There has been active research worldwide to develop the next-generation, i.e., fifth-generation (5G), wireless network. The 5G network is expected to support a significantly large amount of mobile data traffic and a huge number of wireless connections and achieve better cost- and energy-efficiency as well as quality of service (QoS) in terms of communication delay, reliability, and security. To this end, the 5G wireless network should exploit the potential of new developments, including superdense and heterogeneous deployment of cells and massive antenna arrays [i.e., massive
Overview of 5G Wireless Systems

Research on next-generation 5G wireless systems, which aims to resolve several unprecedented technical requirements and challenges, has attracted growing interest from both academia and industry in the past few years. More than 5 billion devices demand wireless connections that run voice, data, and other applications in today’s wireless networks [1]. A study performed by Wireless World Research Forum predicted that 7 trillion wireless devices will be served by wireless networks, for both human and machine-type communications, in 2017. Furthermore, mobile data traffic has increased dramatically over the years, driven mainly by the massive demand of data-hungry devices, such as smartphones and tablets, and broadband wireless applications, such as multimedia, three-dimensional (3-D) video games, e-health, and Car-to-X (Car2X) communications [1]–[3].

This trend will continue, and it is expected that the 5G wireless network will deliver about 1,000 times more capacity than the current fourth-generation (4G) system. In addition, significant improvement in communications QoS is expected in the 5G network.

In particular, real-time support will be the key requirement to realizing many emerging wireless applications. In fact, real time is a highly subjective term and depends on the specific use case. According to [4], a service can be defined as real time when the communication response time (i.e., round-trip latency) is faster than the time constants of the application. Moreover, different use cases require different round-trip latencies of interactions. For instance, the latency requirement for an audio signal is around 80 ms, which can be supported by the current long-term evolution (LTE), with a typical round-trip latency of 25 ms [1], [5]. Although the LTE latency is sufficient for most current services, the number of new use cases that require very small latency is anticipated to rise in the future 5G network, including two-way gaming, virtual and enhanced reality (e.g., networked wearable computed devices), and touch-screen-enabled applications (i.e., a tactile Internet). Among these use cases, the tactile Internet requires more stringent latency, which is on the order of 1 ms [4], [6]. The tactile Internet is predicted to influence our daily lives and transform important socioeconomic sectors such as health care, education, the smart grid, and intelligent transportation.

There is, therefore, a strong need to achieve these technical requirements while improving the cost and energy efficiency of the future wireless network. The exponential growth of mobile data in recent years contradicts the flattening of the revenue of mobile operators. The 5G network should be able to realize cost-efficient wireless technologies (measured in US$/b). Energy-related costs account for a significant portion of the overall operational expenditure of wireless operators [7]. More than 70% of the mobile operator’s electricity bill is due to the radio part of the wireless cellular network [7], [8]. In addition, the telecommunications sector’s contribution of carbon dioxide (CO₂) to global CO₂ emissions has increased rapidly over the last decade, and mobile operators are among the top energy consumers. Thus, apart from spectral efficiency, energy efficiency is a crucial design objective to reduce operating costs for mobile operators as well as to minimize the environmental impact of the wireless domain.

To resolve these challenges, it is essential to adopt a network infrastructure that can efficiently integrate various disruptive wireless technologies and to enable inter-networking of existing and newly deployed technologies. Such development should consider the emerging wireless applications and services in the short, medium, and long terms. In particular, the 5G network should enable us to realize the truly networked society with unlimited access of information for anyone, anywhere, and anytime. It should also allow us to support various smart infrastructures and smart cities that are green, safe, mobile, connected, and informed [1].

The air-interface, spectral efficiency, available spectrum bands, and number of deployed base stations (BSs) are the key contributors to the performance of any network. Toward this end, it is important to realize network densification in multiple dimensions, including deployment of superdense HetNets with different types of cells, multiple radio access technologies (RATs), massive MIMO at BSs and/or user equipment (UE), and exploitation of both the microwave and mmWave frequency bands. The realization of these technologies will lead to the full-scale 5G HetNet, which poses various challenges at the architecture as well as communications and networking levels.

Requirements of the 5G Wireless Network

The 5G wireless network has not been standardized yet. The detailed and exact technical specifications of this network will be available in the near future. However, the following technical requirements are already accepted by the wireless industry and academia [1]–[3].

Coverage and Data Rate

The 5G network should maintain connectivity anytime and anywhere with a minimum user-experience data rate of 1 Gb/s [9]. In general, because the low-mobility UE channels change much more slowly than those of

multiple-input, multiple-output (MIMO) technologies] and utilization of higher frequencies, particularly millimeter-wave (mmWave) frequencies. This article discusses the potential benefits and challenges of the 5G wireless heterogeneous network (HetNet) that incorporates massive MIMO and mmWave technologies.
high-mobility UE, they require more resources for channel state information (CSI) acquisition (i.e., reduced effective data rate). Therefore, the peak data rates required by high- and low-mobility users in the 5G network can be different. The network must also ensure a certain QoS for users traveling at very high speeds (e.g., on high-speed trains traveling at 500 km/h), where the existing networks cannot satisfactorily support users. (The 4G network can support mobility of up to 250 km/h.)

