Computer Networks 106 (2016) 17–48

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

5G roadmap: 10 key enabling technologies

Ian F. Akyildiz, Shuai Nie*, Shih-Chun Lin, Manoj Chandrasekaran

Broadband Wireless Networking Laboratory, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

ARTICLE INFO

Article history:
Received 14 April 2016
Revised 5 June 2016
Accepted 10 June 2016
Available online 14 June 2016

Keywords:
5G
Wireless software-defined network (WSDN)
Network function virtualization (NFV)
Millimeter wave
Massive MIMO
Heterogeneous networks (HetNets)
Device-to-device communications (D2D)
Internet of Things (IoT)
Big data
Mobile cloud computing
Green communications
Radio access techniques

ABSTRACT

The fifth generation (5G) mobile communication networks will require a major paradigm shift to satisfy the increasing demand for higher data rates, lower network latencies, better energy efficiency, and reliable ubiquitous connectivity. With prediction of the advent of 5G systems in the near future, many efforts and revolutionary ideas have been proposed and explored around the world. The major technological breakthroughs that will bring renaissance to wireless communication networks include (1) a wireless software-defined network, (2) network function virtualization, (3) millimeter wave spectrum, (4) massive MIMO, (5) network ultra-densification, (6) big data and mobile cloud computing, (7) scalable Internet of Things, (8) device-to-device connectivity with high mobility, (9) green communications, and (10) new radio access techniques. In this paper, the state-of-the-art and the potentials of these ten enabling technologies are extensively surveyed. Furthermore, the challenges and limitations for each technology are treated in depth, while the possible solutions are highlighted.

1. Introduction

1.1. Motivation towards 5G

Over the past decades, the demand for higher data rates has been continuously growing to satisfy consumers' desire for a faster, safer, and smarter wireless network. The industry of wireless communications has experienced a great evolution in efforts of boosting system performance on data rates, from analog to digital system, from circuit-switched to packet-switched network, from a bulky handheld cellular phone to a smartwatch on wrist [1,2]. However, a significant paradigm shift is required to further strengthen the wireless communication networks, since current wireless systems are facing a bottleneck in spectrum resources which makes it difficult to enhance performance in the limited available bandwidth. The industry has predicted that in order to reach the network-level capability of serving more users with higher data rates, more spectrum resources are required to be allocated for the next generation of wireless communication networks, and current spectrum needs to be utilized more efficiently [3]. Moreover, small cells and ultra-dense network will be deployed to make the network more flexible, and to provide more users with network connectivity at anywhere, anytime. Additionally, new wireless access technologies should be backward-compatible with the existing solutions to obtain the optimal network performance with faster data rates, which ultimately lead to the driving forces toward the fifth generation (5G) wireless communication networks [4]. In this paper, we investigate 10 enabling technologies, namely, wireless software-defined networking, network function virtualization, millimeter wave communications, massive MIMO, ultra-densification, mobile cloud computing, Internet of Things, device-to-device communications, green communications, and radio access techniques, as shown in Fig. 1. Although some of the key technologies have been researched for LTE-Advanced networks, the exponential increase in data rates is propelling a major wireless network architecture paradigm shift toward the wireless software-defining networking (WSDN) and network function virtualization (NFV), which will fundamentally help reconfigure and solve the open problems in the 5G networks. Note that there is no particular priority order for the technologies in this paper.

1.2. Organization of this paper

As the research process of 5G communication system will remain active and keep growing in the coming years, this paper

* Corresponding author.
E-mail addresses: ian@ece.gatech.edu (I.F. Akyildiz), shuainie@ece.gatech.edu, shuainie@gatech.edu (S. Nie), slin88@ece.gatech.edu (S.-C. Lin), manoj@ece.gatech.edu (M. Chandrasekaran).

http://dx.doi.org/10.1016/j.comnet.2016.06.010
1389-1286/

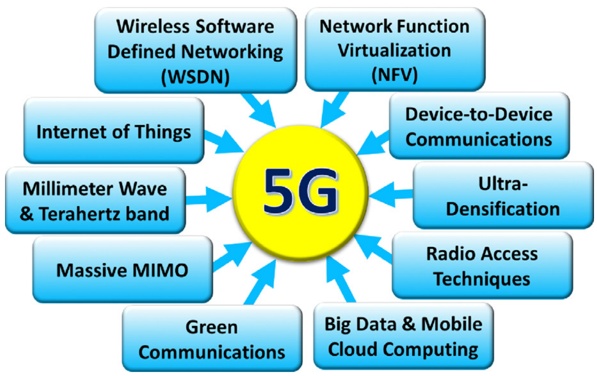

Fig. 1. The 10 key enabling technologies for 5G.

aims to motivate readers to foresee the revolutionary techniques that will shape the next generation wireless communication network. The paper is organized in the following structure. The requirements for 5G are elaborated in Section 2, which covers the demands of high data rates, low latency, cost-efficient energy consumption, high scalability, improved connectivity, and high network security. Based on the system requirements and user demands that the 5G wireless networks are expected to satisfy, the 10 key enabling solutions are explained in detail in Sections 3–12, where we cover their strengths and challenges, as well as the open problems needed to be resolved before deployed in the next generation wireless networks. The global research activities on 5G are summarized in Section 13. In Section 14, we conclude this paper.

2. Requirements for 5G

In conventional cellular networks, mobile phones were practically the only type of device expected to be supported. With the proliferation of Internet and its numerous applications, there was the problem of handling several classes of traffic to meet the different QoS requirements of diverse applications like video streaming, data, VoIP calls, etc. A similar situation is arising now with the need to support several types of devices and applications with drastically varying QoS requirement to provide better experience to the user. Unlike previous generations of cellular networks, 5G cellular network is envisioned to support a multitude of devices and applications like smartwatches, autonomous vehicles, Internet of Things (IoT), and tactile Internet. In particular, according to the International Telecommunication Union (ITU), there are three types of service scenarios to be supported in 5G, which are mobile broadband services, massive machine type communications, and ultra-reliable and low latency communications, respectively [5]. The various types of devices and application scenarios need more sophisticated networks that not only can support high throughput, but also provide low latency in data delivery, efficient energy consumption scheme, high scalability to accommodate a large number of devices, and ubiquitous connectivity for users. In this section, we describe these requirements and explain how they can be met by potential solutions.

2.1. High data rates

The metric of data rate has been the most important evaluation factor over generations of wireless communication networks. With the advent of mobile Internet and services such as high-definition (HD) video streaming, pervasive video and video sharing, virtual reality available on mobile phones, as well as the proliferation of tablets and laptops which are accessible to wireless networks, increasing the data rate of cellular network is becoming an inevitable market driving force. Although the current maximum data rates can support HD video streaming which requires 8–15 Mbps, there are applications like ultra-HD 4K video streaming, high definition gaming, and 3D contents, which require even higher data rates at around 25 Mbps to provide a satisfactory experience to users.

With these emerging applications demanding higher data rates, 5G networks are expected to have the peak data rate of around 10 Gbps which is a 100-fold improvement over current 4G networks [6]. Besides increasing the maximum data rate, the cell-edge data rate, as the worst case data rate users experience, should also be improved to 100 Mbps, which is a 100 times improvement over 4G networks at cell edge. The maximum data rate is an optimum estimate that a user can experience. In fact, the affects of intercell interference and transmission loss make the maximum value hardly achievable. Therefore edge data rate level becomes more important from the perspective of network engineering, as this data rate must support around 95% of users connected to the network. Another metric based on data rate that characterizes the network is the area capacity, which specifies the total data rate the network can serve per unit area. According to its definition, the unit of area capacity is normally bits per second per unit area. This metric is expected to increase 100 times in 5G compared to 4G network. This demand for increase in data rate can be met by techniques such as millimeter wave communications, massive MIMO systems, and wireless software-defined networking, etc.

2.2. Low latency

The roundtrip latency of data plane in the LTE network is around 15 milliseconds (ms) [7]. However, for the recently emerging applications such as tactile Internet, virtual reality, and multiplayer gaming that 5G networks are expected to support, the latency should be upgraded to an order of magnitude faster than current network, at around 1 ms [8]. For instance, tactile Internet is a recently developed application where the wireless network is used for real-time control applications [9]. The latency required for such applications is determined by the typical interaction for steering and control of real and virtual objects without creating cybersickness. The expected latency that would make these applications feasible is around 1 ms [9]. Although current smartphones have touch screens as the main interface, future devices will integrate various other interfaces like haptic, visual and auditory input and feedback, which will provide a new way of interacting with online environment for applications in virtual reality, healthcare, gaming, and sports, etc. These applications require real-time interactions with the user and any delay in the system will cause degradation to the user experience, thus latency is a crucial factor in 5G.

Another application that 5G networks are expected to support is the machine type communications (MTC) where the devices communicate with each other automatically [10]. This type of communication also requires extremely low latency for applications like vehicle-to-vehicle communication. The METIS project proposes that traffic efficiency and safety should be a typical application test case where the latency is critical in system evaluation. One typical scenario presented in METIS is intelligent traffic systems in which vehicles require timely exchange of data to avoid accidents [11]. The requirement of low latency will also improve the user experience for currently existing applications such as multi-user gaming. This demand in latency reduction requires technological innovation in waveform design as well as a flexible architecture in the higher layers of the network, which can be addressed by wireless software-defined networking.

2.3. Low energy consumption

The 5G networks are expected to support the IoT devices [12] which are basically some sensors that gather information about an environment and transmit it to a central server. These devices are mostly low-power, low-cost devices with lifespans as long as several years. Since these devices are not always connected to the base station and are only switched on occasionally, their battery life cannot afford the process of synchronization with the base station every time, as the synchronization step costs more energy than that of actual data transmission. This specific case in IoT requires that the radio access technique for 5G support loose or no synchronization. Moreover, this type of service also puts constraints on the computational power for decoding, the length of header, and packet forwarding scheme, etc.

With the increasing number of connected smart devices, the number of base stations required to support these devices will also escalate. Because of the deployment of small cells, the base station will be densified. This foreseeable trend demands the base stations to be energy efficient since even a small improvement in energy efficiency will translate to huge energy savings in large scale.

2.4. High scalability

To support increasing amount of mobile devices that connect to the wireless network and communicate with each other, network scalability becomes an important factor in design of the next generation wireless communications network. The increase in number of devices is further aggravated by the myriad of IoT devices and vehicle-to-vehicle communication technologies that are expected to surge in the 5G cellular network. Fueled by this smart equipment proliferation, it is expected that the number of devices connected to the cellular network will grow to 50 billion by 2020 [13]. Consequently, a highly scalable network that can efficiently accommodate this upsurge in number of devices is required. High scalability is also critical to performance of current and emerging applications, such as the IoT services, autonomous vehicles, etc. In the case of autonomous vehicles, prompt communications among them at high traffic densities necessitate the scalability of cellular network [14].

In fact, the scalability of a network requires full-scale upgrade in all network layers. On the physical layer, there should be enough frequency spectrum resources to support high volume of signaling and data transmission. The network infrastructure should also be able to control the transmitted power adaptively for channel estimation and to minimize interference. Scalability also influences the design of efficient media access control (MAC) protocols to accommodate large amount of connected devices. Scheduling and multicasting protocols based on geographic locations of users can greatly reduce the latency and increase the spectral efficiency. At the network and transport layers, high scalability requires networks to deploy intelligent routing algorithms for huge groups of users to provide fast and reliable connections. For users with high mobility, the vehicular network should also satisfy the scalability by providing efficient and reliable handoffs along the directions of the moving routes, as well as designing dynamic routing algorithms based on the user movement prediction [15]. The scalability in networks will be achieved by changes in all layers from radio access to core networks, using wireless software defined networks and network function virtualization.

2.5. Improved connectivity and reliability

Apart from the aforementioned requirements, coverage and handover efficiency should also be improved for a better user experience, particularly when millimeter wave spectrum is exploited. With the increase in density of the base stations and the number of devices connected, as well as the introduction of femtocells and picocells, the number of handovers that the base station should handle will increase by at least two orders of magnitude. To support this demand, novel handover algorithms and techniques that provide improved coverage in cell edge areas are required. Another related issue is the authentication and privacy concerns related to the handover [16]. The delay to contact the authentication server for each handover will be hundreds of milliseconds which would be intolerable for 5G applications. Also, given the use of higher frequency bands in millimeter wave, the transmission range of signals is greatly reduced. Hence, maintaining connectivity becomes a great challenge for 5G. For mission-critical services, the requirements on high reliability as well as connectivity should always be guaranteed.

2.6. Improved security

The security aspect of wireless network recently attracts high attention, especially after 2015, when the applications of mobile payments and digital wallet became popular [17]. In retrospect of the previous generations of systems, the general purpose of security is to protect basic connectivity and maintain user privacy. However, since the 5G system will ultimately face the dramatically increasing data traffic in the entire network, the requirement of security of 5G should not only be limited to providing trustworthy connectivity to users, but also improving the security on the whole network, addressing concerns on authentication, authorization as well as accounting, developing novel encryption protocols, and safeguarding cloud computing and management activities. For example, the security concerns are increasing since the introduction of near field communication (NFC) technique, which not only enables close proximity data transmission, but also may cause identity leaks. The 4G networks were not able to develop a unified standard to protect users' personal information, which will be fully addressed in the 5G networks.

In particular, as the IoT will come to its prime time in 5G networks, the processes of authentication, authorization, and accounting (AAA) for interconnected devices should be granted with fine-grained protecting mechanisms. Network operators, device manufacturers, as well as standardization bodies should work together on designing services, products, and protocols that can substantially protect the users subscribed to the 5G wireless network. Most importantly, as the Internet has become one of the indispensable infrastructures of the society similar to power grids, there should be enhanced federal regulations on liabilities and obligations on wireless network security.

The aforementioned requirements are summarized in Table 1, with their specifications and the associated enabling solutions. The 10 enabling technologies will collaborate to shape the 5G network architecture. The radio access network architecture will be enhanced with the deployment of wireless software-defined networking, network function virtualization, ultra-densification, device-to-device communications, millimeter wave, massive MIMO, and new radio access techniques, while the core network will evolve with key roles played by wireless software-defined networking, network function virtualization, Internet of Things, green communications, big data and mobile cloud computing. In the following sections, all the key technologies are elaborated in detail.

3. Wireless software-defined networking

One of major challenges for 5G communication networks is the design of *flexible network architectures*, which can be realized through the software-defined networking (SDN) paradigm [18]. As SDN has emerged primarily for data center networks and for the

Table 1
Key requirements and enabling solutions for 5G wireless networks.

Requirements	Specifications	Enabling solutions
High data rates	10 Gbps peak data rate; 100 Mbps cell edge data rate; Enhancing mobile broadband services.	Millimeter wave communications; Massive MIMO; Ultra-densification.
Reduced latency	1 ms end-to-end latency	D2D communications; Big data and mobile cloud computing.
Low energy	1000 times decrease in energy consumption per bit; Enhancing massive machine type communications.	Ultra-densification; D2D communications; Green communications.
High scalability	Accommodating 50 billion devices	Massive MIMO; Wireless software-defined networking; Mobile cloud computing.
High connectivity	Improving connectivity for cell edge users	Ultra-densification; D2D communications; Wireless software-defined networking.
High security	Standardization on authentication, authorization and accounting	Wireless software-defined networking; Big data and mobile cloud computing.

Table 2
Existing WSDN comparison [33].

WSDN	Architecture	Network scalability	Community
SoftAir [20]	SD-RAN & SD-CN	High	Academia/industry
SoftCell [21]	SD-CN	Low	Academia/industry
Cloud-RAN [22]	SD-RAN	Low	Academia/industry
ProgRAN [23]	SD-RAN	Low	Industry
SK Telecom [24]	SD-RAN & SD-CN	Low	Industry
DOCOMO [25]	SD-RAN	Low	Industry
CONTENT [26]	SD-RAN & SD-CN	Low	European Union
OpenRoads [27]	SD-WiFi; programmable flow tables	High	Academia/industry
OpenRadio [28]	Programmable data plane	Low	Academia
CloudMAC [29]	SD-MAC in WANs	Low	Academia
Odin [30]	SD-MAC in WANs	Low	Academia
ADRENALINE [31]	SD-CN	Low	Industry
SoftRAN [32]	SD-RAN	Low	Academia/industry

next-generation Internet [18,19], its main ideas are (i) to separate the data plane from the control plane and (ii) to introduce novel network control functionalities based on network abstraction. Towards this, by utilizing the concept of SDN, we have proposed a new architecture for wireless SDN (WSDN), called SoftAir [20], as well as its management tools.

In the literature, several studies [20–32] have explored integrating SDN with 5G network from academic and/or industrial perspectives. The comparison of these WSDN solutions is summarized in Table 2, and more details can be found in [33]. It is worth to note that several common crucial problems exist in all of these solutions (except SoftAir [20]), which significantly degrade the network scalability. Specifically, first, none of these designs have a complete architectural solution for 5G systems, but only cover partial designs such as software-defined core network (SD-CN), software-defined radio access network (SD-RAN), etc. On the other hand, SoftAir has scalable software-defined planning that brings ubiquitous software-defined design across core and access networks and has seamless OpenFlow incorporation that enables unified centralized management for the control plane. Second, priori arts, such as in the well-known Cloud-RAN [22], have limited scalability due to their coarse-grained fronthaul network decoupling, as I-Q data needs to be transmitted across processing servers. This will place a huge burden on the optical transport network, especially when dealing with advanced wireless technologies. Instead, SoftAir provides fine-grained fronthaul network composition that eliminates the fronthaul bottleneck and remarkably increases system capacity. Last but not the least, while missing in existing solutions, SoftAir equips a complete management kit, including control traffic balancing, network virtualization, traffic classifier, and carrier scheduling, maximizing the entire system performance.

To the best of our knowledge, SoftAir is the first comprehensive solution suite for 5G cellular systems that accelerates the innovations for both hardware forwarding infrastructure and software algorithms, enables efficient and adaptive resource sharing, achieves maximum spectrum efficiency, encourages the convergence of heterogeneous networks, and enhances energy efficiency. In the following subsections, we first briefly overview the key concepts of both SoftAir architecture and its detailed management tools that both serve as the foundation for WSDN design in 5G systems. Then, we introduce critical open problems in WSDN domain. As WSDN and NFV are the fundamental building blocks of the enabling technologies in 5G systems, rest of the 10 key technologies in this paper closely relates to these two technologies, and the connections will be explicitly explained later in the respective sections.

3.1. WSDN architecture

The overall architecture of SoftAir for WSDN in 5G systems is shown in Fig. 2, which composes of a data plane and a control plane. The data plane, which includes SD-RANs and a SD-CN, is an open, programmable, and virtualizable forwarding infrastructure. The control plane mainly consists of two components, which are network management tools and customized applications from service providers or virtual network operators. Three key elements of scalable SoftAir architecture are introduced as follows, which serve as the fundamental building block for general WSDN.

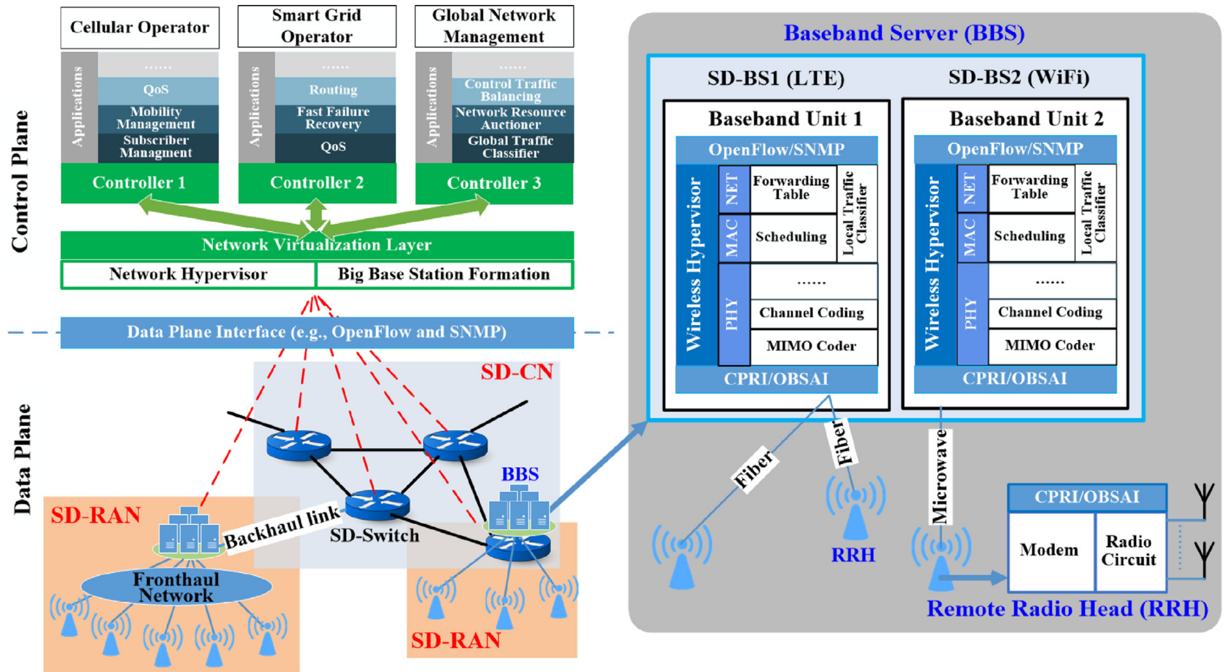

Fig. 2. Network architecture of SoftAir [20].

Scalable software-defined planning: To increase network scalability, SoftAir decouples control and data planes for both SD-RANs and the SD-CN. As shown in Fig. 2, a SD-RAN consists of software-defined base stations (SD-BSs) that jointly form a baseband server (BBS) and connect with numerous remote radio heads (RRHs). Moreover, by interconnecting physical links, the SD-CN contains software-defined switches (SD-Switches) that have flow tables to route traffic. The control logic of SD-RANs (e.g., PHY/MAC/network functions) and the SD-CN (e.g., network management and optimization tools) is implemented in software on general purpose computers, network servers, and remote data centers as high-performance controllers. Specifically, the network operating system in the control plane collects information about network states such as device status, link utilization, link delays, etc. The information collected is used to create a high-level network abstraction that helps the computation of optimum forwarding rules and efficient resource allocations. These control policies are then fed back to SD-BSs and SD-Switches for execution through standard (southbound) interfaces, e.g., OpenFlow [34] and SNMP (Simple Network Management Protocol). Also, the control plane (or controllers) provides the network abstraction to applications, such as mobility management, security, load balancing, etc., running upon it through (northbound) interfaces. Hence, SoftAir establishes a ubiquitous software-defined planning from access networks, the core network, to the Internet, effectively realizing multi-controller scenarios and optimum managements for large-scale wireless systems.

Fine-grained fronthaul network decomposition: Inspired by the concept of NFV, SoftAir supports fine-grained fronthaul network decomposition through dedicated base station NFV. As in shown Fig. 2, to enhance network flexibility, SoftAir simultaneously realizes the physical-, MAC-, and network-layer function virtualization for RAN and thus forms SD-RAN. Here, the proposed SD-RAN follows a distributed RAN architecture, and each SD-BS is split into hardware-only radio heads and software-implemented baseband units. These RRHs are connected to the baseband units on BBS through fronthaul network (fiber or microwave) using standardized interfaces, such as CPRI (Common Public Radio Interface) or OBSAI (Open Base Station Architecture Initiative) interface.

While the distributed RAN architecture has recently received significant attention for WSDN from both industry and academia (e.g., Cloud-RAN [22]), it faces two fundamental limitations. First, they cannot achieve scalable PHY/MAC-layer function virtualization. Second, they do not support NFV as the SD-CN. One major consequence of realizing all physical layer functions in the remote server is that whole I-Q plane data needs to be transmitted to the server, placing a huge burden on the fronthaul network. Towards this, SoftAir adopts a new fine-grained fronthaul network decomposition architecture by leaving partial baseband processing at the RRH (e.g., modulation/demodulation), while implementing the remaining baseband functions (e.g., source coding and MAC) at the BBS. Hence, the decomposition not only preserves sufficient flexibility offered by the distributed RAN architecture, but also eliminates the possible fronthaul network bottleneck. In short, with reduced data rate requirements, SoftAir offers excellent cooperative gain and evolvability by allowing the aggregation of a large number of technology-evolving RRHs at BBS through the diverse, cost-efficient, CPRI-supported fronthaul network topologies, and over different fronthaul mediums.

Seamless OpenFlow incorporation: Unlike existing distributed RAN architecture [22], SoftAir incorporates OpenFlow protocol into SD-BSs as shown in Fig. 2. Specifically, SD-RANs implement an OpenFlow interface for each SD-BS by utilizing Open vSwitch (OVS) [35], which is an OpenFlow-capable software switch that can easily be realized in BBS. With OVS, each SD-BS can interpret, exchange, and respond to OpenFlow protocol messages. Equipping SD-BSs with OpenFlow capabilities provides a unified interface to control and manage base stations with different wireless standards, thus leading to a multi-technology converged RAN that allows smooth transitions among different radio access technologies. Adopting the common OpenFlow interface on both SD-Switches and SD-BSs also promises the transparent interconnections between SD-CNs and SD-RANs and allows the unified management over the entire SoftAir.

3.2. WSDN management tools

To enable promising features and to maximize the overall system performance, SoftAir supports cloud orchestration that automates the configuration, coordination, and management of software and software interactions in the cloud environment [20]. This is the first time that a complete management kit is investigated in detail for WSDN.

