KEY TECHNOLOGY ADVANCEMENTS DRIVING MOBILE COMMUNICATIONS FROM GENERATION TO GENERATION

Contributors

Bernhard Raaf
Platform Engineering Group, Intel Corporation

Michael Faerber
Mobile and Communications Group, Intel Corporation

Biljana Badic
Platform Engineering Group, Intel Corporation

Valerio Frascolla
Platform Engineering Group, Intel Corporation

This article reviews all the generations of technologies deployed in wireless telecommunication systems, starting from 1G (various national analog standards), over 2G (continental standards) and 3G (almost worldwide accepted standards), finally focusing on 4G (the first worldwide accepted standard) also known as LTE (Long Term Evolution). Special attention is given to the pivotal influence of WiMAX (Worldwide Interoperability for Microwave Access), it being the first standard to spearhead the 4G technology, and how WiMAX both paved the way and caused acceleration to ensure a timely development, standardization, and introduction of LTE.

Regulatory and spectrum aspects from the environment and business perspectives, operator and manufacturer consolidation, globalization of standards, challenges to roll out networks and the necessity to cope with existing access networks are some of the key topics discussed.

Finally the article describes the transformation of the LTE technology from the initial version, also called 3.9G, to a true 4G standard, also called LTE-Advanced. Further developments and some insights on what might happen even beyond are provided as well, such as enhancements towards heterogeneous networks and new advanced interference mitigation techniques.

Introduction: Can't We Make Do with the Current Generation, Once and for All?

Why do we need a succession of generations of communications standards, just to continue to communicate? Why are we scraping investments into the old generation each time and accepting massive costs and engineering efforts going into the billions both on vendors' and operators' sides, just to install a replacement one?

A senior researcher once hoped that if LTE was just tuned a little bit in the next release, one could prove that it provides the best possible performance on the physical layer, making any further optimizations redundant and assuring operators' CFOs that they invested in the right technology, once and forever. Unfortunately, his research revealed that LTE had not approached such an optimum yet, and even if it had, changing requirements (user expectations, deployment constraints, and so on) would immediately call for other optimizations, new features, and eventually a new generation.

"Each generation is characterized by pushing the air interface closer to the theoretical Shannon limit…"

Each generation is characterized by pushing the air interface closer to the theoretical Shannon limit, at least within the technical viability of the moment. With Moore's law still holding, after some time we achieve technical and commercial feasibility of functions that engineers ten years ago just dreamed about implementing into a chip.

"With Moore's law still holding, after some time we achieve technical and commercial feasibility of functions that engineers ten years ago just dreamed about implementing into a chip"

Now, as it seems we have to live with this endless succession, let's see what has driven it in the past, what is driving it now, and where it may lead us in the future. First we recap the existing generations of wireless standards, starting from analog 1G over digital but voice-centric 2G and more broadband multiservice 3G towards fully packet-optimized 4G. We then discuss which advanced features are currently being worked on. Finally we use the findings from the previous generations to draw conclusions about general trends and speculate about future generations.

How the Generations of Wireless Standards Evolved in the Past

Two-way radio systems were known and popular for decades, but the shared channel for many users provided no privacy at all and the propagation conditions limited the services to local use. The vision was to provide a two-way communication link to a vehicle, with the standard of a telephone connection, that is, simultaneous bidirectional voice connection and an exclusive transmission channel that offers privacy. The usage scenario as a car phone, or for other vehicles that could carry the hardware, is the reason that still today many scientific publications are published in the IEEE transactions on vehicular technology.

The First Generation, a Patchwork of Analog Systems

The enabling technology breakthrough allowing first generation (1G) cellular radio networks to become a user-friendly, easy-to-handle service was the microprocessor. This allowed storing programs in a cellular phone for executing procedures in idle and connected mode and running communication protocols in the background, reducing the user interaction to operate the system to a level of simplicity like using a landline phone.

"The enabling technology breakthrough allowing first generation (1G) cellular radio networks to become a user-friendly, easy-to-handle service was the microprocessor"

In the early days, the first mobile telephone systems could be named heterogeneous networks, although this term would have a very different meaning from what it does today. The heterogeneity emerged from a wide range of national requirements set by the local operators, country-specific frequency allocations, and local types of standards.

AMPS (Advanced Mobile Phone System)[1] in the United States was the first system starting commercial operation in the early 1980s, and showed all the autonomous characteristics required to provide a user-friendly mobile telephone system. In northern Europe different national telecommunication authorities teamed up to create a standard called NMT (Nord Mobile Telephone). This standard was later adopted by some other EU countries. However, local frequency variations prevented international roaming and also reduced the economy of scale. In Germany a system called C450 was introduced, and in the UK Total Access System (TACS), which was a derivate of the US AMPS system, in Japan also known as Japanese Total Access

"The first-generation cellular systems started as car phone systems, [compared to the predecessors,] the size of the devices shrunk from suitcase to shoebox size, thanks to high integration and the use of microprocessors"

System (JTACS), and again incompatible by national requirements. These first-generation cellular systems started as car phone systems, compared to the predecessors, the size of the devices shrunk from suitcase to shoebox size, thanks to high integration and the use of microprocessors. The first handheld devices (still weighing almost a kilogram) were becoming available in the second half of the 1980s. Common to all these systems was that voice transmission was made by narrowband FM transmission in FDMA-FDD (Frequency Domain Multiple Access Frequency Division Duplex) mode. Signaling solutions of first-generation cellular were based on low rate binary FSK (Frequency Shift Keying) signaling, control mechanisms were mainly network centric, and signaling mostly downlink. This technology decision was a logical choice for several reasons:

- Voice is analog, so analog modulation is most straightforward with the fewest intermediate steps.
- Voice signals flow steadily, and FDMA allows transferring them immediately and steadily, not requiring any buffer or storage that is hard to implement in analog technology.
- When sent over a small bandwidth, symbols used for signaling take longer than typical echoes due to reflections (multipath) and therefore don't suffer from inter-symbol interference.
- This way digital processing, not easily available in those times, could be avoided (for voice) or kept at a minimum (for signaling).

