Towards 5th Generation Cellular Mobile Networks

Jian A. Zhang, Peng Cheng, Andrew R. Weily and Y. Jay Guo

CSIRO Digital Productivity and Services Flagship

Summary

Cellular mobile networks have enabled ubiquitous communications and largely changed the way we live and work. At the same time, the network itself has been undergoing significant changes in the process of meeting our ever increasing demands on data rate and quality of service. In this article, we show the path of the evolution in both standards and techniques, and provide our vision for the future of the cellular networks. We review the evolution of international standards for cellular mobile networks in the last two decades, describe how the network layout has been migrating from rigid cellular architecture to random and dense small cells, and provide an in-depth discussion on potential enabling techniques for the next generation (5G) cellular networks, particularly massive MIMO and multiband base-station antennas.

Introduction

Cellular networks have been undergoing significant changes in the process of meeting our ever increasing demands on data rate and quality of service. Since its invention in 1980s, the cellular network has evolved from the first generation to the 4th generation, and is advancing towards the 5th generation. Some of the major changes, which have driven the development of cellular network and have vastly improved its overall performance, are listed below:

- From analogue to digital networks. The first generation cellular network was purely analogue, and digital systems have been used since the second generation;
- From narrow band to broadband communications. This change resulted from the demand for higher data rate, and led to fundamental changes of modulation and multiple access techniques;
- From circuit switched to packet switched. Packet switching based on the internet protocol (IP) has been dominating since the 3rd generation;
• From rigid cellular architecture to random and dense small cells. This change becomes apparent in the 4th generation and is becoming increasingly important in boosting system capacity;

• From hierarchical to flat network structure. Network architecture is becoming simpler to reduce processing complexity and backhaul requirements; and

• From pure base-stations for macro-cells to heterogeneous networks containing various base-station and access devices such as small cells, remote radio heads, and distributed antenna systems.

In this article, we track these changes and demonstrate the evolution of cellular networks in both standards and technologies, and provide our vision for its future. We will first review the evolution of international standards for cellular mobile networks, highlighting their key techniques. We then investigate how the network layout has been migrating from rigid cellular architecture to random and dense small cells, and discuss techniques that are being used to deal with problems associated with such layout variation. We will also provide in-depth discussions for potential enabling techniques for the next generation (5G) cellular networks, particularly massive multiple-input multiple-out (MIMO) and multiband base-station antennas.

Evolution of Mobile Standards from 2G to 5G

Mobile systems have evolved from the first generation to the current fourth generation, and are evolving towards the fifth generation. Mobile standards have evolved approximately at the pace of one generation every five years since the 2nd generation. The evolution of mobile standards from 2G to 5G is depicted in Fig. 1 with key features highlighted. In Table 1, key parameters of some main mobile systems are also listed. For more details, interested readers are referred to (Ghosh et al. 2010).
2G

2G mobile systems were introduced in early 1990s, and are still being used today. They are mainly based on circuit-switched data transmission, and support voice services. Their evolved systems also support packet-switched and data services such as SMS. Typical 2G systems include the Global System for Mobile Communications (GSM), IS-95 code division multiple access (CDMA), and IS-136 time division multiple access (TDMA) systems. Among them, GSM is the most widely deployed.

3G

3G mobile systems provide much higher data rates and better quality of service than 2G systems by using larger bandwidth and primarily code division multiple access (CDMA) techniques. Among the many 3G standards, wideband CDMA (WCDMA), CDMA-2000 and TD-SCDMA are most widely used. WCDMA and CDMA-2000 are evolutions of GSM and IS-95 CDMA, respectively, while TD-SCDMA is a new standard proposed for operation in unpaired spectrum using time division duplexing. Due to the popularity of GSM, WCDMA is still the most popular 3G system in the world.

The first WCDMA standard was published in 1999. Its air interface is based on wideband CDMA, where a single data symbol is spread to multiple samples of chips by multiplying with...
a pseudo-random sequence. Alamouti space-time coding is also introduced to achieve transmit diversity. At the network layer, WCDMA retains the basic architecture of GSM/GPRS networks, with network elements enhanced for 3G capabilities. Starting from 2002, high-speed packet access (HSPA) was introduced to achieve higher-speed downlink and uplink packet access by using new techniques such as multicode transmission, short transmission time interval, adaptive modulation and coding, fast dynamic scheduling and hybrid automatic repeat request.