Latency
The latency requirement is usually more difficult to achieve compared with that of the data rate as it demands that the data be delivered to the destination within a given period of time. For the 5G network, the end-to-end latency requirement will be on the order of 1–5 ms [2], [3], [6].

Connected Devices
The future 5G network is expected to incorporate a massive number of connected devices, which may reach up to 100 times that of the current wireless network. Potential use cases in this regard are wearable computing, machine-type communications, wireless sensors, and the Internet of Things [1], [7]. Importantly, these connected devices may have different requirements in terms of communication rate, delay, and reliability.

Multiple RATs
The 5G network will not be developed to replace current wireless networks but rather to advance and integrate the existing network infrastructures with new ones. In the 5G network, the existing wireless technologies, including global system for mobile communications, third-generation (3G), high-speed packet access, LTE and LTE-advanced, and Wi-Fi technologies, will continue to evolve and be integrated into a unified system [1], [7].

Energy and Cost Efficiency
5G wireless technologies must be designed to achieve significantly better cost efficiency (measured in US$/b) to address mobile operators’ concerns about revenue flattening. In particular, the energy efficiency (measured in b/J) of the 5G network may need to be improved by a factor of 1,000 compared with that achieved by current wireless technologies [1].

5G Wireless HetNet Based on mmWave and Massive MIMO
Network densification through the massive deployment of cells of different types, such as macro-, micro-, pico-, and femtocells, is a key technique to enhance the network capacity, coverage performance, and energy efficiency. This densification approach has been adopted in the existing wireless cellular networks, particularly 3G and 4G LTE systems, which essentially results in a multilayer cellular HetNet [8], [10]. Wireless HetNets may also comprise remote radio heads (RRHs) and wireless relays, which can further boost the network performance. Relaying and multihop communications are expected to be among the central elements of the 5G wireless architecture (in contrast to the existing LTE system, where multihop communications have been considered as an additional feature) [3]. In general, radio resource management for HetNets plays a crucial role in achieving the benefits of this advanced architecture. Specifically, the development of a resource-allocation algorithm that efficiently utilizes radio resources, including bandwidth, transmission power, and antennas, while mitigating intercell and interuser interference and guaranteeing acceptable QoS for active users, is one of the most critical issues. In addition, the design and deployment of reliable backhaul networks that enable efficient resource management and coordination are also very important.

The massive MIMO and mmWave technologies provide vital means to resolve many technical challenges of the future 5G HetNet, and they can be seamlessly integrated with the current networks and access technologies. The deployment of a massive number of antennas at the transmitter and/or receiver (massive MIMO) can significantly enhance the spectral and energy efficiency of the wireless network [11], [12]. In a rich scattering environment, these performance gains can be achieved with simple beamforming strategies, such as maximum ratio transmission (MRT) or zero forcing (ZF) [11]. Moreover, most of today’s wireless systems operate at microwave frequencies below 6 GHz. The sheer capacity requirement of the next-generation wireless network would inevitably demand us to exploit the frequency bands above 6 GHz, where the mmWave frequency range (30–300 GHz) can offer a huge segment of spectrum that is still underutilized [1], [2], [13]. Most importantly, as the mmWaves have an extremely short wavelength, it becomes possible to pack a large number of antenna elements into a small area, which consequently helps realize massive MIMO at both the BSs and UE.

In particular, mmWave frequencies can be used for outdoor point-to-point backhaul links or for supporting indoor high-speed wireless applications (e.g., high-resolution multimedia streaming). In fact, mmWave technologies have already been standardized for short-range services in IEEE 802.11ad. However, these frequencies
have not been well explored for cellular applications. Some potential reasons are the high propagation loss, penetration loss, and rain fading, and these frequencies are easily absorbed or scattered by gases [14]. The massive deployment of small cells, such as pico- and femtocells in the future 5G HetNet, renders the short-range mmWave technologies very useful. Therefore, the mmWave frequencies can be considered one of the potential technologies to meet the requirements of the 5G network.

There are many possibilities to enable the 5G wireless HetNet incorporating the massive MIMO and mmWave technologies. One such 5G network architecture is shown in Figure 1, where we demonstrate how massive MIMO and mmWave technologies can be used in different parts and for different communication purposes. The architecture of Figure 1 employs both mmWave and microwave frequencies. To determine the operating frequency bands of different communications in the architecture of Figure 1, several factors may need to be considered, such as the regulatory issues, application, channel, and path loss characteristics of various frequency bands. In general, path loss increases as the carrier frequency increases. This observation leads to the utilization of microwave frequencies for long-range outdoor communications.