"A major shortcoming of the first-generation systems has been the lack of security"

A major shortcoming of the first-generation systems has been the lack of security. The FM transmission could be easily eavesdropped and besides listening to the conversation, the signaling messages could be tapped. This enabled the reading of a user's identity. Typically user and phone, i.e. the user's equipment, were considered identical and fraud was easy to do. Another shortcoming of the first-generation mobile systems was that the handover procedures had been based on field strength measurements of the base stations. In networks with an increasing traffic density, the downlink controlled handover procedure required high reuse distances and thus reduced the spectral efficiency. The other clear shortcoming of the first-generation systems was that roaming was not possible. Even when a system like NMT was used in Austria, another NMT user from Sweden couldn't use the Swedish phone in Austria.

2G, All Digital and Multinational

When the first-generation mobile systems had been introduced and were showing commercial success, the planning of the second generation started. In Europe the national telecommunication monopolists founded the "Groupe Speciale Mobile" (GSM) within the Conférence Européenne des Administrations des Postes et des Télécommunications (CEPT). The design of this future standard was initially entirely in the hands of network operators, and the industry was only informed about the technical progress from time to time in the form of a technical bulletin. Later industry observers were allowed to participate in the meetings, but without the permission to provide input documents or being part of the discussion. Changes in policy changed the working environment, and the liberalization of the telecommunication market required a change of the standardization. The EU founded ETSI (European Telecommunication Standards Institute), where the GSM standard[2] was further developed in the "Special Mobile Group" (SMG). The term GSM was getting a brand name for the system concept as such, held by the GSM Association, formerly known as the MoU group. This was a group of operators signing a commitment to introduce GSM systems within a given timeframe, and to allow international roaming. This created a large common market for this upcoming system, which was designed from scratch as a digital system. It introduced a split between the user identity, stored in the SIM (Subscriber Identity Module) card, and the phone's hardware, then called User Terminal (UT) or Mobile Terminal (MT) and later User Equipment (UE), as the equipment may transfer data instead of making phone calls. It introduced encryption, and roaming was a must. Network infrastructure was getting more efficient by a TDMA (Time Division Multiplexing Access) system, comprising eight traffic channels on a 200 kHz carrier. For the phone, the standard makers made a big bet on the progress of signal processing, higher integration of functions, and a common market providing an economy of scale.

"The MoU group created a large common market for GSM, which was designed from scratch as a digital system"

Digital processing completely changed paradigms of the first generation: the modulation was no longer analog but digital, and besides, audio signals were compressed by digital algorithms before transmission, typically processing the speech signal in blocks. Consequently steady transmission didn't offer any advantages anymore, and consequently TDMA was introduced. This required buffers, but in digital that requires just a small amount of memory (RAM, Random Access Memory), which as a side effect allows different processing steps to work independently. The instantaneous date rate was increased thanks to the digital design that made equalizers feasible to numerically compensate for inter-symbol interference. Compared to analog designs, many processing steps could be implemented in comparatively cheap and small processors of full custom integrated circuits.

"Compared to analog designs, many processing steps could be implemented in comparatively cheap and small processors of full custom integrated circuits."

In the United States similar activities had been started in ANSI/ATIS (American National Standard Institute/Alliance for Telecommunication Industry Solutions), also aiming at a TDMA-oriented digital system. The difference from Europe was that in the United States AMPS was widely introduced and covered a large national market. Therefore, the aim was to create a digital AMPS (D-AMPS or IS-136), with a backward compatibility to the existing AMPS. This was intended to allow multimode phones, operating in analog mode in areas where D-AMPS was not yet established. This non-disruptive deployment was of course a risk-minimizing approach for the mobile network operators, but put constraints on the to-be-designed digital standard. Such constraints had been to retain the channel bandwidth of 30 kHz, limiting the system to only three TDMA channels per carrier. Furthermore, the backward-compatible operation with AMPS required the use of the analog systems signaling concept. In the United States a homegrown competition emerged by the concept to apply CDMA (Code Division

Multiple Access) transmission technology to mobile cellular networks. The origin of the concept stems from Qualcomm, thus initially a company concept, which received support by some operators, and then was conveyed to US standardization. The concept was later known as IS-95 or simply as "CDMA." CDMA product availability emerged in the first half of the nineties.

Therefore the US market faced the challenge to have two competing local standards for the 2G digital mobile radio, and GSM also started to press into the markets of the United States. This situation caused a patchwork of standards in the United States, which is good for competition and as an innovation driver. However, it confuses the end user in his or her buying decision, and it fragments the phone market.

"Soon it became clear that GSM was not only suited for Europe, even though the commitment of the European countries to deploy this system on a continental scale ensured the critical initial impetus."

In Europe the market was clearly set by regulation to GSM, and in the first half of the nineties the system operation started. Soon it became clear that GSM was not only suited for Europe, even though the commitment of the European countries to deploy this system on a continental scale ensured the critical initial impetus. Therefore GSM was confidently redubbed "Global System of Mobile communication." But sometimes GSM expressed the early operator's desperate desire "God Send Mobiles" because when the first networks got deployed, phone availability was a bottleneck. The significantly increased complexity, compared to the analog predecessors, required quite a high degree of testing for type approval. Fortunately, after overcoming these problems, the market started growing beyond all expectations. International roaming capability, increased privacy and security, and reasonable prices, partly due to economy of scale and partly due to hardware subsidization, made the system very attractive for consumers.