4G

The first long term evolution (LTE) standard was released in 2009 as 3GPP release 8 (http://www.3gpp.org/specifications/releases/72-release-8). Enhanced versions since release 10 in 2011 are known as LTE-Advanced standards. Although LTE does not meet the requirements of ITU for 4G wireless standard such as 1 Gbps peak data rate for pedestrian users, it, together with LTE-Advanced, are referred to as 4G systems due to several major changes they share, compared to 3G systems.

The major differentiating techniques between 4G and 3G include orthogonal frequency division multiplexing (OFDM); channel dependent multiuser resource scheduling; multi-antenna; and IP-based flat network architecture. OFDM has many advantages for broadband communications, such as low complexity equalisation in dense multipath channels, simple implementation complexity and flexibility in supporting multiple access. Since mobile users can have uncorrelated multipath channels, it is possible to allocate frequency subcarriers to users with the lowest attenuation at these subcarriers to achieve maximum diversity. Multi-antenna techniques, MIMO in particular, are also extensively used in 4G, including transmit diversity, beamforming, spatial multiplexing and multiuser MIMO (also known as spatial division multiple access, SDMA). The application of these techniques gives a large improvement in the data rate and reception performance. Besides the major changes in air-interface, the other radical aspect of 4G is the flat radio and core network architecture, which means fewer nodes and a less hierarchical structure for the network. In such a flat architecture, all services are supported on the IP packet network.

Compared to LTE, LTE-Advanced achieves higher data rates, spectral efficiency and spectral efficiency per coverage area. In particular, it uses wider bandwidth through Carrier Aggregation (CA); enhanced multi-antenna techniques such as downlink MIMO up to 8x8; multiuser MIMO, and uplink MIMO up to 4x4; heterogeneous network (HetNet) and Coordinated MultiPoint (CoMP) techniques. Carrier aggregation combines multiple channels, ranging from initially two 10 MHz channels to 45 being defined, to enable a cellular network to boost the data rate. The concept is also being extended to facilitate
aggregation across adjacent cells on the same carrier for load balancing (known as MultiFlow), and aggregation across paired and unpaired spectrum (LTE FDD and TDD). A HetNet is formed of a mixture of macro-, micro- and pico- base stations serving areas of different sizes. Different cells collaborate through CoMP techniques to mitigate interference and improve network performance. Small cells become the future of cellular evolution, as will be discussed in detail in next Section.

Currently, LTE-Advanced is being further improved to address the mobile data capacity challenges, such as extension to unlicensed spectrum bands in particular the 5 GHz ISM band, WiFi offloading, and LTE for broadcasting (Qualcomm 2014).

5G

Currently, 5G mobile networks are still a vision with the technical specifications yet to be written. It is estimated that 5G may be introduced in the late 2010s. Five disruptive technologies for 5G are described in Boccardi et al. (2014), including device-centric architectures, millimetre wave, massive MIMO, smarter devices and native support for machine-to-machine communications. The following techniques, in our opinion, could become the foundations of 5G:

- Small cell and centralised processing architecture such as the cloud radio access network (C-RAN), which combines wired and wireless networks and can significantly increase system capacity and provide great flexibility in dealing with network dynamics;

- Massive MIMO techniques which use a very large number of antennas to suppress interference and increase spatial multiplexing capabilities for both single and multiple users;

- Smart base-station multiband antenna array and millimetre wave hybrid array techniques, which are low-cost and efficient solutions to implementing small cell and massive MIMO technologies.