In mmWave frequency bands, different frequencies have distinct behaviors. For example, in general, the oxygen molecule \(O_2\) naturally absorbs electromagnetic energy at 60 GHz to a much higher degree than in the regions 30–160 GHz. This absorption weakens (attenuates) 60-GHz signals over distance significantly; thus, the signals cannot reach far-away users. This makes 60 GHz suitable for high-data-rate and secure indoor communications. Hence, the selection of operating frequency depends on several factors, such as application, different absorptions, and blockages. Given these factors, however, there is a general consensus that mmWave frequency bands (30–300 GHz) can be useful for backhaul links, indoor, short-range, and line-of-sight (LOS) communications.

In general, the deployment of multiple antennas at the transmitter and/or receiver improves the overall performance of a wireless system. This performance improvement is achieved when the channel coefficients corresponding to different transmit–receive antennas experience independent fading. For a given carrier frequency, such independent fading is exhibited when the distance between two antennas is at least \(0.5 \lambda\), where \(\lambda\) is the wavelength [12], [15]. Thus, for a fixed spatial dimension, the number of deployed antennas increases as the carrier frequency increases, which consequently allows a large number of antennas to be deployed at mmWave frequencies. Moreover, the deployment of massive MIMO can be realized in different transportation systems such as trains and buses even at microwave frequency bands since sufficient space is available to do so (Figure 1).

**Figure 1** The potential 5G HetNet network architecture incorporating massive MIMO and mmWave technologies.
In recent years, 3-D and full-dimensional (FD) MIMO techniques have been promoted to increase overall network efficiency as they allow cellular systems to support a large number of UE using multiuser MIMO techniques. Thus, the massive MIMO and mmWave systems of the considered architecture can also be designed to be either 3-D or FD. On the other hand, this architecture can support coordinated multipoint (CoMP) transmission, where BSs are coordinated using either fiber or wireless backhauls. In addition, Figure 1 incorporates a cell virtualization concept, where the virtual cell can be defined as either the network (network-centric) or the users (user-centric) and can also be integrated as part of the cloud radio access network [16], [17].

**Research Challenges**

The 5G network presents significantly enhanced requirements compared with those in existing wireless networks. While the wireless HetNet incorporating the massive MIMO and mmWave technologies is expected to enable us to meet the technical requirements of the 5G network, there are various challenges to tackle. In the following sections, some key challenges and potential solutions are discussed.

**Network Planning and Traffic Management**

Considering the future mobile data traffic demand for the 5G wireless HetNet, network planning is an important issue to address. In fact, mobile operators would only optimize the placement and planning of certain network nodes (BSs), i.e., the numbers and locations of macro-, micro-, and pico-BSs. This optimization must be performed considering that a large number of low-power femtocells are unplanned and installed quite arbitrarily by the end users. Moreover, the fact that a large amount of mobile data traffic can be offloaded from one RAT to another tier of the same RAT or a different RAT must be considered in the network planning. This design task can be even more complicated since the network capacity offered by the indoor infrastructure may not be easy to estimate and can vary over time.

Hence, existing planning tools for site locations—such as using an evenly spaced lattice structure with a hexagonal grid—are unlikely to be good solutions for all network tiers, even though they may still be reasonable for the macro-BSs. In the current cellular network, the site locations and zoning of the BSs are determined from limited but expensive field tests, while link and system-level modeling consider the carrier frequency, propagation environment, and antenna characteristics. However, the usefulness of this site acquisition approach may need to be verified for the 5G wireless HetNet.

Because the 5G network can be operated at both microwave and mmWave frequencies, the optimal frequency-management and planning strategy is also not clear. This is especially true because the cellular network has asymmetric uplink and downlink traffic demands. Furthermore, the mmWave frequencies have high penetration losses, which are easily absorbed or scattered by gases [14]. Consequently, the frequency-management strategy at mmWave frequency bands may perhaps be location dependent (e.g., more mmWave frequency bands can be assigned when the surrounding environment has fewer buildings). In summary, the network planning and optimization of the 5G network should consider several unique parameters, such as high cell density, extremely time-varying traffic, and the fact that various frequency regions have quite different propagation characteristics.

**Radio Resource Management**

The capacity of future 5G networks would increase significantly with considerably higher cell densification, usage of higher mmWave frequencies, and the advanced massive MIMO technologies. However, this HetNet architecture poses several challenges for radio resource management. At mmWave frequency bands, the channel vector of each UE would have the specular characteristic, which is the result of very few scatterers. Consequently, the beam pattern of a given UE is concentrated mainly around the direction of the LOS path. Thus, UE having different LOS paths would minimally interfere with each other. However, since the signal power of mmWave communication systems experiences high attenuation and the transmission bandwidth of mmWave systems is fairly large, noise will be a limiting factor for mmWave frequency communications.