"A common denominator for all 2G systems in the beginning had been provision of voice services"

A common denominator for all 2G systems in the beginning had been provision of voice services. The systems essentially provided a single service (voice), and a cell edge was easily defined by definition of 98 percent service availability (voice, of course) at the cell edge. This was simple for planning network layouts, especially in the early years of the mobile radio network design, where good coverage was to be achieved with a minimum number of base stations. This network deployment principle has largely influenced the mobile network operators' philosophy of how to run a network efficiently by macro base stations. Therefore the operator loves features that can be used in the existing deployments and can avoid increasing the base station density. However, with the evolution towards a 5G concept, it is now commonly understood that the future traffic demands require a high number of small cells and well-designed heterogeneous network algorithms and procedures. To get to this point today, the 2G systems needed to implement the driver for this evolution, the efficient provision of data to mobile users.

Short message service (SMS) was a surprise package in the GSM system concept. Originally intended to enable operators to send small informational messages to their customers, it evolved into a two-way end user communication system. Gaining market share by the sheer simplicity of its interface, with no formatting/carriage return function (no dependency on display format) and a limited size of the message, all you need is the phone number of your communication partner. This made it a huge success. Especially since the operators enabled the sending of messages between different networks. It was indeed at first a highly successful data service, although the end user paid quite a lot per single bit, relative to the price of a 154-digit text message. In GSM the standards development of General Packet Radio Service (GPRS) started in 1994, and built the foundation of the Packet Switched (PS) domain in mobile network architectures. GPRS featured different modulation and coding schemes (MCS), but missed a practical automatic link adaptation. Packet Scheduling had been introduced, as well as sharing the same radio resource for a multiplicity of users, but the available data rates still remained quite low. Time slot concatenation was introduced, but practical phone implementations did not use the theoretical possible maximum rates. Therefore in the second half of the nineties, the work on EDGE (Enhanced Data Rates for GSM Evolution) was started. EDGE introduced a new modulation format to the GSM concept to enable higher data rates under good radio conditions. It features as well an automatic link adaptation, adjusting the MCS to the channel quality. HARQ (Hybrid Automatic Repeat Request) schemes were introduced as well. Using the new modulation scheme and timeslot concatenation, the 2G system was able to provide serious data rates for the first time.

"EDGE introduced a new modulation format to the GSM concept to enable higher data rates…"

"Using the new modulation scheme and timeslot concatenation, the 2G system was able to provide serious data rates for the first time."

After the foundation of the 3rd Generation Partnership Program (3GPP), GSM standardization remained for a while in ETSI, because some of the partners simply had no GSM systems in operation. Therefore there was no interest to deal with GSM; it was seen as an unneeded complexity. However, in reality there was a need for solutions for GSM/UMTS (Universal Mobile Telecommunications System) handover, and core network aspects had to be treated in parallel in ETSI and 3GPP groups. This was not only an inefficient way of working, it bore as well a high risk for failures, conflicts, and ambiguities in the standard. Thus, GSM was later integrated into the 3GPP as the GERAN (GSM EDGE Radio Access Network) group, with a promise to the non-GSM-using standards bodies that the group would be dissolved and integrated to RAN for GSM maintenance. This was now more than ten years ago, and the GERAN group still exists, and new features are made for GERAN release after release.

3G, High Data Rates and Multiple Services in Parallel

UMTS work was started in 1996 in ETSI, but was quickly turned into a global undertaking, by ARIB (Association of Radio Industries and Businesses) and ATIS expressing their strong interests for getting a global standard. ARIB was quite ahead in their regional system concepts, because the market pressure in Japan for 3G was very high due to the shortcomings of the 2G Japanese systems. This was all influenced as well by the timetable of the ITU (International Telecommunications Union), which required proposal deliveries in 1999 for creation of a set of standards fulfilling International Mobile Telecommunications (IMT) requirements for the year 2000, called IMT-2000. ITU competition was there as well, by the IS-136/EDGE operators and supporters in the US, and by the evolution of IS-95 CDMA towards CDMA 2000. The CDMA2000 supporters later founded the 3GPP2 organization, to have a global counterpart organization to the 3GPP.

The side conditions of the UMTS release 99 development had been really difficult. After going through a quite competitive and time-consuming concept discussion in Europe throughout 1996 and 1997, two years had been left for the entire system design. Furthermore the standardization had been moved from regional standardization organizations to the 3GPP, which also did not make the work initially any simpler. But in the end, 3GPP managed to create the standard release 99, to deliver all required input to ITU, to get the IMT-2000 credentials, and keeping the time plan as well, to allow Japan to start its regional FOMA (Freedom of Mobile Multimedia Access) system, based on WCDMA (Wideband code division multiple access).[3]

"…In the end, 3GPP managed to create the standard release 99, to deliver all required input to ITU, to get the IMT-2000 credentials, and keeping the time plan as well…"

Technologically speaking, UMTS remained digital, but intended to increase peak data rates significantly compared to 2G. This was only possible by using more frequency bandwidth and higher symbol rates. TDMA was scraped again to support high data rates by utilizing the transmission medium for 100 percent. But most services don't require the full data rate, so another multiplexing strategy had to be used. While equalizers were introduced for 2G, their complexity increased exponentially with bandwidth, making them quickly infeasible. Therefore CDMA (Code Division Multiple Access) was introduced, together with the rake-receiver, a new kind of equalizer, easy to implement with simple operations (just add and subtract).

