessed content is replicated and stored closer to the client, instead of being kept only in the content provider server. The concept of a information centric network (ICN) can also be brought to wireless networks [25], such that this content is mirrored in other radio network elements, such as another BS or even other UE [26]. This approach allows for a significant reduction of the backhaul necessity, which can be a bottleneck, particularly in very dense networks, besides reducing the end-to-end data latency.

One of the main challenges for this technique is how to decide which information should be stored and at which element, as there is a clear trade-off between the cost of local storage and backhaul expenses. In order to reduce the efficiency and storage costs of this technique, distributed coding techniques can be used to allow pieces of information to be stored in different places and still be recovered. Possible solutions come from regenerating codes or locally repairable codes, some of which are also based on network coding [27].

E. Device-to-Device Communications

In traditional cellular networks, the communication between different mobile terminals must also be carried through the fixed infrastructure, even if the communication is made between two terminals located at a short distance from each other. In the device-to-device (D2D) concept, also known in 3GPP as proximity services (ProSe), the communication can be made directly by a communication link between these two terminals [28]. The radio resources are the same as in the overlaying wireless network, and these resources are still managed and assigned by the network, and even though a direct link between terminals is provided, there is still traffic signalling between the terminals and the BS.

In LTE, this technique is standardized in Release 12, however, with the increase in machine-to-machine (M2M) communications predicted with IoT, one can expect that this type of direct link will become more frequent. Therefore, 5G networks should already be designed with D2D as an essential part of the system, and not just as an additional feature.

The use of D2D links, on account of the proximity, results in an increase in the transmission rate and in a reduction in the transmit power, with the consequent energy consumption saving.

The large-scale usage of D2D presents however some challenges. Despite a reduction in the transmit power, the interference scenario will become more complex, since besides the downlink and uplink interference, there will also be UE-to-UE mutual interference. To minimize the impact, interference cancellation and management techniques are needed [29]. Also, many D2D applications require the deployment of discovery algorithms, particularly taking into account mobility, high traffic variability and the use of directional links with beamforming [30]. D2D discovery is the the function that allows two UEs in the same vicinity to discover each other and create a communication link. For example a small shop may want to send advertisements to passers, so the small shop UE constantly sends its discovery signal, so the others users that are close to the shop may detect it and allow the UE to share the advertisement media.

F. Multi-Hop Communications

Even if a UE is not capable of directly connecting to a BS, the connection could still be established by direct communication links with nearby terminals. Besides allowing a power reduction and data rate increase, multi-hop communications permits the provision of services to users outside the BS coverage area. Other than that, it is an essential
technique for critical communications and emergency services [31], that must have a connection when part of or even the whole network is out of operation.

The intermediate nodes may be either fixed relay BS stations or other UE terminals. One of the main challenges is the creation of efficient routing algorithms, that allow the selection of the best route with a minimum amount of signalling overhead, taking into account the network mobility and that relay nodes may be switched on and off dynamically. Incentive mechanisms are also needed such that users share their resources among themselves, despite reducing the available bandwidth and battery life. Additionally, security and privacy issues have to be addressed.

**G. Decoupling of User and Control Planes**

In current cellular networks the user and control planes, i.e., the actual user data and the signalling, respectively, are carried in the same link. However, in order to minimize the difficulties with mobility management and signalling in ultra-dense networks, one could split them [32]. In other words, the exchange of control information can be made with a given BS, whereas the user data exchange can be made with another one. Typically, signalling should be exchanged with wide-coverage macro cells, making mobility management easier, while data transmission can be made, depending on the user conditions, through small cells, with a higher data rate and lower energy consumption, and/or with highly directive beams obtained with massive MIMO. The resource management and resource allocation would be in charge of the larger cells, and the small cells would be basically high-capacity wireless modems. By doing this, even if the connection to the small cell is lost or the beam interrupted, the UE would still be connected to its management entity, and connection would not be lost. The smaller cells, which do not contain management and signalling mechanisms, are also known as phantom cells [32].

**H. User Equipment Energy Efficiency**

Energy efficiency (EE) will be a key performance indicator for 5G and distinguish it from the current cellular technologies. The improved EE is demanded by both mobile network operators, who seek to reduce their operating expense, and by the users, who currently experience short smartphone battery life. In addition, new IoT/MTC use cases expect sensor battery life approaching 10 years. However, as discussed in [33] and section VI, 5G will also pose new challenges to the user equipment in terms of processing and implementation complexity, physical area, and power consumption. The reason is that the UE has to support up to 10 Gb/s, several hundred MHz of bandwidth, at least 4x4 MIMO, and extensive use of carrier aggregation. Fortunately, the increased data rate, achieved using the higher bandwidth, MIMO and other techniques as discussed previously, may also help reduce the ON time of the UE, because it can finish the data transfer faster. Furthermore, a short and optimized 5G frame structure, which may be combined with more efficient waveforms will also reduce the ON time, which is key to improve the overall EE [34]. Combining the short ON time with fast synchronization and efficient discontinuous reception sleep mode has been shown to improve the battery life 5-15 times as compared with LTE, depending on the activity pattern of the UE [35].

**I. Cellular Sleep Mode**

Energy saving must be considered, not only at user terminals, but also in the network as a whole. The discontinuous reception (DRX) and discontinuous transmission (DTX) concepts are already widely employed in the handsets, that can switch off their modems temporarily when not in use. However, most of the energy in a cellular network is consumed by its base stations, regardless of the number of active connections, so that the best solution to reduce energy consumption is to switch off idle base stations [36]. This is especially relevant in situations where the number of UE terminals connected to a BS is small, making more likely that the BS become idle at any given time.