An in-depth review of these techniques is further provided in each of the following three sections.
### Table 1 Key parameters of main mobile standards

<table>
<thead>
<tr>
<th></th>
<th>GSM</th>
<th>WCDMA</th>
<th>LTE/LTE-Advanced</th>
<th>5G (Possibility)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Frequency Bands</strong></td>
<td>850/900 MHz, 1.8/1.9 GHz</td>
<td>850/900 MHz; 1.8/1.9/2.1 GHz</td>
<td>New spectrum: 700 MHz; 1.5/2.6 GHz</td>
<td>New spectrum at millimeter wave band from 28 to 95 GHz</td>
</tr>
<tr>
<td><strong>Channel Bandwidth</strong></td>
<td>200 KHz</td>
<td>5 MHz</td>
<td>1.4, 3, 5, 10,15 and 20 MHz</td>
<td>Up to a few hundred MHz</td>
</tr>
<tr>
<td><strong>Modulation</strong></td>
<td>GMSK</td>
<td>QPSK (+ 16QAM and 64QAM for HSPA)</td>
<td>QPSK, 16QAM and 64QAM</td>
<td>Higher modulation 256QAM</td>
</tr>
<tr>
<td><strong>Channel Coding</strong></td>
<td>Convolutional Coding</td>
<td>Convolutional Coding</td>
<td>Convolutional, Turbo and LDPC Coding</td>
<td>Low density parity check (LDPC) coding</td>
</tr>
<tr>
<td><strong>Multiple Access Scheme</strong></td>
<td>TDMA, FDMA</td>
<td>CDMA, TDMA</td>
<td>OFDMA, SC-FDMA</td>
<td>OFDM</td>
</tr>
<tr>
<td><strong>Peak Data rate</strong></td>
<td>107 Kbps (GPRS); 384 Kbps (EDGE)</td>
<td>384-2048 Kbps; 3.6-14.4 Mbps (HSPA Downlink)</td>
<td>LTE: 300 Mbps (Downlink); 75 Mbps (Uplink); LTE-Advanced: 1 Gbps (Downlink); 500 Mbps (Uplink)</td>
<td>Tens of Gbps</td>
</tr>
<tr>
<td><strong>Latency (ms)</strong></td>
<td>700 (GPRS)</td>
<td>&lt;200</td>
<td>&lt;30</td>
<td>lower latency</td>
</tr>
</tbody>
</table>

### Evolution of Cellular Layout

The cellular concept was a breakthrough in mobile communications. By dividing a large area into several smaller areas and replacing a single high-power transmitter with several lower-power ones, system capacity can be significantly increased. Automatic handover is applied to support mobile users moving between different cells. The cellular layout has experienced significant changes since its invention.

Traditional cellular networks (1-3G) have a regular cellular architecture, as shown in Fig. 2a. Each base-station is located in the centre of the cell, and each cell covers a similar area. Each cell uses a fraction of the total available spectrum and the same frequencies can be reused in sufficiently separated base stations. In such a cell architecture, the capacity of each cell is similar and hence it is inefficient in supporting a “hotspot” where the capacity demand is much larger than the average. Users at the edge of a cell can also have inferior performance due to increased signal attenuation and co-channel interference.

To satisfy non-uniform capacity requirements and to increase system capacity, the cell size becomes smaller, and the shape of cells become less regular. In particular, HetNet cells
overlap with each other, and one large cell can even contain several smaller cells, as shown in Fig. 2b where two smaller cells are located within the cell outlined in red. Such a HetNet cell architecture can efficiently support “hotspots”, adapt to different capacity requirements and largely increase the overall system capacity. However, co-channel interference becomes a major problem. 3GPP Release 11 and beyond have proposed many efficient approaches to support HetNet and solve its interference problem, such as CoMP and Range Expansion.

CoMP is designed to enable neighbouring cells to collaborate with each other, particularly in supporting users at the edge of the cell. Depending on the scheduling complexity and the capacity of the backhaul links between base stations, three types of CoMP techniques can be applied, including coordinated scheduling and coordinated beamforming, joint transmission and transmission point selection (Lee et al. 2012).

Range Expansion refers to a set of interference management tools that enable the association of more users to small cells to increase the overall network capacity (Lopez-Perez et al. 2011). Two key techniques, enhanced Inter Cell Interference Coordination (eICIC) from the network side and Interference Cancellation (IC) from the device side are the core of Range Expansion. Cells use eICIC to coordinate resource scheduling to minimize co-channel interference, and mobile devices use advanced IC techniques to connect to small cells even at very low signal to noise ratio (SNR).