The outside conditions, like ITU deadlines and regional availability needs, made the UMTS standard available early. Furthermore, the huge commercial success of the 2G systems in the second half of the nineties made investors eager to repeat this success on an even bigger scale. This created a strong hype on 3G systems. Regulatory bodies started to auction the spectrum for 3G, and in anticipation that 3G would be a kind of Sampo (a mill in Finnish mythology that made flour, salt, and gold out of thin air), excessive prices were paid by new and existing operators to get hold of 3G spectrum.

"…Excessive prices were paid by new and existing operators to get hold of 3G spectrum."

The demand for 3G systems was not as high in all regional markets as in Japan. Although operators had invested heavily in spectrum, the commercial rollout was slow, again limited by phone availability, market demand, and the faltering economy in the beginning of the 2000s. The following release, now coined Release 4, to cut dependency to finish on targeted calendar years, was largely a "repair" release, used to get essential corrections in the standard. With Release 5 the HSDPA (High Speed Downlink Packet Access) concept became part of the standard, introducing important technical enhancements for data service support in downlink. In the following release, improvements for the UL were introduced (HSUPA, High Speed Uplink Packet Access), which turned the WCDMA into the High speed packet access (HSPA).[4] Voice and video was kept on the DCH (Downlink Channel), but HSPA enabled efficient provision of packet data over the 3G air interface.

In the US in CDMA 2000 there had been similar activities to improve the data part as CDMA2000 Evolution (EV), in combination with voice as EV-DV, or data only as EV-DO. The competition between the WCDMA and CDMA 2000 based concepts worked as a driver for important system enhancements. The IS-136/EDGE finally disappeared from the markets, whereas GERAN filled this gap.

"The competition between the WCDMA and CDMA 2000 based concepts worked as a driver for important system enhancements"

The markets recovered from the recession, and an increasing need for mobile data connections boosted the market for data dongles. The Internet use on the move and the need for wide area broadband coverage created a huge increase in the market. However, looking back to Release 99, it took almost ten years from standards availability to real market relevance.

WiMAX and LTE, the Path towards 4G

Two things happened in the mid-2000s. ITU was calling for another deadline to define more advanced requirements, called IMT-Advanced, and with the cumulated experience from 3G, the idea emerged to design a system concept for IMT. This was also pushed by the emerging of WiMAX (Worldwide Interoperability for Microwave Access), an IEEE (Institute of Electrical and Electronics Engineers) standard activity. IEEE and the Wi-Fi certification group continuously worked and enhanced the WLAN (wireless local area network) standard, making the users free of having computers and notebooks connected by cables to a LAN. Logically this leads to the idea to adapt this successful technology to wide area usage. It is worth remembering that WiMAX has a suite of solutions, many taking care of wireless backhaul solutions; only one variant was a mobile WiMAX. But this mobile version is today often thought of as synonymous with WiMAX.

WiMAX, emerging from the IEEE world, was based on a different network architecture than 3GPP and introduced a new air interface based on Orthogonal Frequency Division Multiple Access (OFDMA). Many network operators had been frustrated by the complexity of operating a 3G system and were happy to see a fresh approach. The 3GPP community realized the threat coming from this competing technology, and this accelerated the work on 3G evolution and removed many acceptance barriers on new features and architectures. Previous networks placed an additional aggregation point between the base station and the core network: the Base Station Controller (BSC) in 2G, and the Radio Network Controller (RNC) in 3G, respectively. These were removed by absorbing their functionalities with the remaining units. Furthermore all services were based on packet transmission rather than a continuous stream of bits, making the corresponding circuit switched domain redundant, which was until then still existing in parallel to the packet switched domain. Multi-antenna support became a day 1 requirement for the phone, enabling higher peak data rates by MIMO (Multiple Input Multiple Output), simplifying channel structures and state engines, and last but not least, enabling cheaper network rollout, because two antennas help to provide more reliable service at the coverage edge, providing decent coverage with fewer base stations. To avoid the impression of a disruption to 3G, the term 4G was avoided in the

"WiMAX, emerging from the IEEE world, was based on a different network architecture than 3GPP and introduced a new air interface based on Orthogonal Frequency Division Multiple Access (OFDMA)"

beginning and the concept was therefore called simply Long Term Evolution (LTE).[5] WiMAX served in that phase as a very efficient incubator (almost a blueprint) for the LTE design, but failed to get a foothold in the 3GPP standardization or to influence the global standard itself. There was an attempt to promote WiMAX as the TDD (Time Division Duplex) solution for LTE. However, in 3G the TDD developed to a technical field dominated by Chinese organizations, and there was little willingness to leave this field to a different community. Eventually the attempt to align the numerologies of the emerging LTE FDD and a WiMAX TDD was coming possibly too late. It was decided that the LTE uplink should be based on SC-FDMA (Single Carrier FDMA) for PAPR (Peak-to-Average Power Ratio) reasons, but politically it raised the bar for WiMAX to make its way into a 3GPP standard.

"…OFDMA allows picking the best frequencies for a user's signal to support both time- and frequency-domain scheduling."

OFDMA was used instead of CDMA because the rake receiver got less efficient at the highest data rates and because OFDMA allows picking the best frequencies for a user's signal to support both time- and frequency-domain scheduling. The former had already been introduced in HSDPA, and there had even been attempts to use OFDMA in conjunction with UMTS, but in the end it was felt that the advantages of OFDM wouldn't justify a complex, hybrid system.

WiMAX made its market attempts but did not manage to grow into a relevant size. CDMA2000 lost connection to the 4G evolution, when key companies removed their experts from the 3GPP2 standards group and moved them to 3GPP.