On the other hand, microwave channels are formed from several scatterers arising from different directions. Consequently, the beam-forming pattern designed for one UE would create strong interference for all other UE. For this reason, interference is a major limiting factor for microwave communication systems [6]. Thus, interference management is very crucial for microwave frequency bands. In particular, interuser and intercell interference becomes more challenging to manage in the dense environment. This is because the cross-tier interference between macrocells and small cells can only be mitigated efficiently if heavy coordination between the network tiers is performed by using high-speed and reliable backhauls, which are difficult to realize in practice.

Simple interference-management methods based on orthogonal spectrum allocation approaches can be applied [2] but are not so efficient for the dense 5G HetNet operating in the limited microwave spectrum. In fact, the U.S. Federal Communications Commission found that more than 80% of the microwave-frequency spectrum bands are not utilized efficiently. One way to improve the
spectral efficiency of these bands is to adopt a more flexible management approach of available spectrum using cognitive radio (CR) technology. In particular, the spectrum utilization of the 5G network can be improved by dynamically detecting the unused spectrum (i.e., white spaces) using different spectrum-sensing algorithms. Then, interference-free transmissions between UE and small-cell (macrocell) BSs can be enabled using the detected white spaces [18]. Different nonorthogonal spectrum-management approaches have also been recently developed, but most of them require cooperation among BSs of different tiers. Thus, it is better for the future 5G network to rely on autonomous interference-management techniques that enable low-complexity, distributed solutions and possibly CR technologies.

Furthermore, in recent years, 3-D and FD MIMO beamforming techniques have been promoted to increase the spectral and energy efficiency of MIMO systems, since these techniques would allow cellular systems to support a large number of UE using multiuser MIMO techniques both for single-cell and CoMP multicell systems. On the other hand, different requirements of the 5G network can be in conflict in such a way that improving one will degrade the others. To efficiently exploit all available resources, the future 5G network design must handle the existence of multiple objectives and inherent tradeoffs among them [19]. In particular, backhaul constrained multiobjective CoMP FD MIMO transceiver design presents an interesting design example with unique challenges.

To this end, the massive MIMO systems provide large degrees of freedom, which not only considerably increases the system spectral and energy efficiency but also fundamentally eases the interference management. For instance, in a multicell setup, the massive MIMO capacity gains can be achieved with simple and uncoordinated beamforming schemes, such as MRT or ZF (i.e., whenever there is no pilot contamination) [11]. However, since complete removal of pilot contamination is not a trivial task [note that one can efficiently mitigate the effect of pilot contamination (see the “Low-Cost CSI Acquisition and Beamforming for Massive MIMO” section)]. CoMP transmission could be employed to reap the full benefits of massive MIMO systems. However, enabling CoMP transmission for massive MIMO systems may require significant coordination between different cells, which may be practically infeasible as the number of UE and antennas deployed in the system is likely very large. In contrast, certain degrees of freedom (i.e., antenna elements) offered by the massive MIMO can be leveraged to mitigate or cancel multicell and/or cross-tier interference via suitable beamforming techniques. Fully exploiting massive MIMO to mitigate the interference among different cells with negligible BS coordination and handling multiobjective designs is a nontrivial challenge.

**The Deployment of Multiple Antennas at the Transmitter and/or Receiver Improves the Overall Performance of a Wireless System.**

The low-power BSs of the 5G HetNet can be installed and managed by both wireless operators (such as pico-BSs) and end users (such as femto-BSs). Unlike the operator-installed, low-power BSs, the low-power BSs deployed by the end users are installed quite arbitrarily. In addition, the unplanned low-power and indoor BSs can significantly outnumber the planned ones, while the low-power BSs would be the main capacity drivers of the future 5G network, since more than 70% of the traffic arises from the indoor environment. Therefore, these low-power nodes should possess self-configuration, self-optimization, and self-healing functionality [i.e., self-organized network (SON) characteristics] in resource and interference management so that efficient and scalable ultradense network deployment can be achieved. In this regard, ensuring cost-effective customer satisfaction will be the critical challenge of the SONs. On the other hand, the resource-allocation approach of the 5G network, unlike the traditional approach (which relies mainly on bit error rate), may need to consider different context information, such as application, environment, and required QoS in terms of the latency, bit error rate, and minimum data rate requirements. The context considers the user location, mobility, other proximity devices, resolution, central processing unit, and battery level of the device. How to deploy the SON that can enable a cost-effective, context-aware resource allocation approach is a challenging problem.

The 5G network is expected to achieve a minimum rate of 1 Gb/s in an energy-efficient manner. One potential way to achieve this goal is to adopt the virtual-cell (soft-cell) concept, which has been proposed for the dense wireless HetNet [20]. Specifically, radio resources available to a group of heterogeneous cells (macro-, micro-, and femtocells) are considered a single resource pool to serve each UE, which, in turn, views the group of cells as a big macrocell. Furthermore, the idea of the phantom cell, which was originally proposed by NTT DoCoMo, can be employed to better manage the control and data planes at different frequencies and nodes [21]. Specifically, the control and data planes are split so that critical control data are transmitted over reliable microwave links between the UE and macro-BSs, while high-speed data communications between UE and small-cell BSs are realized over mmWave frequency bands. Consequently, reliable and stable communications can be maintained while reaping the benefits of mmWave bands. (Sporadic interruptions of the data plane during the handover process between two small cells become less harmful.)
The 5G Wireless Network Has Not Been Standardized Yet.