The uncoupling between the control and user planes makes the implementation of BS sleep mode easier, as just the phantom cells can be switched off, whereas the signalling keeps being carried through the larger cells. This means that a BS may be switched on and off at a very short time scale, and the usage of sleep mode remains transparent to the end user.

**J. Uplink and Downlink Decoupling**

Because of the asymmetry between the modems in a BS and in a UE, in addition to decoupling the control and user planes, downlink (DL) and uplink (UL) could also be decoupled [37]. In this context, both links are basically treated as separate network entities and a UE can connect to different serving nodes in the UL and DL. As a consequence of the better sensitivity and also higher transmit power at the BS, the best server for the downlink may be a macro cell, while for the uplink it may be a small cell, particularly for users at the cell edge. This decoupling can also take into account traffic loads in both link directions.

**K. Vehicular Communications**

Along with new trends in the automotive industry towards autonomous driving, vehicular communications are likely to become one of the main applications in 5G networks. Vehicle-to-vehicle (V2V), and vehicle-to-infrastructure (V2I), i.e., roadside infrastructure, are envisioned, but other forms of communications are also possible, as between vehicles and pedestrians. These different links may have different requirements and characteristics, but they all can be embraced under what is called V2X communications. This is a potential application for D2D communications [38], for instance in that different vehicles exchange information to coordinate traffic flow or to avoid accidents.

However, vehicular communications represent novel challenges for wireless networks [39]. On one hand, as critical data for traffic safety is handled, stringent requirements in
terms of latency and reliability are needed. On the other hand, the high mobility of vehicles brings difficulties for the physical layer, on account of the Doppler effect and adaptive beamforming, and for the medium access control (MAC) layer, in terms of dynamic resource allocation, mobility management and routing.

L. Moving Networks

The combination of multi-hop and vehicular communications concepts bring the so-called moving networks [40]. For example, one can have a small cell inside a moving train, bus or even airplane, connected through a wireless backhaul to a fixed BS, i.e., the small moving cell may be seen as a moving relay node.

This approach has some inherent advantages. The short distances in the small cell allow higher data rates and a smaller battery consumption. Mobility management is also made easier, since the local small cell may exchange the handover signalling with the network, instead of each UE by itself, reducing the control overhead.

Nevertheless, considering mobility, providing a reliable high-rate wireless backhaul link from the moving cell to the fixed network is a big challenge.

M. Wireless Backhaul

As a consequence of network densification and the emergence of moving cells, the deployment of optical fiber links between all BSs and the core networks is out of the question. With this in mind, the deployment of wireless backhaul for a great number of cells will be needed [41]. Preferably, highly directional links should be employed, such as free-space optical (FSO) or millimeter-wave links. Possibly, these backhaul links may also occupy part of the spectrum used also by the access networks, requiring intelligent radio resource allocation schemes. The latency and reliability of these links, which are prone to blocking and fading, are also important issues to be considered.

N. Machine-Type Communications

One of the main foreseen applications of 5G networks is the communication among machines, known as machine-type communications (MTC) [42], which, as stated in Section I, is an essential enabler for the IoT. This kind of application shall influence significantly the design of the new network architecture. In particular, current networks are based on a model in that each BS controls just a few UE terminals inside its coverage area, which are constantly exchanging control information with the network, e.g., location updates, connection setups and periodic registrations. For MTC applications, however, it is likely that a much larger amount of wireless terminals will be connected to each cell, that, for energy-saving reasons, will only communicate sporadically. This new setup implies the necessity of a new approach in the wireless network signalling protocol, as this shall support a large number of devices in an energy-efficient manner and occupying the minimum possible bandwidth.

Furthermore, the emergence of MTC has also some impacts on the physical layer, as exchanged messages are likely to be short. Current synchronization requirements, which demand time and resources, and, consequently, represent a large overhead, will need to be relaxed. Short packets, with possibly very low error rate requirements, also impose some challenges to channel coding. These issues will be discussed in Section IV.

O. Context Awareness and Self Organizing Networks

Context awareness can help an efficient radio resource management and network configuration, and it is being underexploited in existing networks. Context awareness [43] is much more than just location information, even though this is also useful. A trove of data will be available, such as user trajectory, traffic patterns, type and capability of terminals, coverage information and used applications. It is a difficult multi-dimensional problem, but recent advances in data mining and machine learning will make the use of this data feasible in the near future.

Besides that, 5G networks are expected to be characterized by a large heterogeneity in services and transmission technologies and by network topologies with a large number of nodes. This makes planning and configuration impossible to human operators only. The concept of self-organizing networks (SON) is already beginning to be applied in LTE networks, for instance with automatic neighbour relations (ANR) and other features, but in 5G SON is going to be unavoidable and it will be extended to the whole node configuration. Many challenges are yet to be solved, such as development of efficient algorithms that work in short and long time scales, i.e., that respond rapidly to punctual changes in the network but also deal with long-term trends, like traffic volume; focus in energy efficiency; and not only reactive but also proactive algorithms. Besides that, proper key performance indicators (KPI) must be defined, representing end-to-end performance of a great variety of user needs, such as energy consumption, latency, reliability or user throughput. KPIs definition may vary depending on the different applications requested by network users. Again, advances in big data may help us overcome these challenges [44].