Small Cell techniques are the future of mobile communications. Cells are becoming denser and the cell setup is becoming more random, as show in Fig. 2c. A base station is deployed primarily on demand without considering location optimisation. It can even be randomly deployed by users rather than the operator, in an integrated wired and wireless network such as the mobile over FTTx (MoF) networks (Zhang et al 2013). As discussed in (Zhang 2013), among many enabling techniques for such dense and random networks, self-organised networks (SON) and centralised processing are two critical ones.

SON techniques enable small cells to be deployed like plug-and-play devices. Each new small cell can obtain its neighbouring cell information via either self-calibration or a central controller, and automatically complete its resource configuration with little manual intervention. Centralised processing is an efficient solution to the severe co-channel interference problem in such dense small cells. A centralised processing network such as C-RAN can use distributed MIMO techniques to increase spatial multiplexing gain and achieve spatial division multiplexing access, hence solve co-channel interference and largely increase spectrum efficiency. To exploit the power of centralised processing, small cells can be recast as relaying nodes, and many baseband processing functions can be shifted from small cell base stations to a central point. The central baseband processing can also flexibly adapt to various channel and network dynamics using powerful general purpose processors.
To meet the increasing demand for wireless broadband services from fast-growing mobile users over the next decade, a potential technology termed massive MIMO (Rusek et al. 2013) (also known as large-scale MIMO, full-dimension MIMO, or hyper-MIMO) has emerged to further reap the benefits of utilising multiple antenna technology and promises orders-of-magnitude improvements in spectral-efficiency over 4G LTE-Advanced. With a large number of antennas (possibly hundreds or even thousands) at the wireless transmitter, the massive MIMO technology can not only improve link reliability, but also increase the radiated energy efficiency due to significant array gains (Ngo et al. 2013). Combining massive MIMO technology with time- and frequency-division multiplexing, the base station offers the possibility to transmit multiple messages for multiple terminals on the same time-frequency resource using SDMA, providing much higher multiplexing gains. The active terminals being served are much less than the antenna elements at the base-station in a massive MIMO system. Such disparity helps make the simplest precoding-decoding algorithms nearly optimal, and also enables flexible user scheduling and resource allocation.

While conventional MIMO techniques have been extensively studied both in the literature and practice, massive MIMO is still an emerging field where rigorous performance analysis and feasible engineering methodologies are limited. In this section, we present an overview of opportunities and challenges for massive MIMO.

**Benefits of Massive MIMO**

Employing a large number of antennas in a massive MIMO system brings a lot of benefits, some of which are specified as follows.
• **Massive MIMO can largely increase the link reliability and the data rate.** Theoretically, in a point-to-point MIMO link with $N$ transmit and $M$ receive antennas, the probability of link outage is proportional to $SNR^{-NM}$, which indicates that the link performance can be greatly improved with increased number of antennas. On the other hand, the achievable rate scales as $\min(N, M) \log (1+SNR)$, which offers the possibility of achieving high data rates using large $N$, $M$ without requiring an increase in bandwidth. The gains in point-to-multipoint multiuser systems will be more considerable, as such systems allow the transmitter to communicate with several co-channel terminals on the same time-frequency resource simultaneously.

• **Massive MIMO makes the channel matrix deterministic.** When the dimension of the random channel matrix is large, the distribution of its singular values becomes less sensitive to the actual distribution of the entries of the channel matrix. This means that the randomness of the channel observed in the small array regime starts to appear deterministic. Typically, the channel matrix tends to be well conditioned and the users are effectively spatially separated. This variation results in several advantages in large dimensional signal processing. For example, both theoretical analysis and measurements conducted in practice verify the asymptotic orthogonality between channels of different users. Therefore, simple linear processing techniques such as maximum-ratio combining for the uplink and maximum-ratio transmission for the downlink come close to the optimal performance bounds.

• **Massive MIMO can be constructed with inexpensive, low-power units.** The aim of massive MIMO is to offer all the benefits of conventional MIMO, but on a large scale, via simple, cheap and low-power hardware implementation. With massive MIMO, expensive power amplifiers are substituted by many low-cost low-power units. In terms of energy efficiency, this practice shows significant advantages over the classical MIMO architecture.