In addition, macro-BSs are utilized to ensure wide-area coverage to maintain good connectivity and mobility. On the other hand, in a massive deployment of small cells, separating the control and data planes will improve the energy efficiency of the network [21]. These design approaches advocate the user-centric principle, and much more research is expected to realize it in practice.

Cell Association and Mobility Management

In the dense HetNet with multiple tiers and multiple RATs, such as cellular technologies of different generations, Wi-Fi, and Worldwide Interoperability for Microwave Access (WiMax), each UE can have several cell-association options during the lifetime of its active sessions. Cell association should be designed to efficiently utilize the network resources and provide acceptable QoS for UE. In general, small cells, such as picocells, femtocells, and Wi-Fi, are preferable to serve low-mobility UE, whereas high-mobility UE needs to be served by the macro-BS to avoid frequent handover. Furthermore, macro-BSs should also fill the coverage holes of the network to reduce the call-drop probability. For instance, in a HetNet network, a typical pico-BS has a maximum cell radius of around 200 m, and if one tries to associate UE on a high-speed train traveling at 250 km/h to this BS, handover may need to take place in a few seconds, which is quite undesirable.

The development of an optimal cell-association metric considering different factors, such as signal quality, interference, traffic loads, data-offloading capability, and mobility, is one of the key research issues. Moreover, the employed metric should enable decentralized implementation with low signaling overhead. Traditional association metrics based on signal strength or signal-to-noise ratio would not be sufficient for the future 5G HetNet. Obviously, the performance of the 5G network can be improved further if each UE can be associated with more than one node as in the virtual cell. Moreover, cell association must be jointly designed with mobility management so that low- and high-speed UE can be treated differently. In fact, high-speed UE is usually inside vehicles, which should be served by macro-BSs to avoid frequent handoffs. Furthermore, since vehicles can enable us to deploy massive MIMO, the reliability of the UE can be improved significantly.

Offloading UE to another tier of the same RAT or to another RAT offers an effective strategy to meet the requirements of the future 5G HetNet. In this respect, a fraction of user traffic can be rerouted either to another tier of the same RAT or to another RAT. Cell range expansion is one practical approach to offload traffic from macro-BSs to low-power nodes. To further improve the offloading capability, operators can deploy their own Wi-Fi access points (APs) to relieve the traffic congestion. These approaches may not be an ideal solution for the future network.

To achieve the requirements of the 5G network, the offloading approach for the 5G network may need to exploit the advantages of all RATs. In a typical metropolitan environment where several collocated RATs exist, discovering the best RAT for traffic offloading is not trivial. This is because different RATs can be administered by different operators (in the case of Wi-Fi AP, for example). On the other hand, as inter-RAT offloading utilizes the resources of other RATs, the offloading revenue should be shared between the RAT operators fairly. This offloading approach is particularly useful to disseminate nonprivate and delay-tolerant information, such as bulk data files, news, software, and files generated by laboratory experiments and sensor networks, which accounts for 64% of the current world mobile data traffic. Furthermore, the offloading method may need to consider location, user, and time-dependent RAT availabilities; incorporate both licensed and unlicensed bands at mmWave frequencies such as Wireless Gigabit Alliance (WiGIG); and consider the seamless service delivery switching time constraint between different RATs, which is around 10 ms [1]. Designing a cell-association and traffic-offloading algorithm by integrating all these issues is a challenging problem.

Backhaul Design

Backhaul links are required for data and signaling exchanges between BSs and between the core and access networks (i.e., the system of BSs). In general, high-capacity and reliable backhaul links support different types of traffic and cooperation between different BSs, which consequently improves the user experience and overall throughput of the network. There has been active research on the advanced CoMP transmission and reception techniques for the LTE-based system over the past few years. For the future dense HetNet, the deployment of an efficient backhaul network supporting coordination and signaling among BSs of different types is required.