P. Coexistence

It is envisioned that 5G will not be a single technology, but a set of different technologies, adapted to the distinct user needs, that can be seamlessly integrated. In particular, nowadays there are two main technology families in wireless communications, IEEE 802.11 and 3GPP, and it is envisaged that in the near future they will have to coexist, especially in unlicensed frequency bands. In LTE standardization there are already different approaches to this issue, LTE-U (unlicensed) and LTE-LAA (Licensed Assisted Access) [45], and the main challenge here is to guarantee that none of the different technologies will suffer disproportionately when they have to coexist. With the emergence of 5G, the rapid increase of the number of multi-technology terminals, and the evolution of
both IEEE 802.11 and 3GPP, coexistence will be a key feature to make the best use of available unlicensed spectrum.

Q. Other Topics

This section has addressed some of the main techniques, reported in the literature, that may help to achieve the 5G technology goals from a radio architecture perspective. For the sake of simplicity, some other important issues such as privacy, security, authentication [46] and core network architecture [47], among others, have not been addressed in this overview.

IV. AIR INTERFACE SOLUTIONS

As mentioned in the introduction, future 5G networks present challenging goals, such as an increased user capacity, higher throughput per user, high reliability, very low latency, and, in particular, the large-scale use of machine-type communications. The architectural changes summarized in Section III may help reaching these goals, but it is also very likely that some substantial improvements will be needed in the lower layers of the air interface, namely physical (PHY) and MAC layers [48]. Achieving the high throughput expected for 5G is a challenge that will require also enhancing the overall system data rate, by means of enhancing the spectral efficiency and increasing the system bandwidth.

However, there is a physical limitation in the spectral efficiency. The Shannon-Hartley theorem [49] states that the maximum amount of information that can be sent through a communication channel depends on the signal-to-noise ratio (SNR). Several transmission techniques can be used either to improve signal quality, therefore increasing system capacity, or to get us closer to Shannon capacity. This section will detail some of the solutions, considered as candidates to take part in 5G.

A. New waveforms

Orthogonal frequency-division multiplexing (OFDM) is currently the preferred waveform for wideband wireless communication systems, being used in Wi-Fi [50], digital TV [51], WiMAX [52], LTE and LTE-A [53], among others. This is a consequence of some OFDM characteristics, particularly a relatively low-complexity equalization, which is independent of the number of multipaths, and, when in multiple access (OFDMA), the scheduler flexibility. However, a new waveform may be needed in order to meet some of 5G application requirements [52].

Applications in sensors networks or in MTC present a different scenario, where some elements may be dormant for a long period of time and wake up to transmit small pieces of information. Moreover, many of these devices may be indoors. In these cases, there will be a degradation in the synchronization between these devices and the network, and the OFDM subcarriers will be likely to lose the orthogonality in the reception. Losing the subcarriers orthogonality can cause an increase in the number of lost packets and retransmissions, or in latency, if a resynchronization is started. Therefore, other waveforms, that are more robust to the loss of synchronization can be a preferable alternative to OFDM for this kind of scenario.

Another disadvantage of OFDM is power leakage outside the useful bandwidth, which makes a guard band between neighboring channels necessary. In LTE, for instance, 10% of the bandwidth is reserved for null subcarriers (guard band), e.g., in a 20 MHz channel, only 18 MHz effectively transmit information [53]. Also another challenge for OFDM implementation is the high Peak-to-Average Power Ratio (PAPR) that can degrade the transmitted signal, due to the non-linearity of power amplifiers. To overcome this problem it is necessary the usage of computationally intensive techniques or deploying power amplifiers with large linear dynamic range, but both solutions may be prohibitive for small sensors that are energy and cost constrained [54].

With these aspects in mind, some waveforms have been proposed to allow a more efficient spectrum usage, while addressing the weaknesses presented by OFDM:

- FBMC (Filter Bank Multicarrier) [55], defines a waveform where every symbol transmitted in each subcarrier passes through a specifically designed filter. OFDM can be seen as an FBMC system, where the selected filter is the rectangular window in the time domain. The rectangular window results in a sinc function in the frequency domain, a waveform with high-power side lobes, but with orthogonal subcarriers. In FBMC other filter prototypes may be selected, aiming at a mitigation of the power leakage, sometimes at the cost of loss in orthogonality.

- UFMC (Universal Filtered Multicarrier) [56], is an extension of FBMC, and the main difference is that the filtering is applied over a block of subcarriers.

- GFDM (Generalized Frequency Division Multiplexing) [57], has an approach similar to OFDM, but based in transmission blocks also in the time domain, with a cyclic prefix added for each time-domain block, reducing the overhead, and with a tail-biting digital filter to reduce cyclic prefix duration [57].

- ZT-DS-OFDMA (Zero-Tail DFT-Spread OFDMA) [58], is a waveform that adds a low-power suffix and prefix, and, differently from OFDM, these are added inside the symbol duration, which makes it possible to have different guard intervals for different users, without the need of an additional cyclic prefix after the FFT. Furthermore, the use of both a prefix and a suffix leads to a lower power leakage.

Due to the higher complexity and loss of orthogonality related to some of these proposed waveforms, it is possible that a mixed approach is adopted in future systems. OFDM may be used in some parts of the spectrum, for a given set of applications and other waveforms, with lower power leakage and more robust to synchronization errors, may be used in narrow areas of the available spectrum.