"…Today the dominant standards group is 3GPP and the pacemaking standard is LTE, including LTE-Advanced"

This created the situation of today, where the dominant standards group is 3GPP and the pacemaking standard is LTE, including LTE-Advanced. The lack of competition can create a kind of comfort zone, which may have a negative impact on the technical evolution. Competition and ITU deadlines called for decisions, what is part of a release or not. 3GPP faces the risk of slowing down, developing endless ramifications, and taking functional steps that are too large.

Transformation of LTE to LTE-Advanced and Beyond

Large scale deployments of 4G networks based on the first releases of the LTE systems (Release 8 and Release 9) have been ongoing since 2010. The first evolution of LTE towards LTE-Advanced[7] was defined in 3GPP Release 10, which was finalized in 2010. Release 10 introduced several important improvements[6]:

- Transmission bandwidths over 20 MHz and spectrum flexibility through carrier aggregation: two or even up to five standard LTE carriers of same or different bandwidth can be bundled together and operated like a single carrier. Besides higher data rates, this approach also allows the utilization of several small chunks of spectrum in different bands. That eases gradual re-farming of legacy bands because the legacy systems' bands can be absorbed incrementally rather than having to cannibalize a big fraction immediately.

- Enhanced multi-antenna transmission with flexible reference-signal design, allowing up to eight antennas in downlink, both at the base station and optionally even at the phone, and up to four transmit antennas at the phone, as explained in the next section.
- Relaying to allow deployment of small nodes at the cell edge being fed by base stations via the LTE air interface, using either the same band as used by the mobiles or a different one. Previously installing a new cell to serve an area of insufficient coverage required not only installation of a small, typically cheap base station, but also a data connection to the core network. In the worst case that meant digging cable ditches, an inconvenient and expensive enterprise. Microwave links can ease that burden but require dedicated, matching equipment to be added at both the new and the existing site. Here relaying can offer an advantage, by utilizing the existing LTE air interface to link newly installed small base stations with existing stations, so called donor eNBs. This backhaul link can use a different frequency band than the band used to communicate with phones, requiring two receivers, two transmitters and duplexers at the relay. Another option is to use the same band, simplifying deployment and allowing cheaper relay implementation. In this variant the relay cannot communicate with the phones continuously but needs to devote part of the time for the backhaul link. In order not to confuse them by the absence of signals, these times are advertised as MBSFN (Multicast-broadcast single-frequency network) subframes (see Bachl et al.[8]). Even old LTE phones are prepared to at least ignore such subframes without any harm.

  Unfortunately, the capacity gain provided by the relays is somewhat offset by a loss of capacity from the donor cell, therefore this technology has not yet been widely deployed. It is expected to become more relevant when cells get even smaller and operate at higher frequencies. A similar approach on the phone side is device-to-device communication (see Roessel et al.[9], Zaus and Choi[10]).
- Enhanced Inter Cell Interference Coordination (eICIC) allowing more liberal deployments of Heterogeneous Networks (HetNet), as shown in the section "Coordination across Base Stations in 4G Systems."

"relaying can offer an advantage, by utilizing the existing LTE air interface to link newly installed small base stations with existing stations"

MIMO Transmission in 4G Systems

MIMO transmission is used in 4G systems to increase the overall bitrate through transmission of two or more different data streams on two or more transmit antennas using the same resources in both frequency and time, separated only through the use of a different reference signal and received by two or more receive antennas. The usage of multiple antennas at transmitter and/or receiver sides improves reliability and increases spectral efficiency and spatial separation of users. From LTE (Release 8 and 9) to LTE Advanced (Release 10 and 11), ten different MIMO Transmission Modes (TM) have been defined, which differ in number of layers (rank), used antenna ports, type of reference signal, Cell Specific Reference (CRS) or Demodulation Reference Signals (DMRS), number of users supported, and precoding type. Since the performance of MIMO transmission depends on various factors, no

"The usage of multiple antennas at transmitter and/or receiver sides improves reliability and increases spectral efficiency and spatial separation of users."

single mode provides superior performance in all usage scenarios. The leading MIMO modes are transmit diversity and spatial multiplexing (closed and open loop transmission) because they are robust to implement and can deliver the promised 4G data rates. One of the MIMO transmission techniques that has been extensively investigated in standardization as well as in research communities is multiuser MIMO transmission because of several benefits over single-user MIMO. In a multi-user scenario multiple users share the same time and frequency resources by exploiting the spatial diversity of the propagation channel. Because a single user's phone typically only has few antennas in a small volume, which are thus often highly correlated, multiple users have more antennas in total and the channels are typically much less correlated due to larger separation. Therefore, Multi-User MIMO(MU-MIMO) transmission is much more robust with respect to propagation environments and achieves a high spatial multiplexing gain even with the small number of antennas at the phones.

"…Multi-User MIMO(MU-MIMO) transmission is much more robust with respect to propagation environments and achieves a high spatial multiplexing gain even with the small number of antennas at the phones."