Studies in 2012 showed that the backhaul of the existing cellular network comprises 70% copper, 10% fiber, and 15% wireless backhaul. And it would be preferable to deploy more wireless-over-wired backhaul technologies, although wired backhaul using optical fiber and Internet protocol (IP) cables would still be needed in the 5G network. In the existing wireless backhaul, large physical aperture antennas are used to achieve the required link gain, which is not economically desirable. Therefore, engineering a cost-effective, reliable, and scalable wireless backhaul network will be one of the fundamental challenges of the future 5G network.
MmWave and massive MIMO technologies provide great opportunities to resolve these challenges. Specifically, one can form a large number of beams to establish point-to-multipoint backhaul links by using massive MIMO [11]. In addition, mmWave offers virtually unlimited bandwidth for short-range backhaul links, which would be sufficient for the future dense HetNet (for instance, around 12.9 GHz of bandwidth is available in the E-band). This is illustrated in the hierarchical (mesh) backhaul network shown in Figure 1. Massive MIMO beamforming, if designed appropriately, can realize mmWave narrow beams that provide a unique opportunity to have scalable backhaul network with minimal (negligible) interference. One can utilize simple resource allocation approaches such as static partitioning (usually in the frequency domain) for the access and backhaul networks. A more efficient approach is to allow the backhaul and access networks to share the same resource pool. By doing so, the deployment flexibility of the network can be improved [15]. There has been some research in this direction, including design, resource allocation, and scheduling mechanisms [22]. However, the unique challenge of the latter approach is that as a low-power node operates on the same frequency for both backhaul and access links, each node may need to allocate backhaul resources dynamically according to its load and the meshed low power nodes. For this reason, reengineering the backhaul may need to be considered, taking into account the time-varying load characteristics of all low-power nodes.

**Low-Cost CSI Acquisition and Beamforming for Massive MIMO**

Beamforming is the key tool to exploit the potential of MIMO systems, which typically requires the availability of CSI at the transmitter and/or receiver. For the future 5G network that employs the massive MIMO at the microwave and/or mmWave frequency bands, CSI estimation and beamforming are the key design issues and strongly impact the network performance. The existing MIMO systems (such as LTE) are equipped with small numbers of antennas $N$ (from 1 to 10). For such a case, the number of radio frequency (RF) chains, digital-to-analog converters (DACs), and analog-to-digital converters (ADCs), which are the most expensive parts of a wireless transceiver, can be the same as that of the number of antennas. However, in a massive MIMO system with 100–1,000 antennas, deploying $N$ RF chains is not practically feasible. (Note that the number of RF chains is the same as that of DACs at the transmitter and ADCs at the receiver. Thus, we simply refer to RF chains as they implicitly include ADCs and DACs in the sequel.) On the other hand, the energy consumption of a MIMO transceiver increases with the number of active RF chains. Specifically, when the transmission bandwidth is very large, the energy consumption of ADCs would be unacceptably high. Thus, the channel estimation and beamforming algorithms should be designed taking into account the constraints on the number of RF chains (i.e., to be much less than $N$) and the finite resolutions of ADCs.

In the conventional channel estimation (i.e., the case of $N$ RF chains), the channels between transmitters and receivers are estimated from orthogonal pilot sequences that are limited in number due to the finite coherence time of the channel. In a massive MIMO setup, as $N \gg K$ (i.e., the number of UE), time division duplex-based communication would be more efficient where the pilot symbols are transmitted from UE [11]. In a multicell setup, however, orthogonal pilot sequences may need to be reused over the cells, which, unfortunately, results in the so-called pilot contamination.

Pilot contamination imposes a fundamental performance bottleneck for the massive MIMO system. (In a traditional multicell multiuser MIMO system, although pilot contamination arises, its effect is negligible or can be alleviated by employing appropriate coordinated BS transmission and CSI acquisition approaches.) Therefore, research on potential techniques to eliminate or mitigate the pilot-contamination effect through efficient CSI estimation and pilot optimization is important [11]. Different approaches have been proposed to reduce pilot contamination, such as eigenvalue decomposition-based channel estimation, successive pilot transmission, time-shifted channel estimation, and a linear combination approach that exploits multipath components [23], [24]. In this regard, it was shown recently in [24] that a maximum of $L$ (i.e., number of multipath components) cells can reliably estimate the channels of their UE and perform beamforming while efficiently mitigating the effect of pilot contamination, where $L$ can be increased just by increasing the transmission bandwidth. Nevertheless, all these approaches can only mitigate the effect of pilot contamination, and how to completely eliminate the effect of pilot contamination is still an open research problem.

When the number of RF chains $N_{RF}$ at the BS is much less than the number of BS antennas, the orthogonal channel estimation approach cannot be used even for a single-cell setup (i.e., without pilot contamination). To this end, there are two potential methods. The first method is to employ a combined time-frequency analog–digital channel estimation approach. In such an
approach, the high-dimensional section of the channel estimation process will be commonly employed for all UE and subcarriers in time-domain and analog form (i.e., depending on $N$). The low-dimensional part (i.e., that depends on $N_{RF}$) can be designed in digital form for each UE and subcarrier. In general, this approach is effective for an environment where the channel parameters have several multipath components (e.g., microwave frequency bands).

The second method is to leverage the direction-of-arrival (DOA) estimation approach from radar signal processing to mmWave channel estimation. The mmWave frequency bands are effective in a LOS environment, and detecting the DOA of the transmitted signal is sufficient in such an environment. In this regard, one can estimate the DOA at microwave frequencies and employ this DOA information for data transmission at mmWave bands. Specifically, this approach is quite interesting when the low-power node is capable of operating at both mmWave and microwave frequency bands, such as a WiGIG AP. [Note that in areas where there are multiple reflected rays, the DOA information of mmWave and microwave frequency bands may be different from each other. For such scenario, an environmental (context)-aware DOA estimation approach can be employed jointly with the CR technology [18].]