B. New Frame Structures

The current LTE frame is divided in 10 subframes of 1 ms each, which imposes a lower bound to the system latency,
since it is the lowest time unit considered by the scheduler. In the same manner, the contention based protocol used in Wi-Fi is also a limiting factor for the development of very low latency applications, as in dense networks, the user spends more time listening to the channel than being able to access it [59]. 5G technologies are expected to present a frame structure that allows a reduced latency when compared to the currently available systems.

A limitation is observed in LTE-A when time division duplexing (TDD) system is used, which is likely to be the preferred mode in dense networks [16]. Currently, the system selects one of 9 predefined sequences of uplink/downlink assignment for the 10 subframes, which can only be changed at relatively long intervals. As system load and types of services being used may vary significantly in different times of the same day, a more flexible allocation of subframes is also expected for the new technologies to be proposed. This would allow each cell to adapt to its users demands. The Flexible TDD technique has been studied as a solution to this problem [60], allowing the cell to decide instantly, which subframes should be assigned to downlink or uplink. Different frame proposals have been made, for instance in [61]–[63], but most rely on very short frames, of less than 0.5 ms, that can be assigned either to uplink or to downlink. Differently from LTE, the signalling and data fields in these frame proposals are likely to be separated in time domain, allowing the receiver to switch off when no data in addressed to it, saving energy. However, a different approach was presented in [64], where the authors propose an in-resource control signalling. The advantages of this approach are that, besides being a flexible and scalable solution, it also allows beamforming for control and data transmission and efficient time-frequency domain ICIC (inter-cell interference coordination).

Another solution being proposed is the full duplex transmission [65], where there is no separation in time or frequency between downlink and uplink, and all subframes may be used in both directions at the same time, and the self interference (SI) has to be canceled by a combination of analog and digital processing. However, SI suppression and cancellation are complex tasks, based in costly hardware design and digital signal processing [66].

Full duplex and Flexible-TDD have the potential to increase system flexibility, therefore enhancing how the system deals with changes in the load conditions [67, 68]. However, the two techniques present significant implementation challenges, because in both cases the additional interference created becomes a major factor to be considered. In full duplex, the self interference is still an open issue that may harm the received signal, while in Flexible TDD, simultaneous downlink and uplink frame assignments for neighboring cells may cause users at the cell edge to suffer strong interference.

C. Interference Mitigation Techniques

To improve the signal-to-interference-plus-noise ratio (SINR), it is necessary to either strengthen the signal or reduce noise and interference power. Increasing the desired signal power is costly and is not always possible due to limitations in the equipment. Moreover, increasing power for a given user, means that the others will suffer from increased interference, which would require them to increase their power too. Hence, in a heavily-loaded scenario there is no gain to be obtained this way.

Currently deployed 4G networks have efficient techniques to deal with the interference, such as hybrid automatic repeat request (HARQ), adaptive modulation and coding scheme (MCS), advanced receivers and network-assisted interference cancellation and suppression (NAICS) [69]. However, as the cells are expected to become much more dense, these techniques alone are not capable to cope with all the interference generated in the network.

Besides the densification itself, some techniques proposed for future generations increase the interference problem, such as Flexible-TDD and full duplex, as discussed in Subsection IV-B. Co-channel interference also presents a problem for in-band Hetnet deployments, which is a deployment where macro and small cells coexist in a geographical area using the same set of frequencies. In this case, the macro cell power tends to degrade the signal quality of the small cell.

Some coordination schemes between network nodes have been proposed in the 3GPP standards and in the literature, aiming at interference mitigation. One of these proposals consists in preventing, by means of coordination, some cells to transmit in a given time window or refraining allocation for users at the cell edge, targeting better interference conditions among neighbor cells. These schemes are an evolution of enhanced inter-cell interference coordination (e-ICIC) presented by 3GPP in LTE Rel. 10 [70]. The side effect is a decrease in the overall system capacity. As presented in Section III-B, CoMP is another technique that uses interference information to obtain transmission performance gains.

5G networks, besides employing some of the techniques already available for LTE, will be likely to rely on interference cancellation techniques, both linear, such as interference rejection combining (IRC) or non linear, such as parallel or serial interference cancellation (PIC/SIC) [71, 72]. Although these techniques are essential to deal with the increased interference in a Flexible-TDD scenario, they rely heavily on the knowledge of the noise-plus-interference covariance matrix. The need to have this knowledge has an impact on the frame design, which must provide a reference signal with good cross-correlation properties and must be synchronized with all transmitters.

In full-duplex systems, the major challenge is to deal with the self-interference created by the transmit circuitry in the receiver end of the same equipment. It is still not clear if full duplex will be commercially available for the next generation, because it requires complex filtering and interference cancellation, which tends to increase the equipment cost. Moreover, the required processing increase and its impact in energy consumption may be prohibitive by 5G requirements.
D. Massive MIMO

Multiple antennas have been largely used in 4G Networks. MIMO systems consist in the adoption of multiple antennas in both receiver and transmitter ends, aiming at improvements in spectral efficiency and robustness. The gain obtained by MIMO systems comes from beamforming, spatial diversity, spatial multiplexing or any combination of them [73].

In spatial diversity, different versions of the same signal are combined at the receiver, improving signal robustness. If the information is partially degraded due to fading in one of the versions, it can still be recovered with the other versions. A necessary condition is that each received version has observed a channel propagation with low correlation. The diversity created by the several versions make the signal more robust to channel impairments and higher modulation and coding rates may be used in the transmission. In a link where the transmitter operates with $M$ antennas and the receiver operates with $N$ antennas, the maximum diversity gain equals $MN$, and represents the number of independent fading coefficients that the received symbol can be averaged over in the detection procedure. The maximum gain is achievable if all the $MN$ coefficients are independent.