To some extent MU-MIMO is supported in Release 8 and 9 but with several practical limitations. Already in Release 8 (Transmission Mode 5) the same codebook optimized for single-user (SU)-MIMO precoding is applied for MU-MIMO, which is suboptimal. Such a configuration results in a performance inferior to Release 8 MIMO modes (Transmission Mode 4 and 6). To keep the feedback overhead low, 3GPP did not dedicate any symbols for MU-MIMO system but MU-MIMO transmission was enabled using SU-MIMO feedback, that is, CQI/PMI/RI (Channel Quality Indicator/Precoding Matrix Index/Rank Indicator) only. This is a rather minimal MU-MIMO transmission scheme, limiting the achievable multiuser gains and thereby the practical implementation of MU-MIMO systems. With User-specific demodulation reference symbols (DMRS) introduced in Release 9, support for MU-MIMO transmission for up to four users rank 1 (orthogonal) or up to two users rank 2 (nonorthogonal) was enabled in Transmission Mode 8. However, the antenna port and scrambling code allocations are wideband and with such a configuration it is not always possible to ensure orthogonality even when only two users are multiplexed in MU-MIMO mode. Neither Release 8 nor Release 9 introduced explicit signaling of the presence/absence of a co-scheduled user on the same resources, that is, the phone does not know whether interlayer interference exists or not. This limitation has significant impact on MU-MIMO detection with conventional (interference-unaware) receivers. As shown in Duplicy et al.[11], interference-aware receivers are required in MU-MIMO to eliminate the residual spatial interference in order to avoid detection error floor. The MU-MIMO implementation challenges, solutions, and how to overcome them are discussed further in Badic et al., "MU-MIMO System Concepts and Implementation Aspects."[12]

"LTE-Advanced has enhanced MIMO transmission and also relaxed some MU-MIMO limitations"

LTE-Advanced has enhanced MIMO transmission and also relaxed some MU-MIMO limitations. New reference signal Transmission Mode 9 is introduced in Release 10, supporting up to 8x8 (MU)-MIMO and new reference signals for CSI measurements together with DMRS. This type of signaling enables switching between SU and MU-MIMO mode without need for the phones to be reconfigured via higher layer signaling. A dual codebook approach is adopted for 8x8 (MU)-MIMO, where one codebook captures wideband and long-term channel properties, while the other captures frequency-selective and/or short-term channel properties.

In order to enable reliable practical deployments of (MU)-MIMO Release 11 studied possible enhancements of DL MIMO and the study continues currently in Release 12.[15] The study focuses on CSI feedback enhancements in order to provide finer spatial domain granularity and to support different antenna configurations for both single and multiuser transmission.

Coordination across Base Stations in 4G Systems

LTE Release 11, which is in the final specification stage, introduced coordinated multipoint transmission and reception (CoMP) with the aim of improving the coverage of high data rates and the cell-edge user experience. The basic concept behind CoMP is the cooperative multiple-input and multiple-output technique, where geographically distributed transmitters and receivers are jointly working with advanced MIMO schemes. In theory, CoMP can provide significant gains in the order of high double-digit percentages for cell-edge user throughput while increasing overall network efficiency.

"In theory, CoMP can provide significant gains in the order of high double-digit percentages for cell-edge user throughput while increasing overall network efficiency."

Even though CoMP techniques have received increasing interest within research and 3GPP communities, their practical implementation still has a long way to go. As for any other closed-loop transmission, the key obstacle in the CoMP realization is the feedback link (CSI) between the devices and network. Release 11 defined a multipoint CSI framework but the framework contains a number of practical constraints. First, CSI feedback is designed without any synchronization between transmission points, that is, PMI and CQI are selected individually per transmission point. Furthermore, to keep feedback overhead low, Release 11 cancelled support for phase alignment between transmission points. Obviously, those constraints restrict the network's ability to fully exploit benefits of CoMP since the transmit signal optimization across all TPs is infeasible. As discussed in Hosemann et al., "Implementation Challenges Facing UE in Coordinated LTE Networks"[13] and Bai et al., "Feedback Generation for Cell-Edge Transmission in Unsynchronized Coordinated Heterogeneous Networks,"[14] nonaligned transmit signals between transmission points in a CoMP network can lead to significant performance losses, which can be up to several decibels.

"…nonaligned transmit signals between transmission points in a CoMP network can lead to significant performance losses, which can be up to several decibels"

Release 11 did not address the specified support of CoMP involving multiple eNBs with non-ideal backhaul but assumed ideal fiber connection between the cooperating points. Due to this limitation, the operators having non-ideal backhaul may not be able to benefit from CoMP operation. As discussed in ADVA[16], interference coordination schemes such as eICIC and CoMP require very short latencies (less than 1 ms) across the backhaul network in order to achieve real-time coordination between base stations. In addition to the low-latency requirement, base station clocks need to be in phase to enable proper operation of coordination in order to achieve highly accurate phase or time-of-day synchronization. With heterogeneous networks becoming reality, the demand for flexible backhaul and tighter synchronization between the cells in time and frequency is increasing. In order to meet those requirements, LTE-Advanced networks and beyond will have to challenge existing backhaul networks with respect to capacity, latency and synchronization performance.[16] A new study item on coordinated multipoint operation for LTE non-ideal backhaul has been agreed to in Release 12 with the aim to evaluate the performance benefits and to identify potential standardization impacts for candidate CoMP techniques involving multiple eNBs with non-ideal backhaul.[17]

Concluding Remarks and Future Directions

"Hardly any other modern technology has provided, generation after generation, such a consistent and fast-paced evolution of services, changing so drastically the way people communicate among themselves, finally making a reality the "always-on, always-available" paradigm"

In the previous sections we have discussed the technical evolution of the mobile cellular communication system, from its first inception to the currently deployed status. Hardly any other modern technology has provided, generation after generation, such a consistent and fast-paced evolution of services, changing so drastically the way people communicate among themselves, finally making a reality the "always-on, always-available" paradigm. One might even dare say it changed the society altogether, that is, the way people behave and, to a certain extent, what they strive for.