Once the channel estimation process has been completed, the next step is to perform beamforming. Again, the conventional MIMO system is based on digital beamforming that is performed at the baseband level, and it requires $N_{RF} = N$, which is very costly. To reduce the implementation cost of beamforming, a combined time-frequency analog–digital approach is one of the most promising ones. This beamforming approach will have both a high-dimensional analog part operating in the time domain and a low-dimensional digital part designed for each subcarrier [13]. The channel estimation and beamforming are interrelated; when the channel estimation is more inaccurate, the performance of the beamformer can become worse. Therefore, the critical challenge is how to jointly and optimally perform channel estimation and beamforming considering the number of RF chains and pilot contamination for the multiuser multicell massive MIMO system operating over the frequency selective channel.

**Case Studies**

In the following sections, we discuss some design and case studies that illustrate how to address some challenges of the 5G MIMO HetNet presented in the “Low-Cost CSI Acquisition and Beamforming for Massive MIMO” section.

**Low-Cost Hybrid Digital–Analog Beamforming for Multiuser Massive MIMO**

We demonstrate the desirable cost and performance tradeoff achieved by the hybrid beamforming for the multiuser massive MIMO system with large-scale antennas at both the transmitter and receiver sides [13]. As detailed earlier, the key advantage of hybrid beamforming is that it requires a smaller number of RF chains, ADCs, and DACs compared with the number of transmitter and receiver antennas, respectively. To investigate the performance achieved by hybrid and digital beamforming, we adopt a geometric channel model with $L$ scatterers, where each scatterer is assumed to contribute a single propagation path between the BS and each UE. Furthermore, the BS and each UE are equipped with a uniform linear arrays antenna with half wavelength spacing. The number of transmit antennas, transmitter RF chains, receiver antennas, and receiver RF chains is set equal to 128, 64, 32, and 16, respectively. The number of receivers is $K = 4$, and the number of multiplexed symbols for each receiver is $S_k = 8$ and $L = 16$. The sum rate of the digital and hybrid beamforming for the downlink multiuser massive MIMO system is shown in Figure 2. As can be seen from this figure, the hybrid beamforming approach of [13] achieves almost the same performance as that of the digital beamforming.

The result in this figure is obtained for the channel parameters discussed in this section. Thus, one could wonder whether the hybrid design of...
[13] can achieve the same performance as that of the digital one for arbitrary channel parameters. In this respect, [25] demonstrates via simulation that the hybrid beamforming design of [13] can experience a significant performance loss (compared with digital beamforming) when the number of scatterers is large. This motivates [25] to develop a hybrid beamforming utilizing a fixed (digitally controllable) paired phase shifters and switches, which achieves closer performance to that of the fully digital beamforming. This article also determines the required number of RF chains (and phase shifters), such that the hybrid design achieves the same beamforming performance as that of the digital one. Moreover, for the parameter setups, the hybrid design requires a maximum of $K_S K$ and $S K$ RF chains at the transmitter and receiver, respectively, to maintain the same performance as that of the digital design [25].

**Mobile Relay with Large-Scale Antenna Array**

We demonstrate the performance gain achieved by implementing large-scale antenna arrays at mobile relays, which can be deployed on public transportation vehicles, such as trains or buses, as shown in Figure 1, to better support highly mobile UE. In particular, we consider two downlink scenarios where a macro-BS communicates with individual UE units directly and through the mobile relay, which are referred as the first and second scenarios, respectively. In the second scenario, the communication between UE inside the vehicle and mobile relay is performed at mmWave frequency bands, where there is almost no capacity limitation. The BS is located at the center of its cell, whereas each UE is located randomly inside the vehicle of dimension $3.10 \text{ m} \times 25.91 \text{ m}$ (Amtrak passenger train) with uniform distribution.

The propagation model between the BS and each UE contains two components. One is the path-loss component decaying with distance, and the other one is the Rayleigh fading random component, which has a zero mean and unit variance. All of the other parameter settings are summarized in Table 1.

We compare the total rates achieved in these two scenarios. For the first scenario, we consider that the macro-BS employs the multiuser block diagonalization beamforming approach. In the second scenario, since the mmWave frequency band transmission is assumed to support a very high rate, the achievable rate is determined by the radio link between the macro-BS and mobile relay. For this scenario, the macro-BS employs singular-value-decomposition beamforming. The sum rate computed by the Shannon capacity formula of these two scenarios is shown in Figure 3. As we can see from this figure, the rate achieved by the second scenario (i.e., with mobile relay) is significantly higher than that of the first scenario (i.e., without mobile relay). In addition, we also see that increasing the number of mobile relay antennas improves the achievable rate, which is expected.