Spatial multiplexing uses the different channels created by each transmitter/receiver antenna pair to transmit different signals in the air interface. In other words, spatial multiplexing reuses the spectrum frequencies creating multiple communication channels, increasing system capacity. Each of these different signals, must be transmitted in an independent channel to enable the usage of spatial multiplexing. Otherwise, it will not be possible to separate the signals at receiver end. The gain in capacity produced by the independent data streams, i.e. the spatial multiplexing gain, depends on the signal to noise ratio in the receiver side and it is limited to $\min(M, N)$ for high SINR. In other words, MIMO techniques may produce up to $\min(M, N)$ independent data streams. The tradeoff between spatial multiplexing and spatial diversity is discussed in [73].

Channel independence is a requirement that imposes a restriction in MIMO adoption. The signals must be uncorrelated, which requires abundant multipath environments and antennas separation. The minimum distance allowed to guarantee antenna decorrelation is a function of the wavelength and it limits the number of antennas that may be deployed in current miniaturized mobile equipments. Currently, LTE-A standards envision up to 8 antennas in downlink transmission and 4 antennas in uplink.

For 5G networks, a massive number of antennas (>>8) is considered as a real option to boost system capacity. Millimeter waves spectrum usage is considered as an enabler for massive MIMO, shortening the minimum distance required between deployed antennas. The adoption of these new spectrum bands will be discussed in Section V.

Massive MIMO has a potential to increase up to 10 times system capacity [74]. However, there are still open challenges in the implementation of such high number of antennas. One important challenge is the hardware implementation of this high number of transmission/reception chains, due to the increased computational complexity and to the amplification of impairment effects in hardware, as it will be discussed in Subsection VI-C.

Another significant challenge is the pilot pollution, which refers to the usage of pilot signals by receivers to perform channel estimation. In order to be able to separate different signal versions, receivers need to estimate the channel for each transmitter, which requires a high number of pilots transmitted. New allocation schemes should be developed preventing that the additional pilot overhead surpasses the gains obtained by the adoption of new antennas.

E. Coding

The error-correcting codes applied in WiFi and LTE, namely low-density parity-check (LDPC) and turbo-codes are known to have very good performances, approaching Shannon's capacity limits [75, 76]. There are, nevertheless, still some improvements to be made in the coding strategies, which will be discussed briefly in this subsection.

Both LDPC and turbo-codes perform well with long codewords. This may not be a problem with wireless broadband, but machine-type communications will consist mostly of short packets with low latency requirements. Furthermore, LDPC and turbo codes are known to have an error floor [77, 78], which makes it difficult to achieve very low error rates without retransmissions, which is a key requirement in many 5G applications. This means that new efficient codes for short packets need to be devised, and even though performance bounds can be obtained [79], exact performance limits for finite-length codes are still unknown.

Another important issue is decoder complexity, particularly for mobile devices. Current iterative decoding algorithms are very demanding in terms of signal processing, and more efficient decoding algorithms and/or codes must be devised for this kind of usage. Some implementation issues will be addressed in Section VI.

A very promising class of codes for 5G is the so-called polar codes [80], which achieve capacity, at the cost of knowing the channel statistics a priori. These codes also have a relatively low decoder complexity, but their application in a scenario with time-varying channels and imperfect channel knowledge remains to be studied in detail.

If instead of considering coding over a single point-to-point link, one thinks about the whole network, there is also possible room for improvement. The data capacity of a multipoint-to-multipoint network, as the wireless network, is still an open issue, but promising performance gains can be obtained with network codes [81], which can be applied in upper layers on the digital signal, but also on the physical layer. There are, however, still some important hurdles for their application, such as, complexity, the need for synchronization and a novel radio resource management.

V. RADIO PROPAGATION CONSIDERATIONS

As introduced in Subsection III-A, current mobile networks are evolving towards ultra-dense heterogeneous deployments, where layers of (outdoor or indoor) small cells are deployed.
on the top of the existing macro base station layout in order to cover small areas, where there is a lack of coverage, or crowded areas, where additional network capacity is needed (hotspots) [82]. In macro-only scenarios, large area coverage was provided mainly from elevated positions. The use of low frequency bands (below 3 GHz), at which propagation loss is small, made it possible to provide both outdoor-to-outdoor and outdoor-to-indoor coverage. However, this will change with the roll-out of small cells, located closer to the end user, and the use of new spectrum (frequency bands between 3 and 6 GHz). Outdoor micro base stations are typically deployed with antennas below rooftops (e.g. lamp post close to street level) in order to provide outdoor-to-outdoor coverage, while femto-cells are deployed in indoor locations, similar to WiFi access points, aiming at providing indoor-only coverage.

When considering urban HetNet scenarios, radio propagation conditions are very diverse, depending on the type of base station. Differently from the classification presented in Subsection III-A, distinct types of cells can be also defined from a radio propagation perspective. This is illustrated in Figure 2.b. In macro cells, the radio signal propagates above rooftops and gets diffraeted from rooftop to street level [83, 84]. Differently, in outdoor small cells, the radio signal is guided along street canyons, due to reflection and scattering on the buildings, and diffraction on the corners [85]. For indoor small cells, propagation is more complicated due to the multiple reflections on the walls, propagation through floors and walls and corridor guiding effects [86].