Mobility-aware control traffic balancing: Rather than exploiting costly out-band control from a separate control channel, the in-band signaling is favored and adopted gradually in practical SDN implementation. To this end, each SD-Switch or SD-BS (i.e., BBS) needs to send control traffic, such as route setup requests for new flows and real-time network congestion status, to the controller(s). Then, based on the real-time control messages and unique mobile traffic features from mobile users, the controller optimizes the best routes for data flows according to dynamically changing traffic patterns and flow QoS requirements and sets up the flow tables of SD-switch or SD-BS along the optimal path via certain secure protocols (e.g., OpenFlow [34]), thus enabling highly efficient data transmissions and superior link utilization.

Resource-efficient network virtualization: To support Infrastructure-as-a-Service (IaaS), network virtualization capacity is critical to enable a wide range of emerging applications, provide network services at the edge, as well as achieve network slicing. While wireless network resources are limited, resource-efficient wireless network virtualization is highly desirable, and SoftAir equips two levels of hypervisors. Network hypervisor focuses on high-level resource management, which determines how to distribute non-conflicting network resource blocks among virtual network operators based on their demands; wireless/switch hypervisor is a low-level resource scheduler that enforces or executes the resource management policies determined by the network hypervisor by employing a variety of wireless resource dimensioning or wireless scheduling. These hypervisors aim to slice network resources and provide differentiated services to different slices.

Distributed and collaborative traffic classifier: SoftAir enjoys great capability to significantly improve spectrum efficiency from a designated distributed, collaborative traffic classifier at SD-BS. This classifier collaborates with a global traffic learner at the network controller(s) to achieve fast, fine-grained, and accurate traffic classification. Specifically, it identifies applications, QoS requirements, and stochastic features associated with each traffic flow and thus preserves satisfaction to mobile users even with highly bursty mobile data traffic from diverse network applications and QoS demands.

Dynamic carrier scheduling: With the aid of the proposed fine-grained fronthaul network decomposition, SoftAir enables dynamic carrier scheduling that exploits the baseband pool from BBS and allocates different carrier resources to RRHs in an energy-efficient manner. In particular, this scheduling scheme computes resource requirements with respect to different users and loads. Then, it allocates carriers dynamically to satisfy the given requirements with minimum energy consumption and thus facilitates green communication in 5G SoftAir.

3.3. Open problems

While SDN is a promising technology for future 5G networks, WSDN is still in the stage of infancy. In the following, we list several essential open problems that need to be addressed before WSDN could be readily adopted.

WSDN scalability: While several SDN solutions have been proposed recently for WLANs in an enterprise or a university, scaling WSDN with respect to SD-CN is a huge challenge that needs to be addressed. Furthermore, the advent of IoT devices also exacerbates the scaling of SD-RAN, as SD-BSs will have uneven and widely varying load differences. Therefore, to overcome the problem of WSDN scalability, three perspectives should be considered. First, based on the network topology, traffic flow statistics, application QoS requirements, and peak flow initiation rate, a capacity analysis of a single controller should be addressed in terms of throughput and latency. Next, an analysis and comparison of centralized and distributed architectures for WSDN controllers should be studied to include signaling overhead characterization and the dimensioning for each architecture. Third, in the distributed case, the optimal number and location of SDN controllers should be determined with regard to the requirements of network state synchronization as well as the corresponding consistency performance. A possible solution towards this can be found in [36], in which an on-demand deployment of local controllers is proposed based on network dynamics.

Control traffic latency: Another critical problem with WSDN is the control traffic latency, which possibly prohibits online and adaptive traffic engineering. In particular, as SD-Switches or SD-BSs have to contact controllers for routing information before forwarding data traffic, the increased latency from control traffic should be addressed carefully, especially for delay-sensitive applications. The latency resulted from OpenFlow switch processing is roughly 10 ms, which measures the time of a packet being sent until acknowledged by an unloaded controller. Another type of latency occurs in ring protection, which is around 50 ms, in order to provide the time of target restoration of a SONET ring. The highest latency, which is shared-mesh restoration, takes about 200 ms under the circumstances of voice calls dropping or ATM circuit rerouting being triggered [37]. To overcome the difficulties, two issues should be considered: mobile feature extraction of SD-RANs and timely control traffic balancing. First, to provide mobility-aware design, based on users' movement and transmission behaviors from applications, the mobile traffic distribution can be well established to capture the unique mobile feature of traffic in SD-RAN. The derived traffic distribution thus serves as an accurate mobile traffic model and will be used in the control traffic balancing design. Second, the most important requirement in the design of control traffic balancing is its computation time. To this end, in [37], the control traffic balancing problem is first formulated as a nonlinear optimization that finds the optimal control traffic forwarding paths and minimizes the average control traffic delay.

Infrastructure-as-a-Service (IaaS) with programmable data plane: Given that 5G network will support different types of emerging applications and services, such as machine-type communication, smart grid, mobile virtual network operators, and over-the-top content services like Netflix, it is essential for WSDN to offer IaaS for all kinds of services complying various service level agreements. Also, as servers that run the control plane software of WSDN will be located in a third-party cloud computing environment, service providers should facilitate IaaS to meet this geo-separation need as well. Moreover, while WSDN comes with the promise of flexibility and configurability, this could be fully exploited only when SD-Switches and SD-BSs can be programmed to execute control policies. In order to accomplish this, both SD-Switches and SD-BSs must provide a rich set of interfaces for the controllers to program. Facing these challenges, resource-efficient hypervisors [38] should be established to slice the infrastructure and to efficiently bring virtual infrastructures. Also, new architectures that speedup switches and base stations need to be developed [39].

Software-defined mobility management and green communication: In 5G wireless systems, mobility management faces significant new challenges, including minimum signaling overhead and QoS guarantees during handoff rerouting. To address these challenges, a new software-defined mobility management framework that ensures proper interaction with legacy protocols and systems should be proposed to leverage the multi-service multi-technology multi-network converged WSDN for low-cost, QoS guaranteed, and seamless mobility services. Moreover, from dynamic carrier scheduling, green communication can be easily realized through network managements of WSDN. Enabling the WSDN controller for resource allocations requires the knowledge of link loads and channel characteristics, specifically the delay characteristics, loss rate, and stability. Given that these parameters are affected by multi-path fading, extensive measurements of all these parameters being sent to WSDN controller will cause a surge in control traffic and need to be examined [40].

Reliable, efficient traffic engineering: To support various and stringent end-to-end QoS requirements of real-time applications simultaneously, WSDN should support QoS-aware routing upon designated network virtualization, thus enabling multiple isolated virtual networks to fulfill respective QoS requirements of users' applications. This means WSDN should not only slice network resources for multiple virtual networks so that they can simultaneously share the same physical network architecture, but also offer a QoS-aware traffic engineering accordingly [41]. On the other hand, to serve best-effort traffic which typically does not have specific QoS requirements (the counterpart of real-time traffic), WSDN should support throughput-optimal collaborative scheduling [42] that maximizes network throughput while reducing delay as much as possible. Throughput-optimal scheduling aims to optimally and collaboratively distribute network resources of a cluster of RRHs among their users based on the statistical properties of their traffic flows so that the overall throughput of the SD-BS is maximized. Specifically, each SD-BS can be optimally formed so that fine-grained coordination among a cluster of RRHs is enabled. Such promising feature along with the built-in traffic classifier on each SD-BS, for the first time, allows performing throughput-optimal collaborative scheduling on a SD-BS.

4. Network Function Virtualization (NFV)

NFV, a complementary concept to SDN, enables the virtualization of entire network functions that were tied to hardware before to run on *cloud infrastructure*. Specifically, in conventional architectures, operators purchase and install proprietary devices to deploy each network function, while specialized hardware is usually very expensive but barely configurable. Because of these constraints, network operators face problems in low agility, leading to longer product cycles. NFV, emerging as a breakthrough in 5G cellular systems, decouples physical hardware and underlying network functions, and let the network functions run centrally on generic cloud servers, thus providing advantages in scalability and flexibility. NFV greatly reduces capital expenses (CAPEX) required to buy hardware devices and saves operating expenses (OPEX) by aggregating resources for virtual network functions that run on a centralized server pool.

One use case of NFV is the IoT. Specifically, as the IoT paradigm and its widespread applications create an explosive number of devices connected to the network, NFV would go a long way to reduce the cost of IoT devices by reducing the functionalities in devices and virtualizing those functions. For example, in [43], a sensor function virtualization that enables distributed intelligence in IoT is proposed through the deployment of modular blocks. More details of interaction between IoT and NFV will be discussed in Section 9. As being another use case of NFV, the usage of proprietary equipments in mobile core networks causes inflexible deployments and time consuming, costly network re-planning. Instead, NFV can make core networks more intelligent, scalable, resilient, and flexible by virtualizing various entities (e.g., MME, PGW, SGW, home subscriber server, etc.) and eliminating geographical

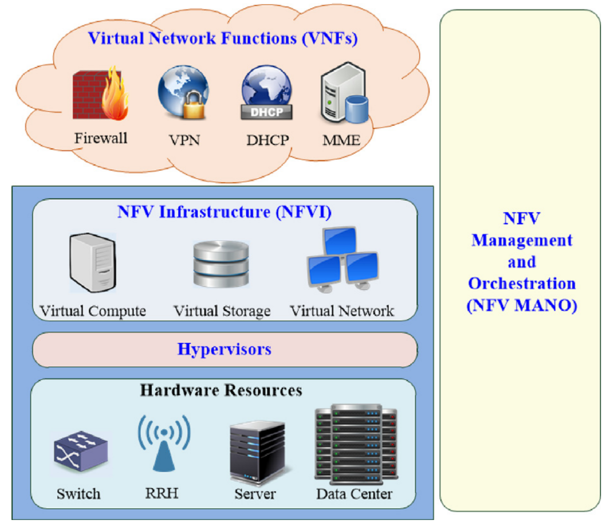

Fig. 3. Overall architecture of NFV.

limitations of the corresponding services. For example, the SGW and PGW can be co-located with the base station, avoiding long-distance links. Also, NFV can remove bottlenecks in core networks by pooling computing resources. Table 3 compares and briefly summarizes these existing NFV solutions.

A summary for the solutions that integrate NFV and WSDN designs can be found in [33]. Note that from its sophisticated architectural design, SoftAir [20] embraces a complete solution for NFV and perfectly brings the synergy for WSDN and NFV. Specifically, regarding NFV (network cloudification), SoftAir introduces a fine-grained fronthaul network decomposition that decreases the volume of data transferred between RRHs and BBS by implementing partial baseband processing at RRHs and remaining processing at BBS. Moreover, as each SD-BS has OpenFlow interface implemented, SD-BSs have a uniform interface across different wireless standards. These OpenFlow capabilities enable vertical mobility where users can roam across SD-BSs with different wireless standards. With these, SoftAir also leverages high performance controllers and optimizes network management to improve the scalability and flexibility. In the following, we first briefly overview the NFV architecture according to ETSI [44] and then introduce several crucial open problems in NFV domain.

4.1. NFV architecture

According to ETSI, the overall architecture of NFV consists of four key elements: NFV infrastructure (NFVI), virtual network functions (VNFs), hypervisors, and NFV management and orchestration (NFV MANO), as shown in Fig. 3 [44,47,48]. Specifically, the main component of NFV is VNFs which are software implementations of network functions, running on a generic cloud infrastructure. VNFs are deployed upon the NFVI that includes virtual computation, virtual storage, and virtual network resources. These virtual resources, created by hypervisors, bring virtualization over physical hardware resources within the network. In particular, hardware resources might include networking (e.g., switches and RRHs), computing (e.g., server), and storage (e.g., data centers) infrastructures. Moreover, the NFV MANO framework controls the provisioning of VNFs, the configuration of VNFs, and the infrastructure they run on. MANO can also chain several VNFs to activate an end-to-end service. We briefly explain each component [49] as follows.

NFV Infrastructure (NFVI): NFVI is the infrastructure platform over which the VNFs are deployed. Specifically, NFVI is the collection of the physical hardware which include the computing,

Table 3
Existing NFV comparison [33].

NFV	NFVI	Hypervisors	NFV MANO
SoftAir [20]	Fine-grained NFV; solve redundant I-Q transmissions	Network and Wireless/Switch hypervisors	Flexible platform for fully or partial centralized architecture (via local control agents)
OpenRoads [27]	Simple decoupling via OpenFlow	FlowVisor [45] and SNMPVisor	Simple extension to WiMAX AP
OpenRadio [28]	Enable multi-core DSP architectures	Abstraction layer for PHY and MAC functions	Focus on hardware design without control plane
CloudMAC [29]	VAPs for IEEE 802.11 stations	Dynamic spectrum use; on-demand AP; downlink scheduling	VAPs with virtual machines and dumb WTPs
Odin [30]	LVAPs for IEEE 802.11 stations	Hidden terminal mitigation; AP and its function hypervisor	No concrete solution for radio heads
ADRENALINE [31]	SDN integrated IT and Network Orchestrator (SINO)	OpenStack [46]	Deploy end-to-end VNFs
SoftRAN [32]	Rough NFV with big-base station abstraction	Antenna hypervisor	No concrete solution

storage and networking hardware and software resources. In general, the NFVI could be geo-distributed over several locations (e.g., customer premises and cloud data center) so that the network connectivity between the locations is also considered as a part of NFVI.

Virtual Network Functions (VNFs): A network function refers to the functional component of network infrastructure that provides a well-defined functional behavior and external interfaces. Some examples of network functions are DHCP servers, firewalls, NAT, gateways, and MME. VNFs are the software implementations of network functions that are deployed over virtual resources created over the NFVI.

Hypervisors: Hypervisors provide the abstraction of virtual resources over physical hardware for VNFs to run on. That is, hypervisors setup the virtualization layer that offers virtual machines (e.g., virtual compute and storage) and virtual networks over physical computing, storage and networking resources. Hypervisors also provide a logical slicing of actual network infrastructure into virtual networks for efficient management.

NFV Management and Orchestration (NFV MANO): NFV MANO provides the framework for managing and orchestrating all infrastructure resources. Specifically, it provides virtual machines, configures the virtual machines and the physical infrastructure, and manages of physical resources for virtual machines. The MANO contains three functional blocks: NFV orchestrator, VNF manager, and virtual infrastructure manager. First, the NFV orchestrator handles on-boarding new network services and VNFs, manages global resources, and validates and authorizes NFVI resource requests. Second, the VNF manager manages the life-cycle of VNF instances and coordinates the configuration and event reporting between NFVI and network management software. Last, the virtual infrastructure manager controls and manages the compute, storage, and network resources of NFVI.

4.2. Open Problems

While bringing network flexibility, NFV is not yet fully mature to be embraced by the industry for 5G systems. There are several technical gaps that need to be closed as described in the following.

Energy-efficient Network Cloudification: One of the major challenges for adopting NFV in 5G systems is to offer network cloudification in an energy-efficient way. Specifically, using hypervisors and moving functionality to the cloud, NFV should support virtual functions from physical-, MAC-, to network-layers regarding energy consumption. In particular, network-layer function virtualization decouples the routing function from hardware routers and centralizes it at network controller(s) through an open network interface (e.g., OpenFlow [34]). MAC/PHY-layer function virtualization exploits software-defined radio technologies to implement resource (e.g., time, frequency, and code) sharing schemes at MAC layer and baseband processing solutions (e.g., modulation, demodulation, channel coding, source coding, and MIMO precoding) at physical layer. Moreover, various device techniques could also be adopted to reduce power consumption, such as the migration to energy-efficient hardware, the reduction of CPU clock speeds, the shutdown of unused components, and energy-efficient function design [49]. Also, reducing the power consumption on data transfer among virtual machines should be greatly explored.

Performance and Portability: Performance is a big concern in virtualization environment, but fortunately the OVS technology [35] is quite flexible to leverage various offload technologies in hardware-level (e.g., Linux NAPI or Intel DPDK). Also, since NFV virtualizes service components (does not change the original telecom service portfolio), OSS/BSS support will be kept as the same. Specifically, VNFs will be run in a generic industry standard server, and these servers for the NFVI must be manufactured without the knowledge of function types that would run on in the future. To this end, given the unacceptance of quality degradation from NFV migration, these generic severs must provide the service comparable to that running on specialized hardware. Moreover, software techniques (like multi-threading, independent memory structure to avoid OS deadlock, and direct access to I/O interfaces) can be explored for higher performance [50]. To evaluate the performance in virtual networks, a phased cloud diagnosis [51] is proposed that collects appropriate statistical data. Another possible approach for solving the performance degradation is to consider the tradeoff between flexibility and performance by implementing some high-performance functions on hardware while virtualizing other network functions.

Management and Orchestration: Although NFV brings network flexibility and configurability, it also requires modifications to the current network management systems and will change the way networks are deployed and managed. First, a rich set of programmable interfaces should be developed that supports complex services for heterogeneous networks [52]. Second, while most of NFV frameworks only define intra-operator interfaces, the inter-operability between operators needs to be investigated. Third, when the management systems are implemented in cloud virtual machines, due to dynamic loads, the virtual machine for current VNF program should be decided on the fly. Forth, as NFV provides the ability to elastically support network function demands, the automation is the key to scale the NFV environment in regard to reliability and availability, stability, resource allocation, and energy efficiency. Last, innovative algorithms to manage the instantiation and migration of virtual functions should be developed to improve the effectiveness of NFV systems.

Security and Privacy: VNFs, which NFV runs on third-party public cloud, correspond to subscriber services and contain personal information about the subscriber. Hence, the security of the information being processed becomes important [53]. Specifically, as the VNFs might be run over several third-party servers, service providers have no direct control of the data, which raises new concerns on data privacy and security. Moreover, the new designed elements, like MANO and NFV, may also bring additional security vulnerabilities to intrusion detection systems [53]. Last, the underlying shared NFVI (i.e., computing, storage, and networking) also generates new security threats.

5. Millimeter wave

As suggested by the Federal Communications Commission (FCC) and many research groups from academia and industry, millimeter wave shows great potential in enabling gigabit per second throughput with its great amount of available bandwidth. Millimeter wave usually refers to the frequency bands at 30–300 GHz, but researchers also prefer to add the nearby centimeter frequency bands from 24 GHz to 28 GHz in discussion. The unlicensed and light-licensed millimeter wave frequency bands provide great amount of available spectrum resources that can support the requirements for high data rate, low latency, and flexible connectivity for a wide range of users.

In October 2015, the FCC proposed the rules to release millimeter wave spectrum bands for 5G services [54]. Specifically, the frequency bands that the FCC proposed to authorize for small cells deployment are the 28 GHz (27.5–28.35 GHz) and the 39 GHz (38.6–40 GHz) bands. Since wireless local area network (WLAN) at 60 GHz has been deployed as IEEE 802.11ad standard and demonstrated great success on high data rates in short range indoor environments, the FCC proposed to release an adjacent 64–71 GHz band for similar indoor short distance operations. This additional bandwidth of 7 GHz will further enhance the data rates for WLAN services. In addition, the FCC promoted a hybrid licensing scheme which leaves the choice of operation to service providers in the 37 GHz band. The hybrid band can help facilitate the development of novel broadband network techniques and also serve as a supplement to traditional cellular networks [55].

In current 4G LTE standards and widely accepted channel models, there are several commonly studied propagation scenarios, namely indoor office, indoor shopping mall, rural, dense urban, urban, and highway, etc. Those scenarios determine the arrangement of wireless communication network infrastructures and configuration of key channel parameters, including the specific electrical and mechanical downtilt angles of the base station sector antenna, the number of uniform linear arrays to install in each sector antenna, the height of base stations, and the projected coverage of a base station sector, just to name a few. In the next generation of wireless system, these legacy cell arrangements will continue to exist and serve the users. The new communication scenarios should include the direct communications among devices at millimeter wave, a millimeter wave multi-user communication scenario, and a vehicular communication scenario for autonomous vehicles at millimeter wave. Fig. 4 depicts the combination of currently supported and predicted scenarios that are discussed below in this section.

5.1. Propagation characterization

Many recent research from both academia and industry can be found on the topic of millimeter wave communication. One research focus is the propagation characterization of millimeter wave frequencies, which is essential in designing the new wireless propagation channel model for 5G network. When the frequency goes up to the range of millimeter wave, the energy absorption caused by atmosphere, rain, and snow becomes increasingly prominent, which causes limitation in signal transmission distance [56]. In particular, the 60 GHz band, which shows a significant 15 dB/km atmospheric attenuation, is only feasible for short range communications [57]. For other candidate millimeter wave bands (e.g. the 28 and 39 GHz bands), the effect of atmospheric attenuation is less dominant yet still stronger than the microwave frequencies, making them only feasible for small range deployments. The field measurements conducted in urban areas in New York City at 28 GHz showed that the line-of-sight (LOS) transmission provides the best signal strength with a pair of high-gain directional antennas [58]. However, in non-line-of-sight (NLOS) scenario when direct propagation is blocked, additional path loss is observed which greatly degrades the received signal strength and further limits the propagation distance. From the same field measurements in the dense urban environment, it is reported that the average NLOS path yields a path loss exponent of 5.76, which is much higher than the path loss exponent of 2 in free space [59]. The blockage effect at millimeter wave frequencies is also inevitable, which can be caused by common objects, such as buildings, vehicles, pedestrians, and even tree branches and foliage [60]. Consequently, the coverage range of a millimeter wave base station is dramatically limited to only a few hundred meters, even with high-gain (more than 50 dBi at transmitter and receiver combined) antennas used at both the transmitter and receiver. An outage study from the measurement campaign in New York City showed that the maximum cell radius is 200 meters at 28 and 73 GHz, and the major blockage effect is caused by the dense buildings in urban environments [61].

5.2. Spatial multiplexing and beamforming

One solution to compensate the high atmospheric attenuation brought by millimeter wave is to use large antenna arrays at the base station. Instead of the omnidirectional antennas used by past and current generations of wireless communication systems, 5G will deploy a large set of narrow-beam directional antennas that can pinpoint to multiple users simultaneously. Therefore, large antenna arrays are essential for serving a wide range of users in highly populated areas.

Among all antenna array techniques, spatial multiplexing and beamforming are two most popular approaches in research [62]. Normally, spatial multiplexing utilizes the multiple antenna elements to transmit multiple independent and separately encoded data streams. While each data stream travels through different propagation channels, the receivers also equipped with multiple antennas will be able to receive and reconstruct the original data transmitted with better spectral efficiency [63]. Spatial multiplexing is mostly effective when the propagation paths are rich and can support multiple data streams for transmission, and also when the channel has high signal-to-noise ratio (SNR) which will not affect the signal strength when the original signal has to be split into multiple data streams [64]. Beamforming, on the other hand, combines antenna array elements at the base station adaptively and utilizes the beamforming weights multiplied on each element to control the directions to which the data streams are transmitted [65]. In this way, specific users will be directed to receive the data while others will not be interfered. Beamforming performs well when the channel has limited power or low SNR, as a result of the combination of multiple antenna elements coherently [66,67]. As for millimeter wave, since the wavelength becomes much shorter, the physical size of antenna will accordingly be smaller. The hardware prototype design for both spatial multiplexing and beamforming are carried out in various research groups. In one millimeter wave beamforming prototype tested by Samsung Electronics in South Korea at 28 GHz carrier frequency, researchers use a 32 × 32 antenna arrays as the RF transceiver configuration. Their test

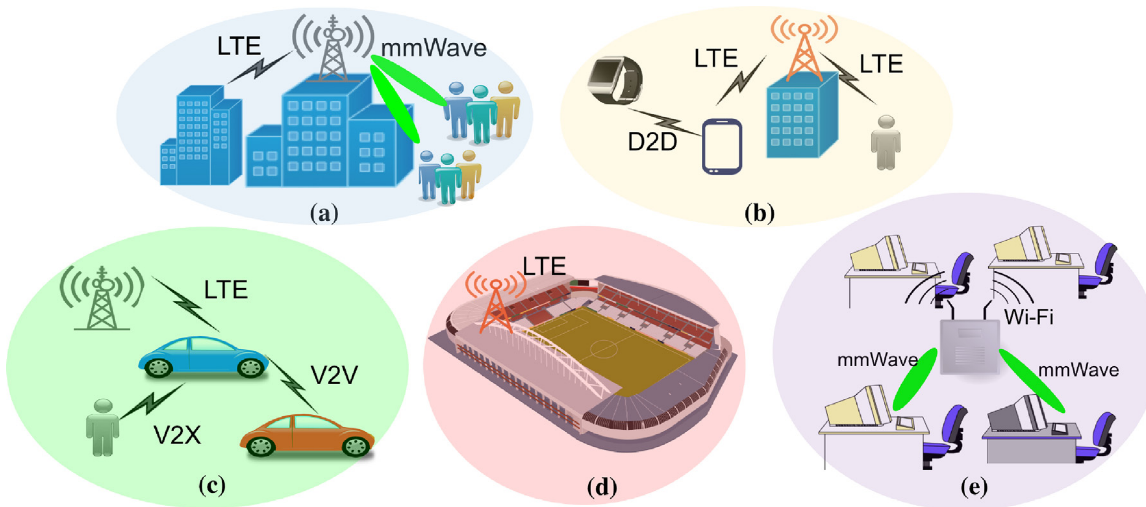

Fig. 4. The vision for 5G wireless communication scenarios. (a) Urban environment with high user density; (b) device-to-device communications; (c) vehicular communications among autonomous cars and users; (d) dense population in a stadium; and (e) indoor millimeter wave communications.

results demonstrate great potential of millimeter wave beamforming in the small cell deployment scenario [68].