On the one hand, during the last twenty years the mobile devices introduced at each generation have improved a lot the way people gather information, conduct business, and interact among themselves, according to several aspects: being available everywhere at every time, even in far-fetched locations; communicating quickly to counteract unforeseeable problems; delivering in a real-time manner the latest news as soon as it happens; sharing huge data or video information, thus allowing work while on the move or remotely, but together with colleagues using a common virtual room; creating lots of new business opportunities via new applications running on mobile phones, such as augmented reality games[18] or location-based services, which were unthinkable just ten years ago.

"These aspects, among others, show that people are more and more giving mobile devices a new and unexpected "value," very different from the reason of their initial conception: make people communicate"

On the other hand, those same devices are changing society in some unpredicted ways as strange new phenomena are popping up, for example the "alone together" syndrome[19]: Couples or groups of friends cluster together, but instead of communicating directly among themselves, each one uses his or her own device, side by side but lost in his or her own virtual world of bits. Or the frantic quest to buy the newest device, be it the latest tablet or the thinnest notebook, especially among youngsters: they queue for days outside, in front of the shop, where the desired announced new product will soon be launched. These aspects, among others, show that people are more and more giving mobile devices a new and unexpected "value," very different from the reason of their initial conception: make people communicate.

What can be said about the generation coming right after 4G? It might help to survey what it has taken, so far, to launch a new generation. Which novelties have spurred such a "change of pace" that it is reasonable to talk about a new generation of mobile devices? Surely that includes the appearance of new applications and corresponding requirements, like the data-transfer capability in a world of voice-only devices.

In general, the development of a novel technology follows a common path: from the research phase (birth in academia), through the innovation phase (growing the technology's potential by startups and early innovators), up to the real exploitation via "productization" (matured by the industry), culminating in the development and launch of novel products.

Another aspect is the availability, at affordable prices of enabling technologies, for example the capability to concentrate a certain processing power in a specific form factor, allowing exploitation of previously too power-hungry algorithms.

It is interesting to note that a new generation often doesn't introduce a new feature for the very first time. Most often, it was tried already in the previous generation. In fact in that previous generation, due to legacy limitations or intrinsic constraints, such pioneering implementations of a new feature often suffered from some inconsistencies and therefore couldn't get to a widespread deployment. This doesn't necessarily create a blocking point for a new feature, as this way it is possible to gain experience to pave the path for a better implementation in the subsequent generation. Eventually, in the next generation the intended new feature can be implemented more consistently, mainly because a new generation is designed around that feature, rather than implementing it as an add-on to an existing environment. Such an example are data calls, which are not only possible in isolation but in parallel to voice, a feature introduced by UMTS.

"It is interesting to note that a new generation often doesn't introduce a new feature for the very first time"

Looking back, each generation introduced a new multiple access scheme (1G: FDMA, 2G: TDMA, 3G: CDMA, 4G: OFDMA), therefore one could conclude it characterizes a generation. Indeed an access scheme is a very fundamental aspect that cannot be changed easily within a generation. However, the introduction of a new access scheme is neither a sufficient nor a necessary condition to call for a new generation: Enhancements in several areas of the technology are needed for such a big leap forward.

"…the introduction of a new access scheme is neither a sufficient nor a necessary condition to call for a new generation:"

Another aspect to take into consideration could be the attitude towards what is considered the "mainstream technology" during a lifetime of a specific generation: if the verb of a new technology is spread around, more and more people try to leverage it and in some cases it manages to become "hype" no one dares to resist. But not every "hyped" technology actually makes it into the standards, as can be seen from another cornerstone of digital communication, the coding/decoding schemes: While 1G basically didn't encode at all, 2G introduced convolutional codes, and 3G chose the turbo codes that were researched shortly before. When 4G was defined, the Low Density Parity Check (LDPC) codes were just a hot topic in academia. However, instead of using the latter, the turbo codes were slightly modified to allow for a parallel processing, thus compensating for the main advantage of LDPC codes, which therefore didn't make it into 4G.

In conclusion, what can be said about the specific technical content of the generation after 4G, namely 5G, with a reasonable level of certainty? At the moment, to be very frank, not very much. The focus of the wireless community has so far been on tuning the 4G technology and on deploying it (still ongoing though in many countries) and therefore it's really too early to write any stable assertions.

"…as long as wireless communication is an important aspect of our lives and as long as new devices and applications change (and typically increase) the underlying requirements, there will be both an evolution within the established generations and, roughly every decade, a revolutionary approach leading to the birth of a new generation"

We can speculate that some technologies currently being discussed in research communities—say millimeter waves or tighter coordination schemes of more and more sites and even more diverse services and applications—might be very good candidates. Moreover, as long as wireless communication is an important aspect of our lives and as long as new devices and applications change (and typically increase) the underlying requirements, there will be both an evolution within the established generations and, roughly every decade, a revolutionary approach leading to the birth of a new generation. In fact this new generation will be in a dilemma between two targets: both to be reasonably compatible to the already deployed technology (legacy) following an evolutionary path and to allow for newly introduced revolutionary breakthroughs, in order to provide a timely solution for the most pressing problems at hand.

We all look forward to that future.

References

[1] Arrendondo, G. A., J. C. Feggeler, and J. I. Smith, "Advanced Mobile Phone Service," *The Bell System Technical Journal*, Vol. 58, No. 1, January 1979

[2] Mouly, M. and M.-B. Pautet, "The GSM System for Mobile Communications," *Telecom Publishing*, June 1992

[3] Toskala, Antti and Harry Holma (eds.) *WCDMA for UMTS: Radio Access for Third Generation Mobile Communications*, 3rd Ed. (Hoboken, NJ: Wiley, 2004).

[4] Toskala, Antti and Harry Holma (eds.) *HSDPA/HSUPA for UMTS: High Speed Radio Access for Mobile Communications* (Hoboken, NJ: Wiley, 2006).