By the massive deployment of small cells, mobile operators are slowly migrating their coverage and capacity provision strategies from outdoor-to-indoor (typical in the sparse macro deployments) to outdoor-only or indoor-only. However, these strategies could be different from country to country due to penetration loss particularities of different building materials. Some modern materials, such as metal coating layers for thermal energy isolation, used mainly in cold countries, have been proven to be a huge obstacle for radio waves when penetrating into buildings [87] and represent a problem for operators in outdoor-to-indoor coverage. Anyhow, this increased attenuation can also be considered as an advantage for the operator for deploying co-channel outdoor and indoor small cells, as walls may act as a natural barrier to interference. In warmer countries, this may represent a problem at some point in the future, with energy regulations aiming to achieve zero-energy consumption inside buildings.

As a consequence of the ongoing evolution towards HetNets, which combine different types of outdoor, outdoor-to-indoor and indoor deployment strategies in terms of cells and spectrum; a number of new propagation scenarios arise. In terms of spectrum, the scarcity in the traditional frequency bands (below 6 GHz) will strengthen the use of new frequencies; while in terms of 5G technology, new techniques such as massive MIMO with large antenna arrays, vehicular or M2M communications will start to be developed. The combination of these new applications, new spectrum and deployment techniques will lead to an enormous set of unexplored scenarios [8]. All the different possibilities should be explored soon, in order to predict the behavior and performance of the future radio systems, and, therefore, radio channel measurement, analysis and modeling will be an essential part of the future 5G development.

### A. Spectrum

Looking at spectrum utilization, traditional cellular systems are typically deployed in the lower frequency bands, below 6 GHz, and operate on relatively narrow bands (specially 2G and 3G). The situation changed a bit with the scalable approach of 4G LTE and LTE-A, that allows system bandwidths of 1.4-20 MHz in single band and up to 100 MHz with carrier aggregation techniques.

To illustrate how the situation looks nowadays, Table I presents an overview of the spectrum owned by a European operator (by the end of 2014, and only considering access network) [88].

<table>
<thead>
<tr>
<th>Technology</th>
<th>Frequency Allocations</th>
<th>Total Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td>2G</td>
<td>(2x9 MHz) @ 900 MHz</td>
<td>40 MHz</td>
</tr>
<tr>
<td></td>
<td>(2x11.8 MHz) @ 1800 MHz</td>
<td></td>
</tr>
<tr>
<td>3G</td>
<td>(115 MHz) @ 1900 MHz</td>
<td>35 MHz</td>
</tr>
<tr>
<td></td>
<td>(2x1 MHz) @ 2100 MHz</td>
<td></td>
</tr>
<tr>
<td>4G</td>
<td>(2x20 MHz) @ 800 MHz</td>
<td>100 MHz</td>
</tr>
<tr>
<td></td>
<td>(2x10 MHz) @ 1800 MHz</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2x20 MHz) @ 2600 MHz</td>
<td></td>
</tr>
</tbody>
</table>

As it can be deduced from the table, this operator has a total amount of spectrum of approximately 175 MHz. The problem is that this bandwidth is not contiguous, therefore, it cannot be easily combined into large amounts of effective system bandwidth. In order to satisfy the 5G requirements, a larger amount of spectrum is needed. Since re-farming is not an easy solution, as users with devices compliant only with legacy systems would require those to still be available, 5G systems will look at higher frequencies for new large contiguous bandwidth opportunities [89].

As depicted in Figure 3, 5G systems are expected to exploit not only the traditional spectrum below 6 GHz, but also higher bands such as the cm-wave (in this paper, defined between 6 and 30 GHz) and mm-wave (30-300 GHz) frequency bands [16, 90, 91]. There are approximately 2.5 GHz of cm-wave spectrum available between 6 and 28 GHz, while in mm-wave bands, non-occupied spectrum opportunities are even larger: 4 GHz around 38 GHz, 10 GHz between 70 GHz and 86 GHz and approximately 3 GHz around 90 GHz.

### B. Channel Modeling

There is a huge interest from both industry and academia in accurately characterizing radio propagation in frequency bands above 6 GHz, with focus on the future 5G scenarios. The first attempt was the recently finalized METIS project [8], which acted as an enabler for other activities such as COST IC1004 [92], ETSI mmWave ISG [93] and NIST 5G Millimeter Wave Channel Model Alliance [94], among others. Apart from these
ongoing activities, it is expected that interested groups are likely to collaborate in the future on developing a unified channel model for 3GPP [95]. The initial 5G channel modeling requirements were set by [8], and most of the posterior activities have adhered to them. They include:

- Spatial and temporal consistency: typical channel models do not provide an accurate representation of LOS and NLOS transitions, which translates into discontinuities when simulating, for example, non-stationary scenarios with fast mobility.
- Wideband characterization: as explained in the previous subsection, current cellular wireless systems operate over a limited bandwidth of at most 100 MHz (3×20 + 1×40 MHz) [96]. 5G aims to operate over larger bandwidths (maybe 200 MHz, 300 MHz, 500 MHz... a few GHz), and, therefore, it is necessary to understand and model the potential channel variations across the total bandwidth.
- 3D extension (for both outdoor and indoor scenarios): elevation was typically not considered by existing models, or otherwise, elevation correction factors were not very accurate. This lack of information starts to matter when considering ultra-dense deployments with a massive number of small cells where accurate models are needed in order to compute interference correctly. MIMO techniques, particularly 3D beamforming, where 3D geometry is one of the key factors, represent, of course, another relevant need for an accurate 3D modeling.
- Need for different scenarios: as introduced at the beginning of the section, the range of different propagation scenarios is enormous. For instance, for V2X and M2M scenarios it may include high speed users, with shadowing from other users and potentially crossing tunnels.
- Applicability in all frequency bands, from 2 GHz up to 60 GHz: apart from many different scenarios, the model should be consistent and applicable within a wide range of frequencies, including cm-wave and mm-wave bands.
- Massive MIMO with large antenna arrays should be considered: spherical (not-plane) wavefronts and axis non-stationary cluster evolutions over large antenna arrays must be accurately modeled, in order to be able to quantify the realistic gains from massive MIMO in real-world scenarios. Due to complexity, but also reality, this will probably be considered only for mm-wave frequencies.

5G channel modeling will need to be more complete, presenting new challenges. One of them is to perform measurements and simulations to fully validate the developed models, and, therefore, in order to evaluate them, there should exist a trade-off between accuracy, complexity and flexibility. To cope with this issue, the 5G channel modeling considers different approaches. A stochastic approach, that relies on a geometry-based statistical channel model further developed from for example WINNER/3GPP; a deterministic approach, based on ray-tracing models applied using a simplified 3D geometric description of the propagation environment; and a hybrid approach, that combines the stochastic model with extra features from the map-based model.

As an illustration on how this multi-option modeling approach will be useful for future 5G system simulations, consider a micro cell scenario: it will be possible to simulate the mean path loss with a statistical model and, then, apply the 3D information in order to compute consistent shadow fading, blockage loss and LOS/NLOS transitions. Furthermore, path delay and phase evolution could be also extracted from advanced 3D simulations, and will allow to analyze different MIMO antenna array schemes.

Focusing now on measurements for modeling development/validation, Figure 4 shows some of the different approaches and challenges. The first approach is based on extensive drive-test measurements covering large distances, and resulting in empirical statistical models. This a fair-enough approach for the “old” sparse networks operating...
in low frequency bands. The second approach is oriented to geometrical models, which, if adequately calibrated, can be more accurate than the statistical models. It requires more measurement planning and is more time consuming, since it is necessary to consider the different geometrical parameters (distances, heights,...). This approach is, as an example, extensively used in the simulation of Manhattan-grid micro cell scenarios [97]. The third approach puts more emphasis on modeling the different propagation mechanisms (reflection, scattering, diffraction, transmission) independently, instead of on the combined end-to-end path loss itself, and is therefore more recommendable/suitable for calibration of ray-tracing tools, for example.

Measurements at higher frequencies are inherently more accurate (due to the imposed narrow-beam antenna patterns that increase the angular resolution), but also more time consuming, since it is necessary to put some effort in performing alignment procedures and scanning the correct directions of arrival [98]. Due to technological limitations, measurement equipment for higher frequencies is still quite expensive. Especially at higher frequencies, at which the large bandwidths and massive MIMO will be used, not only path-loss oriented measurements, but also a lot of channel sounding is still required in order to characterize the channel impulse response, and this is an accurate but very time-consuming controlled measurement.

VI. IMPLEMENTATION CHALLENGES

At the present time, it is unknown what technology solutions will be implemented in 5G networks. However, some techniques such as Massive MIMO and beamforming, together with new waveforms and the use of higher frequency bands will be the key enablers for even higher data rates. Those techniques require fast algorithm processing, very selective filters, precise clock generation and better printed circuit board design. Design of intelligent antennas is also a key area for 5G development [99, 100]. This section will describe some of the technological challenges for realization of some 5G techniques in hardware.

A. Digital Signal Processing

Advanced modulation schemes, channel coding and signalling rely on complex numeric algorithms and require a very high computing power. The digital signal processors responsible for implementing these algorithms should be of very high performance, consume little power and should occupy a small area on the printed circuit board for machine type communication devices for example. The silicon industry has invested in higher number of processing cores in the same chip and the use of parallelism to increase processing capacity. It is now common to see mobile devices equipped with 8 processor cores. The density of transistors on these processors increase at a rate that reaches the predictions of Moore’s Law [101]. However, the estimates presented in [33] show that the baseband processing complexity may be so high that the power consumption will exceed 3 W in 2020. The reason is the turbo decoding complexity and the use of wide bandwidths, resulting in complex FFT, channel estimation, and equalization. However, due to the slow CMOS technology node evolution the mobile terminal power consumption is first expected to be similar to todays LTE devices in 2027 [33].

In the radio base stations industry the trend is to use reconfigurable platforms, like field-programmable gate arrays (FPGA), due to their versatility, rapid introduction to the market of new technologies and incorporation of high-speed buses. On the other hand, some manufacturers are betting on the versatility of software defined radio (SDR) systems running on massively parallel platforms. These platforms consist of a processor containing up to hundreds of digital signal processing (DSP) elements, connected in a
topology similar to a mesh (Network on Chip) [102, 103]. These platforms enable the application of parallel processing techniques to increase the computing capacity with low power consumption and are a bet for applications like ultra-dense networks.