5.3. Terahertz (THz) band for beyond 5G

Besides millimeter wave, there is also increasing attention focusing on even higher frequencies in the Terahertz (THz) band. The THz band spans the frequencies from 0.3 THz to 3 THz, with a broad overlap with millimeter wave frequencies. Although the IEEE 802.15 Terahertz Interest Group was chartered in 2008 to explore the feasibility of Terahertz band communications, there is yet no regulation on this extremely high frequency bands, which makes it a “no man’s land” [69]. However, the THz band offers great amount of spectrum resources, which show the great potential to support data rates of more than 100 Gbps or even 1 Tbps [70]. Additionally, with the advances in semiconductor transceiver design at high frequencies, the THz band transceiver architecture also can be backward compatible with the millimeter wave bands, making the THz band the next frontier in research of wireless communication networks, and creating a paradigm shift from traditional large scale transceiver design to nanoscale architecture.

Current research has covered various topics such as channel modeling, transceiver device design, antenna architecture design, waveform design, network and transport layer design, etc. [69–73]. New research areas are being developed, such as Wireless NanoSensor networks and the Internet of NanoThings [74,75], and the novel concept of Ultra-Massive MIMO by 1024 $\times$ 1024 is being introduced as a means to increase communication distance [76]. Hence, the THz band is envisioned as one of the key enablers for ultra-high-speed short range communications.

5.4. WSDN enabled millimeter wave radio access network architecture

With the foreseeable deployment of millimeter wave which means cell size will shrink greatly, the challenge is brought by coordination among millimeter wave base stations with a dramatically increased density. Hence, the WSDN offers a flexible and efficient approach for network resource management of the RAN for millimeter wave communications in SoftAir, as illustrated in Fig. 2 [20].

In the coordination of millimeter wave base station, the SD-BS can manipulate the adaptive antenna arrays at the RRH in order

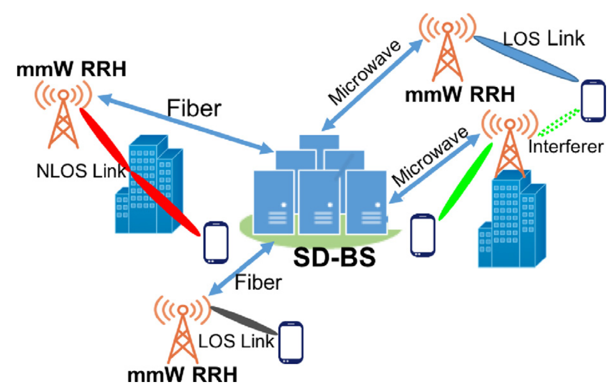

Fig. 5. The architecture of WSDN-enabled millimeter wave radio access network.

to maximize the gain for specific user while minimizing the interference to nearby RRHs. The channel state information (CSI) estimated by the SD-BS can determine the hosting scheme (i.e. the corresponding RRHs) and the transmission mode (i.e. legacy LTE or millimeter wave) for satisfying QoS. Since the millimeter wave propagation links can be constantly changing in channel conditions such as LOS probability, received signal strength, level crossing rate, etc., the SD-BS and its associated RRHs will make dynamic power level adjustments to ensure the system performance. Millimeter wave backhaul is also very promising in supporting ultra-high speed wireless backhaul compared to fiber. WSDN architecture with millimeter wave backhaul links can solve the deployment restricts that fiber faces. The centralized and scalable management of the SD-BS will efficiently improve the performance in RAN and reduces the cost of CAPEX/OPEX. As shown in Fig. 5, the millimeter wave RAN is enabled by WSDN.

5.5. Open problems

Millimeter wave shows many attractive characteristics on large amount of available spectrum resource and acceptable propagation behavior in LOS environment, which makes it promising in serving the 5G technological revolution. The advantage of broad bandwidth offers great potential for developing new air interface technologies. The large antenna arrays will also bring high spectral efficiency, high throughput, and additional channel gain when deployed along with millimeter wave system. However, there are still some technical obstacles that must be resolved before moving from the prototype to commercial market. The problem of high power consumption of large number of signal processing components is still an open issue. Also the challenge of channel estimation and prediction for adaptive power control also need solutions. In this subsection, we discuss the open problems of millimeter wave communications and potential solutions.

3-D channel modeling: One open challenge calls for a uniform 3-D channel model to cover all types of communication scenarios. So far, there is no widely accepted conclusive model that can comprehensively characterize the wide millimeter wave frequency bands. Research groups have been promoting their developed channel models in hope for standardization body adoption. The 3-D channel model proposed by a research group in New York University is based on measurements conducted in dense urban environment, which is considered where the 5G requirement of high data rates will mostly need to satisfy [77]. The U.S. National Institute of Standards and Technology (NIST) are supporting a collaborative research effort called “5G mmWave Channel Model Alliance” to collect insights to systematically characterize mmWave channels. In Europe, a model specifically design for the 60 GHz frequency bands named “Millimeter-Wave Evolution for Backhaul and Access (MiWEBA)” proposed a quasi-deterministic model for a wide variety of communication scenarios, and took the backhaul links into model considerations [78].

In general, channel modeling approaches including both statistical and stochastic methods are considered, which also have their distinct advantages and drawbacks. The statistical models are created based on huge amount of data collected from actual field measurements, which is the reflection of the real-world propagation channel, but the data collection process is extremely time-consuming and normally vary a lot with different geographical locations. The accuracy of the data is also limited by the measurement equipment as well as the post-processing methods. On the other hand, the stochastic model is based on a step-by-step procedure following the probability distributions of channel model parameters given the environment geometry. No exhaustive labor is required to complete the channel modeling, but the stochastic model is not convincing unless validated using either field measurement data or the widely accepted ray-tracing method. For the 5G communication system, a hybrid model, which is the combination of both statistical and stochastic model, is preferred. The hybrid model should consist accuracy, spatial consistency, and environment scalability [11]. In the hybrid channel model, the first step can be implemented by the stochastic model where key channel parameters can be characterized by probability distributions. Once the model is completed in various environments, field measurements can be conducted to verify the stochastic model and provide modifications in case of discrepancy.

Since in real world environment objects are constantly moving, the transition between LOS and NLOS will be frequent and unpredictable, especially at higher frequencies. In current 3GPP channel models, the probability of LOS is modeled as a function of separation distance in 2D and the height of the UE in only two environments, indoor and outdoor [79]. However, the actual probability of LOS is more complicated, and should not be limited in two static environments. Therefore, the millimeter wave channel model should also consider the dynamic signal strengths and develop a flexible link transition model between LOS and NLOS. Also, comparing with current microwave frequencies, millimeter wave demonstrates high penetration loss for concrete walls, making the signal difficult to receive by an indoor receiver from an outdoor base station [80]. However, this can be a benefit since it eliminates the interference from nearby cells. Hence in terms of small cell system design and base station setup, the same frequency bands can be reused by indoor access points and outdoor small cell base stations.

Dynamic power control: Another open problem of using millimeter wave for communication is how to dynamically control the propagating signal strengths and track the signals in order to adapt to actual channel environments. As mentioned before, at millimeter wave frequency the NLOS transmission will cause a big drop in signal strengths for $15 \sim 40$ dB as compared to LOS, the channel condition thus varies significantly with surrounding environments, walking pedestrians, and traffic. In order to deploy millimeter wave for 5G network, researchers need to dynamically track and control the power of the signal so that users can experience satisfying high data rate services while base stations have cost-effective energy consumption.

Cell search: In the 4G LTE systems, the cell search process is designed based on omnidirectional transmissions of base stations, which allows the user equipments (UEs) to establish links to the serving cell. However, in millimeter wave communication, directional transmission will be deployed in order to compensate the high propagation loss in omnidirectional transmission scenario. But the directional transmission scheme causes complexity issues to initial cell search, as both the UEs and BSs will search and locate suitable transmission space to establish links. One proposed directional cell search process provides a solution by letting the BS broadcasts synchronization signals at random directions [81]. The detection process is based on a generalized likelihood ratio test and the results demonstrated that digital beamforming outperforms analog beamforming in cell discovery.

User scheduling and congestion control: Under the circumstances of high path loss and blockage effect in millimeter wave communications, a key solution is to utilize the gain in highly directional beamforming to establish a multihop relaying network architecture. Since the multihop architecture provides link diversity and cell range extension, more users can be connected to the BSs. However, in the air interface system design, an open problem comes to user scheduling and congestion control scheme. A possible solution is to deploy a dynamic duplex scheme to allow maximum resource allocated for users, as in contrast of static duplexing schemes (i.e. TDD and FDD) in current 4G network [82]. However, more solutions are still needed as well as field measurements validations.

Hardware limitations: One of the open challenges is the high cost and energy consumption of RF hardware. Due to the frequency characteristic difference between millimeter wave and legacy microwave, the transceiver equipment designed for lower frequencies is no longer suitable for millimeter wave frequencies. As the wavelength decreases, the antenna arrays used at both base station and mobile terminals will be replaced accordingly. The good thing is, with the smaller wavelength at millimeter wave, the physical size of antennas will shrink, allowing more antennas to be built in the same area as before. However, the downside comes to the equipment of RF chains after the antennas. In current wide-deployed digital transceiver architecture, each antenna is connected to an RF chain, which consists of power amplifiers (PAs), analog-to-digital converters (ADCs) and digital-to-analog converters (DACs). The power consumption of the large number of PAs, ADCs, and DACs is huge, making this conventional architecture not feasible for millimeter wave antenna array architecture. And the manufacturing cost for large number of antenna elements per base station also prohibits the deployment of traditional architecture. In another word, from an energy efficient and cost-effective perspective, the number of RF chains should be less than the number of antenna elements. A hybrid beamforming architecture is promoted which could take advantage from both analog and digital beamforming to achieve higher throughput [83].

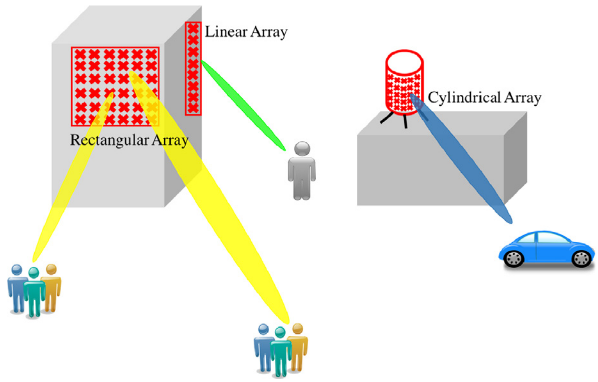

Fig. 6. Three types of layout of massive MIMO base station antenna array: linear, rectangular, and cylindrical arrays.

Adaptive beam-steering technique: Another one of the important open challenges in millimeter wave is to support beam-steering where the beams can be dynamically steered to follow user mobility as well as to overcome blockage effects. This adaptive beam-steering technique involves designing intelligent prediction algorithms, rapid beam switching methods, as well as overcoming the hardware constraints for beamforming.

6. Massive MIMO

In addition to millimeter wave communications, massive MIMO is proposed as another key enabling solution for 5G networks. Different from the multiuser MIMO (MU-MIMO) deployed in the 4G system in which only several tens of antenna components are built on base stations and user terminals, massive MIMO designs hundreds of antennas at base stations to further increase capacity and system throughput, as well as to provide more degrees of freedom [84]. Massive MIMO utilizes spatial multiplexing (previously discussed in Section 5.2) and time-division duplexing (TDD) to serve multiple users on the time-frequency resources. Fig. 6 demonstrates three possible types of layout of massive MIMO antenna arrays at base station, namely linear, rectangular, and cylindrical arrays, respectively. The different layouts of massive MIMO base stations are designed for different deployment scenarios. For example, the linear and rectangular antenna arrays can be installed on the outer surface of high-rise buildings to serve indoor users with various heights and outdoor pedestrian users. The rectangular antenna array can also serve multiple users using spatial multiplexing technique. For moving users, the cylindrical arrays can be deployed to track the users' moving routes.

The benefits of massive MIMO include a huge increase in spectral efficiency, a reduced latency, and a scalable air interface structure. Since there are many more antenna elements at the base station than those at mobile terminals, the signal pre-coding procedures will be simplified, resulting an enhanced spectral efficiency [85,86]. The reduced latency is achieved by the law of large numbers embedded in the nature of massive MIMO which eliminates the frequency dependence of the channel and the frequency selective fading will not affect the signal strength. Consequently, each transmitted data stream will no longer suffer from channel distortion and can arrive at the receiver with reduced latency. Due to the large number of base station antenna components, the actual air interface structure architecture can be designed to have flexibility and scalability depending on the real-time data traffic conditions. For example, if a cell served by a massive MIMO base station only connects to several user terminals, not all base station antenna elements will be necessarily active, so the energy consumption per each individual antenna element can be reduced to its minimum value that still satisfies the operating requirement. If heavy data traffic occurs in one massive MIMO cell range, all the components will be active to provide fast data transmission to all users. The benefit of scalability of massive MIMO is quite attractive to 5G architecture researchers and will play a significant role in the 5G network design.

6.1. Time-division duplexing in massive MIMO

As discussed in Section 5.2, massive MIMO is operated on the principle of spatial multiplexing which requires that the propagation channel has rich paths. The channel estimation is dependent on the number of orthogonal pilots sent by the base station to the user terminal, which is proportional to the number of antenna elements at the base station. Therefore, the frequency-division duplexing (FDD) mode for current LTE systems is not feasible for massive MIMO channel estimation, since it requires that both the uplink and downlink channel knowledge should be known to the base station in order to achieve optimal efficiency. In the case of massive MIMO where several hundreds of antenna elements are configured at the base station, the downlink channel estimation will become a bulky task for user terminals to estimate all the pilots from the base station. Hence the channel estimation becomes computationally cumbersome and is not realistic as the user terminal has the nature of mobility. The solution is by utilizing the reciprocity of the uplink and downlink in TDD for channel estimation [85,86]. The user terminals will send pilot waveforms to base stations for uplink channel estimation, and then the base station can relatively easily perform downlink estimation based on the received uplink pilots. The detailed quantitative analysis can be found in [87–89]. Some studies propose a technique called joint spatial division and multiplexing (JSDM) to exploit the correlation of channel vectors and demonstrate the feasibility of FDD systems in achieving comparable performance as its counterpart of TDD in massive MIMO [90].

6.2. Full-dimension MIMO

One of the scalable implementation of massive MIMO is full-dimension MIMO (FD-MIMO), which takes the spatial multiplexing in the elevational plane into consideration and utilizes the 3D channel model to achieve better performance [91]. In traditional spatial channel modeling standards such as 3GPP [92], only signals propagating in the azimuth direction are considered, while the vertical beam pattern and antenna downtilt angles are normally fixed. This type of antenna array configuration has been proven to be inadequate [93,94]. Recent research realizes the importance of providing one extra degree of freedom in the elevation domain. For example, in some cases where users are distributed at multiple heights in a building (e.g. a stadium or a shopping mall), the FD-MIMO can dynamically steer the vertical beams toward the users. Hence, FD-MIMO shows great potential in massive MIMO application which utilizes the full 3D spatial channel. With the extra degree of freedom brought by the antenna beams arranged in the elevation direction, the system capacity can be increased by 30% [95].

6.3. WSDN-based massive MIMO coordination scheme

Since the deployment of large scale antenna arrays in massive MIMO will yield huge amount of spatial domain information about the channel knowledge, the estimation of channel state information requires timely update and quick computations. WSDN provides an alternative for dynamic spatial domain radio resource processing [96]. Traditional BS can only collect partial channel state information due to the distance limitation, whereas the interference from other cells are normally neglected, which can be problematic. By the use of RRH and SD-BS, the WSDN can collect channel knowledge with a widely extended coverage, thus enabling intelligent interference cancellation and spatial domain utilization.

6.4. Open problems

In this subsection, several open issues in massive MIMO are described and possible solutions are presented.

Near-field vs. far-field channel models: In cellular systems, it is usually assumed that signals, which are considered as plane waves, propagate in the far-field region. This assumption is valid for current and legacy generations of wireless systems because of the relatively low carrier frequencies. According to the definition of Fraunhofer distance which determines the boundary between the near and far fields [97], we know that with higher frequency (i.e. smaller wavelength) in millimeter wave and larger dimension of the antenna, the Fraunhofer distance will increase and thus the near-field region expands. In that case, the basis of plane wave propagation will not be exclusively valid for propagation channel characterization. To overcome this challenge, a distance-dependent channel model that considers the distinct effects in both near and far fields should be developed in order to fully understand the channel characteristics with huge antenna array aperture and small wavelength.

Pilot contamination: With the base station requirement of hundreds of antenna components to be installed to serve multiple users at the same time, the challenges of synchronization and multiuser interference suppression occur. In small-size MU-MIMO systems, the problem of synchronization is solved by performing a calibration process through exchanging pilot waveforms between base station antennas and a so-called "reference" antenna [98,99]. This calibration process is very sensitive to the layout of the reference antenna and not feasible for the large scale massive MIMO case. In traditional MIMO there exists the problem of pilot contamination, which is caused by the reuse of limited pilot waveforms in multiple cells in channel estimation. But the effect of pilot contamination is much serious in massive MIMO. Study shows that with incremental antenna elements without upper bound, the problem of pilot contamination can lead to severe interference among cells when each cell serves the maximum number of user terminals [86].

Several solutions are proposed to mitigate the effect of pilot contamination. One is to decrease the pilot reuse factor, which can raise the distance between two cells utilizing the same frequency band. The technique of fractional frequency reuse (FFR) proposed to maintain the original pilot reuse factors in cell edges and set an additional pilot waveform to each cell's center, in order to decrease the interference among cells [100]. Another solution is to utilize the blind pilot decontamination technique to avoid using pilot waveforms in channel estimation [101]. This technique is based on the eigenvalue decomposition of channel's sample covariance matrix, in the theory of random matrix. The third solution looks at new precoding algorithms, such as pilot contamination precoding (PCP) [102]. This algorithm allows multiple base stations with the same pilot waveform to collaborate constructively and therefore cancel out the interference caused by pilot contamination.

Very large antenna array design: As shown in Fig. 4, various massive MIMO base station layouts can facilitate diverse user scenarios. However, the popular open issue comes to the design of such very large antenna arrays. Since the physical space for base station equipment is often limited, research has been focused on how to fit hundreds of antenna elements into a limited area without sacrificing system performance and reducing the complexity of the signal processing hardware. Up to now, three commonly discussed architectures are analog, digital, and hybrid arrays. In the analog array architecture, all signal processing components are located at the central processing unit and thus can save space in the base station setup; however, the distance between the antenna and signal processing units introduces unneglectable signal loss and noise, as well as interference. In the digital array architecture, there is a complete RF signal processing chain after each antenna element, which reduces the unfavorable noise and interference, but the cost for large number of RF components constrains market application, also the physical size for the signal processing chains is insurmountable. Considering the above tradeoffs, the hybrid architecture seems an optimal solution. In the hybrid architecture, the digital baseband processing is done at the central processing unit and the RF beamforming only requires a series of phase shifters which does not need complicated circuitry [103]. However, the implementation of the large antenna array architecture still holds challenges.

When the antenna elements are placed close to each other, there is an inevitable electromagnetic interaction (i.e. mutual coupling) among them [85]. One solution is to mitigate the effect of coupling via impedance matching RF circuits with multiple ports [104]. However, as the number of antenna elements increases and the spacing shrinks, the influence of mutual coupling will become prominent and hence new approaches should be considered.

Distributed massive MIMO: The concept of distributed massive MIMO is brought to address the issue of antenna array form factor constraints, by utilizing the formation of arbitrarily large number of antenna elements opportunistically in both transmission and reception [105]. With the flexible scheme in antenna array form factors, higher transmission and reception directivity can be achieved, as well as the increase in spatial and spectral efficiency. However, the challenge resides in the synchronization of RF chains in massive MIMO antenna arrays. If digital beamforming is deployed, there should be a synchronization signal sending to each assigned RF chain to establish the array. The algorithms of maximum ratio combing (MRC) and minimum mean square error (MMSE) beamforming are being discussed in [106]. The complexity of the algorithms and the power consumption for the arbitrary array synchronization are still calling for better solutions.

Energy consumption and cost: The most energy-consuming part in a massive MIMO system is the RF signal processing components, which include ADCs, DACs, PAs and some other necessary electric components. The manufacture of microwave components is already matured so the cost will not be a major issue. However, if massive MIMO is to be used at millimeter wave frequencies, the design and manufacturing will set higher requirements and thus cost will increase [86]. Testbeds of millimeter wave massive MIMO system should be verified carefully before mass production, and some key performance indicators (KPIs) (e.g. energy efficiency) should be taken into evaluation.

7. Ultra-densification

The traditional macrocell network architecture has served as the basic network architecture in the 4G system. Estimations have shown that the monthly mobile data traffic around the world will exceed 15 exabytes by the year of 2018 [11]. Such tremendous growth in bandwidth demand comes from the accelerated development of new smartphone, tablets, wearable devices, and other types of new devices that requires network access and desires decent connectivity for data roaming and video streaming. With the unprecedented diversity of mobile devices, applications, and services, the 5G cellular network calls for high flexibility at all network layers. However, the traditional high-tower mounted base stations cannot keep up with this increase. In order to accommodate the ever-growing user terminals and offload the congested

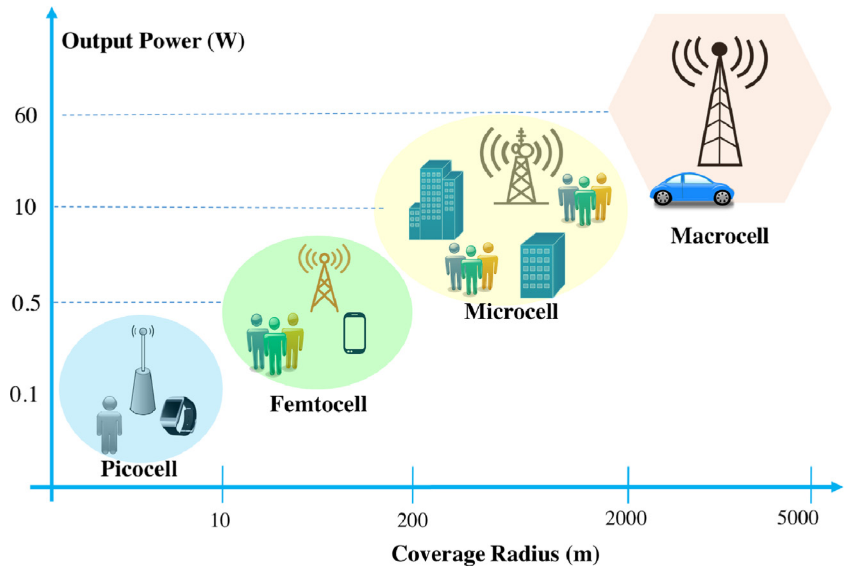

Fig. 7. Ultra-densification as a combination of microcell, picocell, and femtocell.

data flow in 4G base stations, two major solutions favorable to many are spatial densification and spectral aggregation, which can create more layers of cells and manage the shared spectrum resources wisely to increase capacity. To be more specific, spatial densification includes the idea of heterogeneous network (HetNet), which has been brought to attention in the 4G LTE system. The group of smaller elements for HetNets including picocells, femtocells, and distributed antenna elements will serve a significant role in enhancing system performance, as shown in Fig. 7. On the other hand, spectral aggregation takes the challenges of aggregating the non-continuous fragments in bandwidth resource, and utilizes spectrum sharing to coordinate the licensed as well as unlicensed spectrum in order to achieve higher spectrum usage.

7.1. Spatial densification

A key performance indicator in network performance is the outage probability. Outage probability evaluates the quality of service of a cell on the quantitative basis of cumulative distribution function of the signal-to-noise-and-interference ratio (SINR). Under the premise that high SINR also leads to high data throughput, the users in the center of a cell coverage will tend to have higher data rates than the users at the edge of a cell. And in the multiuser case, according to the Round-Robin (RR) scheduling algorithm, the user with higher SINR will always be assigned to reach system efficiency. However, it leads to unfairness as the users at the cell edge with lower SINR might not be served at all, causing unbalanced network resource allocation and service degradation for edge users. In HetNets, the users originally located at the edge of traditional microcells will be covered by picocells or femtocells, depending on the cell size and environment, thus the distance between the serving base station to the user is shrunk with higher SINR at the receiver, which in turn decreases the outage probability.

The idea of small cell is a realization of spatial densification in network architecture planning, which utilizes low-power nodes to provide a cost-efficient alternative to reduce the load factor for traditional urban microcells. The benefits of ultra-densification over spatial densification are twofold. First, it allows more capacity in the network. With multiple layers of cells in one area, a user can be assigned to a cell that can provide the best channel throughput. Hence the data traffic can be dispersed to different microcells, picocells, and femtocells. In this way the network capacity can be improved. Second, it enables a more flexible network architecture planning. In traditional network planning hexagonal shaped cells are considered in almost all scenarios. Numerous results have proven that in real-world network hexagonal cells may not necessarily serve as the best option [107–110]. For 5G network, the cell size and shape should be able to have configuration flexibility in order to achieve both better spectrum and energy efficiency. For instance, the concept of moving cells and nomadic nodes will provide small cells with high mobility. Together with D2D communications which will be discussed in later sections, the small cell BS can be mounted on top of trains, cars, and buses, which allows users in close proximity to have network connectivity [11].

7.2. Spectral aggregation and spectrum sharing

In addition to spatial densification where users are associated to different cells to increase network capacity, the spectral aggregation provides an alternative to condensing the network through the means of spectrum resource management. At the early stage of wireless communications, due to the limitations in signal processing techniques and the fact that frequency spectrum resources were not envisaged as scarce resources, some spectrum resources experienced an extravagant deployment. For example, the analog TV used to cover a wide spectrum in the VHF to UHF range in the 20th century. However, with the invention and widespread of digital TV which takes much fewer bandwidth for signal transmission, the analog TV gradually resigned globally in late 2000s, leaving the white space spectrum underutilized. This precious spectrum resource can provide effective broadband access and has shown better performance than current Wi-Fi network. In 2010, the FCC opened this spectrum white space for public use, and then the IEEE standardized this spectrum as IEEE 802.11af for spectrum sharing [111,112].