![Figure 3](image-url)  
**Figure 3** A comparison of the sum rates achieved with and without mobile relay. $M$ denotes the number of antennas at the mobile relay.
These results confirm that a large-scale antenna system (massive MIMO) can lead to a significant performance gain when utilized appropriately.

**Dual-Band Small Cells**

We now discuss a potential design for a small-cell network that operates in both microwave frequencies and mmWave frequencies [15]. In general, microwave frequencies have more favorable propagation properties but are more limited in bandwidth compared to the mmWave frequencies. Thus, a dual band small-cell with a prioritized and layered cell-association strategy, as shown in Figure 4, can exploit the advantages of these different frequency bands. In this design, the coverage area is divided into three regions, where the UE in the inner region (i.e., radius $a$) is served by the small cell at mmWave/microwave frequency bands. And the UE in the outer region (i.e., radius $> b$) is served by the macrocell at microwave frequency bands. Here, the outer region operates at microwave frequency bands, since these bands allow long-range communications to support mobility. The parameters $a$ and $b$ can be chosen or optimized adaptively based on different context information, as discussed in the “Mobile Relay with Large-Scale Antenna Array” section.

The middle region (i.e., with radius $b$) is analogous to the expanded range considered in the LTE standard, where its radius can vary from one design criteria to another [10]. This dual-band small-cell design isolates the mmWave and microwave frequency UE (i.e., the mmWave UE inside the radius $a$ does not experience any interference from the macro-BS, which is desirable). However, how to choose the best radius for different regions again depends on specific design objectives, which present interesting open research problems.

**Conclusions**

This article discusses the 5G wireless HetNet integrating massive MIMO and mmWave technologies. We also presented a potential HetNet architecture that employs both microwave and mmWave frequencies. Various design and technical challenges for this network architecture were described. We then presented three case studies addressing some of the challenges and showing their benefits. Overall, this article lays out the potential architecture and points out the underlying research challenges that must be addressed to meet the technical requirements of the future 5G network.

**Author Information**

**Tadilo Endeshaw Bogale** (tadilo.bogale@emt.inrs.ca) received his B.Sc. degree in electrical engineering from Jimma University, Ethiopia. From 2004 to 2007, he was with Ethio Telecom, Addis Ababa, Ethiopia. He received his M.Sc. and Ph.D. degrees in electrical engineering from Karlstad University, Sweden, and the University Catholique de Louvain, Louvain la Neuve, Belgium, in 2008 and 2013, respectively. In 2014, he was a post-doctoral researcher at the Institut National de la Recherche Scientifique (INRS), Montreal, Canada. Since November 2014, he has been working as a joint postdoctoral researcher with INRS and the University of Western Ontario, London, Canada. Currently, he is working on assessing the potential technologies to enable the future fifth-generation (5G) network. Specifically, his research focuses on the exploitation of massive multiple-input, multiple-output (MIMO) and millimeter-wave (mmWave) techniques for the 5G network. His research interests include hybrid analog–digital beamforming for massive MIMO and mmWave systems, pilot contamination reduction for multicell massive MIMO.
systems, spectrum sensing and resource allocation for cognitive radio networks, robust (nonrobust) transceiver design for multiuser MIMO systems, centralized and distributed algorithms, and convex optimization techniques for multiuser systems. He organized a workshop on cognitive radio for 5G networks, which was colocated with the 2015 EAI International Conference on Cognitive Radio Oriented Wireless Networks (CROWNCOM). He acted as a session chair for the IEEE International Communications Conference, the Conference on Information Systems and Sciences, and CROWNCOM. He has also served as a technical program committee member for different international conferences, such as the IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), CROWNCOM, and IEEE Vehicular Technology Conference. He recently delivered a tutorial at PIMRC 2015 on the 5G network.

Long Bao Le (long.le@emt.inrs.ca) received his B.Eng. degree in electrical engineering from Ho Chi Minh City University of Technology, Vietnam, in 1999, his M.Eng. degree in telecommunications from the Asian Institute of Technology, Bangkok, Thailand, in 2002, and his Ph.D. degree in electrical engineering from the University of Manitoba, Canada, in 2007. He was a postdoctoral researcher at the Massachusetts Institute of Technology from 2008 to 2010 and the University of Waterloo from 2007 to 2008. Since 2010, he has been with the Institut National de la Recherche Scientifique, Université du Québec, Montréal, Canada, where he is currently an associate professor. His current research interests include smart grids, cognitive radio, radio resource management, network control and optimization, and emerging enabling technologies for 5G wireless systems. He is a coauthor of the book Radio Resource Management in MultiTier Cellular Wireless Networks (Wiley, 2013). He is a member of the editorial boards of IEEE Transactions on Wireless Communications, IEEE Communications Surveys and Tutorials, and IEEE Wireless Communications Letters. He has served as a technical program committee chair/cochair for several IEEE conferences, including the IEEE Wireless Communications and Networking Conference, the IEEE Vehicular Technology Conference, and the IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications.

References