B. Transceiver Design

As explained in Section V, the scarcity of spectrum in traditional cellular frequency bands will make it necessary to look for higher frequency spectrum in order to find larger amounts of contiguous bandwidth. The use of bandwidths up to 3 GHz is expected for the high data rate transmission of data in the backhaul, or even to end users on the downlink. However, the use of such wide bands imposes some technological challenges [104]. One such challenge is the feasibility of direct sampling, where samples are directly converted to carrier frequency by the digital-to-analog converter (DAC) for RF transmission, and vice versa. This technique reduces the complexity and cost of the transceiver since there is no need to use multiple filtering and mixing stages, and enables the use of advanced techniques such as carrier aggregation, using bands that are located in different places in the spectrum. For direct sampling to work, the digital-to-analog converters must operate at extremely high sampling rates. Today, it is possible to find analog-to-digital converters (ADC) on the market able to sample close to 5 GHz [105], but the resolution and power consumption are still challenges for its use in mobile devices, for example.

The ADC challenge was also examined in [33] where the ADC evolution was evaluated based on literature studies covering more than 20 years. The average annual improvement was noted to be about 34% which results in a 100 MHz ADC with 12 effective bits consuming less than 10 mW in 2020. The main challenge is thus that the combination of carrier aggregation and MIMO requires up to 32 ADCs to accommodate 400 MHz bandwidth with 4x4 MIMO. This results in high power consumption and a very challenging RF transceiver design.

C. Implementation of Massive MIMO and Beamforming

Massive MIMO is one of the key techniques that enables significant gains in data rate and increased robustness in the radio link, as seen in Subsection IV-D. With the use of multiple antennas (it’s envisioned that more than 100 will be used for mmWaves) [8], it is possible to radiate energy in a more directional manner towards the user, reducing system interference and increasing capacity [106].

But with such a high number of antennas, various technical challenges arise to incorporate Massive MIMO into future products. For example, multiple RF transceivers are used in the BS and they must be perfectly synchronized. With extremely wide bands and high carrier frequencies considered for 5G, the clocks for the transceiver systems must have high precision and low jitter. Transceiver design is by itself complicated. The filtering elements that meet the tight transmission mask requirements are very expensive and hard to implement in reduced circuit space. This makes the design more complicated and more expensive.

The BSs usually have a system module where the radio control system and baseband algorithms run. With Massive MIMO the system module should communicate with numerous RF modules [107]. This communication requires extremely fast and reliable buses for the system to operate properly. With the expected rates of 5G, new techniques and protocols used on the bus must be considered to grant the modularity of the hardware and its resistance to failure.

For the MIMO technique to work properly the radio-base station should have indicators of channel state for each antenna, which increases the required channel feedback. The processing of this information and downlink precoding need an enormous processing power to achieve the real-time requirements of the system. As discussed earlier, the industry has evolved greatly in the development of systems for digital signal processing, and also some suboptimal algorithms are being proposed for implementation of this new technology in the currently available hardware. It is also possible that future massive MIMO systems will consist of a combination of analog components, like phase shifters, with digital signal processing.

D. Other Challenges

Besides the previous discussion, some other challenges for implementation of potential 5G techniques can be cited [100, 104, 108]:

- Power amplifier non-linearities for ultra wide bands (>5 GHz).
- For high frequency applications, the oscillators frequency deviations are more apparent, making the RF imperfections more notable, especially in synchronous systems such as OFDM.
- Fitting a high number of antennas in hand-held devices is challenging.
- Printed circuit boards (PCB) with high frequency buses are difficult to project. Phenomena such as cross-coupling, electromagnetic emission and stability of digital ports must be taken into account for PCB design.
- Ensuring fast ramp up/down when devices exit/enter low power sleep mode.

VII. Final Remarks

5G will probably become a reality at the end of this decade (by 2020) and a lot of research effort is being undertaken all over the world to define the technologies that will be employed to make 5G a reality. In this paper we presented an overview of some promising techniques and of the challenges that telecommunications R&D engineers will face in the next few years. We hope to have pointed out some of the trends that will motivate researches to dig deeper into some of the many topics still open for discussion.

REFERENCES


Mads Lauridsen received his BSc, MSC EE, and PhD in Wireless Communications from Aalborg University in 2007, 2009, and 2015 respectively. Since 2009 he has been employed in the Wireless Communication Networks section at Aalborg University. He is currently an industrial PostDoc at Nokia Bell Labs, Aalborg. His research interests include LTE-M, NB-IoT, and Ultra Reliable Low Latency Communication and Massive Machine Type Communication for 5G, and mobile terminal energy consumption.

Rafhael Amorim received his BSc and MsC in EE from Universidade de Brasília in 2009 and 2012, respectively.

Robson Vieira received the B.S. degree in Electrical Engineering from the Federal University of Goiás, Brazil, in 1999, M.S. and the Ph.D. degree in Electrical Engineering from the Catholic University of Rio de Janeiro, Brazil, in 2001 and 2005 respectively. From March 1999 to March 2005, he worked as a researcher assistant at CETUC, Rio de Janeiro, Brazil. From 2005 to 2010 he worked at INDT as a Telecommunication Specialist Researcher and his research were focus on standardization in the topics M2M, 802.16m and TVWS. From 2010 to 2016 he worked as a R&D technical manager and Head of Wireless at INDT, leading the Connectivity Research Team and driving the Wireless strategy area. Since 2016, he is working at Ektrum, a technology-based start-up focused on delivering high-quality concepts. He is author of 3 book chapters and over 30 papers in international conferences and journals as well as of 9 international patents. His research interests include IoT, 802.11 standards, LTE and 5G.