One way of integrating the spectrum required by new applications into current network layout is through spectrum sharing techniques with the help of cognitive radio technologies [113,114]. In LTE-A, the approach of licensed assisted access (LAA) is deployed to boost small cell coverage, where the licensed channels are primary carriers, serving voice calls in both uplink and downlink. And the unlicensed spectrum available in the network will serve UEs when they demand high-throughput data services. The unlicensed secondary users can obtain the dynamic spectrum resource availability and thus make decisions on spectrum usage, without causing interference or channel collision to licensed primary users. Additionally, this idea of dynamic spectrum sharing can improve the scalability of the entire wireless network. In a real-world scenario, when an emergency alert needs to be spread from one cell to a large area, the current available spectrum resources will be deployed for that specific purpose, without causing interference to occupied spectrum or collisions to busy channels.

7.3. WSDN-enabled ultra-densification

The idea of ultra-densification calls for the design of a scalable, efficient, and flexible radio access network (RAN) for the 5G system. The approach of logical separation between the control and data planes in WSDN is considered promising to realize such requirement. Under the premise of control-data separation architecture in WSDN, there will be two types of base stations, namely control base stations (CBS) and data base stations (DBS) [115]. The CBS manages cell coverage and resource allocation, whereas the DBS executes the signal transmission and reception, within the control of CBS. In this way, all UEs within the cell range, no matter busy or idle, are connected to the CBS, but only the busy UEs are also connected to DBS for data transmission. The separation of the control and data planes in WSDN can therefore improve the spectral efficiency and reduce the energy consumption of signaling at the base station.

Though appealing in theoretical analysis, the control-data separation architecture faces a few challenges before full deployment. The backhaul between the CBS and DBS requires low latency and high data rate transmission to reflect the UE activity in a timely manner. However this increases the system overhead, which brings the question on how to design an efficient signaling mechanism between CBS and DBS [116].

Another challenge is the authentication of new UEs admitted to another small cell upon handoff. Since in current wireless network architecture, the authentication servers are often located far away from the local RRH and BS, the round trip time for authentication request and confirmation is normally up to the scale of hundred milliseconds, which is unbearable for the 5G requirement on latency [117]. Even with some simplified authentication algorithms such as mutual authentication [118], the packet overhead is still large which makes the authentication process less efficient.

Regarding the above open problems, WSDN can be utilized for centralized orchestration, as shown in Fig. 8, which coordinates all handoff requests and confirmation messages, distributes forwarding instructions to SD-Switches, and assigns RRH to admit new UEs to its cell range, as described in detail in Section 3.2. In this way, the solution provided by WSDN-enabled network architecture can help resolve the latency issue in ultra-densification.

WSDN will also serve as an optimal solution for the challenge of inter-cell interference, one of the significant issues existing in HetNets. Since the RRHs are controlled in a centralized manner by the SD-Controllers which can adaptively manage the radio power of adjacent cells, the inter-cell interference will be greatly mitigated.

7.4. Open problems

Although many benefits will make ultra-densification an indispensable technique for realizing 5G, we still need to tackle the challenges before integrating the novel network architecture.

Intra-/inter-layer interference management and cancellation: The source of interference can be intra-layer interference and inter-layer interference. The intra-layer interference consists of uplink and downlink interference, which originate from the uplink pilot waveforms from user terminals to each small cell and the

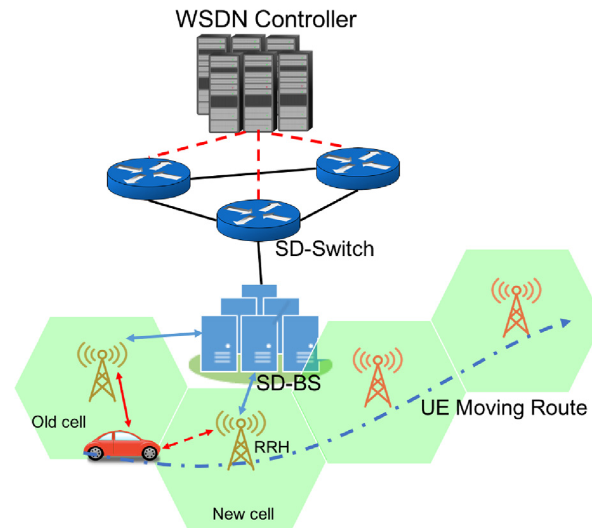

Fig. 8. The centralized authentication process enabled by WSDN controller in ultra-densified small cells.

downlink signal carrying data packets, respectively. On the other hand, the inter-layer interference, in a broad sense, comes from all types of signaling and packet streaming from active user terminals and all available cell towers under the macrocell coverage [119–121]. Even though the small cells use low-power nodes for smaller coverage area to suppress intra-layer interference, the inter-cell interference generated by the higher power cells would still cause service degradation. There are two solutions available to cope with the interference, which are the smart power management and inter-cell interference coordination (ICIC). Although intelligent power management research and ICIC solutions are available for deployment in current LTE network, more advanced and comprehensive solutions are necessary for network with new proposed 5G techniques such as massive MIMO and high frequency systems [120,122,123].

User scheduling and mobility management: In ultra-dense networks, other challenges include the user scheduling and handoff among cells. In the current 4G networks, the user will normally be associated to the cell that provides the best connection. However, in case of the HetNet, a single user may find more than one cell available for data transmission and reception, on the norm of SINR. And the BS will also likely find more than one user equipment that indicates the best channel condition to serve. In this case, both parties in communication link should take more metrics into consideration. The power efficiency, which can be expressed as the output signal power divided by the input signal power, can be assessed as another KPI [124]. When the SINR is not the sole channel condition estimator, the power efficiency can be introduced to the algorithms of cell association. However, when the user is in motion, the signal strength may vary and thus the user might experience handoff from one cell to another to remain connected. In HetNets, users may find more than one cells available to handoff, therefore an intelligent handoff scheme in HetNet is required to address this problem.

Costs for small cells: At the early stage of small cell's deployment in the 4G network, bulky transceiver equipment, complicated power amplifier and filtering circuits, and heavy cables were behind each small cell base station in the backhaul connections. Both the energy consumption and costs set limitation on the potentials of small cells. An estimation shows that the cost of backhaul network occupies 30% of the total cost in wireless networks [125]. The optic fiber is widely deployed in backhaul connection due to its high bandwidth and robustness compared to the wireless option.

However, with the 5G projection of large number of small cells installed in dense urban areas, the fiber infrastructure needs to be planned efficiently to minimize the cost and also maintain the network resilience. One solution proposed is using WSDN (as describe in Section 3) to assist a hybrid backhaul architecture, which contains both optic fiber and millimeter wave connections [126]. The simulated backhaul network was proven to provide cost efficiency in dynamic resource allocation. However, this promising approach still needs prototype validation before commercial deployment.

Management and pricing of small cells: Since the main idea of small cell is to provide higher throughput for a smaller covered area, then for one macrocell coverage there should be many more small cells to be deployed, which will induce the management issue of small cell installation and maintenance. Network operators are reluctant to increase their budget on OPEX of small cells, and end-users are prone to choose the less expensive Wi-Fi service which have similar functionalities, rather than having the costly small cells. A moderate solution is applying WSDN and NFV on small cell management, which can be done remotely and with much lower OPEX [20]. The SDN controller will configure the packet forwarding schemes for a group of small cells in similar operating conditions, thus reducing the manual configuration time and resource.

8. Big data and mobile cloud computing

Facing higher network throughput, the data storage techniques will also experience evolution in 5G. Traditional data storage on local devices will no longer be capable to handle the exponentially increasing data cache especially when the user wants to download huge files or programs, or even stream HD-videos. In recent years the cloud storage has become popular because of the convenient and on demand service. Users can upload data to cloud servers through the Internet and therefore save the local storage on their devices. In 5G networks, the cloud storage will become an inevitable part, and the mobile cloud computing will become the major method for data computing on a higher level.

Additionally, the sheer amount of data generated by mobile devices and networks also intensify the importance of big data analytics. Currently, most process of data analytics is performed on groups of physical machines, using the platform named Apache Hadoop (i.e. in-house Hadoop). The platform always requires infrastructure to be installed and maintained manually, and also lacks scalability if the already-installed platform cannot handle higher demands, which leads the reconfiguration process to become costly and inefficient. In contrast to the in-house Hadoop, the big data in cloud offers advantages on cost, efficiency, and scalability. The big data analytics will be even more efficient and robust, through the advantages of software-defined switching and control brought by WSDN.

The enhancements brought by big data and mobile cloud computing to 5G are flexibility and high efficiency in data management. Since more and more users choose to share files on cloud storage servers, mobile cloud computing offers data management on the cloud servers. On the contrary to cellular system where the downlink plays a major role on system performance analysis, mobile cloud computing considers the uplink quality and the backhaul network. The mobile cloud computing requires a stable connection to the cloud server, which in turn sets high requirements for backhaul data gateway to provide traffic-smart routes [127].

8.1. Smart data management

One of the promising techniques of mobile cloud computing is computation offloading, which is a process that allows the remote server, or cloud, to share part of the computational responsibility that is traditionally assumed to be fully accomplished by the mobile device. Since more and more sophisticated applications are installed and running on mobile devices with limited battery life and storage [128], the sheer amount of computation will significantly drain the battery which will be unfavorable to users. Especially for those smart devices with small size and weight, such as smart watches and bracelets, the demands for high speed data computation, large amount of local storage, and decent standby time, will be some of the most competitive features that attract market. Therefore, one benefit of mobile cloud computing is to help increase the energy efficiency of the mobile device [129,130].

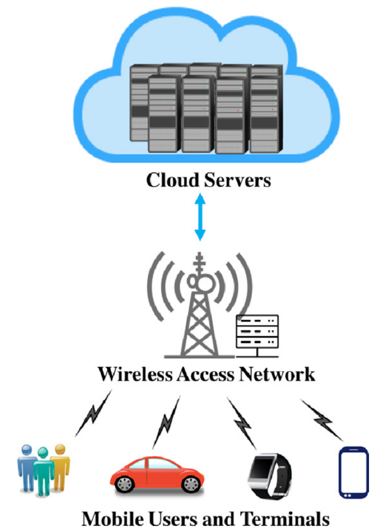

Fig. 9. The architecture of mobile cloud computing, in which mobile users and terminals can offload local data computation using the remote cloud servers.

As shown in Fig. 9, the mobile cloud architecture is fairly straightforward and can be divided into three major parts, namely the cloud server, the mobile access point, and the mobile device. Similar as the cellular network, the mobile device will connect to the telecommunication network via the base station or access point. The telecommunication network is further connected to the Internet, which allows the mobile users to access the cloud server through Internet connection [131].

8.2. Mobile edge computing

Another important technique being developed for 5G cloud computing is the mobile edge computing (MEC), which provides the advantages of low latency and high bandwidth at the edge of RAN [132]. The initial idea is proposed by the European Telecommunications Standards Institute (ESTI), which utilizes the distributed computing environment in close proximity to UEs in order to reduce network traffic congestion and therefore achieve faster response time. The applications of MEC can be found in real-time content delivery, big data analytics where time-efficiency is critical.

8.3. Open problems

This subsection discusses some open problems related to big data and mobile cloud computing. The imbalance of network resources is one big challenge for mobile cloud computing, and the importance of addressing serious cloud security problems should be emphasized. Since current cloud computing is facing transmission delays in wide area networks, we argue that some new routing schemes should be proposed to solve the open problem.

Limited network resources: The challenges of mobile cloud computing include the current limited network resources and the concern for data security. Since this technique has been available for only a few years, there are many open issues need to be addressed. Even though the network resources should be theoretically equally shared by users and service providers, due to different network standards around the world, users from different service providers might encounter difficulties in changing networks when they travel to other countries. Another critical issue is the data security. With more advanced hacking technologies threatening the Internet around the world, more research focus has been put on network security [127,131]. Current cloud data computing has not yet employed robust security defense techniques [133]. This concern must be addressed before mobile cloud computing becomes a norm in 5G network in order to provide a safe environment for users to enjoy the benefits of 5G.

Cloud security: In order to protect the data in cloud storage from being hacked, the issue of cloud security and privacy should be fully addressed in the 5G wireless systems. In addition to research on anti-hacking technologies, governments should establish regulations on cloud security and take appropriate actions against cyber crimes [134]. For instance, the European Union (EU) has set up a specific agency, the European Network and Information Security Agency (ENISA), to assess the network environment and provide recommendations. Global initiatives should be encouraged to protect the cloud security against attackers, as the network has been shared almost by anyone from anywhere.

Transmission delays: Another limitation of current mobile cloud computing is the delay caused by the data transmission in wide area network (WAN). The transmission delay results from signaling and traffic congestion in a highly crowded network architecture. In order to tackle this limit, the workload of WAN should be separated and released to smaller networks with reduced area of coverage. One proposed method is to merge the idea of small cells with cloud computing [128], which utilizes the small cell base station to gain access to the cloud server, in order to provide faster computation over the cloud with reduced network latency. This mobile computation offloading framework is operated on the joint basis of communication and computing, which brings more promising enabling solutions to the 5G mobile cloud computing techniques.

Worldwide standardization: Since the Internet is a public property shared by users worldwide, the mobile cloud computing will also be adopted to facilitate boundary-less cooperation among different countries, as predicted in the 5G networks. In the US, the National Institute of Standards and Technology (NIST) has initiated the cloud computing program in order to regulate the cloud computing practices, which can support interoperability, portability, and security requirements [135]. The International Telecommunication Union (ITU) also published several recommendations and technical reports in the efforts to standardize and cultivate an ecosystem in the cloud computing [136]. However, the new requirements and technologies emerging in the 5G networks will eventually drive the evolution of standardization in mobile cloud computing.

9. Internet of Things

Internet of Things (IoT) is the key concept that is changing many perspectives in 5G cellular networks landscape. IoT is the network of everyday physical objects, vehicles, appliances, devices, buildings, etc. Some examples of the devices that would be part of the IoT are microwave ovens, washing machines, wearable smart watches, and health monitors. These devices sense some information and it is passed on to a remote server, mostly through the Internet. The server can also remotely issue commands to control the device. The information collected at the server is then processed to gain knowledge about the underlying process. This knowledge is used to build smarter systems like smart homes, smart cities, intelligent transport systems, healthcare systems, etc [137]. With applications in several diverse areas, the advent of IoT will enable a multitude of devices to be connected to the Internet. This paradigm shift from the concept of connected people to connected things predicts that the number of devices in network increase to 50 billion by 2020 [138]. Because of the sheer volume of IoT devices expected to be connected and the plethora of applications they could be used in, IoT devices play a significant part in design of 5G systems and the several parts of the 5G network are being designed with the requirements of IoT in mind. Although some of the technologies discussed in other sections would enable IoT, we present them here separately because of the huge importance of IoT devices and the role they play in development of 5G systems.

9.1. Scalability

The widespread application of IoT and the cheap availability of sensor devices makes it easier and cheaper to deploy many sensors to monitor an area or to perform a task. This is expected to cause an unprecedented surge in the number of devices connected and the density of devices. This outburst will pose several challenges in terms of providing reliable coverage to all the devices. As the number of devices increase and the size of the data they transmit decreases, simplified access and core network techniques will be required to accommodate all the devices and reduce the signaling and connection setup overheads for each packet. Apart from increasing the number of devices a BS needs to support, increase in the number of devices in the network also brings the problem of increased number of handoffs that the BS should be able to handle. This could be solved by adoption of technologies like SDN that would enable smoother handoffs [16].

9.2. Intelligent processing and storage

Although the sensor devices in IoT collect data, it is processed and converted to actionable knowledge and intelligence only at the backend servers. With the data being collected from IoT devices becoming increasingly complex, cloud servers are often deployed to collect the data and process it to get insights. Compared to standard cloud providers, IoT cloud services have different requirements in terms of communication protocols, software layers, and provisioning models [139]. Although elasticity by providing on-demand computational power is one of the basic features of cloud computing, it should be made scalable to the extent of accommodating the enormous number of devices IoT would bring in. Thus, the service level agreements (SLAs) provided by the cloud provider will be critical for time sensitive applications like reporting equipment failure, fire, and other emergency reporting. Most of the existing cloud services come with proprietary interfaces and this will require customization of all IoT devices [140]. This issue is aggravated when multiple service providers occupy different parts of processing in the same IoT network.

Another approach to ease the computational load at the cloud servers is to move the computing service towards the edge of the network. Instead of having a centralized cloud computing facility at some data center, the cloud servers and storage are placed near the end users, typically in the same building. The physical proximity of the computing server to the nodes decreases the delay due to communication. Fog computing also provides location awareness and geographic distribution features to IoT services [141].

Apart from processing the data, storing of the data generated by IoT devices will need smart and efficient solutions. These types of huge data which cannot be handled by ordinary computing hardware are called big data. As discussed in Section 8, although there are big data solutions like Apache Hadoop and SciDB commercially available, they cannot scale enough to meet the needs of IoT [12]. These technologies fall short particularly for applications where real time actions need to be taken. An interesting approach to address this problem is to avoid sending redundant data to the server. Several works on sensor networks have noted that the data collected by sensors is highly redundant [142]. Reducing data redundancy could be achieved using some dimensionality reduction algorithms like principal component analysis (PCA) or by using a smart gateway for the sensor network, which aggregates the data collected by the network before sending to the remote server [143]. This would reduce the transmission load on the cellular network as well as the processing and storage load on the server.

9.3. Network slicing

The 5G networks are expected to serve not only mobile phones, but also more diverse types of devices with different system requirements and functionalities, such as smartwatches, tablets, massive IoT devices, mission-critical IoT devices, etc. In order to provide efficient services to each type of devices, the concept of "network slicing" is being discussed, which offers the freedom for network operators to slice the RAN into multiple end-to-end virtualized network [144]. Each of the sliced networks contains its own radio access, transport, and core network dedicated to service a single type of device. These dedicated slices provide independent network sources so there is no interference among other slices. Network slicing offers a great boost for mission-critical IoT services, as well as achieves a dynamic network resource allocation for different traffic scenarios.

9.4. WSDN and NFV based IoT

With the WSDN and NFV offering dynamic and scalable network configuration capabilities, they are seen as the key technologies that would enable accommodation of 50 billion IoT devices which are expected to be a key part of 5G cellular network by 2020 $\sim$ 2030. WSDN will be able to intelligently route the huge traffic generated by IoT devices through underutilized network resources, providing an increase in performance of the network while spreading the load across several devices in the network. The level of abstraction provided by WSDN will also enable the IoT applications to access the data and control the devices without the knowledge of underlying infrastructure. This makes deployment and management of IoT devices simpler. The WSDN and NFV based IoT could configure and manage the devices dynamically to optimize required performance parameters such as reducing power consumption of devices, reducing the packet delay, etc. The centralized control of data flow offered by WSDN/NFV also enables and simplifies data aggregation.

Another advantage of deploying WSDN and NFV architecture for the IoT network is the improved security. The security of data transmitted is particularly important for mission-critical applications. Since WSDN and NFV centralize the network management and control, they will be able to efficiently handle the network security at different layers. Further, as WSDN and NFV profile each flow differently, they can handle the traffic from different IoT services which have different security profiles. This would ease the unnecessary processing of less secure applications.

9.5. Open problems

Since IoT is one of the crucial application scenarios of 5G networks, supporting them is of paramount importance. The open challenges that are yet to be solved to enable IoT services are discussed here.

Low power consumption: Most of the IoT devices will be standalone sensors equipped with transceivers. These devices are battery powered and they die off when the battery is drained. Hence, the power consumed by the IoT devices has to be kept minimal, in order to extend their lifetime. This makes power consumption at device one of the biggest challenges for supporting IoT devices. Another interesting way to increase the lifetime of devices is by harvesting several forms of energy available in the environment in which the devices are deployed and using it to power the device. For example, sensors deployed over open fields could use solar power, sensors in industrial automation could generate power from vibrations. These new energy harvesting techniques provide a sustainable solution to the power consumption problem. The energy consumption of the device could also be decreased by optimizing the software running on them.

Signaling and control overheads: Most of the IoT devices send data to the server periodically or only when an event occurs and go to sleep after transmission. The devices are not active and connected to the network all the time. Moreover, the data sent by the IoT devices is much shorter compared to the packets sent by conventional applications such as video streaming. So the signaling and control overhead of the frames transmitted becomes significant compared to the size of data in the frame. This makes the overhead in signaling and control signals a major part of the total consumed power[145]. This overhead was reduced by using techniques such as fast transition to idle as proposed in 3GPP Release 11. Approaches like connection-less communication that fits within the 3GPP standards [146] and the use of WSDN/NFV based solutions to reduce the signaling and control overheads should be explored further.

Although the number of IoT devices is going to be mammoth, these devices are not connected to the network at all times. The devices send data to the server once every time period or when an event occurs and go back to sleep again. This is a significant transition from the mode of operation of traditional mobile devices which are connected to the BS at all times. Consequently, the IoT device has to synchronize with the BS every time it switches on before it starts transmitting any useful data. Further, the size of data sent each time the device is turned on is small and hence the power consumed by the synchronization step is significant compared to the power consumed for actual data transmission. So, the radio access technique to be used in 5G cellular networks should be able to support devices with no synchronization or at least support loose synchronization.

Low latency: Several applications that use IoT networks are mission-critical and require very low latency. Applications like autonomous cars, drones, eHealth monitoring, industrial process monitoring are delay-sensitive and require ultra-low latency on the order of 1 ms. This very low latency could be achieved by a combination of several technologies such as new radio access techniques and WSDN.

Interoperability and standardization Due to the diverse range of applications, there are various types of IoT devices manufactured by several vendors and having different capabilities and requirements. An IoT device could range from a simple temperature sensor, which monitors temperature and sending it to a server every hour, to a sensor monitoring critical industrial processes. Among the IoT devices available in market today, most of them deploy their own standards and interfaces for communicating with other nodes and servers. This creates incompatibilities when sensors from different manufacturers are used simultaneously. Further, the interfaces at the cloud servers that collect information also need to be standardized to provide interoperability between cloud services.

10. Device-to-device communications

In conventional cellular networks, the majority of communication occurs between BSs and devices. Even though two users are in their range of direct communication, the link has to be establish through the BS, which is not efficient for real-time services requiring high data rates and low latency. In order to enhance spectral efficiency, the idea of direct device communications was proposed in [147] to initially create multi-hop relays among devices in cellular networks. Since then, several D2D communication use cases have been widely investigated as part of the LTE-A networks, in both licensed cellular spectrum and unlicensed bands. Details about D2D communication classification can be found in a comprehensive survey in [148]. The 5G networks are expected to continue supporting applications that require devices to set up direct communication links with their peers [149]. The applications include file sharing, gaming, and social networking [150].

The D2D communications will improve QoS of the cellular network as relaying can help increase the data rates of the users at cell edge. The other scenario where the devices communicate directly with each other will alleviate the data traffic between nearby devices and the BS. This can offload the BS capacity and simultaneously provide a better user experience. Meanwhile, D2D communications bring in three types of gains – proximity gain due to the reduced distance between transmitter and receiver which allows high data rates at low power consumption; reuse gain provided by the reuse of spectral resources at other locations given that the communication range is shrunk; hop gain obtained by using a single link in D2D while the conventional communication through BS requires both uplink and downlink communications [151].

In the application level, D2D communications will unlock several innovative applications and services. One important application is in the area of public safety and security [152]. In scenarios such as accidents in long tunnels, landslides, and earthquakes, where cellular networks are not available or collapsed, D2D communications can be readily used for rescue and relief operations. D2D communications will be greatly helpful in machine-type communications, especially vehicle-to-vehicle communications where it can be used to get information of accidents and other alerts about routes, such as route finding in self-driving cars. For example, the 3GPP Release 12 added D2D communications to proximity services (ProSe) in order to enhance public safety [153]. D2D communications can also be used to connect several devices like smartwatches with smartphones. Although other technologies such as Bluetooth also connect devices to mobile phones, a unified protocol would enable lower cost and energy consumption of devices. On the perspective of user mobility, D2D communications offer a more reliable environment, where Bluetooth is not designed for high mobility environment. D2D communications will also unlock new types of social networking applications where nearby devices can establish direct connections.

10.1. Types of D2D communication

Several works incorporate D2D communications in the architecture of 5G cellular networks. A two-tier 5G cellular network with the macrocell and device tiers is proposed in [149]. A device tier is introduced in addition to the conventional macrocell tier to enable and manage the D2D communications. Depending on the level of control the BS has on the D2D communications and whether another device is used as a relay, D2D communications can be classified into four types as described below and visually illustrated in Fig. 10.

Device relaying with BS controlled link establishment: A device with poor coverage communicates with the BS by using the other devices to relay its information to the base station. The relay link establishment is partially or fully controlled by the BS. Since the BS coordinates the relay link establishment, the device should still be in the coverage range of the BS. A scenario where this will be utilized is when the device is within the coverage area of BS but the channel quality is poor that higher data rates cannot be supported. But the BS can establish connections with its base data rate and the device can achieve high data rates through the established relay link. This type of relaying helps the devices to achieve higher QoS and improve the battery life.