• **Massive MIMO can simplify the multiple access layer.** Based on the law of large numbers, the deterministic channel makes frequency domain scheduling unnecessary. Each subcarrier in a massive MIMO-OFDM system will approach the same channel gain. Each terminal can be allocated the whole bandwidth, which renders most of the physical layer control signaling redundant.
Challenges of Massive MIMO

Massive MIMO systems with hundreds or thousands of antennas can enable multi-gigabit rate transmissions at high spectral efficiency. However, several technical challenges need to be tackled in realising such large MIMO systems. Some of the challenges include:

- **Acquisition of Channel State Information.** To achieve the multiplexing gains of massive multiuser MIMO systems, it is crucial to employ precoding techniques at the transmitter to suppress co-channel interference. However, deploying multiuser precoding requires each user’s channel state information (CSI) be made available to the transmitter. Acquisition of CSI essentially limits the capacity of any precoding schemes. In the massive MIMO context, when accurate calibration of the hardware chains is performed, the TDD technique is applied to exploit channel reciprocity in order to simplify the procedure of obtaining channel state information. However, this problem becomes more challenging for Massive MIMO-OFDM FDD systems, where the CSI required by the transmitter increases prohibitively with the number of transmit antennas and subcarriers. As such, high CSI overhead of the massive MIMO-OFDM channel is likely to make conventional feedback algorithms inefficient. Novel CSI feedback schemes to reduce the CSI feedback to a large extent are highly desired.

- **Massive Array Hardwares and Architectures.** In practice, the implementation of massive antenna arrays needs some collocated structures such as linear arrays for antenna allocation. When increasing the number of parallel antennas and associated radio frequency (RF) chains, the size and cost-efficiency of individual RF chains become increasingly critical. A potential approach is to use the classical direct-conversion radio (DCR) architecture; however, DCR suffers from in-phase quadrature-phase (I/Q) imbalance, and the impacts of such RF imperfections might be aggregated in the massive antenna array.

- **Pilot Contamination.** Ideally, every terminal in a massive MIMO system is assigned an orthogonal uplink pilot sequence. Due to the lack of an adequate number of orthogonal pilot sequences at the duration of the coherence interval in practice, channel estimates obtained in a given cell will be impaired by pilots transmitted by users in neighbouring cells (Jose et al. 2011). Pilot contamination results in corruption of precoding matrices used by base stations and causes significant inter-cell interference. The inter-cell interference causes the saturation effect, where the system throughput fails to grow with the number of base station antennas, thereby undermining the value of massive MIMO systems. One solution to this problem is to
design clever channel estimation algorithms or employ some new precoding techniques such as pilot contamination precoding schemes.

- **Low Complexity Large-Dimension MIMO Signal Processing.** Low complexity signal processing algorithms for synchronisation, signal detection, precoding, channel estimation, and channel decoding are critical for practical implementation of massive MIMO systems. On the one hand, the favourable action of the law of large numbers can greatly facilitate the precoding design and the signal detection. On the other, signal processing such as channel estimation turns out to be more challenging since the number of training slots needed for pilot transmission grows linearly with the number of transmit antennas (Hoydis et al. 2013).

**Base Station Antennas – Towards Smart and Multiband Arrays**

As the cellular network standards have evolved, so too have the requirements on base station antennas. High performance base station antennas are crucial to the successful operation of any cellular network. Precisely engineered radiation patterns of base station antennas ensure efficient reuse of frequency and enable the channel capacity to be maximised by optimising the coverage of the intended users. There exist certain similarities between 3G and 4G antennas such as (Chen & Luk 2009):

1. the use of dual slant-polarisation oriented in the +/-45 degree direction,
2. a half power beamwidth with a typical value of 65 degrees in the azimuth or horizontal plane,
3. beam shaping in the elevation plane to reduce the sidelobes above the main beam and minimise the interference to adjacent cells,
4. the ability to tilt the main beam in the elevation plane by values in the range of 1 to 10 degrees through electrical or mechanical means, and
5. the need to keep passive intermodulation products low.

However, 4G base station antenna hardware has also increased in complexity compared to 3G equipment, in order to support the advanced features and higher data rates of 4G networks.