[5] Toskala, Antti and Harry Holma (eds.) *LTE for UMTS: OFDMA and SC-FDMA Based Radio Access* (Hoboken, NJ: Wiley, 2009).

[6] Toskala, Antti and Harry Holma (eds.) *LTE Advanced: 3GPP Solution for IMT-Advanced* (Hoboken, NJ: Wiley, 2012).

[7] Mogensen, P. E., T. Koivisto, K. I. Pedersen, I. Z. Kovacs, B. Raaf, K. Pajukoski, and M. J. Rinne, "LTE-Advanced: The path towards gigabit/s in wireless mobile communications," *Wireless*

Communication, Vehicular Technology, Information Theory and Aerospace and Electronic Systems Technology, 2009. Wireless VITAE 2009, Aalborg, Denmark.

[8] Bachl, R., R. Balraj, and I. Groh. "LTE Physical Layer Aspects for eMBMS: Concept and Performance Evaluation," *Intel Technology Journal*, Vol.18. Issue 01, March 2014.

[9] Roessel, S., M. Faerber, B. Raaf, and J. Hausner, "Radio Network Evolution towards LTE-Advanced and Beyond," *Intel Technology Journal*, Vol.18. Issue 01, March 2014.

[10] Zaus, R., H-N. Choi, "Evolving the Network for Machine-to-Machine Communication," *Intel Technology Journal*, Vol.18. Issue 01, March 2014.

[11] Duplicy, Jonathan, Biljana Badic, Rajarajan Balraj, Rizwan Ghaffar, Péter Horváth, Florian Kaltenberger, Raymond Knopp, István Z. Kovács, Hung T. Nguyen, Deepaknath Tandur, and Guillaume Vivier, "MU-MIMO in LTE Systems," *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, Article ID 496763, 2011. doi:10.1155/2011/496763.

[12] Badic, B., B. Balraj, I. Groh, and P. Fazekas, "MU-MIMO System Concepts and Implementation Aspects," *Intel Technology Journal*, Vol.18. Issue 01, March 2014.

[13] Hosemann, M., D. Yoon, B. Badic, R. Balraj, and C. Papadias, "Implementation Challenges Facing UE in Coordinated LTE Networks," *Intel Technology Journal*, Vol.18. Issue 01, March 2014.

[14] Bai, Z., S. Iwelski, B. Badic, G. Bruck, and P. Jung, "Feedback Generation for Cell-Edge Transmission in Unsynchronized Coordinated Heterogeneous Networks," *Intel Technology Journal*, Vol.18. Issue 01, March 2014.

[15] 3GPP "Further Downlink MIMO Enhancement for LTE-Advanced," RP-121416, 2012.

[16] ADVA, "New Backhaul Challenges Emerging with LTE-Advanced," White Paper, ADVA Optical Networking, June 2013.

[17] 3GPP "Coordinated Multipoint Operation for LTE Non-Ideal Backhaul," 3GPP TR 36.874, 2013.

[18] Saenz, Aaron, "Catching Augmented Reality Butterflies on Your iPhone to Earn Free Stuff," http://singularityhub.com/2010/10/26/catching-augmented-reality-butterflies-on-your-iphone-to-earn-free-stuff-video-2/.

[19] Turkle, Sherry, *Alone Together: Why We Expect More from Technology and Less from Each Other* (New York: Basic Books, 2012).

Author Biographies

Bernhard Raaf is IMC Principal IP and Patents, working on improving Intel's wireless patent portfolio since 2012. He graduated (Dipl.-Phys.) from the Technical University Munich, with a thesis at the Max Planck Institute for Extraterrestrial Physics on radar measurements on northern lights. He has contributed to standardization of GSM, UMTS, LTE, and LTE-Advanced. Bernhard received a VTC best paper award for one of over 40 coauthored papers and holds over 100 granted patents, many essential for UMTS/LTE. Email: Bernhard.Raaf@intel.com

Biljana Badic is a system engineer in the Intel Mobile and Communications Group, has been with Intel since 2010, and has participated in development of 4G wireless modems with the particular focus on interference cancellation and mitigation algorithms. Biljana is also actively involved in Intel research activities on 4G and beyond systems. Biljana holds a PhD from Vienna University of Technology, Austria and has published over 50 scientific articles and filed over 10 4G patents. Email: biljana.badic@intel.com

Michael Färber joined Intel Mobile Communications Standards and Advanced Technology (SAT) in October 2012. Michael received his engineering degree (Dipl.-Ing.) from the University of Applied Sciences Darmstadt-Dieburg (Germany), and joined Siemens in 1985. In 1990 he started with active participation in GSM standardization. In 1995 he was elected as ETSI SMG 2 vice chairman. In 1998 he was appointed as EDGE ad hoc group chairman. His UMTS involvement started 1996 in ETSI, until release 99 was finished in 3GPP. From 2007 to 2012 he was with Nokia Siemens Networks. Michael was named Inventor of the Year at Siemens in 1997 (Siemens I&C) and in 1999 (Siemens global), and filed more than 100 patent applications.

Valerio Frascolla earned his MSc in electrical engineering in 2001 and his PhD in electronics in 2004. He worked as research fellow at Ancona University, Italy, then moved to Germany, joining Comneon in 2006 and Infineon Technologies in 2010. Since 2011 he has been funding and innovation manager at IMC, acting as facilitator of research collaborations using Agile methodologies and focusing on the program management of publicly funded projects and innovation activities. He is author of several peer-reviewed scientific publications and has been an invited speaker at international events. Email: Valerio.frascolla@intel.com.