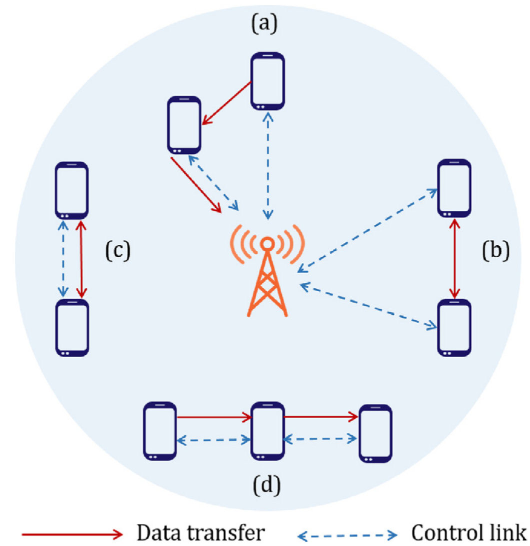

Fig. 10. Device-to-device communication scenarios: (a) Device relaying with BS controlled link establishment; (b) direct D2D communication with BS controlled link establishment; (c) direct D2D communication with device controlled link establishment; and (d) device relaying with device controlled link establishment.

Direct D2D communication with BS controlled link establishment: The transmission and reception devices are within the transmission range of each other and they communicate with each other without the help of BS. However, the BS helps with link establishment between the devices. This will enable several novel social networking applications. This scenario would be advantageous for file sharing over short distances as it can avoid the traffic through the BS and in turn, offloading the load from BS. It would also provide higher data rates as the rate is no longer limited by the connectivity and capacity of the link to the BS.

Direct D2D communication with device controlled link establishment: In this scenario, the source and destination devices are within the communication range of each other and they communicate directly with each other without the help of BS. It should be emphasized here that the BS does not play any role in this scenario. Since the BS is not involved in the setup, this scenario will also enable communications between devices when they are out of the coverage area. This will come in handy during emergencies when the device cannot contact the BS.

Device relaying with device controlled link establishment: The last scenario is where the source and destination devices are not within the transmission range of each other. Under this circumstance, they establish a relay link between each other without the help of BS. Since there is no involvement of BS, the source and destination devices are responsible for the link establishment.

10.2. WSDN and NFV for D2D

The versatile and flexible communication framework provided by D2D can be exploited using WSDN and NFV. The flexibility and configurability that WSDN and NFV provide could be easily leveraged to deploy a new paradigm of communication like D2D.

However, the integration of D2D communication with WSDN and NFV concepts is not explored in detail and this leaves many open problems that should be addressed. One such problem is that WSDN requires the controller to have a view of the entire network, while in many D2D scenarios the decisions are taken locally by the devices. This inherent difference in the requirement of knowledge poses a problem in integrating WSDN and NFV. A hierarchical architecture of mobile cloud formation initiated by the user device is proposed in [154]. The mobile clouds proposed are local clusters of devices that can communicate with the cloud head. The hierarchical architecture enables the distribution of controller functionalities using local and central SDN controllers. The cloud head conducts the formation and management of the devices locally while the central SDN controller supervises the entire network and performs resource allocation and routing functionalities. Due to the strict hierarchical architecture, the scenarios of D2D communications that the architecture proposed in [154] will support are limited. The fact that the WSDN controller has a global view of the network and resource utilization should be exploited intelligently to modify the routing and flow tables at the BS or the user device to enable D2D communication whenever it is feasible.

10.3. Vehicle-to-vehicle communications

Two important use cases of D2D communications are vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications, which can be generalized as the V2X communications. The vehicles communicating with each other as well as with the infrastructure will aid development of several technologies like autonomous cars, automated traffic violation monitoring, etc. The V2X communications have several distinct challenges due to specific deployed conditions. For example, the speed of the vehicles would be several tens of kilometers per hour and they should be able to communicate with the vehicles on the opposite direction during the short time when they cross each other [155]. The high density of traffic movement also increases the node density of vehicles. A detailed list of QoS requirements for the V2X communication scenarios is presented in [156].

Currently two major technologies are being considered for V2X communications, namely IEEE 802.11p and the aforementioned LTE ProSe. Since IEEE 802.11p uses a contention-based medium access technique, its throughput and delay performance degrade at high user density. Furthermore, the range of 802.11p is also limited, requiring more hops, higher protocol complexity, and longer overhead [157]. The LTE ProSe is designed for public safety and consumer applications specifically. Thus, it does not fit the V2X requirements of high mobility, differentiated QoS, and broadcast communications. There are more open challenges in D2D communications that 5G is expected to address to enhance system performance.

10.4. Open problems

The introduction of D2D communication brings in a lot of open problems. The primary challenge is discovering the peer devices and identifying the type of services they support, especially in scenarios where connection is to be established without assistance from the BS. Peer discovery using beacon signals will be time and energy consuming. Another challenge is to make decisions on which service to use among direct D2D communication. Analyzing the QoS that could be achieved using each of these methods and then deciding which mode to use will again not be energy efficient.

Resource management for autonomous operation: In the autonomous operation mode for D2D communications, devices are operated independently, without the synchronization signal from other devices or BS. This operation requires strict resource coordination in order to prevent undesirable interference and unexpected collision. One solution is that devices intended to transmit or receive data can broadcast a busy signal to surrounding devices. The challenge of limited spectrum resources for autonomous operation can be addressed by the future utilization of higher frequency bands.

Spectrum sharing with cellular systems: Currently the D2D communication is operated underlaying the cellular networks, which will inevitably cause interference to the existing cellular networks. A better spectrum sharing and resource allocation scheme is needed to optimize the cooperation between the two types of services. A dynamic power control scheme should also be applied on D2D communication devices. The research in [158] proposed an admission control method on the basis of QoS to realize spectrum resource allocation.

Device pairing and link setup: The discovery of device pair and link setup also bring additional challenges to HetNet operation [159]. In current LTE systems, there are some network-assisted device discovery algorithms discussed. These algorithms either use exhaustive search to locate the pairing device which is computationally costly (i.e. the Kuhn-Munkres algorithm), or only one side of the communication is active in pairing selection [160]. The high computation load drains device battery fast, making it unfeasible for long packet transmission or large file sharing. Therefore some new device pairing and link setup algorithms are proposed. One of methods described in [161] includes two steps in pairing strategy. The first step is to establish a group of devices that are considered admissible candidates, which are active and nearby devices, and the second step is the process of resource matching and power allocation. This method is shown to reduce the computation complexity while increasing the spectral efficiency.

Security and privacy: Since the devices are connected directly and user data is being routed via devices, security and privacy must be maintained. This can be accomplished by assigning a set of trustworthy device group in which the devices are authorized to communicate through. The devices within a trustworthy group can set some encryption between them to prevent other users from gaining access to the data being shared.

Interference management: In traditional cellular networks, the BSs handle the task of interference management among users. Since there is no central controller in D2D scenario, interference management will be an issue. The interference caused can be of two types: with the other D2D communications as well as an ongoing communication with the BS. Approaches like resource pooling [162], admission control and power allocation [158], relay selection and game theoretic techniques like non-cooperative game and bargaining can be used to moderate the interference [163,164].

Resource control: Besides the technical challenges given above, another practical and administrative problem that should be considered is control and pricing of D2D services. For example, sufficient incentive must be provided for the devices that relay other device's data. Otherwise, the relaying device can refuse to do so as it would be a waste of its resources, such as the device's battery and memory.

3-D channel model with spatial consistency: Currently, most wireless communication systems are deploying channel models that are based on the assumption of stationary channel, which shows advantage on time complexity, reliability, and good estimation to channels with fixed links. However, in the scenario of D2D communications with both terminals in movement, there has been an increasing demand on characterizing the rapid time-varying nonstationary channel to increase channel modeling accuracy for the 5G wireless communication systems. In particular, the METIS summarized a number of deficiencies in current channel models, in which the spatial inconsistency of current drop-based channel models requires novel modeling approach.

Some research groups also conducted research and measurements on the time-varying dual-mobility channel. The method of ray tracing is widely used to characterize the ray propagation in a dynamic environment [165–168]. However, the granularity of ray tracing results will be affected greatly by the database of the specific environment and computation capability of workstations. Additionally, the variation of propagation paths and the transition between LOS and NLOS propagation are often neglected. A 3-D time-varying channel model is proposed for the 5G dual-mobility channel which can address the problem of spatial inconsistency [169].

11. Green communications

The number of devices that would be connected to the network is expected to increase 100 times and the data volume is expected to increase over 1000 times in the next decade. While achieving these benchmarks is in itself a challenge, we should also meet these requirements in an affordable and sustainable way. Although the contribution of mobile communications to the CO_2 footprint is less than a percent now, we should try to reduce it further. Operators are already facing that the power bills have become a significant part of the operating expenditure. So, lowering the energy consumption and moving towards green communication alternatives are not only important from an environmental perspective, they are also significant from an economic perspective.

11.1. Network planning and deployment

Energy efficiency of the cellular network can be improved by adopting several network deployment strategies [170]. These strategies can be base station cooperation, different topologies of cells, and distributed antenna systems. Significant improvements in the energy efficiency can also be achieved in heterogeneous cellular networks by using small cells. Currently, small cell base stations are placed at locations to enhance the network capacity as well as keeping the cost of infrastructure and deployment low. Therefore, by choosing the location of the microcells and relays optimally within the range of a macro cell, they can significantly offload the macrocell and produce energy savings while providing a better coverage [171]. These optimizations are especially efficient in areas where extremely high capacity and data rates are needed like offices, shopping malls, subway stations, etc., where the user density is large. Since most of these places are indoor, indoor access points should be deployed so that energy wastage due to wall penetration can be avoided.

Another complication in cellular networks is the high variability of network traffic with time due to the patterns in which everyone tends to access the network at the same time. This causes huge difference between the average and peak hour cellular traffic. Reports indicate that this difference between peak and average traffic is increasing and the peak rate of traffic is expected to grow much faster than the rate of growth of average traffic [13]. This makes network operators deploy more base stations to support the peak hour traffic. This causes unnecessary power consumption and waste when it is not needed. This could be reduced by systematically switching off some of the base stations that are not required to be operating. Based on the traffic pattern, analytical models can be developed to identify optimal BS switch off times [172]. It was also observed that the variation of traffic demands among different network operators serving the same geographical area is significantly different [173]. Hence, the network infrastructure of several network operators could be shared among them to dramatically reduce energy consumption while providing better coverage and capacity. A study of European cellular network operators concluded that a reduction in energy consumption by 35–60% could be achieved by such sharing of infrastructure between network operators [174].

11.2. Harvesting renewable energy resources

Another approach to achieve green communication networks is to harvest the renewable energy resources like solar, wind, vibrations at the BS and use them for its operation, reducing or even eliminating the use of conventional power consumption. A cognitive radio network that not just utilizes the spectral holes for transmission, but also minimizes the energy consumption by opportunistically harvesting energy from ambient sources is presented in [175]. In addition to harvesting natural sources of energy like solar, wind, and vibrations, they also propose to harvest synthetic sources of energy like microwave power transfer. They provide a comprehensive study of the solar energy that can be harvested and whether a BS operating on solar power can be made sustainable. They conclude that by storing the surplus energy received in afternoon and used in evenings, continuous self-sustainability can be achieved during periods of abundant sunlight and even in winter, drawing power from the grid can be avoided for three to six hours a day. Another work [176] incorporated both solar and wind power to power the base station. They also propose to use fuel cell based energy sources for deployment in urban areas where deploying solar and/or wind powered base stations will not be feasible. Besides providing a reduction in energy consumption, use of renewable energy resources will also enable setting up self-sustainable BS in remote locations where power is not available, thereby improving the coverage.

11.3. User-centric design

The traditional cell based coverage of a geographical area does not provide the required elasticity to accommodate the diverse requirements of 5G cellular networks. To overcome this, it is proposed to get rid of BS centric design of cells and move to a user-centric concept of “no more cells” [145]. They propose to retire the cell based design to a user-centric design with amorphous cells, decoupled signaling and data and decoupled downlink and uplink. This enables the small cells within the macrocell to be turned off when they have no traffic. This cannot be done in the traditional cell based network. The decoupling of uplink and downlink will enable more efficient resource allocation. This enables a user to send uplink data through one cell and get downlink data from another cell when the cells are heavily loaded in downlink and uplink respectively.

With the use of SDN that provides separation of control and data plane, this kind of architecture becomes easier to implement. This could be exploited to provide each user with a single radio resource control connection with macro BS and dual data connection with both macro BS and micro BS [177]. The collaboration between macro and micro BS is utilized to minimize control channel overhead and cell-specific reference signals in order to achieve a pure data carrier for small cells. This architecture was shown to provide 90% energy efficiency gain while still achieving a throughput improvement of more than 17%.

11.4. Smaller frame overhead

It is also noted in [145] that the traditional cellular networks are designed for conventional streaming applications. The diverse traffic requirements have exposed the inefficiencies of conventional cellular networks and brought in new challenges. There are several applications that use small sized persistent bursty traffic like instant messaging services. Other types of applications like MTC and D2D communications also send small bursts of data [10]. These applications generate smaller packets that are sent at regular intervals. They cause the mobile device to switch between idle and connected states and it may consume much power. The size of these packets is also small so that this cycle happens very frequently. Another problem with this type of traffic is that these packets are very small and hence the signaling overhead in header data might be significant compared to the actual data size. To combat this scenario, a lightweight radio resource connection state without maintenance overhead for handover and channel status feedback has been introduced in 3GPP Release 11. There are also many modifications proposed to the random access procedures to handle this type of traffic like implementing predetermined dedicated preambles or sending the data in the uplink resource allocated for radio resource control requests. Some contention based methods have also been proposed [178] where the devices directly send packets in a contention-based manner.

11.5. Green metrics

In order to evaluate the energy efficiency of wireless communication systems, a group of green metrics is necessary to be established for the 5G networks. The green metrics are useful in research and development in energy efficient components, standardization of energy efficient equipment manufacturing, and quantified assessment on system performance. Therefore, based on the structure of a wireless communication network, the green metrics can be measured on three levels, namely, component, equipment, and system levels [179].

According to current signal processing architecture in wireless communication networks, the electronic components normally are filters, power amplifiers, A/D converters, and antennas at the RF side. So on the component level, the green metrics can be characterized as the gain of the RF component, radiated efficiency of antennas, or power efficiency in power amplifiers. The measurements for the green metrics in the above components are straightforward, and are normally showed in the component's specification lists.

As to the level of equipment, the performance of green metrics cannot be easily measured and should be evaluated in different environments. Although the equipment consists of numerous electronic components, the green metrics of the equipment are not simple linear additions of those parameters in each component. In the standard operation mode, the equipment should consume less amount of energy than in busy operation mode, while in idle operation mode the energy consumption should be the least. In Europe, the standardized metric in this case is the Energy Consumption Rating (ECR) which can be expressed as the actual energy consumption divided by the effective system throughput, as defined by the ETSI [180].

In current wireless networks, green metrics are evaluated in cellular networks, wireless local area networks, satellite systems, and ad hoc networks [179]. However, since in the 5G visions, ultra-dense networks will be deployed to provide more layers of networks with wider coverage, and new frequency bands will be licensed for cellular networks, there should be another novel set of green metrics for 5G wireless networks. The 5G green metrics should be defined by network layers. Two examples are the signal processing energy efficiency in the physical layer and the modulation and user association energy consumption in the medium access control layers. More research efforts should be encouraged to expedite the standardization of 5G green metrics [181].

11.6. Open problems

In the above discussions of green communications, there are some tradeoffs and challenges also needed to be resolved. In this subsection, we present some open problems in green communications and give potential solutions to them.

Power control in green communications: Although there are several green communication solutions proposed for 5G cellular networks from the point of energy efficiency, it should be verified that achieving energy efficiency does not degrade the performance of networks in terms of data rate and other requirements. For example, in the case of multi-tiered network deployments, the user is not attached to the BS with maximum power. This causes the BS to experience higher interference from other BSs which would affect the received signal quality and hence the data rate that the user could achieve. The feasibility of implementation of some of solutions like harvesting solar and wind energy in dense localities and places where sunlight is not available throughout the year should be studied further.

Energy efficient hardware: Previous and current transceiver equipment and hardware in wireless communication networks are designed to achieve good performance in data throughput and reliability. However, such hardware is normally energy costly. In 5G systems, operators and equipment manufacturers should weigh heavier on the energy efficiency when designing and testing network equipment. The research in [182] examined the energy consumption in both office equipment and portable devices such as laptops and mobile phones. Based on the statistics of global energy consumption, the recommendations are made in perspectives of power management, battery life management, and utilization on energy harvesting to alleviate the issue. A globally adoptable energy saving recommendation can be an alternative to encourage energy saving activities around the world.

Energy efficient network architecture: The energy saving network architecture will be enabled by multiple technologies, which are also discussed in this paper, namely, massive MIMO, ultra-densification, SDN and NFV, D2D communications, and mobile cloud computing. The technology of massive MIMO can reach energy efficiency by the utilization of hundreds of antenna elements for high gains. By separating the control and data planes in BS, idle UEs will not waste energy on keeping connection with BS. The wireless software-defined network architecture will save configuration time and energy in traditional hardware. The approach of enabling direct device connection can further improve energy efficiency in the entire network. Furthermore, a series of reconfigurable and energy-scalable radio network solutions were surveyed in [181], which can also be of important reference to the 5G system design.

Battery technology enhancements: The novel battery technologies will also bring revolutionary changes on the next generation of energy efficient communication networks. Recent electrochemical research on novel energy sources has found that sugar can actually be an excellent energy provider, which offers an order of magnitude more energy than the same weight lithium-ion battery in smartphones [183]. The theory behind the sugar battery is based on using enzymes to extract the energy from sugar, which is similar to human digesting sugar to absorb energy. This prototype on sugar battery is promising in the utilization on sensors and other small devices which can expand their lifespan without adding extra weight. Additionally, unlike the lithium-ion battery which is a limited source and requires professional recycling procedures, the sugar battery is easy to refill and safe to use.

In addition, some plants have been discovered to be equipped with the capabilities of a solar panel. A group of researchers in the University of Cambridge found the Photo Microbial Fuel Cells (Photo-MFCs), which is a type of moss can transfer solar power into electric power [184]. Although the Photo-MFCs are still in the early stage of research and cannot harness energy with acceptable efficiency, the trend for the 5G green communications is clear ahead.

12. Radio access techniques

The different requirements explained in Section 2 urge us to rethink about the radio access technique design. The gigabit speed demand of 5G networks motivates that the underlying radio access technique should also be capable of supporting higher data rates. Since spectrum is a scarce and costly resource, spectral efficiency is a key factor of the radio access technique that would enable the gigabit speeds. Several applications that are envisioned in 5G network, like tactile Internet, require a very low latency of the order of 1 ms. This puts a constraint on the perspective of latency of the radio access technique, so that the lower latency required at higher layers can be attained. The other applications like IoT have scenarios where the devices are not connected to the BS at all times. The power constraints of these devices prevent the devices from having full synchronization with the BS. Hence, the multiple access technique to be adopted should be able to support loose or more preferably no synchronization. As in other wireless systems, interference is a concern while adopting a multiple access technique. The technique we adopt should have low out-of-band (OOB) emissions so that the guard band between other carriers can be minimal. Having a lower interference will require less guard bandwidth and ultimately increase the effective spectral efficiency. The multiple access technique should also efficiently use the multiple antennas that will be present in most of the devices. Hence compatibility with MIMO technology should also be considered while choosing a multiple access technique. Finally, the radio access technique must also be power efficient, so that it does not drain out the power of low power devices.

The major innovations in the radio access techniques can be classified into the introduction of full duplex communications, the choice of multiple access technique, modulation, and channel coding. These innovative improvements on radio access techniques are explained here.

12.1. Full duplex communication

Full duplex communication systems enable the simultaneous transmission and reception of signals in the same frequency bands. This is a significant departure from the traditional half duplex communications that use Time Division Duplexing (TDD) or Frequency Division Duplexing (FDD) in order to achieve duplex communication. The RAN technology of 4G LTE network, Evolved UMTS Terrestrial Radio Access (E-UTRA) supports both TDD and FDD modes of operation [185]. TDD and FDD modes suffer from performance and efficiency issues like out-of-date channel state information, inflexible bandwidth allocation, and requirements of guard intervals and frequency bands [186]. Although full duplex systems overcome these inefficiencies, it was considered infeasible to implement a practical full duplex system due to the very high self-interference from the transmitter blinding the intended received signal. Recent developments in self-interference cancellation techniques has enabled the practical realization of full duplex systems.

The self-interference cancellation techniques developed propose to apply interference suppression at various stages of the receiver chain [187]. Propagation domain interference suppression tries to avoid the input of the RF amplifier being overwhelmed by self-interference, analog domain interference cancellation prevents the analog-to-digital converter from being overwhelmed by self-interference by applying interference cancellation while digital domain interference cancellation attempts to cancel the residue self-interference arising due to the nonlinearities of the analog-to-digital converter and oscillator phase noise. Some works proposed employ a combination of these techniques to achieve superior self-interference cancellation.

The full duplex communication can be used in a couple of ways to increase the capacity of the 5G cellular networks [187]. In the first scenario, the user equipment sends its data to a BS or potentially another device in D2D communication while the BS sends its own data to the same UE at the same time and the UE and BS can decode the data intended for them while simultaneously transmitting its data in the same frequency band. This mode of operation is called bidirectional transmission. The other scenario is the unidirectional transmission similar to the full duplex relaying [188]. Here, the first node transmits its data to the second node which in turn transmits its own data to the third node. Although the full duplex communications seem to double the spectral efficiency that can be realized, the practical throughput gains that are achieved is about 30–40% [186]. This clearly shows that there is significant scope for improvement of full duplex technologies.

Although self-interference cancellation is the primary problem in realizing a full duplex communication, it is not the only issue that needs to be addressed. The whole of the current network stack, was designed with half-duplex communication in mind. This requires us to rethink and redesign these functionalities for full duplex communication system. For example, water filling algorithm can be used for power allocation to maximize the spectral efficiency of half duplex system but, it is not the optimal scheme for full duplex system as it does not take into account the self-interference. The MAC protocols used in earlier systems like Carrier Sense Multiple Access (CSMA) inherently assumes a half duplex system and it cannot be directly employed in full duplex system. This requires development of novel MAC protocols like the protocol proposed in [187]. Similar redesign of higher layer protocols for application in full duplex systems also needs to be explored.

12.2. Multiple access techniques

E-UTRA employs Orthogonal Frequency Division Multiple Access (OFDMA) for the downlink channel and Single Carrier - Frequency Division Multiple Access (SC-FDMA) for the uplink channel to overcome the high peak-to-average power ratio and thereby conserver power [185]. Although these techniques work well for 4G LTE networks, they cause inefficiencies in several scenarios foreseen for 5G networks. This requires us to redesign the multiple-access technique for 5G networks. The several multiple access techniques that are currently being considered for 5G systems are explained below and their advantages and disadvantages are discussed.

12.2.1. Orthogonal Frequency Division Multiplexing

Orthogonal Frequency Division Multiplexing (OFDM) and Orthogonal Frequency Division Multiple Access (OFDMA) have been the popular multi-carrier modulation and multiple access techniques until the 4G LTE [189]. Several advantages of OFDM like ease of implementation, robustness against frequency selective fading channels, compatibility with MIMO, simple structure and efficient implementation using FFT enabled the widespread use of OFDM in several applications like IEEE 802.11 wireless LAN and broadband over power line [190].

Despite all these advantages, OFDM suffers from a lot of disadvantages. Although OFDM is well suited for traditional communication, it does not support several new scenarios foreseen for 5G networks. The orthogonality between subcarriers require perfect synchronization which is difficult to achieve in scenarios like uplink [191] and in downlink when base stations cooperate to send data [192]. Due to the requirement of cyclic prefix, low latency cannot be achieved in OFDM. OFDM suffers from other well known disadvantages like high peak to average power ratio [193], high OOB power emission [194] and pilot contamination problem [195]. The vast literature available on OFDM has tried to overcome these shortcomings. For example, [196] proposes an algorithm to reduce the sidelobes, [195] tries to reduce the pilot contamination problem. All of them try to overcome only one disadvantage of OFDM and there is no solution that solves all the drawbacks mentioned above.

12.2.2. Filter Bank Multi-Carrier

Filter Bank Multi-Carrier (FBMC) transmits the set of data symbols through a bank of modulated filters [197] and it could be seen as a generalization of OFDM, where OFDM filters the complete band while FBMC filters each subcarrier individually [198]. FBMC replaces the sinc-pulse of subcarriers in OFDM with a different shape. This freedom in choosing the filter shape could be exploited and to develop a suitable filter design that satisfies the desired spectral characteristics could be developed. There are two broad classes of FBMC – staggered multi-tone (SMT) and filtered multi-tone (FMT), in which SMT is more popular [199].

FBMC has the advantages of having a very good frequency localization achieved by using different prototype filters. It also gets rid of the use of cyclic prefix thus increases the spectral efficiency of the wireless system. There are efficient implementations of FBMC using FFT and IFFT blocks and polyphase filter structures [197,200]. Compared to OFDM, FBMC is robust to frequency offset [201] and does not have strict symbol synchronization [202].

FBMC requires the use offset-QAM modulation and does not support other types of modulation. Similar to OFDM, FBMC also suffers from a high peak-to-average power ratio [203]. High spectral efficiency offered by FBMC can be realized only when the number of symbols transmitted is very high [198]. In practical scenarios where the transmission time interval is of the order of milliseconds, the filtering introduces ramp-up and ramp-down times at the end of these intervals and thereby reduces the spectral efficiency [204]. This will be a major issue in applications like MTC and IoT where the packet sizes are short. Further, some versions of FBMC support only pulse amplitude modulation [197]. This restriction would prohibit the use of higher order modulation techniques like 64-QAM which is being used in LTE E-UTRA [185], thereby limiting the data rates supported. FBMC systems are more sensitive to errors in channel measurements and FBMC modulation is less efficient than OFDM systems when there are errors in channel state information (CSI) [205]. These imperfections also cause inter-symbol interference (ISI) and inter-carrier interference (ICI) [206] and makes FBMC less compatible with MIMO systems. The advanced wireless communication techniques like spatial multiplexing with maximum likelihood detection and Alamouti Space Time Block Codes cannot be combined efficiently with FBMC [207].

12.2.3. Universal Filtered Multi-Carrier

Universal Filtered Multi-Carrier (UFMC) is a generalization of OFDM and FBMC. While OFDM filters the entire band and FBMC filters each subcarrier individually, UFMC uses filters on a group of subcarriers to reduce the out-of-band emissions [208]. UFMC splits the data stream into several sub-streams and a pulse shaping filter is applied to each of those streams [209]. This reduces the interference from adjacent sub-bands. This aggregation of frequency subbands also increases the bandwidth and thus makes subsymbols shorter in time, making them more suitable for applications that require short packet transmissions [198]. In addition, UFMC also does not need a cyclic prefix thus increases the spectral efficiency. The filter ramp-up and ramp-down time periods at the start and end of the frame give inherent protection against ISI. UFMC is also shown to outperform OFDM under both perfect and imperfect CSI conditions and in the presence of carrier frequency offset (CFO). This makes UFDM more suitable for CoMP transmission where synchronization between transmitters is hard to achieve. UFMC also supports MIMO systems.

In spite of these advantages, UFDM equalization process is very complex. This will be prohibitive for devices with limited computational abilities and are highly power constrained. UFMC is also sensitive to small timing misalignment [210] which makes them unsuitable for applications like MTC and IoT that require loose or no synchronization. In spite of having several improvements over OFDM and FBMC, UFMC does not satisfy all the requirements to become the choice for 5G cellular networks.

12.2.4. Generalized Frequency Division Multiplexing

Generalized Frequency Division Multiplexing (GFDM) was initially proposed in [211] as a waveform for scenarios where there is a high degree of spectrum fragmentation like exploiting spectral holes in TV bands. Due to its favorable properties, it is one of the top contenders for the waveform for 5G cellular networks [212].

GFDM is a block based multiplexing technique where each independently modulated block consists of several subcarriers and subsymbols. These subcarriers are filtered by the prototype filter along with the filter's time and frequency domain shifted versions [212]. This can be viewed as a generalization of OFDM and single carrier frequency domain equalization where OFDM has several subcarriers and only one symbol is transmitted per subcarrier in a block and single carrier frequency domain equalization has only one subcarrier in the whole frequency band and several symbols are transmitted serially. The number of subcarriers and number of subsymbols in each subcarrier, along with the prototype filter structure can be configured to meet the requirements of different applications.

In GFDM, cyclic prefix is needed to be added only once per block. Since the block has several subsymbols transmitted serially, GFDM has higher spectral efficiency compared to OFDM [212]. The requirement of only one cyclic prefix per block could also be exploited by using a long cyclic prefix which could enable applications that require loose or no synchronization [213]. The use of cyclic prefix could also be avoided by multiplying the GFDM block with a window function with has smooth ramp-up and ramp-down at the edge of blocks [212]. GFDM is also compatible with MIMO systems and can be combined efficiently with Alamouti space-time codes. Due to the similarity of GFDM waveform to that of OFDM, most of the vast literature developed for synchronization and timing estimation like [214] can be readily adapted for use in GFDM systems.

The main disadvantage of GFDM system is the computational complexity. Since GFDM cannot use the FFT/IFFT blocks for modulation and demodulation like OFDM and FBMC systems, GFDM systems have a huge computational complexity of encoding and decoding. This might be a forbidding computational requirement at several IoT devices. GFDM is also more sensitive to carrier frequency offset compared to OFDM [215]. Due to the inherent non-orthogonality, GFDM waveform has higher bit error rate compared to OFDM waveforms [216]. Although, there are interference cancellation schemes proposed [217], the use of successive interfere cancellation technique will further make it computationally more expensive.

12.2.5. Non-Orthogonal Multiple Access

Distinct from the other multiple access techniques discussed before, Non-Orthogonal Multiple Access (NOMA) does not provide an alternative waveform design. Instead, it provides a new dimension of multiplexing, power domain, which could be used in conjunction with other multiple access techniques [218]. NOMA exploits the difference in channel gain between the users and uses it to multiplex the data of different users [219].

Table 4
Comparison of multiple access techniques.

Multiple access technique	Advantages	Disadvantages
OFDM	• Computationally simple • Simple modulation and demodulation • High compatibility with MIMO • Well studied	• Requires perfect synchronization • Requires cyclic prefix • High OOB emission • High PAPR • Pilot contamination
FBMC	• Frequency localization • Efficient implementation • Robust to center frequency offset • Supports loose synchronization	• High PAPR • Supports only offset-QAM modulation • Spectrum efficiency achieved only when length of data is large • Sensitive to errors in CSI • Not compatible with MIMO
UFMC	• Frequency localization • Shorter symbols • Soft protection from ISI • Robust to CSI errors • MIMO compatibility	• Complex equalization • Sensitive to timing misalignment
GFDM	• Spectral efficiency • MIMO compatibility • ISI protection • Loose synchronization	• Computational complexity • Sensitive to CFO • High BER
NOMA	• New domain of multiplexing • Compatible with other multiple access techniques • Improves spectral efficiency	• Computational complexity • Requires perfect synchronization

NOMA will be well suited in practical wide area deployments where there are several users spread across the coverage area with some users with higher channel gain and some others with poor channel gain. This will help pairing the users efficiently so that the NOMA schema could be utilized effectively. To make this pairing, we need only the rough channel state estimate at the transmitter and not the finer CSI that would be needed in the receiver for decoding the data. This data could be used to decide whether or not to use NOMA and if we use, which users should be paired together [220].

NOMA inherently uses successive interference cancellation (SIC) for decoding and it comes with several drawbacks. The first one being the complexity. A user has to decode the other user's data first and then decode its own data. This doubles the computational complexity required at the receiver and this could be prohibitive for low power IoT devices. SIC also comes with the error propagation in decoding where the error in decoding the first user's data will cause error in decoding second user's data as well [221]. SIC will also require the two symbols transmitted at the same time-frequency block to be perfectly synchronized and this will be an issue in the uplink scenario.

12.3. Open problems

A summary of the advantages and disadvantages of multiple access techniques discussed in this section are listed in Table 4. The several multiple access techniques proposed here provide an improvement over OFDM and other conventional multiple access techniques. Most of them support the several scenarios that are new and are envisioned to be supported by 5G cellular networks. But, there is no clear winner that could be chosen conclusively. Although full duplex communication systems look set to be deployed, there are several aspects of it which need further stud and research. The open problems in the development of radio access techniques for 5G network are discussed below.

Simultaneously accommodating all devices: Although multiple access techniques like UFMC and GFDM have the advantage of being adaptable to different scenarios by varying the parameters involved, they do not support those scenarios simultaneously. For example, an IoT device and a mobile phone may be in the same cell range simultaneously. The IoT device will want to send short packets without being synchronized with the base station while the mobile phone may be streaming a HD video and it will require very high data rates and it will be ready to synchronize with the base station perfectly so that it can attain maximum data rate. These two scenarios cannot be supported by these multiple access techniques simultaneously. This is because all the minimum resource unit that could be allocated to a user has the same shape in the time-frequency lattice representation. Although LTE E-UTRA supports this to some level by providing flexible spectrum allocation and supporting variable bandwidths from 1.4 MHz to 20 MHz [185]. However, this level of configurability is also very limited and we need a multiple access technique where the size and shape of the time-frequency block allocated to each user can be configured individually so that each device attached to the base station can satisfy their own requirements, without being constrained by the other devices in the network.

A first step in this direction could be the Software Defined Multiple Access (SoDeMA) as envisioned in [222]. SoDeMA is said to have the flexibility where all the NOMA and orthogonal multiple access schemes coexist in a system and different types of devices in the network can use different part of the spectrum using the multiple access technique that would be the best fit for them. For example, the IoT device can use GFDM with a long cyclic prefix while the mobile device streaming video can use some orthogonal multiple access scheme that supports high data rates and the schemes to use can be configured adaptively based on the user requirement and the load on the base station.

Computational complexity: The multiple access technique to be adopted should also be computationally simple so that devices with low computational power can implement them easily. With the 50 billion IoT devices that are expected to be connected to 5G network, having a complex decoding requirement will affect the lifetime of these devices. Thus, low complexity is a necessary requirement for any candidate multiple access technique.

Robustness: The vast majority of the devices in a 5G cellular network will be low cost IoT devices. Due to their low cost, they cannot afford to have high precision clocks and oscillators. This error in oscillator may lead to CFO and timing misalignment. Further, going through a synchronization step every time the device is connected to the network will create overhead in power consumption. This makes it necessary for the multiple access technique to be robust against the imperfections in timing and frequency misalignments.

Replacing OFDM: Despite of all the problems discussed above, OFDM remains one of the simplest, efficient and widely deployed multiple access technique. Although there are several alternatives proposed for replacing OFDM, the industry and the standard bodies would be reluctant for a radical change in the multiple access technique. The multiple access technique adopted for 5G would be an extension or enhancement over the OFDM, at least for the first releases.

Harmonization of radio access techniques: As explained in [223], the 5G radio access technique will be a protocol harmonization of several multiple access and modulation techniques across different bands, services and cell types. The important issue in use of such a technique is the co-existence of several of these techniques. Although the user equipment and the radio access techniques should be as similar as possible across the frequency bands used, some of the techniques like narrowband beamforming for millimeter wave communications are tailored to be used in some of these bands. This makes the techniques disparate and makes harmonization difficult to achieve. In addition to the harmonization of waveform design, the harmonization is also required at higher layers. For example, the MAC protocol deployed should be independent of the frequency bands or waveforms used at the physical layer, providing the abstraction needed.

13. Global 5G activities

With the pursuit of unprecedented high data rates and capacity in wireless communication networks, the research activities for 5G are going on around the world at research institutions, standard bodies, service operators, and telecommunication companies, with the focus ranging from new channel models, signal multiplexing techniques, large scale antenna arrays, to the core network design. We hereby provide an overview of the global research activities.

13.1. The United States

In the United States, great research efforts in both academia and industry have been dedicated to the innovations in 5G. In order to establish a flexible and self-adaptable network architecture that can maximize capacity and spectrum efficiency, the development of SoftAir, a novel design based on WSDN, has been initiated in 2015 in the Broadband Wireless Networking (BWN) Laboratory in Georgia Institute of Technology. The great potential of SoftAir can enable a transformation from hardware-based network toward a low-cost, flexible, and programmable software-based network [20].

At New York University, University of Southern California, and University of Texas at Austin, groups of researchers are exploring the great spectrum resources at millimeter wave frequency, which will become an indispensable supplement to address today's concern on spectrum crunch [58]. Since the deployment of antenna arrays has proven to boost capacity for current LTE system, researchers in University of Texas at Austin are also proposing the development of massive MIMO, which is large scale antenna arrays that can serve multiple users at the same time with high data rates [224].

The AT&T research groups put their major focus on the key technologies of SDN and NFV to improve cost efficiency and enhance network performance. They also partnered with Ericsson and Intel to initiate 5G trials to implement new network architecture and prepare for commercial deployment [225].

Another industrial research effort in the US, Qualcomm, has visioned that 5G will continue to enhance mobile broadband services at frequency bands currently being utilized (i.e. below 6 GHz), and support short range for extremely high throughput at millimeter wave [226].

13.2. Europe

In Europe, the 3rd Generation Partnership Project (3GPP), which is the leading collaborating force in the standardization process of LTE and LTE-A, is focusing on enhancing the current standards towards the 5G era. The 3GPP Release 14 and 15, which are being developed and expected to be finalized by 2020, are promising to provide unprecedented system capacity which is one of the key requirements of 5G systems. Release 15 is also expected to be submitted to ITU as the first 5G standard in Europe.

Additionally, the 5G Infrastructure Public Private Partnership (5GPPP) is another European initiative on collaborations among network operators, infrastructure providers, and researchers. The projects under the collaboration of 5GPPP cover a wide range of 5G networks. The Mobile and wireless communications Enablers for the Twenty-twenty Information Society 5G, also known as METIS 2020 Project, is a leader in the research for 5G networks [11]. METIS initiated its discussions in late 2012 and covers a wide range of new concepts and horizontal topics. They proposed a new series of channel models based on extensive measurements for special scenarios in shopping malls, D2D, and V2V communications. In particular, the spatial consistency for D2D communications needs to be accounted for in the channel modeling for 5G, which is a deficiency of current channel models [92,227–229]. The second phase of the METIS project (METIS-II) was initiated in July 2015 and has a length of 24 months. The METIS-II aims to foster the standardization of the 5G RAN architecture and functionality design [230]. The 5G millimeter-wave RAN is the focus of the “mmMAGIC” project, which covers the wide frequency range of 6 $\sim$ 100 GHz for radio access technology design [231].

As to the aspect of waveform design which is another fundamental issue in 5G system, a collaborative project in Europe named as 5th Generation Non-Orthogonal Waveforms for Asynchronous Signaling, in acronym of “5GNOW”, aims at boosting capacity and coverage via novel waveform design and scalable air interface framework [232]. Moreover, as energy efficiency has become a global concern when more sophisticated computation and frequent signaling have caused excessive energy consumption, the “5GrEEn” project dedicates to provide a sustainable solution for environmentally friendly development in the 5G network [170]. Another project issued by 5GPPP is the 5G novel radio multi-service adaptive network architecture (5G-NORMA), which aims to develop an adaptive and efficient 5G mobile network architecture with the capability of dynamic network resource sharing among operators [233].

In order to design the 5G backhaul and frontaul networks, 5GPPP also initiated the “5G-Xhaul” and “5G-Crosshaul” projects, which are targeted to develop a dynamically reconfigurable backhaul/fronthaul network with cognitive control plane for small cells and C-RANs. Other efforts within 5GPPP include “CogNet” for intelligent network management, “COHERENT” for coordinated control and spectrum management for 5G HetNets, etc [234].

13.3. Asia

In Asia, South Korea, Japan, and China are the leading forces in 5G research activities. Samsung in South Korea announced its successfully tested millimeter wave technique for 5G in 2014, with a peak rate of 7.5 Gbps at the carrier frequency of 28 GHz, which is over 30 times faster than the current 4G LTE network [235]. Korea Telecom aims to provide millimeter wave-based 5G network services for 2018 Winter Olympic Games. In Japan, NTT DoCoMo dedicates the 5G research on small cells, which offers higher spectral density for a smaller cell coverage area [25]. They target the 2020 Summer Olympic Games to launch their 5G commercial services. The four major telecom operators, Korea Telecom, NTT DoCoMo, SK Telecom, and Verizon, are joining forces on the 5G Trial Specification Alliance, with the focus on developing a common and extendable platform for 5G networks. The Chinese government initiated the IMT-2020 Promotion Group as early as 2013. The research groups in China are advocating massive MIMO for next generation of wireless communications. The Academy of China Mobile is developing a prototype of 128 antenna array elements for performance analysis [236]. Huawei's recent demonstration on 5G end-to-end network slicing implementation reveals the efficiency and flexibility that network slicing can achieve to enable multiple network applications on a unified physical network architecture [237].

14. Conclusion

This paper presents an in-depth overview of the ten key enabling technologies which will be landmarks in 5G systems. Specifically, the requirements of the next generation cellular network are first explained and how these ten technologies would aid the realization of these requirements is examined. Furthermore, the strengths and limitations faced by these technologies are investigated in detail, in which the open problems that need to be explored and resolved before the commercialization of these solutions are fully summarized.. Finally, the current research activities being undertaken around the world are discussed. Although some network operators start to promote their "5G services" which are actually based on LTE-Advanced technologies, we predict that the 5G cellular standardization will be completed by the early years of the next decade and commercial services would be launched by mid-late 2020s. Although drastic paradigm shifts have never been easy and require massive upgrading of network infrastructures, the need for better services, which motivated such transitions in the past, will continue to occur in the future.

Acknowledgment

The authors would like to thank David M. Gutierrez-Estevez, Ravikumar Balakrishnan, Elias Chavarria Reyes, Won Yeol Lee, Ozgur Baris Akan, Chong Han, and Wolfgang Gerstacker for their valuable comments and suggestions to improve the quality of the paper.

References

[1] I.F. Akyildiz, D.M. Gutierrez-Estevez, R. Balakrishnan, E. Chavarria-Reyes, LTE-Advanced and the evolution to Beyond 4G (B4G) systems, Phys. Commun. 10 (2014) 31–60. http://dx.doi.org/10.1016/j.phycom.2013.11.009.
[2] I.F. Akyildiz, M. Pierobon, S. Balasubramaniam, Y. Koucheryavy, The Internet of Bio-Nano Things, IEEE Commun. Mag. 53 (3) (2015) 32–40, doi:10.1109/MCOM.2015.7060516.
[3] A. Ghosh, T. Thomas, M. Cudak, R. Ratasuk, P. Moorut, F. Vook, T. Rappaport, G. Maccartney, S. Sun, S. Nie, Millimeter-wave enhanced local area systems: a high-data-rate approach for future wireless networks, IEEE J. Sel. Areas Commun. 32 (6) (2014) 1152–1163, doi:10.1109/JSAC.2014.2328111.
[4] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, P. Popovski, Five disruptive technology directions for 5G, IEEE Commun. Mag. 52 (2) (2014) 74–80, doi:10.1109/MCOM.2014.6736746.
[5] ITU-R, IMT Vision Framework and overall objectives of the future development of IMT for 2020 and beyond, Recommendation ITU-R M.2083-0, Technical Report, International Telecommunication Union, 2015.
[6] E. Hossain, M. Hasan, 5G cellular: key enabling technologies and research challenges, IEEE Instrum. Meas. Mag. 18 (3) (2015) 11–21, doi:10.1109/MIM.2015.7108393.
[7] 3GPP TS 25.913, Requirements for evolved UTRA and evolved UTRAN, 2009.
[8] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, J. Zhang, What will 5G Be? IEEE J. Sel. Areas Commun. 32 (6) (2014) 1065–1082, doi:10.1109/JSAC.2014.2328098.
[9] G. Fettweis, The tactile Internet: applications and challenges, IEEE Veh. Technol. Mag. 9 (1) (2014) 64–70, doi:10.1109/MVT.2013.2295069.
[10] H. Shariatmadari, R. Ratasuk, S. Iraji, A. Laya, T. Taleb, R. Jantti, A. Ghosh, Machine-type communications: current status and future perspectives toward 5G systems, IEEE Commun. Mag. 53 (9) (2015) 10–17, doi:10.1109/MCOM.2015.7263367.
[11] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, H. Tullberg, M. Uusitalo, B. Timus, M. Fallgren, Scenarios for 5G mobile and wireless communications: the vision of the METIS project, IEEE Commun. Mag. 52 (5) (2014) 26–35, doi:10.1109/MCOM.2014.6815890.
[12] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, M. Ayyash, Internet of things: a survey on enabling technologies, protocols, and applications, IEEE Commun. Surveys Tuts. 17 (4) (2015) 2347–2376, doi:10.1109/COMST.2015.2444095.
[13] Cisco visual networking index: forecast and methodology, 2014–2019, CISCO White Paper (2015).
[14] R. Di Taranto, S. Muppirisetty, R. Raulefs, D. Slock, T. Svensson, H. Wymeersch, Location-aware communications for 5G networks: how location information can improve scalability, latency, and robustness of 5G, IEEE Signal Process. Mag. 31 (6) (2014) 102–112, doi:10.1109/MSP.2014.2332611.
[15] K.-T. Feng, C.-H. Hsu, T.-E. Lu, Velocity-assisted predictive mobility and location-aware routing protocols for mobile Ad Hoc networks, IEEE Trans. Veh. Technol. 57 (1) (2008) 448–464, doi:10.1109/TVT.2007.901897.
[16] X. Duan, X. Wang, Authentication handover and privacy protection in 5G hetnets using software-defined networking, IEEE Commun. Mag. 53 (4) (2015) 28–35, doi:10.1109/MCOM.2015.7081072.
[17] A. Ruiz-Martinez, Towards a web payment framework: State-of-the-art and challenges, Electron. Commerce Res. Appl. 14 (5) (2015) 345–350. Contemporary Research on Payments and Cards in the Global Fintech Revolution. http://dx.doi.org/10.1016/j.elerap.2015.08.003.
[18] I.F. Akyildiz, A. Lee, P. Wang, M. Luo, W. Chou, A roadmap for traffic engineering in sdn-openflow networks, Comput. Netw. J. 71 (2014) 1–30.
[19] S. Jain, A. Kumar, S. Mandal, et al., B4: experience with a globally-deployed software defined WAN, in: Proceedings of the ACM SIGCOMM Conference, SIGCOMM'13, August 2013, pp. 3–14.
[20] I.F. Akyildiz, P. Wang, S.C. Lin, SoftAir: a software defined networking architecture for 5G wireless systems, Comput. Netw. 85 (C) (2015) 1–18.
[21] X. Jin, L.E. Li, L. Vanbever, J. Rexford, SoftCell: scalable and flexible cellular core network architecture, in: Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies, ACM, 2013, pp. 163–174.
[22] J. Wu, Z. Zhang, Y. Hong, Y. Wen, Cloud radio access network (C-RAN): a primer, IEEE Netw. 29 (1) (2015) 35–41, doi:10.1109/MNET.2015.7018201.
[23] ARGELA (2016) http://www.argela.com.tr/progran/.
[24] SK Telecom, SK Telecom 5G White Paper: SK Telecom's View on 5G Vision, Architecture, Technology, and Spectrum, Technical Report, 2014.
[25] N.T.T. DOCOMO, DOCOMO 5G White Paper: 5G Radio Access: Requirements, Concept and Technologies, Technical Report, 2014.
[26] Project CONTENT FP, 2012--2015. http://cordis.europa.eu/fp7/ict/future-networks/.
[27] K.-K. Yap, R. Sherwood, M. Kobayashi, T.-Y. Huang, M. Chan, N. Handigol, N. McKeown, G. Parulkar, Blueprint for introducing innovation into wireless mobile networks, in: Proceedings of the Second ACM SIGCOMM Workshop on Virtualized Infrastructure Systems and Architectures, ACM, 2010, pp. 25–32.
[28] M. Bansal, J. Mehlman, S. Katti, P. Levis, OpenRadio: a programmable wireless dataplane, in: Proceedings of the First Workshop on Hot Topics in Software Defined Networks, ACM, 2012, pp. 109–114.
[29] P. Dely, J. Vestin, A. Kassler, N. Bayer, H. Einsiedler, C. Peylo, CloudMAC - an OpenFlow based architecture for 802.11 MAC layer processing in the cloud, in: 2012 IEEE Globecom Workshops (GC Wkshps), 2012, pp. 186–191, doi:10.1109/GLOCOMW.2012.6477567.
[30] L. Suresh, J. Schulz-Zander, R. Merz, A. Feldmann, T. Vazao, Towards programmable enterprise WLANS with Odin, in: Proceedings of the First Workshop on Hot Topics in Software Defined Networks, ACM, 2012, pp. 115–120.
[31] Centre Tecnologic Telecomunicacions Catalunya (CTTC), SDN/NFV Cloud Computing Platform and Core Network for 5G Services, Technical Report.
[32] A. Gudipati, D. Perry, L.E. Li, S. Katti, SoftRAN: software defined radio access network, in: Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking, ACM, 2013, pp. 25–30.
[33] I.F. Akyildiz, S.-C. Lin, P. Wang, Wireless software-defined networks (W-SDNs) and network function virtualization (NFV) for 5G cellular systems: an overview and qualitative evaluation, Comput. Netw. 93, Part 1 (2015) 66–79. http://dx.doi.org/10.1016/j.comnet.2015.10.013.
[34] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, J. Turner, OpenFlow: enabling innovation in campus networks, ACM SIGCOMM Comput. Commun. Rev. 38 (2) (2008) 69–74.
[35] B. Pfaff, J. Pettit, K. Amidon, M. Casado, T. Koponen, S. Shenker, Extending networking into the virtualization layer, ACM Workshop on Hot Topics in Networks, 2009.
[36] A. Aissioui, A. Ksentini, A. Gueroui, T. Taleb, Toward elastic distributed SDN/NFV controller for 5G mobile cloud management systems, IEEE Access 3 (2015) 2055–2064, doi:10.1109/ACCESS.2015.2489930.
[37] S.-C. Lin, P. Wang, M. Luo, Control traffic balancing in software defined networks, Comput. Netw. (2015). http://dx.doi.org/10.1016/j.comnet.2015.08.004.
[38] S.-C. Lin, P. Wang, M. Luo, Jointly optimized qos-aware virtualization and routing in software defined networks, Comput. Netw. 96 (2016) 69–78.

[39] P. Bosshart, G. Gibb, H.-S. Kim, G. Varghese, N. McKeown, M. Izzard, F. Mujica, M. Horowitz, Forwarding metamorphosis: fast programmable match-action processing in hardware for sdn, in: ACM SIGCOMM Computer Communication Review, vol. 43, ACM, 2013, pp. 99–110.
[40] S. Tomovic, M. Pejanovic-Djurisic, I. Radusinovic, SDN based mobile networks: concepts and benefits, Wireless Pers. Commun. 78 (3) (2014) 1629–1644.
[41] A.X. Porxas, S.C. Lin, M. Luo, QoS-aware virtualization-enabled routing in software defined networks, IEEE ICC'15, 2015.
[42] S.C. Lin, P. Wang, I.F. Akyildiz, M. Luo, Delay-based maximum power-weight scheduling with heavy-tailed traffic (2016). Submitted for publication in IEEE/ACM Transaction on Networking.
[43] F. Van den Abeele, J. Hoebeke, G.K. Teklemariam, I. Moerman, P. Demeester, Sensor function virtualization to support distributed intelligence in the internet of things, Wireless Pers. Commun. 81 (4) (2015) 1415–1436.
[44] ETSI GS NFV 003 V1.2.1: Network Functions Virtualisation (NFV); Terminology for Main Concepts in NFV, 2014, (ETSI Industry Specification Group (ISG) NFV).
[45] R. Sherwood, G. Gibby, K.K. Yapy, G. Appenzellery, M. Casado, N. McKeowny, G. Parulkar, FlowVisor: A Network Virtualization Layer, Technical Report, 2009.
[46] OpenStack: The Open Source Cloud Operating System https://www.openstack.org/.
[47] ETSI GS NFV 002 V1.2.1: Network Functions Virtualisation (NFV); Architectural Framework, 2014a, (ETSI Industry Specification Group (ISG) NFV).
[48] ETSI GS NFV-MAN 001 V1.1.1: Network Functions Virtualisation (NFV); Management and Orchestration, 2014b, (ETSI Industry Specification Group (ISG) NFV).
[49] R. Mijumbi, J. Serrat, J. Gorricho, N. Bouten, F. De Turck, R. Boutaba, Network function virtualization: state-of-the-art and research challenges, IEEE Commun. Surveys Tuts. 18 (1) (2016) 236–262, doi:10.1109/COMST.2015.2477041.
[50] H. Hawilo, A. Shami, M. Mirahmadi, R. Asal, NFV: state of the art, challenges, and implementation in next generation mobile networks (vEPC), IEEE Netw. 28 (6) (2014) 18–26, doi:10.1109/MNET.2014.6963800.
[51] M. Ghasemi, T. Benson, J. Rexford, RINC: Real-Time Inference-based Network Diagnosis in the Cloud, Technical Report, Princeton University, 2015.
[52] R. Mijumbi, J. Serrat, J.-L. Gorricho, S. Latre, M. Charalambides, D. Lopez, Management and orchestration challenges in network functions virtualization, IEEE Commun. Mag. 54 (1) (2016) 98–105, doi:10.1109/MCOM.2016.7378433.
[53] B. Han, V. Gopalakrishnan, L. Ji, S. Lee, Network function virtualization: challenges and opportunities for innovations, IEEE Commun. Mag. 53 (2) (2015) 90–97, doi:10.1109/MCOM.2015.7045396.
[54] F.C.C. 15-138, Notice Of Proposed Rulemaking, Use of Spectrum Bands Above 24 Ghz For Mobile Radio Services, 2015.
[55] N.O.I. FCC 14-154, In the Matter of Use of Spectrum Bands Above 24 GHz For Mobile Radio Services, 2014.
[56] Z. Qingling, J. Li, Rain attenuation in millimeter wave ranges, in: 7th International Symposium on Antennas, Propagation EM Theory, 2006, pp. 1–4, doi:10.1109/ISAPE.2006.353538.
[57] H. Xu, V. Kukshya, T. Rappaport, Spatial and temporal characteristics of 60 GHz indoor channels, IEEE J. Sel. Areas Commun. 20 (3) (2002) 620–630, doi:10.1109/49.995521.
[58] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, F. Gutierrez, Millimeter wave mobile communications for 5G cellular: it will work!, IEEE Access 1 (2013) 335–349, doi:10.1109/ACCESS.2013.2260813.
[59] M. Samimi, K. Wang, Y. Azar, G. Wong, R. Mayzus, H. Zhao, J. Schulz, S. Sun, F. Gutierrez, T. Rappaport, 28 GHz angle of arrival and angle of departure analysis for outdoor cellular communications using steerable beam antennas in New York City, in: 2013 IEEE 77th Vehicular Technology Conference (VTC Spring), 2013, pp. 1–6, doi:10.1109/VTCSpring.2013.6691812.
[60] T. Rappaport, S. Deng, 73 GHz wideband millimeter-wave foliage and ground reflection measurements and models, in: 2015 IEEE International Conference on Communication Workshop (ICCW), 2015, pp. 1238–1243, doi:10.1109/ICCW.2015.7247347.
[61] S. Nie, G. Maccartney, S. Sun, T. Rappaport, 28 GHz and 73 GHz signal outage study for millimeter wave cellular and backhaul communications, in: 2014 IEEE International Conference on Communications (ICC), 2014, pp. 4856–4861, doi:10.1109/ICC.2014.6884089.
[62] S. Sun, T. Rappaport, R. Heath, A. Nix, S. Rangan, MIMO for millimeter-wave wireless communications: beamforming, spatial multiplexing, or both? IEEE Commun. Mag. 52 (12) (2014) 110–121, doi:10.1109/MCOM.2014.6979962.
[63] Q. Li, G. Li, W. Lee, M. il Lee, D. Mazzarese, B. Clerckx, Z. Li, MIMO techniques in WiMAX and LTE: a feature overview, IEEE Commun. Mag. 48 (5) (2010) 86–92, doi:10.1109/MCOM.2010.5458368.
[64] E. Torkildson, C. Sheldon, U. Madhow, M. Rodwell, Millimeter-wave spatial multiplexing in an indoor environment, in: 2009 IEEE GLOBECOM Workshops, IEEE, 2009, pp. 1–6.
[65] A. Molisch, M. Win, J. Winters, Capacity of MIMO systems with antenna selection, in: IEEE International Conference on Communications., 2, 2001, pp. 570–574, doi:10.1109/ICC.2001.937004.
[66] A. Molisch, M. Win, MIMO systems with antenna selection, IEEE Microw. Mag. 5 (1) (2004) 46–56, doi:10.1109/MMW.2004.1284943.
[67] Y.-S. Choi, A. Molisch, M. Win, J. Winters, Fast algorithms for antenna selection in MIMO systems, in: IEEE 58th Vehicular Technology Conference., Vol. 3, 2003, pp. 1733–1737, doi:10.1109/VETECF.2003.1285322.
[68] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, F. Aryanfar, Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results, IEEE Commun. Mag. 52 (2) (2014) 106–113, doi:10.1109/MCOM.2014.6736750.
[69] I.F. Akyildiz, J.M. Jornet, C. Han, Terahertz band: next frontier for wireless communications, Phys. Commun. 12 (2014) 16–32.
[70] I. Akyildiz, J. Jornet, C. Han, TeraNets: ultra-broadband communication networks in the terahertz band, IEEE Wireless Commun. 21 (4) (2014) 130–135, doi:10.1109/MWC.2014.6882305.
[71] C. Han, A.O. Bicen, I.F. Akyildiz, Multi-ray channel modeling and wideband characterization for wireless communications in the terahertz band, IEEE Trans. Wireless Commun. 14 (5) (2015) 2402–2412, doi:10.1109/TWC.2014.2386335.
[72] C. Han, A.O. Bicen, I.F. Akyildiz, Multi-wideband waveform design for distance-adaptive wireless communications in the terahertz band, IEEE Trans. Signal Process. 64 (4) (2016) 910–922, doi:10.1109/TSP.2015.2498133.
[73] O. Momeni, E. Afshari, High power terahertz and millimeter-wave oscillator design: a systematic approach, IEEE J. Solid-State Circuits 46 (3) (2011) 583–597, doi:10.1109/JSSC.2011.2104553.
[74] I.F. Akyildiz, J.M. Jornet, Electromagnetic wireless nanosensor networks, Nano Commun. Netw. 1 (1) (2010) 3–19.
[75] I.F. Akyildiz, J.M. Jornet, The internet of nano-things, IEEE Wireless Commun. 17 (6) (2010) 58–63, doi:10.1109/MWC.2010.5675779.
[76] I.F. Akyildiz, J.M. Jornet, Realizing ultra-massive MIMO communication in the (0.06-10) terahertz band, Nano Commun. Netw. 3 (4) (2016) 46–54. http://dx.doi.org/10.1016/j.nancom.2016.02.001.
[77] M.K. Samimi, T.S. Rappaport, Statistical channel model with multi-frequency and arbitrary antenna beamwidth for millimeter-wave outdoor communications, in: IEEE Global Telecommunications Conference (GLOBECOM) Workshop, 2015, pp. 1–7.
[78] A. Maltsev, A. Pudeyev, I. Karls, I. Bolotin, G. Morozov, W. Keusgen, R. Weiler, M. Danchenko, A. Kuznetsov, Quasi-deterministic approach to mmwave channel modeling in the FP7 MiWEBA project, Wireless Communications and Networks, Fraunhofer Heinrich Hertz Institute, Germany, 2014.
[79] 3GPP TR 36.873, Study on 3D channel model for LTE, 2014.
[80] H. Zhao, R. Mayzus, S. Sun, M. Samimi, J. Schulz, Y. Azar, K. Wang, G. Wong, F. Gutierrez, T. Rappaport, 28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York city, in: 2013 IEEE International Conference on Communications (ICC), 2013, pp. 5163–5167, doi:10.1109/ICC.2013.6655403.
[81] C. Barati, S. Hosseini, S. Rangan, P. Liu, T. Korakis, S. Panwar, T. Rappaport, Directional cell discovery in millimeter wave cellular networks, IEEE Trans. Wireless Commun. 14 (12) (2015) 6664–6678, doi:10.1109/TWC.2015.2457921.
[82] J. Garcia-Rois, F. Gomez-Cuba, M. Riza Akdeniz, F. Gonzalez-Castano, J. Burguillo-Rial, S. Rangan, B. Lorenzo, On the analysis of scheduling in dynamic duplex multihop mmWave cellular systems, IEEE Trans. Wireless Commun. 14 (11) (2015) 6028–6042, doi:10.1109/TWC.2015.2446983.
[83] T. Kim, J. Park, J.-Y. Seol, S. Jeong, J. Cho, W. Roh, Tens of Gbps support with mmWave beamforming systems for next generation communications, in: 2013 IEEE Global Communications Conference (GLOBECOM), 2013, pp. 3685–3690, doi:10.1109/GLOCOM.2013.6831646.
[84] E. Larsson, O. Edfors, F. Tufvesson, T. Marzetta, Massive MIMO for next generation wireless systems, IEEE Commun. Mag. 52 (2) (2014) 186–195, doi:10.1109/MCOM.2014.6736761.
[85] F. Rusek, D. Persson, B.K. Lau, E. Larsson, T. Marzetta, O. Edfors, F. Tufvesson, Scaling up MIMO: opportunities and challenges with very large arrays, IEEE Signal Process. Mag. 30 (1) (2013) 40–60, doi:10.1109/MSP.2011.2178495.
[86] T. Marzetta, Noncooperative cellular wireless with unlimited numbers of base station antennas, IEEE Trans. Wireless Commun. 9 (11) (2010) 3590–3600, doi:10.1109/TWC.2010.092810.091092.
[87] J. Jose, A. Ashikhmin, P. Whiting, S. Vishwanath, Channel estimation and linear precoding in multiuser multiple-antenna TDD systems, IEEE Trans. Veh. Technol. 60 (5) (2011) 2102–2116, doi:10.1109/TVT.2011.2146797.
[88] P. Viswanath, D. Tse, Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality, IEEE Trans. Inf. Theory 49 (8) (2003) 1912–1921, doi:10.1109/TIT.2003.814483.
[89] H. Yin, D. Gesbert, M. Filippou, Y. Liu, A coordinated approach to channel estimation in large-scale multiple-antenna systems, IEEE J. Sel. Areas Commun. 31 (2) (2013) 264–273, doi:10.1109/JSAC.2013.130214.
[90] J. Nam, A. Adhikary, J.-Y. Ahn, G. Caire, Joint spatial division and multiplexing: opportunistic beamforming, user grouping and simplified downlink scheduling, IEEE J. Sel. Topics Signal Process. 8 (5) (2014) 876–890, doi:10.1109/JSTSP.2014.2313808.
[91] Y.-H. Nam, B.L. Ng, K. Sayana, Y. Li, J. Zhang, Y. Kim, J. Lee, Full-dimension MIMO (FD-MIMO) for next generation cellular technology, IEEE Commun. Mag. 51 (6) (2013) 172–179, doi:10.1109/MCOM.2013.6525612.
[92] D.S. Baum, J. Hansen, J. Salo, An interim channel model for beyond-3G systems: extending the 3GPP spatial channel model (SCM), in: 2005 IEEE 61st Vehicular Technology Conference, vol. 5, IEEE, 2005, pp. 3132–3136.
[93] P. Almers, E. Bonek, A. Burr, N. Czink, M. Debbah, V. Degli-Esposti, H. Hofstetter, P. Kyösti, D. Laurenson, G. Matz, et al., Survey of channel and radio propagation models for wireless MIMO systems, EURASIP J. Wireless Commun. Netw. 2007 (1) (2007). 56–56

[94] X. Lu, A. Tolli, O. Piirainen, M. Juntti, W. Li, Comparison of antenna arrays in a 3-D multiuser multicell network, in: 2011 IEEE International Conference on Communications (ICC), 2011, pp. 1–6, doi:10.1109/icc.2011.5963126.
[95] X. Cheng, B. Yu, L. Yang, J. Zhang, G. Liu, Y. Wu, L. Wan, Communicating in the real world: 3D MIMO, IEEE Wireless Commun. 21 (4) (2014) 136–144.
[96] S. Sun, B. Rong, R.Q. Hu, Y. Qian, Spatial domain management and massive MIMO coordination in 5G SDN, IEEE Access 3 (2015) 2238–2251, doi:10.1109/ACCESS.2015.2498609.
[97] C.A. Balanis, Antenna Theory: Analysis and Design, Wiley-Interscience, 2005.
[98] R. Rogalin, O. Bursalioglu, H. Papadopoulos, G. Caire, A. Molisch, A. Michaloliakos, V. Balan, K. Psounis, Scalable synchronization and reciprocity calibration for distributed multiuser MIMO, IEEE Trans. Wireless Commun. 13 (4) (2014) 1815–1831, doi:10.1109/TWC.2014.030314.130474.
[99] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, L. Zhong, Argos: practical many-antenna base stations, in: Proceedings of the 18th Annual International Conference on Mobile Computing and Networking, Mobicom, 2012, pp. 53–64.
[100] I. Atzeni, J. Arnau, M. Debbah, Fractional pilot reuse in massive MIMO systems, in: 2015 IEEE International Conference on Communication Workshop (ICCW), 2015, pp. 1030–1035, doi:10.1109/ICCW.2015.7247312.
[101] R. Muller, L. Cottatellucci, M. Vehkapera, Blind pilot decontamination, IEEE J. Sel. Topics Signal Process. 8 (5) (2014) 773–786, doi:10.1109/JSTSP.2014.2310053.
[102] A. Ashikhmin, T. Marzetta, Pilot contamination precoding in multi-cell large scale antenna systems, in: 2012 IEEE International Symposium on Information Theory Proceedings (ISIT), 2012, pp. 1137–1141, doi:10.1109/ISIT.2012.6283031.
[103] D. Ying, F. Vook, T. Thomas, D. Love, Hybrid structure in massive MIMO: achieving large sum rate with fewer RF chains, in: 2015 IEEE International Conference on Communications (ICC), 2015, pp. 2344–2349, doi:10.1109/ICC.2015.7248675.
[104] J. Wallace, M. Jensen, Mutual coupling in MIMO wireless systems: a rigorous network theory analysis, IEEE Trans. Wireless Commun. 3 (4) (2004) 1317–1325, doi:10.1109/TWC.2004.830854.
[105] U. Madhow, D.R. Brown, S. Dasgupta, R. Mudumbai, Distributed massive MIMO: algorithms, architectures and concept systems, in: IEEE Information Theory and Applications Workshop (ITA), IEEE, 2014, pp. 1–7.
[106] K. Truong, R. Heath, The viability of distributed antennas for massive MIMO systems, in: 2013 Asilomar Conference on Signals, Systems and Computers, 2013, pp. 1318–1323, doi:10.1109/ACSSC.2013.6810508.
[107] J. Andrews, Seven ways that HetNets are a cellular paradigm shift, IEEE Commun. Mag. 51 (3) (2013) 136–144, doi:10.1109/MCOM.2013.6476878.
[108] N. Zhang, N. Cheng, A. Gamage, K. Zhang, J. Mark, X. Shen, Cloud assisted HetNets toward 5G wireless networks, IEEE Commun. Mag. 53 (6) (2015) 59–65, doi:10.1109/MCOM.2015.7120046.
[109] M. Munoz, C. Rubio, A new model for service and application convergence in B3G/4G networks, IEEE Wireless Commun. 11 (5) (2004) 6–12, doi:10.1109/MWC.2004.1351676.
[110] A. Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T. Thomas, J. Andrews, P. Xia, H. Jo, H. Dhillon, T. Novlan, Heterogeneous cellular networks: from theory to practice, IEEE Commun. Mag. 50 (6) (2012) 54–64, doi:10.1109/MCOM.2012.6211486.
[111] W. Webb, On using white space spectrum, IEEE Commun. Mag. 50 (8) (2012) 145–151, doi:10.1109/MCOM.2012.6257541.
[112] A.B. Flores, R.E. Guerra, E.W. Knightly, P. Ecclesine, S. Pandey, IEEE 802.11 af: a standard for TV white space spectrum sharing, IEEE Commun. Mag. 51 (10) (2013) 92–100.
[113] R. Menon, R.M. Buehrer, J.H. Reed, On the impact of dynamic spectrum sharing techniques on legacy radio systems, IEEE Trans. Wireless Commun. 7 (11) (2008) 4198–4207.
[114] Z. Ji, K. Liu, Cognitive radios for dynamic spectrum access-dynamic spectrum sharing: a game theoretical overview, IEEE Commun. Mag. 45 (5) (2007) 88–94.
[115] A. Mohamed, O. Onireti, M. Imran, A. Imran, R. Tafazolli, Control-data separation architecture for cellular radio access networks: a survey and outlook, IEEE Commun. Surveys Tuts. 18 (1) (2016) 446–465, doi:10.1109/COMST.2015.2451514.
[116] T. Zhao, P. Yang, H. Pan, R. Deng, S. Zhou, Z. Niu, Software defined radio implementation of signaling splitting in hyper-cellular network, in: Proceedings of the Second Workshop on Software Radio Implementation Forum, ACM, 2013, pp. 81–84.
[117] D. He, C. Chen, S. Chan, J. Bu, Secure and efficient handover authentication based on bilinear pairing functions, IEEE Trans. Wireless Commun. 11 (1) (2012) 48–53, doi:10.1109/TWC.2011.110811.111240.
[118] J. Cao, H. Li, M. Ma, Y. Zhang, C. Lai, A simple and robust handover authentication between HeNB and eNB in LTE networks, Comput. Netw. 56 (8) (2012) 2119–2131, doi:10.1016/j.comnet.2012.02.012.
[119] T. Novlan, R. Ganti, A. Ghosh, J. Andrews, Analytical evaluation of fractional frequency reuse for heterogeneous cellular networks, IEEE Trans. Commun. 60 (7) (2012) 2029–2039, doi:10.1109/TCOMM.2012.061112.110477.
[120] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. Sukhavasi, C. Patel, S. Geirhofer, Network densification: the dominant theme for wireless evolution into 5G, IEEE Commun. Mag. 52 (2) (2014) 82–89, doi:10.1109/MCOM.2014.6736747.
[121] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, D. Malladi, A survey on 3GPP heterogeneous networks, IEEE Wireless Commun. 18 (3) (2011) 10–21, doi:10.1109/MWC.2011.5876496.
[122] S. Geirhofer, P. Gaal, Coordinated multi point transmission in 3GPP LTE heterogeneous networks, in: 2012 IEEE Globecom Workshops, 2012, pp. 608–612, doi:10.1109/GLOCOMW.2012.6477643.
[123] I. Hwang, B. Song, S. Soliman, A holistic view on hyper-dense heterogeneous and small cell networks, IEEE Commun. Mag. 51 (6) (2013) 20–27, doi:10.1109/MCOM.2013.6525591.
[124] D. Lopez-Perez, M. Ding, H. Claussen, A. Jafari, Towards 1 Gbps/UE in cellular systems: understanding ultra-dense small cell deployments, IEEE Commun. Surveys Tuts. 17 (4) (2015) 2078–2101, doi:10.1109/COMST.2015.2439636.
[125] BellAir Networks, Cell site backhaul over unlicensed bands, Technical Report, 2012.
[126] D. Bojic, E. Sasaki, N. Cvijetic, T. Wang, J. Kuno, J. Lessmann, S. Schmid, H. Ishii, S. Nakamura, Advanced wireless and optical technologies for small-cell mobile backhaul with dynamic software-defined management, IEEE Commun. Mag. 51 (9) (2013) 86–93, doi:10.1109/MCOM.2013.6588655.
[127] R. Kaewpuang, D. Niyato, P. Wang, E. Hossain, A framework for cooperative resource management in mobile cloud computing, IEEE J. Sel. Areas Commun. 31 (12) (2013) 2685–2700, doi:10.1109/JSAC.2013.131209.
[128] S. Barbarossa, S. Sardellitti, P. Di Lorenzo, Communicating while computing: distributed mobile cloud computing over 5G heterogeneous networks, IEEE Signal Process. Mag. 31 (6) (2014) 45–55, doi:10.1109/MSP.2014.2334709.
[129] N. Fernando, S.W. Loke, W. Rahayu, Mobile cloud computing: a survey, Future Gener. Comput. Syst. 29 (1) (2013) 84–106.
[130] K. Kumar, J. Liu, Y.-H. Lu, B. Bhargava, A survey of computation offloading for mobile systems, Mob. Netw. Appl. 18 (1) (2013) 129–140.
[131] M. Barbera, S. Kosta, A. Mei, J. Stefa, To offload or not to offload? The bandwidth and energy costs of mobile cloud computing, in: 2013 Proceedings IEEE INFOCOM, 2013, pp. 1285–1293, doi:10.1109/INFCOM.2013.6566921.
[132] M. Patel, B. Naughton, C. Chan, N. Sprecher, S. Abeta, A. Neal, et al., Mobile-Edge Computing Introductory Technical White Paper, Mobile-edge Computing (MEC) industry initiative (2014).
[133] S. Subashini, V. Kavitha, A survey on security issues in service delivery models of cloud computing, Journal of Network and Computer Applications 34 (1) (2011) 1–11.
[134] Y. Chen, V. Paxson, R.H. Katz, What is New About Cloud Computing Security? Technical Report, EECS Department, University of California, Berkeley, 2010.
[135] P. Mell, T. Grance, NIST SP 800-145, The NIST Definition of Cloud Computing, 2011.
[136] C.A. Ardagna, R. Asal, E. Damiani, Q.H. Vu, From security to assurance in the cloud: a survey, ACM Comput. Surv. 48 (1) (2015) 2:1–2:50.
[137] O. Vermesan, P. Friess, Internet of Things: Converging Technologies for Smart Environments and Integrated Ecosystems, River Publishers, 2013.
[138] D. Evans, The Internet of Things - How the Next Evolution of the Internet is Changing Everything, CISCO White Paper (2011).
[139] H.-L. Truong, S. Dustdar, Principles for engineering iot cloud systems, IEEE Cloud Comput. 2 (2) (2015) 68–76, doi:10.1109/MCC.2015.23.
[140] Integration of cloud computing and internet of things: a survey, Future Gener. Comput. Syst. 56 (2016) 684–700. http://dx.doi.org/10.1016/j.future.2015.09.021.
[141] F. Bonomi, R. Milito, J. Zhu, S. Addepalli, Fog computing and its role in the internet of things, in: Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing, MCC '12, 2012, pp. 13–16.
[142] C. Manoj, A.K. Jagannatham, Optimal prediction likelihood tree based source-channel ml decoder for wireless sensor networks, IEEE Signal Process. Lett. 21 (2) (2014) 135–139, doi:10.1109/LSP.2013.2294794.
[143] M. Aazam, E.-N. Huh, M. St-Hilaire, C.-H. Lung, I. Lambadaris, Cloud of things: integration of IoT with cloud computing, in: Robots and Sensor Clouds, Springer, 2016, pp. 77–94.
[144] NGMN Alliance, NGMN 5G white paper, Next Generation Mobile Networks, White paper(2015).
[145] C.-L. I, C. Rowell, S. Han, Z. Xu, G. Li, Z. Pan, Toward green and soft: a 5G perspective, IEEE Commun. Mag. 52 (2) (2014) 66–73, doi:10.1109/MCOM.2014.6736745.
[146] R.P. Jover, I. Murynets, Connection-less communication of iot devices over lte mobile networks, in: Sensing, 12th Annual IEEE International Conference on Communication, and Networking (SECON), 2015, pp. 247–255, doi:10.1109/SAHCN.2015.7338323.
[147] Y.-D. Lin, Y.-C. Hsu, Multihop cellular: a new architecture for wireless communications, in: Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings of IEEE INFOCOM, vol. 3, 2000, pp. 1273–1282, doi:10.1109/INFCOM.2000.832516.
[148] A. Asadi, Q. Wang, V. Mancuso, A survey on device-to-device communication in cellular networks, IEEE Commun. Surveys Tuts. 16 (4) (2014) 1801–1819, doi:10.1109/COMST.2014.2319555.
[149] M. Tehrani, M. Uysal, H. Yanikomeroglu, Device-to-device communication in 5G cellular networks: challenges, solutions, and future directions, IEEE Commun. Mag. 52 (5) (2014) 86–92, doi:10.1109/MCOM.2014.6815897.
[150] B. Bangerter, S. Talwar, R. Arefi, K. Stewart, Networks and devices for the 5G era, IEEE Commun. Mag. 52 (2) (2014) 90–96, doi:10.1109/MCOM.2014.6736748.
[151] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklos, Z. Turanyi, Design aspects of network assisted device-to-device communications, IEEE Commun. Mag. 50 (3) (2012) 170–177, doi:10.1109/MCOM.2012.6163598.
[152] G. Fodor, S. Parkvall, S. Sorrentino, P. Wallentin, Q. Lu, N. Brahmi, Device-to-device communications for national security and public safety, IEEE Access 2 (2014) 1510–1520, doi:10.1109/ACCESS.2014.2379938.

[153] X. Lin, J.G. Andrews, A. Ghosh, R. Ratasuk, An overview of 3GPP device-to-device proximity services, IEEE Commun. Mag. 52 (4) (2014) 40–48, doi:10.1109/MCOM.2014.6807945.
[154] M. Usman, A.A. Gebremariam, U. Raza, F. Granelli, A software-defined device-to-device communication architecture for public safety applications in 5G networks, IEEE Access 3 (2015) 1649–1654, doi:10.1109/ACCESS.2015.2479855.
[155] S. Mumtaz, K.M.S. Huq, M.I. Ashraf, J. Rodriguez, V. Monteiro, C. Politis, Cognitive vehicular communication for 5G, IEEE Commun. Mag. 53 (7) (2015) 109–117, doi:10.1109/MCOM.2015.7158273.
[156] 5G Automotive Vision (2015) https://5g-ppp.eu/wp-content.
[157] D. Jiang, L. Delgrossi, IEEE 802.11p: towards an international standard for wireless access in vehicular environments, in: IEEE Vehicular Technology Conference, 2008, pp. 2036–2040, doi:10.1109/VETECS.2008.458.
[158] D. Feng, L. Lu, Y. Yuan-Wu, G. Li, G. Feng, S. Li, Device-to-device communications underlaying cellular networks, IEEE Trans. Commun. 61 (8) (2013) 3541–3551, doi:10.1109/TCOMM.2013.071013.120787.
[159] W.H. Chin, Z. Fan, R. Haines, Emerging technologies and research challenges for 5G wireless networks, IEEE Wireless Commun. 21 (2) (2014) 106–112, doi:10.1109/MWC.2014.6812298.
[160] C.-H. Yu, K. Doppler, C. Ribeiro, O. Tirkkonen, Resource sharing optimization for device-to-device communication underlaying cellular networks, IEEE Trans. Wireless Commun. 10 (8) (2011) 2752–2763, doi:10.1109/TWC.2011.060811.102120.
[161] L. Wang, H. Wu, Fast pairing of device-to-device link underlay for spectrum sharing with cellular users, IEEE Commun. Lett. 18 (10) (2014) 1803–1806, doi:10.1109/LCOMM.2014.2351400.
[162] G. Fodor, E. Dahlman, G. Mildh, S. Parkvall, N. Reider, G. Miklos, Z. Turanyi, Design aspects of network assisted device-to-device communications, IEEE Commun. Mag. 50 (3) (2012) 170–177, doi:10.1109/MCOM.2012.6163598.
[163] B. Zhou, H. Hu, S.-Q. Huang, H.-H. Chen, Intracluster device-to-device relay algorithm with optimal resource utilization, IEEE Trans. Veh. Technol. 62 (5) (2013) 2315–2326, doi:10.1109/TVT.2012.2237557.
[164] E. Hossain, D. Niyato, Z. Han, Dynamic Spectrum Access and Management in Cognitive Radio Networks, 1st Edition, Cambridge University Press, New York, NY, USA, 2009.
[165] S. Jaeckel, L. Raschkowski, K. Borner, L. Thiele, QuaDRiGa: a 3-D multi-cell channel model with time evolution for enabling virtual field trials, IEEE Trans. Antennas Propag. 62 (6) (2014) 3242–3256, doi:10.1109/TAP.2014.2310220.
[166] T. Zwick, C. Fischer, W. Wiesbeck, A stochastic multipath channel model including path directions for indoor environments, IEEE J. Sel. Areas Commun. 20 (6) (2002) 1178–1192, doi:10.1109/JSAC.2002.801218.
[167] C.-C. Chong, C.-M. Tan, D.I. Laurenson, S. McLaughlin, M.A. Beach, A.R. Nix, A novel wideband dynamic directional indoor channel model based on a Markov process, IEEE Trans. Wireless Commun. 4 (4) (2005) 1539–1552, doi:10.1109/TWC.2005.850341.
[168] A.O. Kaya, L.J. Greenstein, W. Trappe, Characterizing indoor wireless channels via ray tracing combined with stochastic modeling, IEEE Trans. Wireless Commun. 8 (8) (2009) 4165–4175, doi:10.1109/TWC.2009.080785.
[169] S. Nie, C. Han, I.F. Akyildiz, A 3-Dimensional Time-Varying Channel Model for 5G Wireless Communication Systems, 2016. Manuscript submitted for publication
[170] M. Olsson, C. Cavdar, P. Frenger, S. Tombaz, D. Sabella, R. Jantti, 5GrEEn: towards green 5G mobile networks, in: 2013 IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), 2013, pp. 212–216, doi:10.1109/WiMOB.2013.6673363.
[171] H.Y. Lateef, M. Dohler, A. Mohammed, M.M. Guizani, C.F. Chiasserini, Towards energy-aware 5G heterogeneous networks, in: Energy Management in Wireless Cellular and Ad-hoc Networks, Springer, 2016, pp. 31–44.
[172] M. Marsan, L. Chiaraviglio, D. Ciullo, M. Meo, Multiple daily base station switch-offs in cellular networks, in: 2012 Fourth International Conference on Communications and Electronics (ICCE), 2012, pp. 245–250, doi:10.1109/CCE.2012.6315906.
[173] M. Aftab Hossain, R. Jantti, C. Cavdar, Energy saving market for mobile operators, in: 2015 IEEE International Conference on Communication Workshop (ICCW), 2015, pp. 2856–2861, doi:10.1109/ICCW.2015.7247612.
[174] M. Marsan, M. Meo, Network sharing and its energy benefits: a study of European mobile network operators, in: 2013 IEEE Global Communications Conference (GLOBECOM), 2013, pp. 2561–2567, doi:10.1109/GLOCOM.2013.6831460.
[175] S. Raza Zaidi, A. Afzal, M. Hafeez, M. Ghogho, D. McLernon, A. Swami, Solar energy empowered 5G cognitive metro-cellular networks, IEEE Commun. Mag. 53 (7) (2015) 70–77, doi:10.1109/MCOM.2015.7158268.
[176] L.-C. Wang, S. Rangapillai, A survey on green 5G cellular networks, in: 2012 International Conference on Signal Processing and Communications (SPCOM), 2012, pp. 1–5, doi:10.1109/SPCOM.2012.6290252.
[177] X. Zhang, J. Zhang, W. Wang, Y. Zhang, C.-L. I, Z. Pan, G. Li, Y. Chen, Macro-assisted data-only carrier for 5G green cellular systems, IEEE Commun. Mag. 53 (5) (2015) 223–231, doi:10.1109/MCOM.2015.7105669.
[178] D. Lin, G. Charbit, I.-K. Fu, Uplink contention based multiple access for 5G cellular IoT, in: 2015 IEEE 82nd Vehicular Technology Conference (VTC Fall), 2015, pp. 1–5, doi:10.1109/VTCFall.2015.7391184.
[179] T. Chen, H. Kim, Y. Yang, Energy efficiency metrics for green wireless communications, in: 2010 International Conference on Wireless Communications and Signal Processing (WCSP), IEEE, 2010, pp. 1–6.
[180] ECR Initiative, Network and Telecom Equipment - Energy and Performance Assessment, Technical Report, 2010.
[181] C. Han, T. Harrold, S. Armour, I. Krikidis, S. Videv, P.M. Grant, H. Haas, J.S. Thompson, I. Ku, C.-X. Wang, et al., Green radio: radio techniques to enable energy-efficient wireless networks, IEEE Commun. Mag. 49 (6) (2011) 46–54.
[182] P. Somavat, V. Namboodiri, et al., Energy consumption of personal computing including portable communication devices, J. Green Eng. 1 (4) (2011) 447–475.
[183] K. Moskvitch, Sweet Success for Bio-Battery, 2014. http://www.rsc.org/chemistryworld/2014/01.
[184] K. Simona, Moss-Powered Future, 2014. https://ecopostblog.wordpress.com/2014/03/09.
[185] 3GPP, Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN): Overall description Stage 2, Release 13 (2016) http://www.3gpp.org/DynaReport/36300.htm.
[186] N.H. Mahmood, G. Berardinelli, F.M.L. Tavares, P. Mogensen, On the potential of full duplex communication in 5g small cell networks, in: 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), 2015, pp. 1–5, doi:10.1109/VTCSpring.2015.7145975.
[187] X. Zhang, W. Cheng, H. Zhang, Full-duplex transmission in PHY and MAC layers for 5G mobile wireless networks, IEEE Wireless Commun. 22 (5) (2015) 112–121, doi:10.1109/MWC.2015.7306545.
[188] H. Alves, R.D. Souza, M.E. Pellenz, Brief survey on full-duplex relaying and its applications on 5G, in: 2015 IEEE 20th International Workshop on Computer Aided Modelling and Design of Communication Links and Networks (CAMAD), 2015, pp. 17–21, doi:10.1109/CAMAD.2015.7390473.
[189] T. Hwang, C. Yang, G. Wu, S. Li, G. Li, OFDM and its wireless applications: a survey, IEEE Trans. Veh. Technol. 58 (4) (2009) 1673–1694, doi:10.1109/TVT.2008.2004555.
[190] IEEE Standard for Broadband over Power Line Networks: Medium Access Control and Physical Layer Specifications, IEEE Std 1901–2010 (2010) 1–1586, doi:10.1109/IEEESTD.2010.5678772.
[191] M. Morelli, C.-C. Kuo, M.-O. Pun, Synchronization techniques for orthogonal frequency division multiple access (OFDMA): a tutorial review, Proc. IEEE 95 (7) (2007) 1394–1427, doi:10.1109/JPROC.2007.897979.
[192] P. Banelli, S. Buzzi, G. Colavolpe, A. Modenini, F. Rusek, A. Ugolini, Modulation formats and waveforms for 5G networks: who will be the heir of OFDM?: an overview of alternative modulation schemes for improved spectral efficiency, IEEE Signal Process. Mag. 31 (6) (2014) 80–93, doi:10.1109/MSP.2014.2337391.
[193] H. Ochiai, H. Imai, On the distribution of the peak-to-average power ratio in OFDM signals, IEEE Trans. Commun. 49 (2) (2001) 282–289, doi:10.1109/26.905885.
[194] J. Van De Beek, F. Berggren, Out-of-band power suppression in OFDM, IEEE Commun. Lett. 12 (9) (2008) 609–611.
[195] J. Zhang, B. Zhang, S. Chen, X. Mu, M. El-Hajjar, L. Hanzo, Pilot Contamination Elimination for Large-Scale Multiple-Antenna Aided OFDM Systems, IEEE J. Sel. Topics Signal Process. 8 (5) (2014) 759–772, doi:10.1109/JSTSP.2014.2309936.
[196] S. Pagadarai, R. Rajbanshi, A.M. Wyglinski, G.J. Minden, Sidelobe suppression for OFDM-based cognitive radios using constellation expansion, in: 2008 IEEE Wireless Communications and Networking Conference, IEEE, 2008, pp. 888–893.
[197] B. Farhang-Boroujeny, OFDM versus filter bank multicarrier, IEEE Signal Process. Mag. 28 (3) (2011) 92–112, doi:10.1109/MSP.2011.940267.
[198] F. Schaich, T. Wild, Y. Chen, Waveform contenders for 5G - suitability for short packet and low latency transmissions, in: 2014 IEEE 79th Vehicular Technology Conference (VTC Spring), 2014, pp. 1–5, doi:10.1109/VTCSpring.2014.7023145.
[199] F. Schaich, T. Wild, Waveform contenders for 5G–OFDM vs. FBMC vs. UFMC, in: 2014 6th International Symposium on Communications, Control and Signal Processing (ISCCSP), 2014, pp. 457–460, doi:10.1109/ISCCSP.2014.6877912.
[200] A. Sahin, I. Guvenc, H. Arslan, A survey on multicarrier communications: prototype filters, lattice structures, and implementation aspects, IEEE Commun. Surveys Tuts. 16 (3) (2014) 1312–1338, doi:10.1109/SURV.2013.121213.00263.
[201] T. Fusco, A. Petrella, M. Tanda, Sensitivity of multi-user filter-bank multicarrier systems to synchronization errors, in: 2008 3rd International Symposium on Communications, Control and Signal Processing, 2008, pp. 393–398, doi:10.1109/ISCCSP.2008.4537257.
[202] G. Wunder, M. Kasparick, S. ten Brink, F. Schaich, T. Wild, I. Gaspar, E. Ohlmer, S. Krone, N. Michailow, A. Navarro, G. Fettweis, D. Ktenas, V. Berg, M. Dryjanski, S. Pietrzyk, B. Eged, 5GNOW: challenging the LTE design paradigms of orthogonality and synchronicity, in: 2013 IEEE 77th Vehicular Technology Conference (VTC Spring), 2013, pp. 1–5, doi:10.1109/VTCSpring.2013.6691814.
[203] A. Skrzypczak, P. Siohan, J.-P. Javaudin, Analysis of the peak-to-average power ratio for OFDM/OQAM, in: IEEE 7th Workshop on Signal Processing Advances in Wireless Communications, 2006, 2006, pp. 1–5, doi:10.1109/SPAWC.2006.346413.
[204] M. Schellmann, Z. Zhao, H. Lin, P. Siohan, N. Rajatheva, V. Luecken, A. Ishaque, FBMC-based air interface for 5G mobile: challenges and proposed solutions, in: 2014 9th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), 2014, pp. 102–107, doi:10.4108/icst.crowncom.2014.255708.
[205] M. Payaro, A. Pascual-Iserte, M. Najar, Performance comparison between FBMC and OFDM in MIMO systems under channel uncertainty, in: European Wireless Conference, 2010, pp. 1023–1030, doi:10.1109/EW.2010.5483521.

[206] I. Estella, A. Pascual-Iserte, M. Payaro, OFDM and FBMC performance comparison for multistream MIMO systems, in: Future Network and Mobile Summit, 2010, pp. 1–8.
[207] R. Zakaria, D. Le Ruyet, A novel filter-bank multicarrier scheme to mitigate the intrinsic interference: application to MIMO systems, IEEE Trans. Wireless Commun. 11 (3) (2012) 1112–1123, doi:10.1109/TWC.2012.012412.110607.
[208] Y. Chen, F. Schaich, T. Wild, Multiple access and waveforms for 5G: IDMA and universal filtered multi-carrier, in: IEEE 79th Vehicular Technology Conference (VTC Spring), 2014, pp. 1–5, doi:10.1109/VTCSpring.2014.7022995.
[209] V. Vakilian, T. Wild, F. Schaich, S. ten Brink, J.-F. Frigon, Universal-filtered multi-carrier technique for wireless systems beyond LTE, in: IEEE Globecom Workshops (GC Wkshps), 2013, pp. 223–228, doi:10.1109/GLOCOMW.2013.6824990.
[210] X. Wang, T. Wild, F. Schaich, Filter optimization for carrier-frequency- and timing-offset in universal filtered multi-carrier systems, in: IEEE 81st Vehicular Technology Conference (VTC Spring), 2015, pp. 1–6, doi:10.1109/VTCSpring.2015.7145842.
[211] G. Fettweis, M. Krondorf, S. Bittner, GFDM - generalized frequency division multiplexing, in: IEEE 69th Vehicular Technology Conference, 2009, pp. 1–4, doi:10.1109/VETECS.2009.5073571.
[212] N. Michailow, M. Matthe, I. Gaspar, A. Caldevilla, L. Mendes, A. Festag, G. Fettweis, Generalized frequency division multiplexing for 5th generation cellular networks, IEEE Trans. Commun. 62 (9) (2014) 3045–3061, doi:10.1109/TCOMM.2014.2345566.
[213] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen, S. Brink, I. Gaspar, N. Michailow, A. Festag, L. Mendes, N. Cassiau, D. Ktenas, M. Dryjanski, S. Pietrzyk, B. Eged, P. Vago, F. Wiedmann, 5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications, IEEE Commun. Mag. 52 (2) (2014) 97–105, doi:10.1109/MCOM.2014.6736749.
[214] A. Awoseyila, C. Kasparis, B. Evans, Improved preamble-aided timing estimation for OFDM systems, IEEE Commun. Lett. 12 (11) (2008) 825–827, doi:10.1109/LCOMM.2008.081054.
[215] A. Aminjavaheri, A. Farhang, A. RezazadehReyhani, B. Farhang-Boroujeny, Impact of Timing and Frequency Offsets on Multicarrier Waveform Candidates for 5G, arXiv preprint arXiv:1505.00800(2015).
[216] A. Farhang, N. Marchetti, L. Doyle, Low complexity modem design for GFDM, IEEE Trans. Signal Process. PP (99) (2015) 1–1, doi:10.1109/TSP.2015.2502546.
[217] R. Datta, G. Fettweis, Z. Kollar, P. Horvath, FBMC and GFDM interference cancellation schemes for flexible digital radio PHY design, in: 14th Euromicro Conference on Digital System Design (DSD), 2011, pp. 335–339, doi:10.1109/DSD.2011.48.
[218] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, K. Higuchi, Non-orthogonal multiple access (NOMA) for cellular future radio access, in: IEEE 77th Vehicular Technology Conference (VTC Spring), 2013, pp. 1–5, doi:10.1109/VTCSpring.2013.6692652.
[219] A. Benjebbour, Y. Saito, Y. Kishiyama, A. Li, A. Harada, T. Nakamura, Concept and practical considerations of non-orthogonal multiple access (NOMA) for future radio access, in: 2013 International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS), 2013, pp. 770–774, doi:10.1109/ISPACS.2013.6704653.
[220] M.-R. Hojeij, J. Farah, C. Nour, C. Douillard, Resource allocation in downlink non-orthogonal multiple access (NOMA) for future radio access, in: 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), 2015, pp. 1–6, doi:10.1109/VTCSpring.2015.7146056.
[221] A. Benjebbour, A. Li, Y. Saito, Y. Kishiyama, A. Harada, T. Nakamura, System-level performance of downlink NOMA for future LTE enhancements, in: 2013 IEEE Globecom Workshops (GC Wkshps), 2013, pp. 66–70, doi:10.1109/GLOCOMW.2013.6824963.
[222] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I, Z. Wang, Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends, IEEE Commun. Mag. 53 (9) (2015) 74–81, doi:10.1109/MCOM.2015.7263349.
[223] I. Da Silva, E. Authors, S.E. El Ayoubi, O.M. Boldi, Ö. Bulakci, P. Spapis, M. Schellmann, H. ERC, J.F. Monserrat, T. Rosowski, et al., 5G RAN architecture and functional design, METIS II White Paper.
[224] A. Alkhateeb, O. El Ayach, G. Leus, R. Heath, Channel estimation and hybrid precoding for millimeter wave cellular systems, IEEE J. Sel. Topics Signal Process. 8 (5) (2014) 831–846, doi:10.1109/JSTSP.2014.2334278.
[225] AT&T, AT&T Unveils 5G Roadmap Including Trials In 2016. http://about.att.com/story/unveils_5g_roadmap.html.
[226] Qualcomm, 5G - Vision for the next generation of connectivity, White Paper, 2015 https://www.qualcomm.com/invention/technologies/5g.
[227] Y.J. Bultitude, T. Rautiainen, IST-4-027756 WINNER II D1. 1.2 V1. 2 WINNER II Channel Models.
[228] J. Medbo, K. Borner, K. Haneda, V. Hovinen, T. Imai, J. Jarvelainen, T. Jamsa, A. Karttunen, K. Kusume, J. Kyrolainen, P. Kyosti, J. Meinila, V. Nurmela, L. Raschkowski, A. Roivainen, J. Ylitalo, Channel modelling for the fifth generation mobile communications, in: 2014 8th European Conference on Antennas and Propagation (EuCAP), 2014, pp. 219–223, doi:10.1109/EuCAP.2014.6901730.
[229] The Institute of Electrical and Electronics Engineers (IEEE), IEEE Standard for Information technology–Telecommunications and information exchange between systems–Local and metropolitan area networks–Specific requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band, IEEE Std 802.11ad-2012 (Amendment to IEEE Std 802.11-2012, as amended by IEEE Std 802.11ae-2012 and IEEE Std 802.11aa-2012) (2012) 1–628, doi:10.1109/IEEESTD.2012.6392842.
[230] METIS II Project, Preliminary Views and Initial Considerations on 5G RAN Architecture and Functional Design, 2015 https://metis-ii.5g-ppp.eu/.
[231] mmMAGIC Project, mmMAGIC: Millimetre-Wave Based Mobile Radio Access Network for Fifth Generation Integrated Communications, 2015 https://5g-ppp.eu/mmmagic/.
[232] G. Wunder, P. Jung, M. Kasparick, T. Wild, F. Schaich, Y. Chen, S. Brink, I. Gaspar, N. Michailow, A. Festag, L. Mendes, N. Cassiau, D. Ktenas, M. Dryjanski, S. Pietrzyk, B. Eged, P. Vago, F. Wiedmann, 5GNOW: non-orthogonal, asynchronous waveforms for future mobile applications, IEEE Commun. Mag. 52 (2) (2014) 97–105, doi:10.1109/MCOM.2014.6736749.
[233] A. Banchs, M. Breitbach, X. Costa, U. Doetsch, S. Redana, C. Sartori, H. Schotten, A novel radio multiservice adaptive network architecture for 5G networks, in: IEEE 81st Vehicular Technology Conference (VTC Spring), 2015, pp. 1–5, doi:10.1109/VTCSpring.2015.7145636.
[234] 5th Generation Public Private Partnership, mmMAGIC: Millimetre-Wave Based Mobile Radio Access Network for Fifth Generation Integrated Communications, 2015 http://5g-ppp.eu.
[235] Samsung Electronics, Samsung Electronics Sets 5G Speed Record at 7.5Gbps, Over 30 Times Faster Than 4G LTE, 2014 http://www.samsung.com/global/business/networks/insights/news.
[236] C.-L. I, C. Rowell, S. Han, Z. Xu, G. Li, Z. Pan, Toward green and soft: a 5G perspective, IEEE Commun. Mag. 52 (2) (2014) 66–73, doi:10.1109/MCOM.2014.6736745.
[237] Huawei, 5G: New Air Interface and Radio Access Virtualization, Technical Report, 2015.

Ian F. Akyildiz received the B.S., M.S., and Ph.D. degrees in Computer Engineering from the University of Erlangen-Nurnberg, Germany, in 1978, 1981 and 1984, respectively. Currently, he is the Ken Byers Chair Professor in Telecommunications with the School of Electrical and Computer Engineering, Georgia Institute of Technology (Georgia Tech), Atlanta, GA USA; the Director of the Broadband Wireless Networking (BWN) Laboratory and the Chair of the Telecommunication Group at Georgia Tech. Since 2013, he is a FiDiPro Professor (Finland Distinguished Professor Program (FiDiPro) supported by the Academy of Science) with the Department of Electronics and Communications Engineering, at Tampere University of Technology, Finland, and the founding director of NCC (Nano Communications Center). Since 2008, he is also an honorary professor with the School of Electrical Engineering at Universitat Politcnica de Catalunya (UPC) in Barcelona, Catalunya, Spain, and the founding director of N3Cat (NaNoNetworking Center in Catalunya). Since 2011, he is a Consulting Chair Professor at the Department of Information Technology, King Abdulaziz University (KAU) in Jeddah, Saudi Arabia. He is the Editor-in-Chief of Computer Networks (Elsevier) Journal, and the founding Editor-in-Chief of the Ad Hoc Networks (Elsevier) Journal, the Physical Communication (Elsevier) Journal and the Nano Communication Networks (Elsevier) Journal. He is an IEEE Fellow (1996) and an ACM Fellow (1997). He received numerous awards from IEEE and ACM. His h-index is 88 and the total number of citations is above 69K due to Google scholar as of March 2015. His current research interests are in wireless sensor networks in challenged environments, 5G Cellular systems, nanonetworks, Terahertz Band, and software defined networks.

Shuai Nie received the B.S. degree in Electrical Engineering from Xidian University in 2012, and the M.S. degree in Electrical Engineering from New York University in 2014. She is a graduate research assistant in the Broadband Wireless Networking Laboratory (BWN Lab), School of Electrical and Computer Engineering, Georgia Institute of Technology. Currently, she is working toward the Ph.D. degree in Electrical and Computer Engineering under the supervision of Prof. Ian F. Akyildiz. Her research interests include the 5G cellular system and telecommunication networks.

Shih-Chun Lin received the B.S. degree in Electrical Engineering and the M.S. degree in Communication Engineering from National Taiwan University in 2008 and 2010, respectively. He is a graduate research assistant in the Broadband Wireless Networking Laboratory (BWN Lab), School of Electrical and Computer Engineering, Georgia Institute of Technology. Currently, he is working toward the Ph.D. degree in Electrical and Computer Engineering under the supervision of Prof. Ian F. Akyildiz. His research interests include wireless sensor networks, software defined networking, large machine-to-machine communication, cognitive radio networks, and statistical scheduling in wireless systems.

Manoj Chandrasekaran received the B.Tech. degree in Electronics and Communications Engineering from Amrita Vishwa Vidyapeetham University, Coimbatore, India, in 2011 and the M.Tech. degree in Electrical Engineering from Indian Institute of Technology Kanpur, India in 2013. From July 2013 to July 2015, he worked as a Software Development Engineer at Microsoft India R&D Pvt. Ltd. Currently, he is pursuing Ph.D. degree under the supervision of Prof. Ian F. Akyildiz in the Broadband Wireless Networking Lab of School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia. His current research is focused on 5G cellular networks.