Two major advances of current state-of-the-art 4G base station antennas are multiband operation and the inclusion of independent multiple arrays in a single panel antenna. Typical requirements for “lowband” frequencies are from 698-960 MHz, a bandwidth of 31.6%, while for “highband” operation the antenna frequency coverage should be 1710-2690 MHz, a
bandwidth of 44.5%. These antenna bandwidths are considerably wider than what was required for 3G systems, particularly for highband operation, and have led to the development of ultra-broadband antenna designs. Such antennas are available in penta-band and hexa-band configurations.

One penta-band example from commercial supplier Kathrein supports dual polarisation (+/- 45 degrees) with one low band and four independent high band arrays and independent downtilt for each array (Kathrein-Werke KG 2014). This antenna is capable of supporting 8 x 8 MIMO from a single panel, or offers the possibility of sharing a single antenna between multiple network operators that use different frequency bands. Integrating multiple antennas into one panel reduces capital expenditure in migrating from 3G to 4G because it simplifies installation and approval. It also reduces operating expenditure since site rental fees are reduced and a single antenna can potentially be shared between different network providers.

In addition to traditional base station antennas, the increased use of small cells in 4G networks requires dedicated small-cell antennas. These must be compact and unobtrusive as they are typically deployed to cover urban areas with very high data traffic. The gain of small-cell antennas is lower than base station antennas but they still have to support the advanced 4G features such as multi-band operation and MIMO capability by including multiple independent antennas.

Though it is too early to say exactly what 5G cellular networks will look like, it is possible to speculate on the most likely technology that will be used based on current research trends. The use of millimetre wave (mmWave) technology is expected to have a major impact on 5G networks (Roh et al. 2014), and will further increase the complexity of 5G base station antennas. A scarcity of spectrum available for cellular communications at microwave frequencies combined with the need to provide increased data capacity has led researchers to look to mmWave bands.

Frequency bands that appear to have the most potential are 28-30 GHz, the unlicensed 60 GHz band that covers various portions from 57-66 GHz that is country specific, E-band 71-76 GHz and 81-86 GHz, and 92-95 GHz (Boccardi et al. 2014). There will of course be significant regulatory changes required to make these mmWave bands available for cellular networks. Though the path loss is significantly higher at mmWave bands, particularly at 60 GHz where the oxygen absorption is large, this can be partially compensated for to some degree by the use of high-gain smart antennas that can steer the beam of radiation in both the azimuth and elevation planes. Due to the small wavelengths at mmWave bands it is possible to design high-gain arrays using a physically small aperture. Recent studies published by researchers from Samsung reported a mmWave prototype communications
system operating at 28 GHz with more than 500 MHz of bandwidth (Roh et al. 2014). This system provided both indoor and outdoor coverage over a radius of a few hundred metres and was able to support data rates greater than 500 Mb/s.

It has been observed in field measurements that mmWave multipath channels are sparse in both temporal and spatial domains (Rappaport et al. 2013). Thanks to the multipath sparsity, a large-scale hybrid array becomes a potentially better solution than a full digital array for mmWave cellular systems. In contrast to the full digital array, which uses a radio frequency front-end and digital baseband for each antenna element, a hybrid array consists of multiple analog subarrays with each subarray having its digital chain, as shown in Fig. 3.

Compared to the full digital array, it can not only provide significant savings in cost and complexity, but also achieve comparable performance. On the one hand, using a beamforming analog subarray such as a phased array can effectively collect or distribute signal energy by adjusting the phase of the received or transmitted signals. On the other hand, using an individual digital processing chain for each subarray can add multiplexing capability to the system. At the same time, it provides significant flexibility in beamforming design and improves the system’s capability in dealing with multipath and multiuser interference.

Figure 3 Schematic architecture of a hybrid array, where each DAC (Digital to analog converter) and ADC (Analog to digital converter) is connected to an analog phased subarray (RF components such as mixers, filters and amplifiers have been omitted to simplify the schematic)
Conclusion

In this article, we have presented the evolution of cellular mobile networks in both international standards and enabling techniques. We reviewed key changes in every generation of mobile standards from 2G to 5G, and presented the migration of cellular layouts from rigid structure to random and dense small cells. We also discussed two potentially key techniques in 5G, massive MIMO and smart multiband antenna and antenna arrays. Overall, we see that a combination of the flexible cells and new spatial multiplexing techniques is a necessary path to ubiquitous broadband communications.

References


Cite this article as: