Publisher
Richard Bowles

Managing Editor
Stuart Douglas

Content Architect
Biljana Badic

Program Manager
Stuart Douglas

Technical Editor
David Clark

Technical Illustrators
MPS Limited

Technical and Strategic Reviewers
Valerio Frascolla
Biljana Badic
Josef Hausner
Kenan Kocagoez
Zijian Bai
Sebastian Wagner
Stefan Franz
Gary Xiong
Zhang Liang
Xiaofeng Wu
Stefan Fechtel
Kenneth Stewart
Vivek Gupta
Ignacio Rodriguez Larrad
Rajarajan Balraj
Erfan Majed

Intel Technology Journal

ISBN 978-1-934053-60-7, ISSN 1535-864X

Intel Technology Journal
Volume 18, Issue 1

Publisher: Richard Bowles
Managing Editor: Stuart Douglas

Library of Congress Cataloging in Publication Data:

Printed in China
10 9 8 7 6 5 4 3 2 1

First printing: January 2014

Notices and Disclaimers

ALL INFORMATION PROVIDED WITHIN OR OTHERWISE ASSOCIATED WITH THIS PUBLICATION INCLUDING, INTER ALIA, ALL SOFTWARE CODE, IS PROVIDED "AS IS", AND FOR EDUCATIONAL PURPOSES ONLY. INTEL RETAINS ALL OWNERSHIP INTEREST IN ANY INTELLECTUAL PROPERTY RIGHTS ASSOCIATED WITH THIS INFORMATION AND NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHT IS GRANTED BY THIS PUBLICATION OR AS A RESULT OF YOUR PURCHASE THEREOF. INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY RELATING TO THIS INFORMATION INCLUDING, BY WAY OF EXAMPLE AND NOT LIMITATION, LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR THE INFRINGEMENT OF ANY INTELLECTUAL PROPERTY RIGHT ANYWHERE IN THE WORLD.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

For more information go to http://www.intel.com/performance

Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

Notice revision #20110804

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

A "Mission Critical Application" is any application in which failure of the Intel Product could result, directly or indirectly, in personal injury or death. SHOULD YOU PURCHASE OR USE INTEL'S PRODUCTS FOR ANY SUCH MISSION CRITICAL APPLICATION, YOU SHALL INDEMNIFY AND HOLD INTEL AND ITS SUBSIDIARIES, SUBCONTRACTORS AND AFFILIATES, AND THE DIRECTORS, OFFICERS, AND EMPLOYEES OF EACH, HARMLESS AGAINST ALL CLAIMS COSTS, DAMAGES, AND EXPENSES AND REASONABLE ATTORNEYS' FEES ARISING OUT OF, DIRECTLY OR INDIRECTLY, ANY CLAIM OF PRODUCT LIABILITY, PERSONAL INJURY, OR DEATH ARISING IN ANY WAY OUT OF SUCH MISSION CRITICAL APPLICATION, WHETHER OR NOT INTEL OR ITS SUBCONTRACTOR WAS NEGLIGENT IN THE DESIGN, MANUFACTURE, OR WARNING OF THE INTEL PRODUCT OR ANY OF ITS PARTS.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined". Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata, which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order.
Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or go to: http://www.intel.com/design/literature.htm

INTEL® TECHNOLOGY JOURNAL
4G COMMUNICATIONS: SYSTEM DESIGN CHALLENGES

Articles

Foreword LTE: The Move To Global Cellular Broadband

Gerhard P. Fettweis,
Vodafone Chair Professor, Technical University of Dresden, Germany

Cellular communications have dramatically changed our world by shaping our modern way of life. Technological advances have radically carved the way for new forms of human interaction. We have seen the transformation happen from the first generation of analog phones (also known as 1G), available only for a minority of users, via 2G systems (namely GSM, D-AMPS, CDMAone or IS-95, PDC) being able to reach and call anyone everywhere and introducing world-wide two-way paging via text messaging. Recently mobile Internet access at our finger tips has become available through 3G (WCDMA and CDMA2000). With LTE, (the Long Term Evolution, also known as 4G, the fourth generation cellular system) for the first time one common standard has been accepted globally. It opens cellular communications to true broadband connectivity and offers enough capacity to enable cost-competitive machine-to-machine (M2M) communications.[1]

This issue of the *Intel Technology Journal* brings together active research directions and development results achieved within the context of LTE. Looking back at the evolution of standards, LTE is not only a first global standard and a new technology step, but it also shapes the value chain and ecosphere of wireless communications hardware suppliers.

Global Standard

Looking back, every decade has given birth to a new standard in digital cellular and hit the market. Starting with the introduction of GSM in the 1990s and 3G in the 2000s, now 4G/LTE is following in the 2010s. However, after the success of 2G systems, starting a new standardization process has been a difficult task for each following standard generation. To create a new standard, at first research needs to be carried out, showing the path to innovation opportunities, which form the base of the new standardization step—a cumbersome process of agreeing on common ground. Finalization of this process is done by product development and roll-out in the field. As each of these stages needs at least a couple of years, the process starts typically a decade before customers have products in their hands.

LTE standardization had its roots nearly a decade ago. As WiMAX was gaining attention with its many MIMO-based innovations, it became clear that 3GPP standardization had to react, or be overrun. However, many 3GPP members were hesitant since 3G was being rolled out and no return on investment for 3G was yet in sight. Therefore the strategy was to rather slow down the introduction of a new standard than to fully embrace the idea, clearly earmarked by choosing for the new technology the name "Long Term Evolution." As the slow start became evident, WiMAX saw its chance and pressed forward so hard that 3GPP had to wake up, and the race for market traction began.

After having experienced the cost of fragmenting the world into multiple cellular worlds through the 2G and 3G generations, cellular operators pushed for one single global standard. In particular the "big three" (Vodafone, Verizon, and China Mobile) tried to jointly understand the differences between the proposals and aligned their decision onto the LTE standard. One reason for the decision was that the TTI framing of WiMAX had a 5-ms frame duration, which was seen as too large for enabling mobile gaming, a possible killer application. In contrast, the planned LTE's TTI framing of 1 ms opened the possibility to significantly reduce the latency over 3G and WiMAX systems.

However, the LTE protocol stack above layer 2 was not optimized for such a low latency and as a consequence today's round-trip ping results are in the order of 20–30 ms, still well above the real-time gaming requirement of a maximum latency of 10 ms. Therefore there is still ample room for improving in the next standardization steps towards the fifth generation of cellular systems.[1]

Technology Step

At the time of its inception, GSM was designed around the capabilities of implementing single-carrier equalization in silicon with a maximum likelihood sequence estimator (MLSE). The MLSE equalizer grows exponentially with the delay spread of the channel.

A key innovation of using spread spectrum in WCDMA was that the so-called RAKE receiver used for detection/demodulation grows only linearly with the number of relevant echoes of the channel.

With the use of OFDM (Orthogonal Frequency Division Multiplexing [OFDM] is a technique in which the subcarriers of a fast Fourier transform (FFT) are used for parallel modulation and transmission of data bits onto the carrier) in LTE, the equalization/detection challenge has become small when comparing it to GSM and WCDMA. As each subcarrier sees a frequency flat channel, the complexity is dominated by the *N*-point FFT/IFFT computation, which grows as *N*·log (N). The size *N* of the FFT needs to be chosen to be large compared to the delay spread of the channel, for example 4x, to minimize the overhead of the cyclic prefix. In conclusion, the result is that the computation complexity per sample (of an *N* point FFT) grows only logarithmically by the delay spread.

The impact of this moderate growth in complexity with the channel situation, and the ability to demodulate frequency flat subcarrier channels, makes OFDM ideal to implement complex multi-antenna transmission schemes like MIMO. A major focus therefore during the design of LTE was to define the right pilot-aided channel estimation, to support advanced MIMO concepts, to create a hybrid ARQ (Automatic Repeat-reQuest), and scheduling to capture the channel capacity per user at its best.

From a technology point of view, the OFDM modulation chosen is a very good platform to enable the implementation of further complexity at the benefit of system performance, mostly measured in capacity, and in the fairness of capacity distribution among users in a cell. The capacity limiting factor in cellular systems is the interference generated within a cell, or from neighboring cells. The mitigation and/or cancellation of interference have been therefore a main aspect of research over the last decade, to improve LTE towards LTE-Advanced. Most often this is named CoMP (coordinated multi-point) or ICIC (interference cancellation and interference coordination). The signal processing challenge is to take the signals of neighboring cells into account when transmitting and receiving. Initially CoMP[3] and ICIC were two abbreviations for the same idea. Recently the term CoMP has been used more for the specific technique of distributed MIMO signal processing.

The theoretic upper capacity bound of CoMP/ICIC is to transform an interference-limited scenario into a noise-limited scenario. This is very academic due to the need of channel state and signal information. However, on the way towards the bound many interesting ideas have been proposed.

Ecosphere

With the introduction of LTE a major change has happened in the ecosphere of the cellular industry. On the one hand, China, the United States, and Europe have aligned on a single standard, and competing standards have lost.

However, today, in the advent of small cells becoming a next wave of cellular network realization, the LTE base station market is also soon to be transforming itself from a classic infrastructure market of approximately 10 million units per year into volumes comparable with consumer products of 100 million units and above. This opens new opportunities for many new players, be it in the space of (fabless) semiconductors, infrastructure suppliers, network management (SON: self-organizing networks) suppliers, or operators. Therefore, we should be prepared for many forthcoming disruptions, changing the industry once again.

Wireless Future

With each generation of wireless technology the increase in data rate has approximately followed Moore's law of doubling every 18 months, that is, increasing tenfold every five years. Assuming that memory storage increase is the main driver for the so-far-unbounded demand in data rate over wireless, and taking the ITRS Roadmap and 3D integration into account, we can assume this increase to continue until 2030. When projecting this by wireless standards products being introduced into the market, the number in Figure 1 arises.

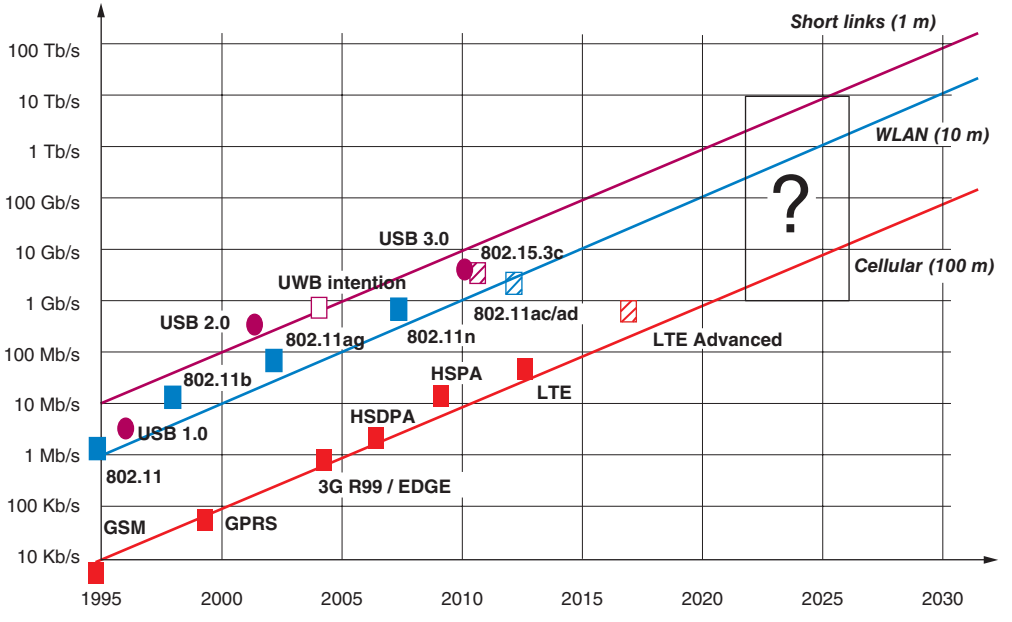


Figure 1: The Wireless Roadmap: Race for Data Rate
(Source: Gerhard P. Fettweis, 2013)

An unimaginable data rate projection of 1–10 Gb/s shows us that major innovations must be accomplished to address the needs of the fifth generation systems, possibly hitting the market ten years from now. Next to cellular, Wi-Fi, at unprecedented data rates around 1 Tb/s, will provide us with a completely new level of connectivity, being a platform for service and application innovations that cannot even be imagined today.

However, not only does the race for data rate need to be addressed, but as mentioned above, the round-trip ping latency reduction to under 10 ms (for example, for gaming) must be achieved as well. When an approximately 1-ms round-trip latency and carrier-grade reliability and availability are both achieved, a step from moving content to controlling real and virtual objects will be enabled, the "Tactile Internet"[4] as sketched in Astély et al.[1] becomes reality.

The "Internet of Things" has not been mentioned so far. It is not a step towards increased broadband, but semiconductor vendors are key enablers to address the challenge of connecting anything with a 10-year battery lifetime. Not only must system concepts such as the extension of transmit intervals by enhancing discontinuous transmission (DTX) cycles be explored, but also lightweight network synchronization must be designed and of new nonorthogonal waveforms like FBMC[5] and GFDM[6] must be implemented. Also, a huge demand for power savings must be addressed by innovating chip architecture, by voltage and frequency scaling, as well as sleep modes and stable wake-up strategies, and by semiconductor technology itself.

LTE is a major step forward in developing technology along the Wireless Roadmap. In the cellular market, the next decade will be dominated by LTE, evolving to LTE-Advanced over time. More than an order of magnitude increase in data rate lies ahead with opportunities that still need to be explored.

Technologies as presented in this edition of the *Intel Technology Journal* mark important ingredients. However, LTE is only a step. Major further innovations in cellular and WLAN wireless communications still lie ahead. Incumbent companies and startups can grab the upcoming market opportunities and/or drive and create opportunities unimagined today.

References

[1] Astély, David, Erik Dahlman, Anders Furuskär, Ylva Jading, Magnus Lindström, and Stefan Parkvall, "LTE: the evolution of mobile broadband," *IEEE Communications Magazine*, Vol. 47, April 2009, pp. 44-51.

[2] Fettweis, Gerhard P., "A 5G Wireless Communications Vision," *Microwave Journal*, December 2012, pp. 24-36.

[3] Marsch, Patrick and Gerhard P. Fettweis (eds.), *Coordinated Multi-Point in Mobile Communications: From Theory to Practice* (Cambridge: Cambridge University Press, 2011).

[4] G.P. Fettweis and S. Alamouti, "5G – Personal Mobile Internet Beyond What Cellular Did to Telephony," accepted for IEEE Communications Magazine, to appear in February 2014.

[5] Farhang-Boroujeny, B., "OFDM Versus Filter Bank Multicarrier," *IEEE Signal Processing Magazine*, Vol. 28, April 2011, pp. 92-112.

[6] Gaspar, Ivan, Nicola Michailow, Ainoa Navarro, Eckhard Ohlmer, Stefan Krone, and Gerhard Fettweis, "Low Complexity GFDM Receiver Based on Sparse Frequency Domain Processing," in Proc. IEEE VTC Spring 2013, Dresden, June 2-6, 2013.

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)"

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

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

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.

"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"

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.

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.

"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"

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.

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

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.

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 Acces*s (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.

FROM FDD TO TDD: CHANCES, CHANGES, AND CHALLENGES IN TDD LTE SYSTEMS

Contributors

John Zhou
Product Engineering Group,
Intel Corporation

Yubao Li
Product Engineering Group,
Intel Corporation

Qingyun Guo
Product Engineering Group,
Intel Corporation

Long Term Evolution (LTE), also named Fourth Generation (4G), is seen as one of the most promising technologies in the radio communication field. It claims significant throughput/capacity improvement compared to 2G/3G using a different radio interface together with core network improvements. In principle, LTE supports both frequency division duplexing (FDD) and time-division duplexing (TDD). In this article, we are focusing on the TDD LTE system where the uplink and downlink shares the same frequency carrier. The brief introduction and the motivation of designing TDD LTE system are provided in the first part of this article; then the detailed comparisons between TDD LTE and FDD LTE, mainly on physical layer, are presented in the second part of the article. In the third part of the article we analyze the difference and indicate the challenges of the LTE TDD physical layer with some possible solutions. The performance benchmarks for different solutions are also given in this part. Finally the article concludes by summarizing the benefits and difficulties of TDD LTE system, with suggestions for further optimization.

Introduction

"Long Term Evolution (LTE), being developed by the Third Generation Partnership Project (3GPP), tries to meet these requirements. It aims to provide improved system capacity and coverage, more robustness, low latency, reduced cost, and flexible bandwidth operations."

Nowadays smart devices are very popular, since they can provide numerous applications to satisfy people's various demands in their daily life. People use mobile phones not only for voice communications, but more and more for entertainment, such as online games, movies, photos, and so on. All these applications need the cellular network to provide better service than with 3G (3rd Generation). Long Term Evolution (LTE), being developed by the Third Generation Partnership Project (3GPP), tries to meet these requirements. It aims to provide improved system capacity and coverage, more robustness, low latency, reduced cost, and flexible bandwidth operations. LTE Release 8[1], which was mainly designed for macro/microcell layout and finalized in 2008, has already supported up to 300 Mbps on downlink and 75 Mbps on the uplink with 20 MHz bandwidth. LTE Release 10/11[2], also named LTE-Advanced (LTE-A), significantly enhanced the capabilities of LTE, achieving better user experience. LTE-A supports the heterogeneous network layout where low-power nodes such as picocells, femtocells, and relays coexist with the macrocells. To achieve better throughput and coverage, technologies such as carrier aggregation supporting bandwidth up to 100 MHz, downlink spatial multiplexing including eight-layer spatial multiplexing and downlink coordinated multipoint transmission, uplink spatial multiplexing up to four layers, are applied. Moreover, the specifications for LTE-A are in the finalization phase, and to further enhance the radio access technologies over

the coming decade, LTE Release 12[3][4][5] (LTE-B) has already been discussed and proposed for future challenges.

In general, LTE supports two kinds of duplexing, frequency division duplexing (FDD) and time-division duplexing (TDD). For an FDD network, the downlink transmission and uplink transmission are separated by different carriers, and the spectrum must be paired. But there is no phase synchronization requirement among the FDD cells, and this means both synchronous networks and asynchronous networks are supported. In practical, an FDD network is developed as an asynchronous network due to the reasons of implementation. For a TDD network, both the downlink transmission and uplink transmission are in the same frequency carrier for a cell. The TDD LTE network is not only welcomed by operators who have already built the TD-SCDMA network since the two networks can coexist with very few modifications, but also operators with TDD band license, such as Softbank in Japan, and Verizon in America. This gives TDD-LTE technology a better chance to be successful in the market, especially in China due to its huge population. From a technical point of view, TDD LTE achieves more flexibility of the spectrum allocation since no paired spectrum is needed. Furthermore, since the uplink and downlink are transmitted in the same frequency and face the same fading, it enables a base station, called Evolved Node B (eNB) in LTE, to use channel reciprocity in downlink spatial multiplexing, thereby reducing complexity of the user end (UE). However, to support the time-division duplexing, some additional requirements have to be satisfied. For example, a guard period (GP) has to be left for supporting RF switching from downlink to uplink in time domain, which may reduce the spectrum efficiency. Furthermore, the base stations have to be synchronized in order to avoid interference between uplink and downlink transmission. All of these changes and challenges for TDD network will be discussed in following sections.

"For an FDD network, the downlink transmission and uplink transmission are separated by different carriers, and the spectrum must be paired."
"For a TDD network, both the downlink transmission and uplink transmission are in the same frequency carrier for a cell."

This article will focus on the physical layer of LTE TDD and it is organized as follows. First, a brief overview of LTE physical layer and the TDD-LTE specified difference compared with FDD LTE are provided in the next section. The overview is a general description of downlink/uplink transmission and orthogonal frequency-division multiplexing (OFDM)/ single-carrier frequency-division multiplexing (SC-FDM) modulation. The TDD LTE specific differences are analyzed from six aspects: duplex mode, synchronization reference signals, downlink and uplink configuration, special subframe, random access, and acknowledgement and hybrid automatic repeat request (HARQ) procedure. In the section "Challenges in TDD LTE Systems," the challenges in TDD LTE systems are analyzed in detail. Note that the challenges in the TDD LTE system are not limited to the automatic gain control (AGC) and synchronization that we presented in this section. The "Conclusion" section summarizes the benefits and difficulties of TDD LTE systems and suggestions for further possible optimizations.

TDD LTE Systems

The TDD LTE and FDD LTE share the same air interface in the physical layer. The major difference between TDD LTE and FDD LTE is in the frame structure. In this section, we first give an overview of the LTE system and then describe the TDD specific differences and their possible effects on system design.

LTE System Overview

This section gives an overview of the LTE system, mainly focusing on the common parts of FDD LTE and TDD LTE networks in the physical layer.

Downlink Transmission

"OFDM is the key point of the downlink transmission in an LTE system. The basic idea for OFDM is separating the high speed transmission bits to numbers of parallel streams, transmitted by a narrow-band subcarrier."

OFDM is the key point of the downlink transmission in an LTE system. The basic idea for OFDM is separating the high speed transmission bits to numbers of parallel streams, transmitted by a narrow-band subcarrier. Usually it is implemented by Inverse Fast Fourier Transformation (IFFT) in practice. After that, a cyclic prefix (CP) is added to the OFDM symbol, which is used to deal with the inter-symbol interference. The length of CP is decided by the delay spread of the channel in the time domain. With the IFFT and CP, the wideband-transmissions under frequency-selective channel turns to numbers of the narrowband-transmissions under flat-fading channels and the signals are protected from the interference from the time domain. This may lead UEs to avoid using the advanced and potentially complex channel equalization. This property of OFDM is quite attractive because it not only helps to simplify the implementation of UEs, but also reduces the power consumption without performance degradation. Furthermore, transmission under flat fading channel is even more important for the UE when the multiple antenna technology is used for spatial diversify and/or multiplexing purposes.

On the other hand, OFDM also has some disadvantages. One is the higher value of peak-to-average power ratio (PAPR) which is mainly caused by combining all the subcarriers together and in some scenarios the PAPR value could be higher than 12 dB. The higher PAPR leads to a higher cost on the linear power amplifier on the transmitter side and higher power consumption due to the lower energy efficiency. The comprehensive discussion/solutions are described by Seung Hee Han and Jae Hong Lee[6] and Tao Jiang and Yiyan Wu[7]. The other disadvantage is the performance sensitivity of the frequency offset. A very small frequency offset, such as 2 to 5 percent of the bandwidth of the subcarrier usually will lead to great performance degradation on the receiver side. Therefore, UEs need to use the complex frequency estimation/compensation algorithm, especially when the crystal oscillator on the UE side is not good enough.

"Although there are some drawbacks for OFDM modulation, it is still used by the LTE network for downlink transmission, due to good performance on the fading channel with simpler channel equalization."

Although there are some drawbacks for OFDM modulation, it is still used by the LTE network for downlink transmission, due to good performance on the fading channel with simpler channel equalization. Generally, a frame in LTE is 10 ms and is divided equally into 10 subframes, which last 1 ms each. Each subframe consists of two slots of length 0.5 ms (7 OFDM symbols for normal CP and 6 OFDM symbols for extended CP). Each grid in the time

and frequency domain is defined as a resource element (RE), and a resource block (RB) is defined as 12 RE in the frequency domain by one slot in the time domain. A typical downlink subframe with normal CP and two TX antenna ports is presented in Figure 1.[8] The cell-specific reference signals (CRS) for two TX antenna ports (R0 and R1) that are broadcast by the base station are also displayed in this figure.

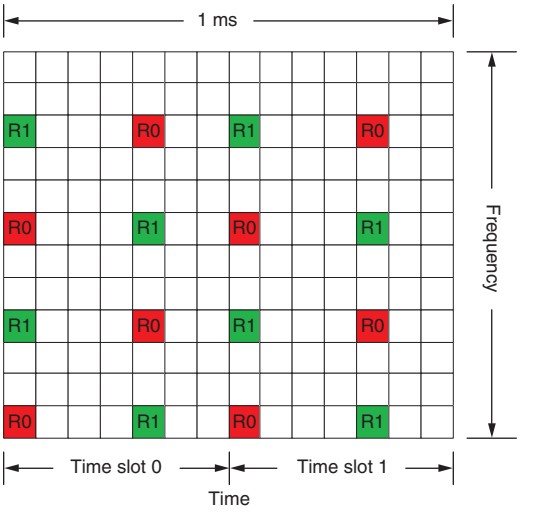


Figure 1: Downlink subframe with normal CP and two TX antenna ports
(Source: Modified from 3GPP 36.211, 2013[8])

Uplink Transmission

OFDM modulation is not suitable for UEs as uplink transmission due to the higher PAPR value described in the previous subsection. It is neither easy for UEs to use potential complex PAPR reduction algorithms nor to apply a linear power amplifier with a wide dynamic range. It will not only increase the cost of the UE but also the power assumption. The prices of UEs are a sensitive factor in markets and the standby time of the UE impacts the user experience quite a lot. Therefore, single-carrier frequency-division multiplexing access (SC-FDMA) is applied in the uplink transmission scheme.

The SC-FDMA has similar numerology to the OFDMA transmission scheme used in downlink, as shown in Figure 2. The main difference between OFDM and SC-FDM is that the constellation symbols are precoded by FFT before mapping to the difference subcarriers. This new added procedure can help to reduce the PAPR of transmitted signals, thereby increasing the coverage of the system and reducing the power consumption of UEs. Furthermore, the eNB (base station) can allocate the various UEs to different subcarriers by N-point IFFT as shown in the figure, thereby reducing or avoiding the interference among UEs.

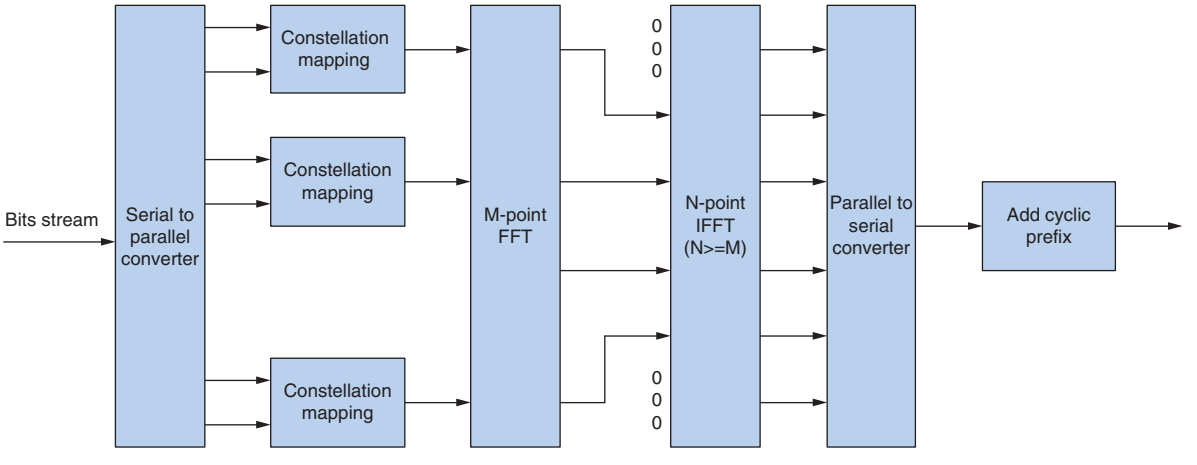


Figure 2: Block diagram for SC-FDMA
(Source: Intel Corporation 2013)

The uplink subframe is also 1 ms and has the similar structure to the downlink subframes. A typical uplink subframe for normal CP is shown in Figure 3. The demodulation reference signal, used for channel estimation and demodulation of uplink data and control, is one symbol in each time slot and is located in the middle of the corresponding time slot for normal CP.[8] In the case of extended

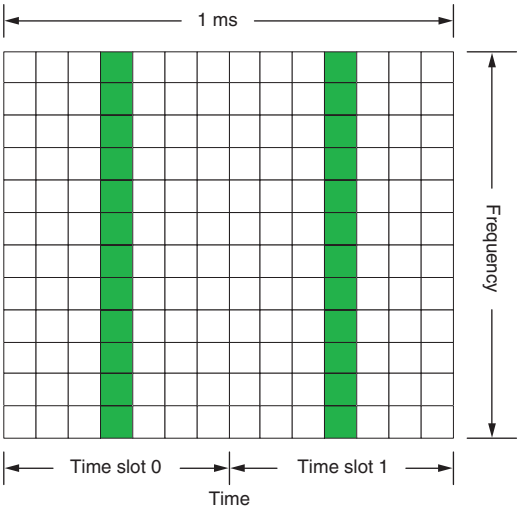


Figure 3: Uplink subframe structure for normal CP
(Source: Intel Corporation 2013)

"The key point is the uplink and downlink transmission using the same frequency carrier but separated by time domain."

CP, where only six symbols are in a time slot, the reference signal is located in the third symbol in a time slot.

TDD and FDD difference

The common properties of TDD LTE and FDD LTE systems are discussed in the previous subsection, and the main differences of these two systems in the physical layer are presented from the standard specification point of view below.

Duplex Mode

As the name of TDD LTE systems makes apparent, the most obvious difference from FDD LTE is the duplex mode. The key point is the uplink and downlink transmission using the same frequency carrier but separated by time domain. The resources in the time domain are allocated to uplink and downlink alternately, and therefore the transmission of uplink and downlink is not continuous.

Compared with FDD-LTE systems, some benefits can be obtained by applying the TDD technologies:

- The TDD network doesn't require the paired spectrum. This may give the operators flexibilities to build their network according to the traffic load and spectrum they bought from the government. It is especially important for small operators who cannot afford enough spectrum to build an FDD LTE network.
- Dynamic adjustment of the resource for uplink and downlink transmission due to the different applications and service type is much easier in TDD LTE networks, since it is not necessary to change the bandwidth of carriers.
- The complexity of channel equalizer from UE side can be reduced because of the channel reciprocity, especially for wide-bandwidth systems, for example, 20 MHz in LTE systems. The UE can get more benefits in multiple antenna scenarios.

On the other hand, there are also some issues for the TDD technologies:

- Strict phase/time synchronization is required by TDD networks in order to avoid interference between uplink and downlink transmission. The phase synchronization requirements for various base station types are from 3 μs to 10 μs.[9] It is also not easy for small cells (femtocells) to satisfy the phase synchronization requirements; Bladsjo et al.[10] provide a detailed discussion on this topic.
- Reduced spectrum efficiency compared to FDD network due to the switch point between uplink transmission and downlink transmission.
- Transmission is not continuous. The downlink transmission and uplink transmission occur alternately.
- The coverage is limited by not only the transmission power but also the guard period (GP) between the downlink transmission and uplink transmission due to the propagation delay.

In general, the major differences in TDD-LTE compared with FDD-LTE systems are all directly or indirectly related to the time-domain duplex mode. For example, the definition of frame structure, the switch point of downlink and uplink transmission, and so on. All of these are presented below.

Downlink and Uplink Configuration

A radio frame is divided equally into 10 subframes and each of them is 1 ms. The subframe is the minimum unit for downlink or uplink transmission in TDD systems except for special subframe in which both downlink and uplink resource is included. According to different service types, seven uplink-downlink configurations are defined in 3GPP TR 36.211.[8] In general, both two switch points and one switch point from uplink to downlink in 10 ms are supported. The subframes 0 and 5 are always for downlink transmission. A typical TDD frame structure for uplink-downlink configuration 1 is presented in Figure 4 where D, S, and U indicate downlink subframe, special subframe, and uplink subframe respectively.

"The subframe is the minimum unit for downlink or uplink transmission in TDD systems except for special subframe in which both downlink and uplink resource is included."

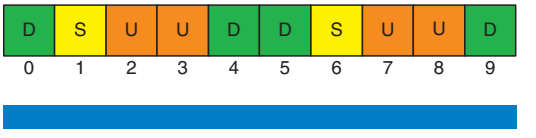


Figure 4: TDD Frame structure for uplink-downlink configuration 1 (Source: Intel Corporation 2013)

The ratios of UL resource and DL resource for each uplink-downlink configuration are shown in Table 1 without considering the special subframe. In the carrier aggression scenario where the serving cells define the various UL and DL configurations for different carrier components working on different frequency band and the UE doesn't support simultaneous reception and transmission, also called half-duplex, the uplink and downlink transmission have to follow the behavior of the primary cell.

Uplink-downlink configuration	Ratio of UL resource to DL resource without special subframe
0	1:3
1	2:2
2	3:1
3	6:3
4	7:2
5	8:1
6	3:5

Table 1: Ratio of UL resource to DL resource in one radio frame without considering special subframe (Source: Intel Corporation 2013)

The purpose of defining these uplink and downlink configurations is to make it possible to dynamically adapt the downlink resource and uplink resource according to the downlink and uplink traffic (difference service type) in real time, also named further enhancements to LTE TDD for DL-UL interference management and traffic adaptation (eIMTA), which is discussed now in 3GPP Release 12. However, because the LTE network usually builds as a lower reuse factor, for example reuse factor 1, two issues need to be kept in mind for the

"The purpose of defining these uplink and downlink configurations is to make it possible to dynamically adapt the downlink resource and uplink resource according to the downlink and uplink traffic (difference service type) in real time, also named further enhancements to LTE TDD for DL-UL interference management and traffic adaptation (eIMTA), which is discussed now in 3GPP Release 12."

uplink-downlink configuration adaptions: one is that if the uplink-downlink configuration for adjacent cells are different, the UEs in the cell edge have to face serious interference between uplink and downlink; the other is that if one cell changes the uplink-downlink configuration, the neighbor cells have to optimize their parameters, for example, transmission power, uplink-downlink configuration, and so on, to mitigate the interference/impact, and then, potentially render the whole network unstable. Therefore, the adjustment of uplink and downlink configuration has to be semi-static and must be done with care.

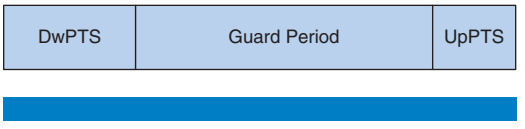


Figure 5: Structure of a special subframe (Source: Intel Corporation, 2013)

"Special subframes are 1 ms and are designed for the switch point from downlink transmission to uplink transmission."

"Another benefit of supporting the special subframe is that it makes it easier for the TDD-LTE system to coexist with currently deployed TDD system, TD-SCDMA."

Special Subframes

Special subframes are 1 ms and are designed for the switch point from downlink transmission to uplink transmission. There are one or two special subframes in 10 ms according to the uplink-downlink configuration. Three areas have been defined in special subframes[8]: downlink pilot time slot (DwPTS), GP, and uplink pilot time slot (UpPTS), shown in Figure 5. The DwPTS and UpPTS are used for downlink transmission and uplink transmission respectively, and the rest is the GP used for UEs switching from downlink transmission to uplink transmission. There are in total 10 different special subframe configurations defined in 3GPP for Release 11[8], and each of the definitions has different length of DwPTS, GP, or UpPTS.

In general, the duration of the GP has to be long enough to handle the propagation delay in the cell edges. Moreover, considering that all the uplink signals have to arrive at the base station at the same time, the GP must cover the roundtrip propagation delay plus the switching time for the UE from downlink to uplink. LTE supports GP ranges from 140–667 μs, and this means that the maximum cell radius is up to 100 km.

Another benefit of supporting the special subframe is that it makes it easier for the TDD-LTE system to coexist with currently deployed TDD system, TD-SCDMA.[2] To avoid the interference between TDD LTE systems and TD-SCDMA systems, the downlink and uplink transmission must be mutually aligned, and this can be achieved through selecting suitable special subframe configuration.

Synchronization Reference Signals

Two types of synchronization signals are defined in LTE systems: one is the primary synchronization signal (PSS) and the other is the secondary synchronization signal (SSS). There is no difference for the synchronization signal generation in TDD LTE and FDD LTE. The only difference for the PSS and SSS between TDD and FDD systems is the location in time domain[8]. For FDD-LTE systems, the PSS is located at the last OFDM symbol of slot 0 and slot 10 (first half of the subframe 1 and 5); the SSS is located at the second last symbol of slot 0 and 10. For TDD-LTE systems, the PSS is located at the third OFDM symbols in subframe 1 and 6, and the SSS is located at the last OFDM symbol of slot 1 and slot 11. The locations of PSS and SSS for both systems in the first half frame are shown in Figure 6.

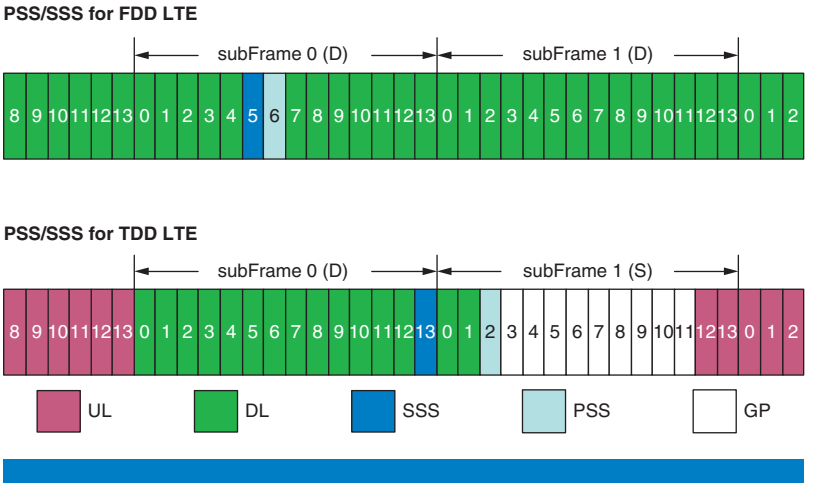


Figure 6: PSS/SSS location in time domain for the first half frame for both FDD LTE and TDD LTE systems in case of normal CP (Source: Intel Corporation 2013)

There are two possible impacts on the different location of synchronization signals:

- The difference of the location of synchronization signals for FDD and TDD LTE network can be used to detect the system is TDD or FDD according to the position of synchronization signals;
- The range of frequency offset estimation is different when applying coherent PSS/SSS frequency offset estimation due to the different distance between PSS and SSS in the time domain. In principle, the range of frequency offset estimation for TDD systems is just about one third of the range of FDD systems.

"The difference of the location of synchronization signals for FDD and TDD LTE network can be used to detect the system is TDD or FDD according to the position of synchronization signals;"

"The random access channel is used by UEs to contact the base station and access the network."

Random Access

The random access channel is used by UEs to contact the base station and access the network. For FDD networks, four random access preamble formats are defined, and at most one random access resource is allocated per subframe.

However, in TDD networks, the uplink resource is quite limited and the random access channel has to be multiplexed by not only time but also frequency domain. This means for one uplink subframe, there may be more than one random access resource allocated. Furthermore, to use the uplink resource in UpPTS, a shorter random access preamble format, format 4, is defined. It can only be used in UpPTS and when the UpPTS length is 4384 Ts and 5120 Ts.[8] Since the length of this sequence of format 4 is much shorter than the format 0 to 3, the performance of channel estimation based on the random access preamble will be degraded.

Acknowledgement and HARQ

The acknowledgement (ACK/NACK) and HARQ in LTE systems are used to improve the throughput performance of UEs. For FDD, since the

radio of uplink resource and downlink resource is 1:1, the ACK/NACK information can be provided by a fixed pattern. However in TDD systems, in most of scenario that downlink to uplink subframe ratio is not 1:1, therefore the fixed pattern for uplink and downlink ACK/NACK is not possible.

"The acknowledgement procedure for the uplink subframes is easy to be achieved since the number of downlink subframe is equal or greater than the number of uplink subframe in one radio frame except for uplink and downlink configuration 0 in which there are 6 uplink subframes and 2 downlink subframes without considering special subframes."

The acknowledgement procedure for the uplink subframes is easy to be achieved since the number of downlink subframe is equal or greater than the number of uplink subframe in one radio frame except for uplink and downlink configuration 0 in which there are 6 uplink subframes and 2 downlink subframes without considering special subframes. Fixed patterns are used for the uplink subframe acknowledgement according to the uplink and downlink configuration/reference uplink downlink configuration when it is not 0; the metric $I_{PHICH,}$ defined in 3GPP TS 213[11], is used to indicate the set of uplink subframes used for uplink transmission for (reference) uplink and downlink configuration 0, and therefore, help to find the corresponding uplink transmission for an acknowledgement.

"The acknowledgement procedure for the downlink subframes is much complex due to the number of uplink subframes is usually less than the downlink subframe except for uplink and downlink configuration 0."

The acknowledgement procedure for the downlink subframes is much complex due to the number of uplink subframes is usually less than the downlink subframe except for uplink and downlink configuration 0. In general, several mechanisms have been defined in the 3GPP standard specification according to different scenarios:

For none carrier aggregation (CA) supported UE:

- ACK/NACK bundling
- ACK/NACK multiplexing

For CA supported UE:

- ACK/NACK bundling
- Physical Uplink Control Channel (PUCCH) format 1b with channel selection
- PUCCH format 3

Determining which mechanisms to use is decided by signaling of the high layer and uplink and downlink configuration. Please refer to 3GPP TS 213[11] for detailed information.

Also due to the non-continuous transmission in TDD networks, it is hard to have a fixed pattern to retransmit the data that is not reliably received by the UE. The retransmission time interval and maximum number of HARQ processes are highly dependent on the uplink/downlink configuration. The details of the HARQ procedure are beyond the scope of this article; please refer to 3GPP TS 213[11] for detailed information.

Challenges in TDD LTE Systems

The differences between TDD and FDD LTE systems are discussed in the previous section from a standard specification point of view. In this section, we look into the difficulties met in system design and partial implementation.

Automatic Gain Control

Automatic gain control (AGC) is a technique to adjust the signal power to a suitable level whose maximum utilizes the dynamic range of RF amplifier, ADC, or bit-width in the digital signal domain.

"Automatic gain control (AGC) is a technique to adjust the signal power to a suitable level whose maximum utilizes the dynamic range of RF amplifier, ADC, or bit-width in the digital signal domain."

Problem Challenges in AGC

A typical AGC scheme applies the proper gain to received signal according to continuously calculating the power of the received signal. This may be suitable for the FDD-LTE system since the downlink transmission is continuous. However, it is not the case for TDD networks, where the uplink signals and downlink signals share the same frequency carrier in different subframes. The power of received signals may change quickly due to the uplink transmission subframe, especially when there is another UE transmitting uplink signals nearby. This requires a large dynamic-range linear amplifier, which is quite expensive, and low power efficiency and quick response AGC in order to adjust the signal power to a suitable level. Otherwise the signals will be saturated or truncated, and will cause certain performance degradation.

One possible solution for TDD LTE system might be that the UE only measures the downlink subframes to calculate the AGC gain, since only the received signals in downlink subframes are useful. The distortion of received signal during uplink subframe will not impact the performance of UEs. The frame timing and uplink downlink configuration have to be known by UEs to apply this AGC scheme. However these conditions cannot be stratified, especially during the initial cell search scenario. In that case the UE doesn't know the uplink downlink configuration and frame timing, which should be detected by the cell search procedure itself. Furthermore, it is known that at least the synchronization signal needs to be at a suitable power level in order to obtain better performance of cell detection.

Therefore, an AGC scheme that can adjust at least the power of the synchronization reference signal to a suitable value without frame timing and uplink downlink configuration information is needed by the TDD-LTE network. A three-step AGC scheme is presented and discussed in the next subsection.

"Therefore, an AGC scheme that can adjust at least the power of the synchronization reference signal to a suitable value without frame timing and uplink downlink configuration information is needed by the TDD-LTE network."

Three-Step AGC

The key concept of three-step AGC is to divide the 1 ms into three equal segments (1/3 ms for each segment), and calculate the signal power and

apply the AGC gain base on each segment. According to the TDD-LTE frame format, there is always a period of downlink transmission (1 ms plus 2 OFDM symbols) before the PSS for any starting point and any assumptions of uplink downlink configuration. And therefore, the AGC gains have already been calculated and applied three times without the uplink interference at the position of primary synchronization signal. It guarantees the primary synchronization signal to be a suitable value. There are two reasons to choose one third of 1 ms as the AGC minimum unit:

- To have a stable average power calculation, at least one cell-specific reference signal symbol (four symbols in 1-ms downlink subframe located at OFDM symbol 0, 4, 7, 11 for normal CP) that is broadcast by base stations for all time and all bandwidth must be included in the minimum AGC unit;
- At least three AGC adjustments are needed to achieve a suitable signal level.

"From the results shown in Figure 7 it is clear that the successful detection rate of initial cell search increases significantly by applying the three-step AGC scheme in all the channel conditions."

The simulation results for performance comparison of initial cell search with the three-step AGC scheme and general AGC scheme are shown in Figure 7. From the results shown in Figure 7 it is clear that the successful detection rate of initial cell search increases significantly by applying the three-step AGC scheme in all the channel conditions.

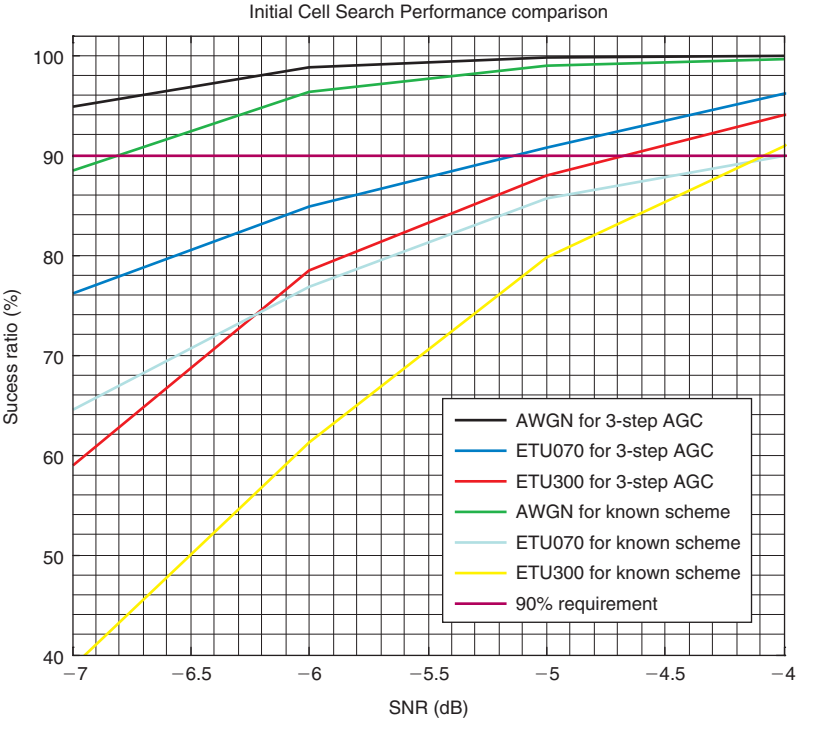


Figure 7: Performance comparisons for initial cell search with/without three-step AGC
(Source: Intel Corporation 2013)

Design Challenges in Cell Detection

Cell search is a fundamental feature for mobile phones in LTE systems, since it is not only the first step for a mobile phone to attach to a network and but it also helps the mobile phone to monitor the neighbor cells, to which the signal could be handed over if the signal qualities of the serving cell drops below certain requirements. The cell search component usually needs to find the basic information of all the available cells, including the timing information, cell ID, CP mode, and so on. The qualities of cell search results have a great impact on the mobility performance of a mobile phone, and this is a key benchmark for a mobile phone vendor. Generally, two indicators, miss-detection rate and false alarm rate, are used to evaluate the performance of cell search. How to reduce the miss-detection rate without increasing the false alarm rate would be an interesting topic.

The main difference of TDD LTE networks and FDD LTE networks for cell search procedure is the phase synchronization requirement in TDD networks. For synchronized networks, the synchronization reference signals (PSS/SSS) disturb each other and it is not easy to be cancelled due to not-so-perfect cross-correlation between PSS/SSS signals. This may lead to performance degradation compared with nonsynchronous networks, such as an FDD LTE system.

In general, the cell search can be divided into two parts: PSS detection, which decides the sector ID, half-frame timing, and a coarse frequency offset estimation; and SSS detection, which estimates the group ID, half-frame order, and CP type. In the following subsection, we discuss the PSS and SSS detection algorithm and their capabilities in TDD LTE systems.

PSS Detection

The PSS is repeated every 5 ms in a fixed position and is generated from a frequency-domain Zadoff-Chu sequence, which has very good self-correlation property in both the frequency domain and time domain. In general, the time-domain cross-correlation is used for PSS detection to find the half-frame timing. A typical PSS detection procedure is presented in Figure 8. To detect the PSS, the cross-correlation between the received signals (after down-sampling, filtering, and normalization) and locally saved three PSS defined by 3GPP, and to find the first *n*th maximum correlation values have to be done

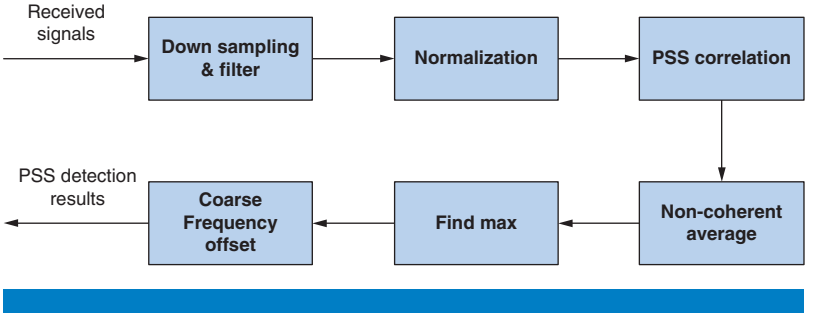


Figure 8: Block diagram for PSS detection (Source: Intel Corporation 2013)

"For synchronized networks, the synchronization reference signals (PSS/SSS) disturb each other and it is not easy to be cancelled due to not-so-perfect cross-correlation between PSS/SSS signals."

"In general, the cell search can be divided into two parts: PSS detection, which decides the sector ID, half-frame timing, and a coarse frequency offset estimation; and SSS detection, which estimates the group ID, half-frame order, and CP type."

for every 5 ms. Furthermore, to increase the detection rate of PSS and reduce the noise impact, the averaging/filtering of correlation results across several 5 ms/half-frame may be necessary.

Note that if the detected cells have the same sector ID and are strictly synchronized in the time domain, only one PSS peak can be detected. This should not be an issue since these cells can be separated during the SSS detection procedure.

In case there is a large frequency offset, for example, more than 7.5 kHz, a partition correlation for PSS detection has to be used. This method can get better performance in a large frequency offset scenario but has lower peak compared with normal correlation when the frequency offset is small.

SSS Coherent Detection Algorithm

Two secondary synchronization signals are located in downlink subframes with time interval 5 ms in one radio frame (10 ms). The aim of SSS detection is to estimate the group ID of a cell and determine the half-frame order and CP mode by detecting which SSS is applied and where it is applied.

The block diagram of SSS coherent detection is provided in Figure 9. The received secondary synchronization signal is compensated by PSS channel estimation results and averaged across several radio frames before cross-correlation with local saved SSSs defined by 3GPP specification.

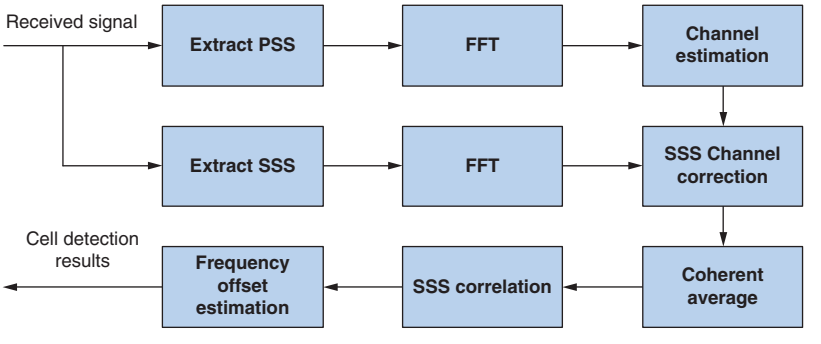


Figure 9: Block diagram for SSS coherent detection (Source: Intel Corporation 2013)

"The key point of this method is that the SSS has to be corrected by the PSS channel estimation result. Therefore, its performance highly depends on the qualities of channel estimation results."

The key point of this method is that the SSS has to be corrected by the PSS channel estimation result. Therefore, its performance highly depends on the qualities of channel estimation results. For high SNR and asynchronous networks, the coherent detection algorithm can achieve a very good detection rate. However, for TDD LTE systems, the qualities of channel estimation using PSS will degrade due to the imperfect cross-correlation among the existing three primary synchronization signals. An even worse scenario is when the detected cells use the same PSS, in which it is very hard for UEs to separate the channel from different base stations. Therefore, the performance degradation can be expected applying the coherent detection algorithm in a TDD-LTE network.

SSS Non-Coherent Detection Algorithm

To increase the performance of SSS detection in synchronized networks, a non-coherent detection algorithm is presented, which is described in Figure 10. The received secondary synchronization signals are correlated with local saved SSS and averaged across several half frames. Moreover, to further improve the performance, a joint PSS/SSS detection metric, which is not only considering the SSS correlation results but also PSS correlation results, is presented.

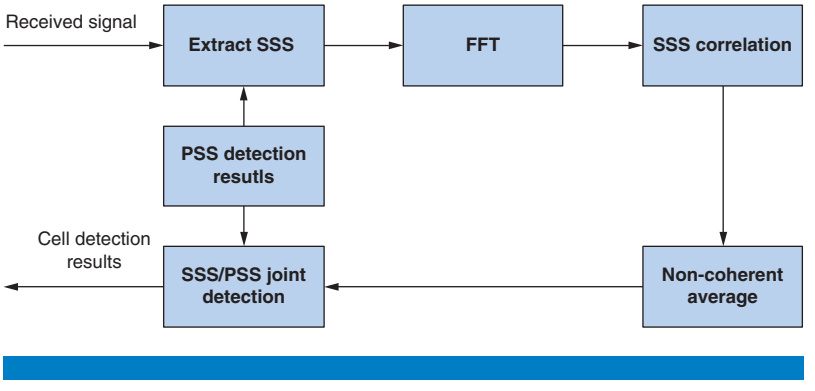


Figure 10: Block diagram for SSS non-coherent detection (Source: Intel Corporation 2013)

Compared with the coherent detection method, the non-coherent method doesn't rely on the channel estimation results, and can therefore perform better in TDD LTE networks while the coherent detection method can achieve better performance in an interference scenario.

"Compared with the coherent detection method, the non-coherent method doesn't rely on the channel estimation results, and can therefore perform better in TDD LTE networks while the coherent detection method can achieve better performance in an interference scenario."

There are still two issues left for this algorithm:

- The SSS correlation has to be done every half frame (5 ms) for each detected PSS peak, and considering the large amount of the local saved SSSs (168 cell ID groups times 2 half-frame order times), the computational complexity is quite high;
- It is not easy to do the fine frequency offset estimation without channel estimation results.

Further Enhancements of Cell Detection

In the previous subsection, we talked about the detection rate of the cell search. Another side of cell search is to reduce the false alarm rate, which is also named ghost cell detection—to filter out the fake cells from the detected cell list.

This issue can be solved via utilizing the natural characteristic of synchronized networks: all cells in the same frequency carrier are synchronized. The timing difference of base station TX in TDD LTE must be less than 3 μs/10 μs.[9] The timing differences of cells in mobile networks in one carrier are mainly depending on the propagation delay of different base stations, which is related to the radius of cells. For a given cell radius, it is easy to calculate the maximal propagation delay and the maximal timing difference. This gives a limitation

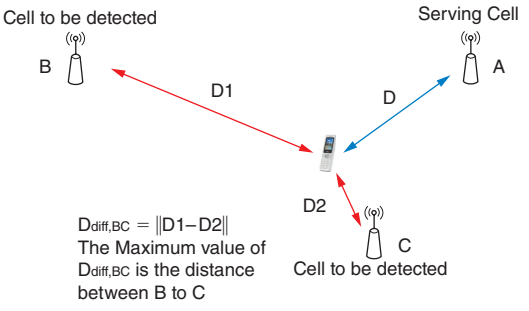


Figure 11: Timing difference of cells (Source: Intel Corporation 2013)

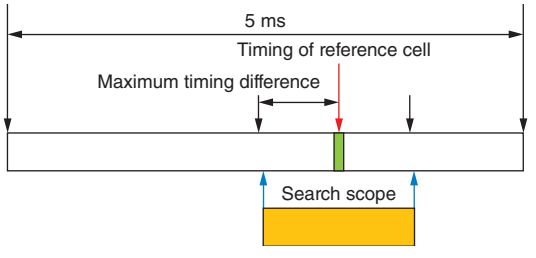


Figure 12: Narrowing search scope according to the timing information of reference cell (Source: Intel Corporation 2013)

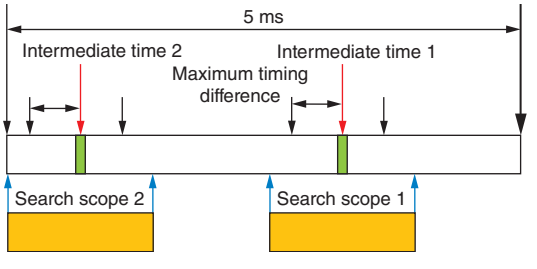


Figure 13: An example of narrowing search scope according to the intermediate search results (Source: Intel Corporation 2013)

of valid cells in a time domain and it can be used to reduce the computational complexity and improve the performance of cell search via narrowing the correlation and search scope—all the detected cells must be in this scope (see Figure 11).

When the timing information of a valid cell is known before the starting of cell search, the search scope can be limited according to the timing information of the known valid cell, shown in Figure 12.

If no such information is provided, the cell search can also use the timing constraints between any detected cells in a synchronized network to reduce the computational complexity and/or memory consumption via separating the PSS detection procedure into two steps (the basic idea is presented in Figure 13):

- Step 1: the cell search processes data as usual across the first several half frames and detects the PSS to obtain the intermediate half-frame time results;
- Step 2: the intermediate results can be used as a kind of timing reference for PSS detection.

The key point of this method is to utilize the timing constraints of cells in synchronized networks to reduce the search scope and filter out the unwanted "ghost cells."

Conclusion

TDD-LTE and FDD-LTE are two kinds of duplex mode supported by the LTE standard specification. This article explained the differences between the two networks in physical layer.

In the first part of this article, the two networks were compared and analyzed from the standard specification point view, and this part can also answer why these changes are needed when adapting from FDD to TDD LTE systems and what we obtain and lose from this adaption.

In the second part of the article we discussed some detailed issues for TDD-LTE networks, including the automatic gain control and cell search. For AGC, a three-step AGC scheme is presented and the simulation result has verified the performance improvement. For cell search, we compared and analyzed the coherent detection and non-coherent detection, and arrived at the conclusion that it is better to use non-coherent detection in TDD-LTE networks. Furthermore, another enhancement for reducing the false alarm rate of cell search is presented.

All in all, although there are various issues existing in the TDD-LTE network, TDD-LTE is still an attractive technology from the operator point of view, due to its flexibility of network deployment. The challenges faced by TDD-LTE in physical layer are not only limited by the aspects described in this article, but also include some other topics. For example, channel estimation under non-continuous downlink transmission leading to UE

throughput degradation, successive interference cancellation applied in cell search and downlink data receive, among other issues, need to be further investigated.

References

[1] Dahlman, Erik, Stefan Parkvall, Johan Sköld, and Per Beming, *3G Evolution: HSPA and LTE for Mobile Broadband*, 2d ed., Academic Press, 2008.

[2] 3GPP TR 36.913 "Requirements for Further Advancement for Evolved Universal Terrestrial Radio Access (E-UTRA)," V11.0.0, 2012/09. (ftp://ftp.3gpp.org)

[3] Intel, "LTE Rel-12 and Beyond: Requirements and Technology Components," 3GPP RSs-120023, June 2012. (http://www.3gpp.org/Future-Radio-in-3GPP-300-attend)

[4] NTT DOCOMO, "Requirements, Candidate Solutions & Technology Roadmap for LTE REl-12 onward," 3GPP RWS-120010, June 2012. (http://www.3gpp.org/Future-Radio-in-3GPP-300-attend)

[5] NSN, "Release 12 and beyond for C^4 (Cost, Coverage, Coordination with small cells and Capacity)," 3GPP RWS-120004, June 2012. (http://www.3gpp.org/Future-Radio-in-3GPP-300-attend)

[6] Seung Hee Han and Jae Hong Lee, "An overview of peak-to-average power ratio reduction techniques for multicarrier transmission," *Wireless Communications*, IEEE, vol. 12, Issue 2, Pages 56–65, 2005.

[7] Tao Jiang and Yiyan Wu, "An Overview: Peak-to-Average Power Ratio Reduction Techniques for OFDM Signals," *Broadcasting, IEEE Transactions on*, vol. 54, Issue 2, Pages 257–268, 2008.

[8] 3GPP TR 36.211 "Physical Channels and Modulation for Evolved Universal Terrestrial Radio Access (E-UTRA)," V11.3.0, 2013/06. (ftp://ftp.3gpp.org)

[9] 3GPP TR 36.133 "Requirements for support of radio resource management for Evolved Universal Terrestrial Radio Access (E-UTRA)," V11.5.0, 2013/07. (ftp://ftp.3gpp.org)

[10] Bladsjo, D., M. Hogan, and S. Ruffini, "Synchronization aspects in LTE small cells," *Communications Magazine*, IEEE, vol. 51, Issue 9, pages 70–77, 2013.

[11] 3GPP TR 36.213 "Physical layer procedures for Evolved Universal Terrestrial Radio Access (E-UTRA)," V11.4.0, 2013/09. (ftp://ftp.3gpp.org)

Author Biographies

John Zhou (john.zhou@intel.com) received his BE and ME degrees from Xi'an Jiaotong University, China, in 1999 and 2002 respectively, and his PhD from the University of Edinburgh, UK, in 2009, all in electrical engineering. From 2011 he joined Intel as system engineer, focusing on the cell search, frequency, and timing synchronization in OFDM systems. His interests are in wireless communications and digital signal processing.

Yubao Li (yubao.li@intel.com) received his Master's degree in electrical engineering from the China Academy of Telecommunications Technology in 2007. He has been with Intel since 2011. Currently he holds a position as a system engineer, working mainly in RF frontend control and DigRF interface related areas.

Qingyun Guo (qingyun.guo@intel.com) received her ME degree from China Academy of Telecommunication Technology in 2003 and majored in telecommunication and information system engineering. After graduating she worked on the TD-SCDMA physical layer for five years and the TDD LTE physical layer for three years at ST Ericsson China. She joined Intel in 2011 as a systems engineer, focusing on LTE firmware. Her responsibilities mainly include defining physical layer use cases and procedures for TDD LTE features and interface definition between subcomponents of the physical layer.

THE EVOLUTION OF THE PROTOCOL STACK FROM 3G TO 4G AND 5G

Contributors

Geng Wu
Mobile and Communications Group
Intel Corporation

Apostolos (Tolis) Papathanassiou
Mobile and Communications Group
Intel Corporation

Qian (Clara) Li
Mobile and Communications Group
Intel Corporation

Mo-Han Fong
Mobile and Communications Group
Intel Corporation

Over the past decade the wireless industry increased capacity a thousandfold, and by doing so not only fueled the explosive growth of mobile communication applications but also shaped the global mobile Internet as we know today. Three key factors contributed to this huge capacity expansion: the dramatic improvement of spectral efficiency in 3G and 4G air interfaces, the unprecedented spectrum allocation for mobile networks in the world, and the continuing evolution of the radio access network and core network architecture.

This article describes the pivotal role of wireless protocol stack design behind these three factors that enabled and supported the evolution from 3G to 4G networks. We review the technology concepts in each generation of protocol stack design, the main challenges and tradeoffs, the performance improvements, and the impact on network architecture. Moving forward, as 4G radio link performance is approaching the theoretical Shannon limit, the focus of the wireless technology development has to shift from spectral efficiency to network efficiency and to energy efficiency. In addition, the proliferation of wearable devices will further transform user experience and impose a new set of more stringent requirements on the protocol stack design and the underlay network architecture, which directly affects the device and network platform design.

"…The evolution of protocol stack design has been enabling the performance breakthrough in each generation of wireless technology."

The General Principles of Air Interface Protocol Stack Designs

In a communication network, the term *protocol stack* refers to the software and hardware implementation of a protocol suite. A *protocol suite* is a set of protocols that are structured into layers to collectively perform specific tasks at different levels amongst connected nodes. The layered structure of a protocol stack allows modularized design and implementation for large scale and complex networks. It also reduces cost for testing, performance optimization, and maintenance, since design changes can be confined to a specific protocol layer or an implementation module without affecting the rest of the network. Since network architecture directly maps protocol stack into logical and physical entities for implementation, an evolution of the network architecture is always associated with protocol stack design changes. In fact, the evolution of protocol stack design has been enabling the performance breakthrough in each generation of wireless technology.

Wireless air interface protocol stack design follows a layered structure. In the vertical direction, each protocol layer performs its layer-specific functions, uses services from the layer below, and provides services to its upper layer. This modularized design methodology is straightforward from an implementation

and maintenance viewpoint, but it has its limitations. In fact, cross-layer optimization has to undo some of strictly layered protocol stack design. H-ARQ in 3G is arguably a classic example of breaking the traditional partition between the physical layer and the data link layer.

To facilitate network implementation and software manageability, the protocol stack is also structured horizontally into a Control Plane and a User Plane. This partition enables more economical and high performance network implementation, since it separates the complex logic and state machine implementation of the Control Plane from the real-time signal and packet processing implementation of the User Plane. In addition, several services have benefited from this partition to achieve high performance; one such example is Push-to-Talk, where Control Plane signaling was accelerated to achieve fast connection setup. Once again, there are limitations to this horizontal protocol stack structure. A good example is the signaling overload in the network due to always-on short data exchanges from mobile applications. There are solutions to address the issue, in particular through network virtualization, which enable dynamic signal processing resource allocation between Control Plane and User Plane workloads. Moving forward, we anticipate further separation between the Control Plane and the User Plane to enable inter-band carrier aggregation in future heterogeneous networks and to improve network energy efficiency by actively turning on and off the User Plane of a base station. We also expect protocol design complexity shifting from the Control Plane to the User Plane to reduce implementation complexity for interworking between networks. One example is to deliver QoS (quality of service) across multiple networks. A traditional way of doing this is for two networks to negotiate with each other through Control Plane functions and set up the User Plane accordingly. The development of such a solution is complex due to different QoS parameters and local policies, and is difficult to scale up when more networks are involved. An increasing trend in the industry is to directly offer QoS handling in the User Plane through techniques such as DPI (Deep Packet Inspection), thereby reducing network dependency on Control Plane-based solutions and reducing complexity. This approach also facilitates future development of more advanced cross-layer optimization since the adaptation to link conditions without Control Plane signaling is usually much faster.

"We expect protocol design complexity shifting from the Control Plane to the User Plane…"

In addition to the Control Plane and User Plane, there is also a less visible OAM (Operations, Administration, and Management) Plane within the network. The OAM Plane is crucial for the operation of a network. In recent years, the industry is looking at expanding the OAM Plane for network self-configuration, in particular in heterogeneous networks. One such example is to use the OAM Plane for passing network loading conditions between networks and to dynamically manage and configure mobile devices.

"…the industry is looking at expanding the OAM Plane for network selfconfiguration, in particular in heterogeneous networks."

When we discuss protocol stack designs, it is worth of mentioning the concept of the logical and the physical channels. The logical channel consists of the information carried over the physical channel. The physical channel is the medium over which the information is carried. This concept has been widely

used since GSM, and has since been one of the fundamental design principles in the 3GPP family of standards. The separation of the information and the physical medium allows significant protocol stack design and implementation reuse when the physical medium evolves from generation to generation. It also provides the flexibility in multiplexing for performance optimization purposes.

"One important but sometimes overlooked issue in protocol stack designs is latency."

One important but sometimes overlooked issue in protocol stack designs is latency. Protocol stack designs focus mainly on the logical operation of a communication network, to ensure a consistent definition of syntax, semantics, and synchronization of communication, and to handle the exceptions when errors occur. With the exception of the physical layer design where tight latency requirement is part of the specifications, upper layers mostly define timers for handling exceptions only. For applications that require very short latency for connection setup, such as Push-to-Talk, the protocol stack design has to be optimized across layers to meet user expectations.

Backward compatibility is another key consideration, since it leverages the existing installed base, allows significant design reuses and interoperability with the existing networks and services at the expenses of added complexity to the protocol stack design. Different standards organizations may choose different approaches. One such example is the design of HSPA in 3GPP where both circuit- and packet-switched data channels were integrated together for strict backward compatibility, while 3GPP2 chose to keep 1xRTT and 1xEV-DO separate at the air interface level but integrate between these two networks.

Finally, although radio link is arguably the most critical and often the most expensive segment in a mobile network, it is nevertheless only part of the end-to-end solution. Traditional design has focused mostly on the optimization within the air interface. In the future, air interface protocol stack designs need to support not only the cross-layer optimizations with upper layers but also the interworking and the integration of other types of air interface connections in a heterogeneous network environment.

"…air interface protocol stack designs need to support not only the cross-layer optimizations with upper layers but also the interworking and the integration of other types of air interface connections…"

Fundamentally, the design and the evolution of the protocol stack are about the tradeoffs amongst performance, scalability, and complexity. The evolution of the protocol stack is essentially a process of continuing optimization among these three competing factors.

3G CDMA and UMTS Protocol Stack Design and Performance Improvement

The 3G wireless systems and standards are mainly comprised of UMTS/HSPA, defined in 3GPP, and cdma2000, defined in 3GPP2. The initial focus of 3G technology is to significantly improve the voice capacity of 2G GSM/TDMA systems. Hence, 3G systems were initially designed (in UMTS and cdma2000/1xRTT) to be voice-centric systems where voice capacity, voice quality, and service continuity during handover are the key design criteria.

3G CDMA Protocol Stack

A significant improvement in voice capacity is brought forth by the introduction of Code Division Multiple Access (CDMA). As opposed to Time Division Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA), the radio resources among UEs are not divided in orthogonal time or frequency domain. Instead, UEs transmit/receive on the same time/frequency resource while the transmit signals are separated by quasi-orthogonal spreading codes. The quasi-orthogonal nature of the spreading codes, in particular frequency in selective/time-dispersive mobile channels, leads to nonzero inter-user interference at the received signal. As the level of interference is highly dependent on the received power of the interfering signal, transmit power control which is used to minimize inter-user interference is a crucial technique in CDMA systems to ensure the minimum possible transmit power is used while maintaining an acceptable received signal-to-interference-plus-noise level (SINR). Power control is particularly important for the case of uplink to mitigate the near-far problem where a UE located close to the base station will transmit at a much lower power than that of a UE located at the cell edge such that the received power of the two UEs at the base station is comparable.

"As the level of interference is highly dependent on the received power of the interfering signal, transmit power control which is used to minimize inter-user interference is a crucial technique in CDMA systems…"

To ensure acceptable voice quality, the received SINR has to be tightly maintained at a level that provides an acceptable voice packet error rate (typically 1–3 percent), regardless of the channel conditions. As a result, fast transmit power control is required, which is on the rate of 1500 Hz and 800 Hz respectively for UMTS and cdma2000/1xRTT.

In addition to power control, another important technique introduced in UMTS and cdma2000/1xRTT is soft handover. Soft handover ensures the transmission link to/from the UE is not interrupted when the UE moves from one base station to another. In the soft handover region, a UE transmits/receives to/from multiple base stations. In the downlink, the signals from multiple base stations are soft-combined at the UE receiver. In the uplink, the signals received at multiple base stations are selectively combined. In order to support soft handover, a centralized layer 2 design is required so that layer 2 is anchored at a centralized entity in the access network. Such a centralized network entity is called the Radio Network Controller (RNC) in UMTS or the Base Station Controller in cdma2000/1xRTT. In the downlink, the RNC/BSC multicasts the layer 2 PDUs to multiple base stations involved in the soft handover operation of a UE. In the uplink, the RNC/BSC performs selective combining of the layer 2 PDUs received from multiple base stations.

The Radio Access Network (RAN) architecture and the air interface protocol layers mapping of UMTS and cdma2000/1xRTT are shown in Figure 1.

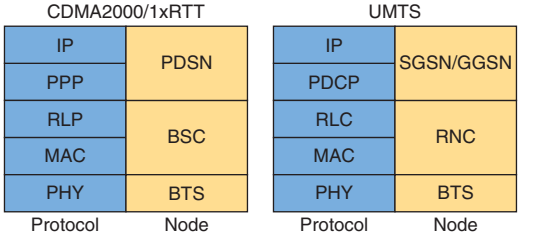


Figure 1: RAN architecture and mapping to the protocol layers for 1xRTT and UMTS (Source: Intel Corporation, 2013)

3G HSPA and 1xEV-DO Protocol Stack

As 3G systems evolve from UMTS to HSPA and from cdma2000/1xRTT to cdma2000/1xEV-DO, the focus is shifted to efficient support of packet data services, which typically are more delay tolerant and whose traffic is bursty in nature. The emphasis is no longer to provide a sustained data rate to the

"…The modulation/coding scheme (MCS) and hence the effective data rate (rather than the transmit power) is adjusted according to the fast fading channel condition,…"

UE as in the case of circuit-switched voice. A new technique called *fast link adaptation* (or *adaptive modulation/coding*) was introduced. In this technique, the modulation/coding scheme (MCS) and hence the effective data rate (rather than the transmit power) is adjusted according to the fast fading channel condition, such that the received SINR for the chosen MCS meets the required SINR for a certain target packet error rate. When the channel is in up-fade, a higher MCS can be chosen, while in down-fade, a lower MCS can be chosen. MCS is chosen at the transmitter based on channel quality indicator (CQI) feedback from the receiver. To combat fast fading, fast CQI feedback of up to 2 ms is used. In addition to fast CQI feedback, fast scheduling is employed at the base station to enable a short roundtrip time (RTT) from when the CQI is received to when the next packet is scheduled to be transmitted. In addition to combating fast fading, fast scheduling can also boost system capacity through multiuser diversity whereby a UE with good channel condition is prioritized by the scheduler.

In addition to fast link adaptation and fast scheduling, another technique called hybrid ARQ (HARQ) is also introduced. HARQ works hand-in-hand with fast link adaptation. As the UE speed increases, the high Doppler of the channel renders the MCS selection to be less accurate even with fast CQI feedback. The packet error caused by inaccurate CQI selection can be mitigated by HARQ retransmission from the transmitter and soft-combining at the receiver. To trigger HARQ retransmission, the receiver sends HARQ positive or negative acknowledgement (ACK or NACK) to the transmitter at every reception of a physical layer packet.

Due to the desired short RTT for fast link adaptation, fast scheduling and HARQ, these operations are implemented at the base station rather than the RNC/BSC. As opposed to UMTS and cdma2000/1xRTT where layer 2 protocols are anchored at the RNC/BSC, in HSPA and cdma2000/1xEV-DO part of the MAC functions (fast link adaption, fast scheduling, HARQ) resides in the base station. Hence the MAC is split between centralized and distributed entities in the RAN.

The RAN architecture and the air interface protocol layer mapping of HSPA and cdma2000/1xEV-DO are shown in Figure 2.

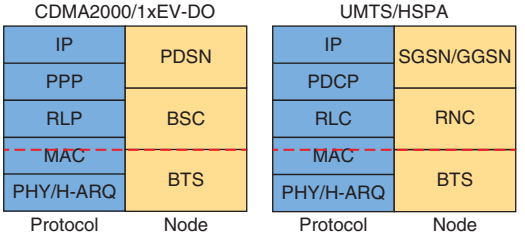


Figure 2: RAN architecture and mapping to the protocol layers for 1xEV-DO and HSPA (Source: Intel Corporation, 2013)

In the case of UMTS/HSPA, both circuit-switched voice and high speed packet data are supported and integrated in the same system sharing the same spectrum. In the case of 1xRTT and 1xEV-DO, they are two separate radio networks that use separate spectra, that is, simultaneous circuit-switched voice on 1xRTT and packet data on 1xEV-DO is not possible.

4G LTE Protocol Stack Design and Network Architecture Evolution

Motivated by the increasing demand for mobile broadband communications requiring higher system spectral efficiency, higher user data rates and better QoS, 3GPP introduced in Release 8 the Evolved Packet System (EPS),

which is the combination of LTE (Long-Term Evolution) and SAE (System Architecture Evolution) defining the radio access network and the network core, respectively. From the radio access point of view, the key radio access technologies of the new system are OFDM (Orthogonal Frequency Division Multiplexing) modulation to increase the system spectral efficiency and MIMO (Multiple-Input Multiple-Output) techniques to increase the system and user data rates. From the network core point of view, an all-IP flat architecture supporting QoS is defined. By the time LTE Release 9 provided minor upgrades to Release 8, the ITU (International Telecommunications Union) had published its IMT-Advanced, also known as 4G, requirements, which revealed that, although LTE Release 8 meets most of those requirements, additional standardization work had to be done to arrive at LTE Release 10 which, through the introduction of advanced beamforming and carrier aggregation techniques, meets all 4G requirements as defined by the ITU.

"…The key radio access technologies of the new system are OFDM (Orthogonal Frequency Division Multiplexing) modulation…"

4G Protocol Stack

The definition of LTE and SAE led to the specification of the Evolved Packet Core (EPC), Evolved Universal Terrestrial Radio Access Network (E-UTRAN), and Evolved Universal Terrestrial Radio Access (E-UTRA), which correspond to the core network, radio access network, and air interface, respectively.

The EPC is a flat, all-IP-based core network that can be accessed through both 3GPP radio access (UMTS, HSPA/HSPA+, LTE) and non-3GPP radio access (such as WiMAX and WLAN) networks. The main components of the EPC are the MME (Mobility Management Entity), the Serving Gateway (S-GW) and the Packet Data Network Gateway (PDN-GW). The MME, among other functionalities, manages the security functions (authentication, authorization, NAS signaling), roaming, and handovers, and also selects the S-GW and the PDN-GW. The S-GW is unique for each UE in the network, is the mobility anchor for both inter-eNB and inter-3GPP system mobility, and performs packet routing and forwarding. The PDN-GW, among other functionalities, provides the UE access to a PDN by assigning an IP address from the PDN to the UE. Further, the evolved Packet Data Gateway (ePDG) provides secure connections between UEs from an untrusted, non-3GPP access network with the EPC by using IPSec tunnels.

The core part of the E-UTRAN is the enhanced Node B (eNB), which provides the air interface with both user and control plane protocol termination for the UE. An eNB can serve one or more E-UTRAN cells with the interface interconnecting the eNBs termed the X2 interface. Figure 3 illustrates the main elements of the 4G protocol stack. The user plane includes the Packet Data Convergence Protocol (PDCP), Radio Link Control (RLC), Medium Access Control (MAC), and Physical Layer (PHY) protocols. The control plane additionally includes the Radio Resource Control (RRC) protocol.

As mentioned above, MIMO is a key technology in LTE, enabling an efficient utilization of multiple antennas implemented at both eNB and UE transceivers and leading to significant spectral efficiency increase through its synergy with

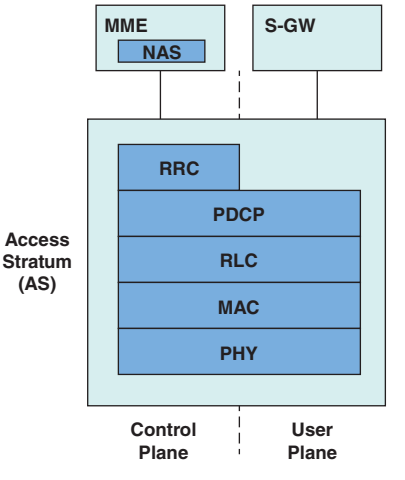


Figure 3: 4G protocol stack (Source: Intel Corporation, 2013)

OFDM modulation. Also, MIMO was shown to be a key technology for LTE-Advanced (LTE Release 10 and beyond) in meeting the IMT-Advanced performance requirements with respect not only to the peak throughput, but also the cell average and cell-edge user throughput.

MIMO and CoMP Protocol in LTE

In order to make the application of MIMO techniques possible, some level of channel state information (CSI)—the level depending mainly on the specific MIMO technique and the duplexing scheme—is necessary at the eNB base station for the system to appropriately adapt to the radio channel varying conditions, including its spatial characteristics. In TDD (Time Division Duplex) systems, where downlink and uplink share the same carrier, CSI for the downlink is typically collected from the uplink because of the TDD channel reciprocity. In FDD (Frequency Division Duplex) systems, where downlink and uplink operate on different carriers, CSI needs to be fed back over the uplink for usage in the downlink. Since the transmission of short-term CSI—also referred to as full CSI—would lead to unacceptable uplink control overhead, long-term and quantized CSI is employed in practice. MIMO in LTE can be considered as an adaptive, multimode framework with the specific MIMO technique(s) selected according to the current system requirements, including radio channel conditions, which are dynamic in nature. In principle, there are three main MIMO concepts in LTE: Single-User MIMO (SU-MIMO), Multi-User MIMO (MU-MIMO), and Cooperative MIMO (typically referred to as CoMP: Cooperative Multi-Point transmission.

SU-MIMO includes both transmit diversity and spatial multiplexing in combination with beamforming and, along with high-order MIMO (high number of antennas), can lead to a dramatic increase of the user data rate whenever possible (from the viewpoints of antenna deployment and channel conditions). Unfortunately, the probability of spatial multiplexing (SM), which is the transmission of more than one data stream to the same user, rarely exceeds 15–20 percent in typical 4G network deployments even with advanced MIMO receivers. MU-MIMO has been shown to overcome the deficiencies of SU-MIMO by allowing for multiplexing data streams from multiple users rather than multiplexing multiple data streams from a single user. In this way, MU-MIMO turns the fundamental problem of SU-MIMO, which is low-rank channels, into an advantage, especially when multiplexing the appropriate set of users based on criteria such as inter-user spatial correlation factors.

The MIMO concept can be extended to multiple cells by enabling multi-cell cooperation, known as CoMP in LTE. CoMP is introduced in LTE Release 11 and enables spectral efficiency benefits in addition to single-cell SU/MU-MIMO for both downlink and uplink in LTE. Depending on the amount of information shared among the cooperating cells, CoMP techniques are classified into joint processing and distributed processing techniques. Joint processing is made possible when cooperating cells share both transmit data and CSI for the UEs served in the CoMP mode. In the downlink, there are two joint processing schemes: JT (Joint Transmission) and Dynamic Point

"Depending on the amount of information shared among the cooperating cells, CoMP techniques are classified into joint processing and distributed processing techniques."

Selection (DPS). In JT, cooperating nodes jointly transmit data to one or more UEs. DPS is a form of fast cell selection in the sense that one or more UEs are served by the best node in the CoMP cluster according to criteria involving for instance the interference situation at the UE side. In the uplink, joint reception (JR) can be considered as the uplink counterpart of the downlink JT scheme. Although joint processing schemes lead to the best system spectral efficiency among the available CoMP schemes, they require high-capacity and low-latency backhaul links among the cooperating nodes. Therefore, alternative CoMP techniques which have less stringent requirements on the backhaul link capacity and latency are studied in 3GPP Release 12.

As already mentioned, beamforming, also termed as precoding, is a key technology in maximizing the benefits of MIMO in LTE. Codebook-based precoding is adopted in LTE Release 8 by defining a common codebook that comprises a fixed set of vectors and matrices and supports various antenna configurations and number of MIMO spatial streams. In this way, a UE can calculate and feedback the so-called PMI (Precoding Matrix Index), based on reference signals transmitted in the downlink, in order to enable downlink precoding at low control overhead. LTE-Advanced adopted closed-loop precoding to realize high-order MIMO spatial multiplexing (up to eight streams for an 8x8 MIMO system) under the assumption of cross-polarized antennas at both the eNB and UE following related, real-world deployment requirements. The resulting codebook can be represented by a product of two matrices, W1 and W2, where W1 quantizes the wideband and long-term channel spatial characteristics and W2 quantizes the frequency-selective and short-term channel spatial characteristics. Therefore, the UE feedback information, in addition to the CQI (Channel Quality Indicator) and the RI (Rand Indicator), includes the W1 and W2 precoding indices in LTE-Advanced. Another critical design component in LTE is the design of efficient reference signals (RS). RS in LTE are designed to minimize both control overhead and performance degradation due to CSI feedback accuracy and channel estimation errors at the UE. LTE employs CRS for each antenna port, which can be used at each UE. To reduce overhead, LTE-Advanced employs UE-specific reference signals, termed DMRS (DeModulation Reference Signal), which are used for channel estimation and demodulation at the UE. Unlike CRS, DMRS is precoded with the same precoding matrix used for data channel transmission, which implies that DMRS cannot be used for calculating CQI. For that purpose, the cell-specific CSI-RS is employed at the UE for CQI, RI, and PMI calculation. Although CSI-RS has similar functionality to CRS, CSI-RS is transmitted at a lower rate than CRS, thus enabling low-overhead closed-loop MIMO operation.

"LTE-Advanced adopted closed-loop precoding to realize high-order MIMO spatial multiplexing…"

"To reduce overhead, LTE-Advanced employs UE-specific reference signals…"

In the last part of this section, we provide representative MIMO and CoMP system-level performance results in order to demonstrate the constantly improving spectral efficiency achieved by multi-antenna technologies in evolving LTE releases. Table 1 summarizes some of our most representative results comparing the performance of LTE Release 8 and Release 10 for both the downlink and uplink. For the results in Table 1, the Urban Microcell (UMi) Test Environment of the IMT-Advanced evaluation methodology is

	LTE Release 8		LTE Release 10	
	Cell average (b/s/Hz/cell)	Cell-edge user (b/s/Hz/UE)	Cell average (b/s/Hz/cell)	Cell-edge user (b/s/Hz/UE)
Downlink, 2x2	1.8	0.050	2.2	0.075
Downlink, 4x2	2.1	0.060	2.6	0.090
Uplink, 1x2	1.0	0.025	1.4	0.040
Uplink, 1x4	1.5	0.040	2.2	0.065

Table 1: Summary of LTE Release 8 and Release 10 MIMO spectral efficiencies for the UMi Test Environment of the IMT-Advanced Evaluation Methodology
(Source: Intel Corporation, 2013)

used, that is, a homogeneous deployment scenario is evaluated. The 2x2 and 4x2 antenna configurations are employed in the downlink, while the 1x2 and 1x4 antenna configurations are assumed in the uplink. The results in Table 1 reveal not only the performance gains offered by using more transmit/receive antennas at the eNB, but also the gains offered by LTE Release 10 compared to LTE Release 8.

Table 2 summarizes some of our most representative results comparing the performance of LTE Release 10, based on MU-MIMO transmission (referred to as non-CoMP in Table 2), and Release 11, based on JT CoMP for the downlink and JR CoMP for the uplink. For the results in Table 2, the CoMP Evaluation Methodology developed in 3GPP RAN1 for heterogeneous (HetNet) deployments is used according to LTE Release 11 technical report TR 36.819. Results are presented for different sizes of the coordination area, which ranges from 1 macro cell (5 nodes) to 21 macro cells (105 nodes). The results in Table 2 reveal the significant performance gains offered by JT and JR CoMP compared to Release 10 non-CoMP, with the CoMP gains becoming even more pronounced as the size of the coordination area increases.

	Downlink 2x2		Uplink 1x2	
	Cell average (b/s/Hz/cell)	Cell-edge user (b/s/Hz/UE)	Cell average (b/s/Hz/cell)	Cell-edge user (b/s/Hz/UE)
Non-CoMP	9.5	0.063	6.3	0.078
CoMP,1 cell	9.9	0.069	8.0	0.115
CoMP,3 cells	10.0	0.078	8.2	0.122
CoMP,9 cells	9.9	0.088	8.3	0.127
CoMP,21 cells	9.7	0.100	8.7	0.164

Table 2: Summary of CoMP (Release 11) vs. non-CoMP (Release 10) spectral efficiencies for the heterogeneous deployment scenario of the CoMP evaluation methodology (TR 36.819)
(Source: Intel Corporation, 2013)

Other 4G Features

In addition to OFDM/MIMO and cooperative networking related features such as CoMP, LTE and LTE-Advanced support a set of new features and capabilities that greatly affect the protocol stack design. In general, the design impact of these features can be grouped into the following categories:

- Expanding system bandwidth through carrier aggregation, where the MAC layer is enhanced to accommodate multiple PHY channels at the same or different frequency bands.
- Improving efficiency for Machine Type Communications (MTC), where changes were introduced at different layers to accommodate the expected large number of machine type devices, to optimize protocol efficiency for short and/or bursty traffic, and to avoid and manage congestions.
- Increasing battery life through enhanced Diverse Data Applications (eDDA), where layer 2 is optimized to conserve energy of the UE.
- Cross-layer optimization for selected key applications, such as the support for dynamic adaptive streaming over HTTP (DASH).[1]

Beyond 4G and 5G Protocol Stack Design Challenges and Innovations

Beyond 4G and 5G are expected to further improve the existing features from previous generations of wireless technologies while introducing new and disruptive capabilities to bring the wireless industry to the next level. In particular, 5G networks are expected to scale up not only in data rate, but also in the number of devices, the level of coordination/cooperation among base stations and devices, and the types of applications. For base stations, the peak to average data rate ratio will be further increased compared to 4G, which makes the sharing of a large signal processing resource pool through Cloud-RAN architecture more beneficial and essential. For user devices, the compute requirements for new applications and services will continue to rise. In certain cases, in particular for wearable devices, applications requirements can significantly outpace the capabilities of the wearable platform. 5G network architecture and protocol stack designs need to tightly integrate application signal processing and wireless communications, and enable a new generation of compute + communication platforms in 5G.[2]

"…5G networks are expected to scale up not only in data rate, but also in the number of devices, the level of coordination/cooperation among base stations and devices, and the types of applications…"

From Spectral Efficiency to Network Efficiency

In 2G to 4G, the key drivers for capacity increase come from air-interface improvement and new spectrum acquisition. We are at a technology turning point as we enter 5G.[3] Over the past 10 years, around 25x more spectrum has been licensed for cellular systems leading to roughly 1 GHz licensed spectrum for use in each region. We have technologies like OFDM, MIMO, and CoMP transmission/reception being developed that significantly improves spectrum efficiency over the air. Looking into the next ten years, considering that the air interface spectrum efficiency has been approaching its capacity limit and the difficulty in getting new spectrum, we expect that

"HetNets with a joint deployment of macro cells, pico cells, underlay ad-hoc networks, and a joint operation of multiple radio access technologies such as GSM, HSPA, LTE, WLAN, and Bluetooth will be the key enabling technique for network efficiency improvement."

more capacity gains would come from the network side. Network architecture improvement, information and communication technology convergence, and the corresponding updates in devices in supporting the network architecture and technology convergence would be the key drivers for the capacity increase.

System-level network efficiency measured in bits/s/Hz/m^2 will be one of the KPIs for the network architecture improvement. To meet the traffic demand in the next 10 years, we will need 40x to 50x increase in network efficiency, and our analysis has shown that 50x to 100x in the next decade is possible.[4] HetNets with a joint deployment of macro cells, pico cells, underlay ad-hoc networks, and a joint operation of multiple radio access technologies such as GSM, HSPA, LTE, WLAN, and Bluetooth* will be the key enabling technique for network efficiency improvement. Figure 4 shows a likely breakdown of contribution factors for the next thousand-fold increase in capacity.

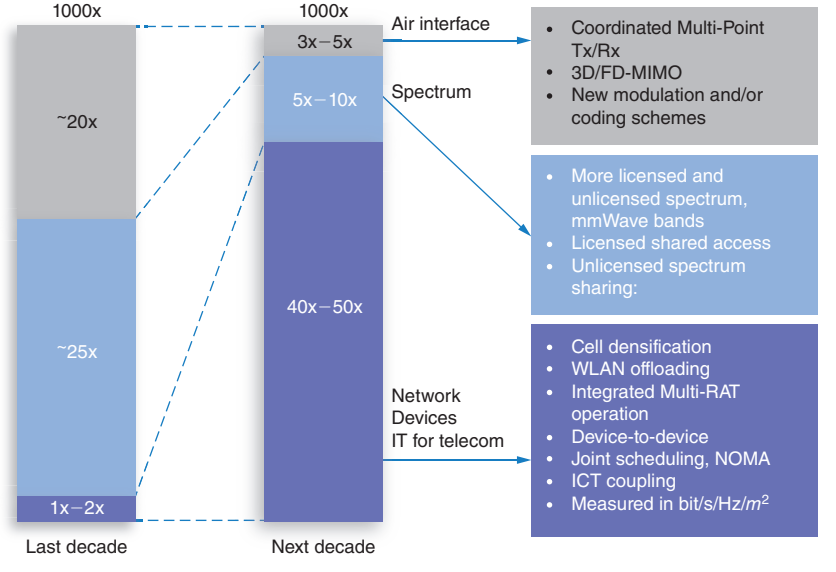


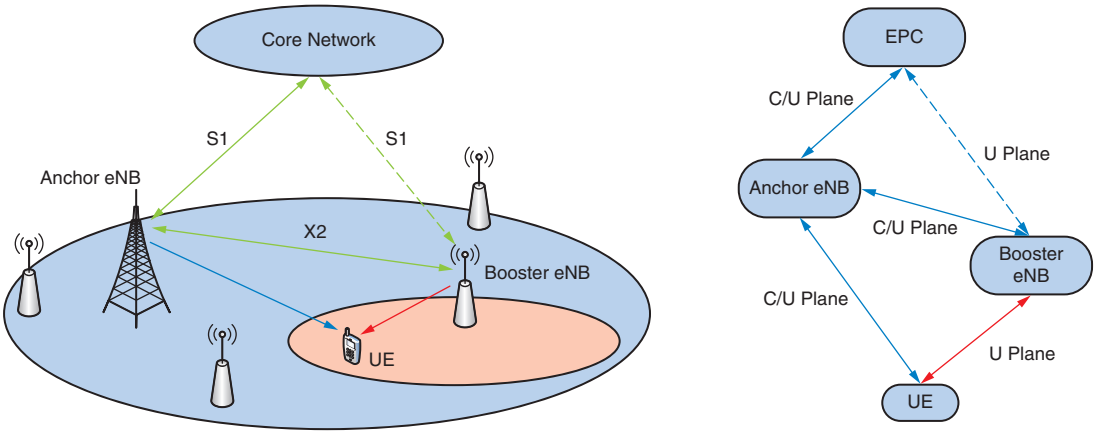
Figure 4: Network efficiency improvements for the next decade (Note: capacity increase baseline is 3GPP Release 10 LTE systems) (Source: Intel Corporation, 2013)

Network protocol designs in HetNets need to accommodate network nodes with different transmission and processing probabilities and ensure efficient joint radio operation among them. We see that anchor-booster based architecture with Control and User plane (C/U Plane) separation for managing communication between radio access network and core network, simple and flexible protocol for underlay D2D proximity communication, and joint network-centric and UE-centric control would be the key for exploiting heterogeneous network efficiency.

Anchor-Booster Architecture

Our study shows that, in theory, network efficiency increases linearly as cell deployment becomes denser.[5] However, cell densification imposes challenges in mobility management and backhaul support, and would cause excessive signaling overhead. In addition, in a network consisting of only small cells, network coverage may be costly to maintain. This is particularly true for systems operating at high frequencies such as the millimeter wave bands. On the other hand, nodes with low transmit power enable dense deployment with tremendous spatial frequency reuse.

Anchor-booster architecture is therefore proposed to use the high transmit power nodes as anchor for coverage and signaling and the low transmit power nodes as booster for data rate boosting.[6] Figure 5 illustrates an application of the anchor-booster based network architecture for LTE and demonstrates one of the options for C/U-plane separation. We can see that the anchor eNB handles all the C-plane signaling for the UE while the booster eNB only handles U-plane data. UE keeps a single C-plane connection to the anchor eNB, but can connect to both anchor and booster eNBs for U-plane data. With C/U plane separation, the anchor eNB and the booster eNB appear to the core network as one node. UE motion across booster cells would not trigger a handover procedure at the core network. The signaling overhead due to frequent handovers among booster eNBs and between booster eNB and the anchor eNB can therefore be avoided.



(a) Anchor booster based architecture in LTE. The dashed line denotes an optional connection

(b) Example of C/U plane separation

Figure 5: Anchor-booster based architecture with C/U plane separation (Source: Intel Corporation, 2013)

Besides its application in LTE, anchor-booster architecture can be applied as a general architecture for heterogeneous networks. To enable multi-RAT joint radio operation, an anchor node can be set up as a multi-RAT controller for multi-RAT selection, radio resource coordination, multi-connection management, and mobility management.

The anchor-booster architecture and the protocol stack design for C/U plane separation also make advanced energy-saving features possible at the network. With an anchor cell providing network coverage and uninterrupted control plane operation, one or many booster cells may completely shut down their power to conserve energy when there is no UE in their respective serving areas.

Network Overlay with Super Cloud-RAN

Existing C-RAN implementation can benefit from the above-mentioned anchor-booster architecture and the protocol stack design to evolve into an advanced Super Cloud-RAN architecture in the future. In today's C-RAN, a large number of remote radio heads (RRH) are connected to a centralized baseband processing unit (BBU), each through a multi-Giga-bps high data rate fiber-optical link running the CPRI protocol, often referred to as the "front-haul" in the industry. It is primarily a macro cell design operating at a single carrier frequency, and may offer content caching and distribution at the centralized site through platform virtualization.[7]

"The protocol stack designed for C/U plane separation in the anchor-booster architecture can be applied to further expand the C-RAN concept into the next generation Super Cloud-RAN architecture."

The protocol stack designed for C/U plane separation in the anchor-booster architecture can be applied to further expand the C-RAN concept into the next generation Super Cloud-RAN architecture. Table 3 provides a high level example of the expected key features in a future Super Cloud-RAN network

	C-RAN	**SUPER CLOUD-RAN**
Radio Access Technology (RAT)	Single	Multi-RAT
Carrier Frequency	Single	Multi-frequencies, Carrier aggregation
RRH/BBU functional partition	Fixed (CPRI links)	Flexible
Anchor cell	BBU/caching/ applications	+ Multi-RAT connection control
Booster cells	RRH	+ small cells/ applications
Multi-network integration	-	Possible centralized connection manager
Support for underlay networks	-	+ Micro-server at booster for services
Content distribution/ Caching	One level	One or two levels
Network energy saving	Eliminate cell site air conditioning	Possible shut down booster sites

Table 3: Example of Super Cloud-RAN key features (Source: Intel Corporation, 2013)

The future Super Cloud-RAN not only offers an efficient platform for cooperative networking within a single layer of a homogeneous network, as C-RAN does today, but it also provides tight integration across multiple layers of networks of potentially different air interface technologies. Through the technique of carrier aggregation, this advanced heterogeneous network design also enables an efficient use of high-frequency bands that are traditionally not suitable for cellular services. Other benefits include network energy saving through active small cell power state management, network application/media processing through platform virtualization, and proximity services by local micro-servers, in particular in the indoor business environments. Like C-RAN, Super Cloud-RAN will continue to be an ideal platform for base station implementation of the massive transceiver that not only manages radio resources across multiple sectors and cell sites, but also aggressively avoids and/or cancels interference among many UEs that are actively communicating in the local network cluster.

Underlay Networks for Wearable Devices

Previous generations of wireless technology focused mainly on addressing the radio link bottlenecks. In 5G the focus is expected to shift towards seamless integration of mobile computing and communication technologies. This shift is driven by several factors. In addition to delivering the required network efficiency and energy efficiency, the anticipated proliferation of machine-type devices[8] and consumer-wearable devices also makes 5G mobile computing and communication integration a necessity.

Take wearable devices as an example. They usually have limited on-board processing power and battery size. If they need extended compute capability and content information, they have to heavily rely on their surrounding local networks and computing platforms to support their applications. The computing requirements of certain wearable devices (such as Google Glass*) may be more demanding than a typical portable device. In fact, the market potential of many wearable devices may be directly tied to their compute capability. Using high performance and very low latency communication links to shift mobile computing workload becomes essential.

This paradigm shift in the next decade from communication to computing and communication calls for the development of the underlay networking technologies. The underlay network is a collection of a large number of small network clusters, where devices are interconnected within the cluster, and the cluster itself is connected to the overlay networks for Internet connectivity and proximity services. An underlay network cluster can be a group of users with devices mostly interconnected through direct device-to-device communications, or one individual user with many wearable devices forming a personal network "cloud". The fundamental differences between an overlay heterogeneous network and an underlay network cluster are that the underlay network cluster usually moves, and that the formation and the dissolvent of a cluster are dynamic and mostly ad-hoc. Table 4 highlights the main characteristics of the underlay networks.

"The fundamental differences between an overlay heterogeneous network and an underlay network cluster are that the underlay network cluster usually moves, and that the formation and the dissolvent of a cluster are dynamic and mostly ad-hoc."

	Infrastructure networks	Underlay networks
Network's own mobility	Fixed	Mobile
UE's role	Terminal	Networking node
Radio link range	Longer range	Mostly proximity links
Resource management	Mostly centralized	Often distributed
Connectivity	Single-hop	Single or more hops
Network association	Traditional	Dynamic/ad-hoc
Radio frequency	Traditional	Co-channel or different
PHY	Cellular/Wi-Fi*	Wi-Fi/short range RAT/ NOMA
MAC	Scheduling or Contention based	Add opportunistic based MAC

Table 4: Main characteristics of underlay networks
(Source: Intel Corporation, 2013)

The protocol stack design for 5G underlay networks poses several new requirements and challenges. Some of the early work has already started in the wireless industry, mainly in the form of D2D (Device-To-Device) communications. The complexity and the performance of future 5G underlay networks will to a large degree depend on the level of coupling with infrastructure networks, in particular if there will be any direct sharing of the radio spectrum between networks. It is expected that very high speed and low latency 5G radio links will be instrumental in connecting the underlay network clusters to the infrastructure network and expanding the mobile computing capabilities.

"It is expected that very high speed and low latency 5G radio links will be instrumental in connecting the underlay network clusters to the infrastructure network and expanding the mobile computing capabilities."

It should be expected that the design within an underlay network cluster itself will also be quite different. Designers will face more stringent requirements in energy saving, unstructured and time-varying network topology, unreliable communication, and scalability demand. Non-orthogonal multiple access (NOMA), efficient neighboring discovery, and opportunistic spectrum access will become essential.

Design Example: Load-Balancing Mobile Association with In-Band Co-Channel Relays

System performance is always an important consideration for a good protocol stack and network architecture design. This is particularly true for the increasingly complex 5G networks where both the over- and underlay networks coexist and share the same radio frequency band. Since the protocol stack design directly affects the network efficiency and the user experience, we increasingly rely on analytical studies to determine the performance bounds to understand general design tradeoffs, and architect the network and design the protocol stack accordingly.

"As an example, we consider the network-centric and device-centric approaches in implementing optimal mobile association with in-band relay nodes (RN)."

As an example, we consider the network-centric and device-centric approaches in implementing optimal mobile association with in-band relay nodes (RN). As the RN transmit power is chosen to be 30 dBm, which is 13 dB to 16 dB

lower than that of the macro eNB, conventional best-power mobile association will associate most of the UEs with the macro eNB, leaving RNs largely underloaded. Li et al.[9] proposed an optimal mobile association framework to balance the traffic load between macro eNB and RN and maximize the overall network efficiency.

In the network-centric approach, we can use a centralized method to solve the optimization problem. Once the mobile association is decided, the macro eNBs and the RNs can be informed and the UEs are admitted. A problem with the centralized approach is that it can only work offline and is not flexible to fit the dynamic change of the network environment.

Using the device-centric approach, since the UE needs to make an independent decision without knowing the overall resource usage, a centralized solution will not work. Instead, we look into a distributed solution using a pricing-based method.[10] The idea is to let each of the macro eNBs and RNs announce a price-per-unit radio resource to the UEs. The UE calculates its cost to the macro eNBs and RNs and chooses the one with the lowest cost to attach. The UE can also choose to connect with both a macro eNB and a RN if the cost is the lowest; this happens when the UE benefits from cooperative diversity. Figure 6 illustrates the proposed pricing-based scheme.

With the price values, the incoming UE calculates the cost of being served by the eNB, the RN, or being cooperatively serviced by both the eNB and RN. The UE can then select the option with the lowest price and trigger the association procedure.

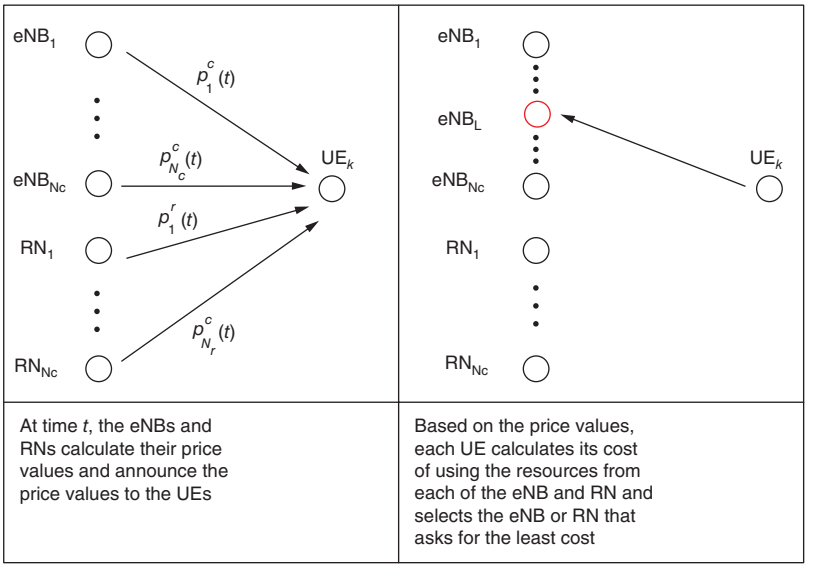


Figure 6: The pricing-based scheme (Source: Intel Corporation, 2013)

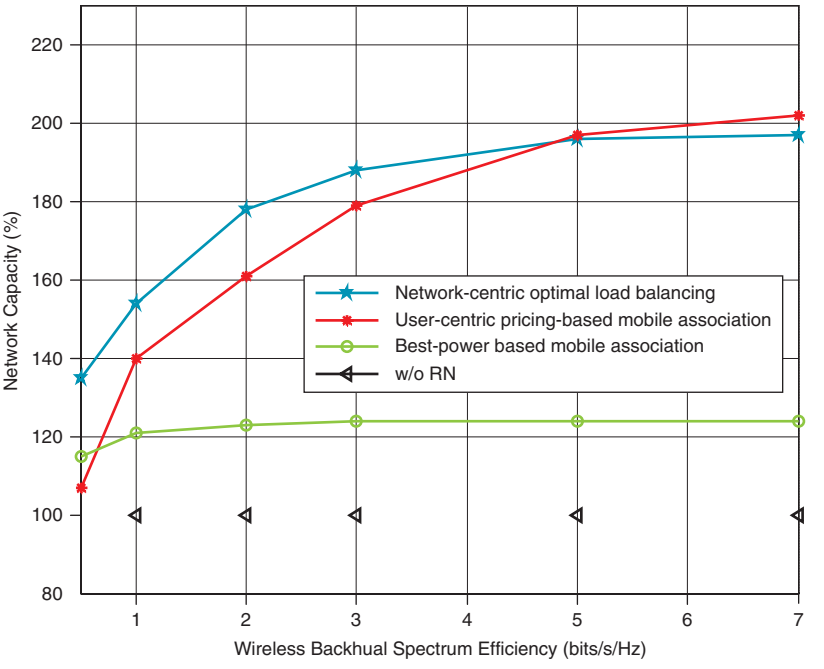


Figure 7: Network spectrum efficiency comparison (Source: Intel Corporation, 2013)

Figure 7 shows the network efficiency by the network-centric and the device-centric load-balancing mobile association schemes under different backhaul spectrum efficiency values. The network efficiency values are expressed as the relative percentage of the network efficiency value of the network without RN. As a comparison, the network efficiency of the best-power based mobile association scheme is also plotted. We can see the gains by both the network-centric approach and the device-centric approach. As the backhaul condition is not included in the pricing function, in cases with low backhaul spectrum efficiency, the UE-centric approach offers a smaller gain than that of the network-centric approach. As the backhaul spectrum efficiency improves, the gain by the UE-centric approach increases and can outperform the network-centric approach because the UE-centric approach reflects the network environment in a timely manner and also takes advantage of the cooperative transmission between the eNB and the RN.

"A hybrid scheme with a joint effort of both the network-centric approach and the UE-centric approach would lead to better performance."

The network-centric approach can be globally optimized but is not flexible to network changes, and the device-centric approach can be implemented online and reflects the instantaneous network change but is only optimized locally. A hybrid scheme with a joint effort of both the network-centric approach and the UE-centric approach would lead to better performance. For example, under the pricing-based mobile-association scheme, instead of using a fixed based price for all the macro eNBs and RNs, the network-centric approach can be used to periodically update and differentiate the base price value for different macro eNB and RNs.

Future Research Directions

Protocol stack design will continue to evolve with wireless network architecture to deliver higher performance at reduced cost and to enable an unprecedented end user experience. Although most of the technology development is expected to be enhancements of existing designs, there are several key areas in which new technologies need to be developed, including underlay networking technologies, mobile computing and communication integration, opportunistic multiple access schedulers, non-orthogonal multiple access schemes, network-centric connection management across multiple radio networks, UE-centric distributed radio resource and connection management, PHY/MAC design for high frequency and mmWave bands, backhaul solutions for ultra-dense networks, co-channel overlay and underlay network coexistence, and interference avoidance and management. Future protocol stack designs also need to consider tighter coupling of different air interface technologies, over both licensed and unlicensed bands, including possible bandwidth aggregation at air interface layers. They also need to be highly scalable in terms of data rate, number of users, network coordination level, and cross-layer optimization for applications.

References

[1] Oyman, O. and S. Singh, "Quality of Experience for HTTP Adaptive Streaming Services," IEEE Communications Magazine, April 2012

[2] Wu, G. and Q. Li, "What is 5G? Devices and Networks for Future Embedded Mobile Internet," Keynote at WWRF-31, Vancouver, Oct. 2013

[3] Wu, G., Q. Li, R.Q. Hu, and Y. Qian, "Overview of Heterogeneous Networks," Chapter 1, *Heterogeneous Cellular Networks* (Hoboken, NJ: Wiley, 2013), ISBN 9781119999126.

[4] Li, Q., "Network Efficiency Improvements for Future Wireless Networks," WWRF-31, Vancouver, Oct. 2013

[5] Li, Q., G. Wu, and R. Q. Hu, "Analytical study on network spectrum efficiency of ultra-dense network," in Proc. IEEE PIMRC2013, pp. 1–5, London, Sep. 2013

[6] Heo, Y., Y. Zhang, S. Sirotkin, C. Cordeiro, and Q. Li, "Anchor-booster Network/Protocol Architecture," Intel internal white paper, Nov. 2013

[7] China Mobile, "C-RAN, The Road Towards Green RAN", Version 2.5, Oct. 2011

[8] Wu, G., S. Talwar, K. Johnsson, N. Himayat, and K. D. Johnson, "M2M: From Mobile to Embedded Internet," IEEE *Communications Magazine*, April 2011

[9] Li, Q., R. Q. Hu, G. Wu, and Y. Qian, "On the optimal mobile association in heterogeneous wireless relay networks," in Proc. IEEE INFOCOM 2012, pp. 1–9, Orlando, Mar. 2012

[10] Li, Q., Y. Xu, R. Q. Hu, and G. Wu, "Pricing-based mobile association for cooperative wireless heterogeneous networks," in Proc. IEEE ICC 2012, Ottawa, Jun. 2012.

Author Biographies

Dr. Geng Wu (geng.wu@intel.com) is the Chief Scientist of MCG Standards and Advanced Technology at Intel, and the head of Intel 3GPP delegation. Dr. Wu has over 20 years of research and development experience in the wireless industry, contributed extensively to 2G, 3G, and 4G air interface technologies and network architecture development. Prior to Intel, he was the director of Wireless Architecture and Standards at Nortel Networks.

Dr. Apostolos (Tolis) Papathanassiou is Senior Principal Engineer and Manager of the Core PHY and Simulations (CPS) group in MCG Standards and Advanced Technology, with his main focus being on LTE PHY standardization. Previously at Intel, he led multiple standardization efforts in ITU-R and IEEE/WiMAX Forum. Before joining Intel, Tolis worked on multiple-antenna PHY technologies for 3G, satellite, and Wi-Fi wireless systems.

Dr. Qian (Clara) Li is a senior research scientist in MCG Standards and Advanced Technology at Intel. She received her BE and MS degrees in information engineering from Nanjing University of Posts and Telecommunications, Nanjing and PhD in communication engineering from Nanyang Technology University, Singapore. Her research interests include information theory, communication theory, cross-layer optimization and RAN evolution. She has served as reviewer of several international journals and as symposium co-chair of numerous IEEE conferences.

Dr. Mo-Han Fong (mo-han.fong@intel.com) is a Principal Engineer and Director of Wireless Technology and Standards in MCG Standards and Advanced Technology at Intel, and the head of Intel 3GPP RAN delegation. Prior to Intel, she was with RIM and Nortel Networks, leading research and technology development on various 3G and 4G wireless systems including LTE/LTE-Advanced, WiMAX/IEEE802.16m, UMTS/HSPA, CDMA/1xEV-DO/UMB. She has over 250 patents granted or pending.

ALGORITHM DESIGN CHALLENGES IN ADVANCED INTERFERENCE MITIGATION AND DOWNLINK MIMO

Contributors

Cecilia Carbonelli
Intel Mobile Communications

Axel Clausen
Intel Mobile Communications

Franz Eder
Intel Mobile Communications

Stefan Fechtel
Intel Mobile Communications

Yeong-Sun Hwang
Intel Mobile Communications

As a result of the growing demand for higher capacity and throughput, new trends have emerged in the wireless cellular communications industry and contributed to the definition of what we call LTE-A (or 4G LTE Release 10/11). Among these, the deployment of homogeneous and heterogeneous networks and the introduction of advanced downlink Multiple Input Multiple Output (MIMO) techniques have significantly impacted the design of baseband algorithms. In this work we discuss the limitations of existing state-of-the art solutions and describe the complexity-performance tradeoffs that need to be considered by LTE-A receivers to support the above-mentioned features in the upcoming LTE-A releases.

Introduction

In the early days of LTE most of the User Equipment (UE) manufacturers' work was focusing on implementing low complexity receivers that could offer competitive performance for Second and Third Generation (2G and 3G) devices within a more efficient packet-switched network architecture. The key technology enablers were at that time the introduction of new MIMO schemes (such as transmit diversity, closed and open loop spatial multiplexing) and a more flexible usage of the available spectrum through orthogonal-frequency division multiple access (OFDMA).[1][2][3]

In 2009, to address the need for increased capacity and higher data throughput, the 3GPP submitted an official proposal for an evolution of LTE to the International Telecommunication Union (ITU) with the aim of meeting the requirements set for International Mobile Telecommunications-Advanced (IMT-A).[4] These requirements included support for a downlink data rate of 100 Mbps and 1 Gbps, respectively, in high and low mobility scenarios, a target peak spectral efficiency of 15 bps/Hz in the downlink, and a total transmission bandwidth of at least 40 MHz and up to 100 MHz.

The new key technology enablers and features introduced by LTE-A (or 4G LTE Release 10 and higher) to fulfill these targets can be grouped as follows:

- Technologies targeting a flexible usage of spectrum:
  - Carrier aggregation
- Technologies targeting increased data rates:
  - UL single-user multiple antenna transmission (SU-MIMO) (up to 4x4)
  - DL multiple-antenna transmission (up to 8x8):

- Technologies targeting an increased capacity and cell coverage:
  - Multi-User MIMO
  - Enhanced beamforming schemes
  - CL-MIMO with an enhanced codebook and an improved CSI feedback mechanism
  - Support of homogeneous and heterogeneous network deployments and enhanced—and further enhanced—inter-cell interference coordination (eICIC/FeICIC) techniques
  - Relaying techniques

In this work we mostly focus on two of the features mentioned above and on their impact on the physical layer algorithm design. Specifically we consider:

- The support of heterogeneous and homogeneous networks and related eICIC (LTE Release 10) and FeICIC (LTE Release 11) techniques
- The introduction of advanced Downlink (DL) MIMO schemes featuring an improved codebook design, up to eight spatial layers, and enhanced beamforming methods

The motivation for this lies in the fact that both these features are being quickly deployed by operators and network vendors, and both have a major impact on the baseband physical layer algorithms.

In order to adequately support the above-mentioned technologies and at the same time ensure backward compatibility with the previous LTE standard releases, most of the existing baseband algorithms need to be analyzed in a more complex framework, possibly enhanced and finally optimized for the newly defined test scenarios.

"In order to adequately support the above-mentioned technologies and at the same time ensure backward compatibility with the previous LTE standard releases, most of the existing baseband algorithms need to be analyzed in a more complex framework, possibly enhanced and finally optimized for the newly defined test scenarios."

For example, in both homogenous and heterogeneous networks, channel estimation, interference mitigation prior to decoding, and CSI feedback generation algorithms face several challenges as the mobile device needs to operate in dynamic scenarios with multiple interfering cells and several unknown parameters (number and type of interferers, scheduling and timing information, varying power levels and channel conditions, and so on).

Similarly, although several antenna array processing techniques are already available in the literature and on the market offering improved coverage and higher data rates, the new LTE-A MIMO schemes and particularly the support of transmission mode 9 (TM9) demands a more accurate channel estimation, an improved handling of spatial layers at the detector, and a more efficient CSI feedback generation mechanism.

In this work we discuss how advanced interference mitigation and DL MIMO techniques are supported in LTE-A receivers, their impact on the complexity of channel estimation, interference mitigation, detection and Channel State Information (CSI) feedback generation, and which limitations practical state-of-the art solutions have with respect to

the optimal implementation. Specifically, we describe the complexity-performance tradeoffs that need to be considered by LTE-A receivers to support the upcoming releases.

This article is organized as follows. In the section "Advanced Interference Mitigation in LTE-A," we discuss eICIC and FeICIC techniques and the algorithm enhancements introduced to support them. The section "Advanced DL MIMO in LTE-A" describes advanced DL MIMO techniques and the algorithm enabling them. Conclusions and future work are outlined in "Final Considerations."

Advanced Interference Mitigation in LTE-A

In this section we provide an overview of the interference mitigation feature in the context of LTE-A and discuss the impact of this feature on the design of efficient baseband algorithms.

Feature Overview

Modern wireless networks face the challenge that the demand for data traffic is increasing dramatically. Network operators need to upgrade their networks to satisfy the demand. Either new spectrum is added and carrier aggregation is applied or the transmission efficiency (bps/Hz) of the spectrum is increased. An increase in transmission efficiency can be achieved in different ways:

- Higher number of spatial streams (see "Advanced DL MIMO in LTE-A")
- Improved receiver algorithms
- Increase of cell density (number of cells per area)

"An increase of cell density—operating with a frequency reuse factor of one—comes at the cost of higher interference."

An increase of cell density—operating with a frequency reuse factor of one—comes at the cost of higher interference. Figure 1 shows the topology of a macro-only homogeneous network on the left side and a heterogeneous network scenario on the right side.[5][6]

The homogeneous network consists of macro cells only, and LTE networks typically use a frequency reuse factor of one, that is, all macro base stations transmit on the same frequency band. The UE faces the strongest interference from neighboring macro cells if it operates at the cell edge. Careful network planning is performed for macro cells to achieve good coverage at every location and to minimize the interference impact at the cell edge.

In the heterogeneous network, small cellular base stations, called pico cells, are placed at hotspots to offload the macro cells as shown on the right part of Figure 1. In this scenario, the UE faces the strongest interference if it operates at the cell edge of the pico cell: it suffers from macro cell interference as well as potential interference from other pico cells. We note that network planning for pico cells might not be as sophisticated as for macro cells.

"We note that network planning for pico cells might not be as sophisticated as for macro cells."

For an operator, there needs to be a business case to introduce a heterogeneous network and to deploy small cells. The commercial benefit increases if the

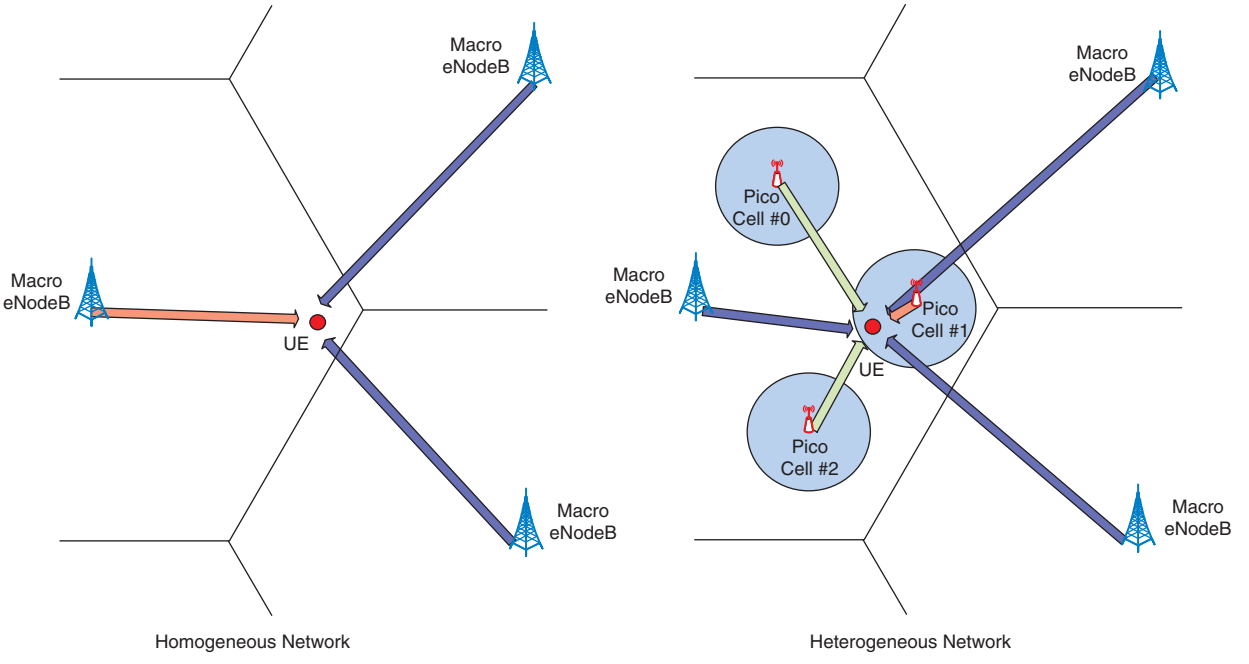


Figure 1: Homogeneous (left) and heterogeneous network topology (Source: Intel Corporation 2013)

coverage of the small cell is increased and more traffic (per small cell) can be offloaded from the macro cell. Therefore, cell range expansion (CRE) or "serving cell association with large bias" has been introduced to increase the chance of UEs to associate with pico cells. Range expansion is typically characterized by associations of a UE with a pico cell although the received signal from the macro cell is up to a cell range expansion value stronger than the received signal of the pico cell. In 3GPP LTE Release 10, the maximum cell range expansion is set to 6 dB, whereas in LTE Release 11 this value is increased to 9 dB.

Figure 2 depicts a macro-pico cell scenario with two pico cells: one pico cell is located at a distance of 100 m from the center of the macro cell; the second pico cell is located at a distance of 250 m from the center of the macro cell. The left graphs depict the coverage area of the pico cells and the right plot shows the signal strength of the macro cell (blue curve), and the two pico cells. In this scenario, cell range expansion of 6 dB increases the diameter of the first pico cell from 46 m to 71 m whereas the diameter of the second pico cell is increased from 71 m to 178 m.

The term heterogeneous network is used as well to describe a second deployment scenario based on femto cells, that is, small, low-power cellular base stations with a range in the order of 10 meters. A femto cell and its service are usually restricted to a closed subscriber group (CSG). A UE that is not part of the CSG faces strong interference from the femto cell and some form of interference coordination is required to ensure reliable operation.

"A UE that is not part of the CSG faces strong interference from the femto cell and some form of interference coordination is required to ensure reliable operation."

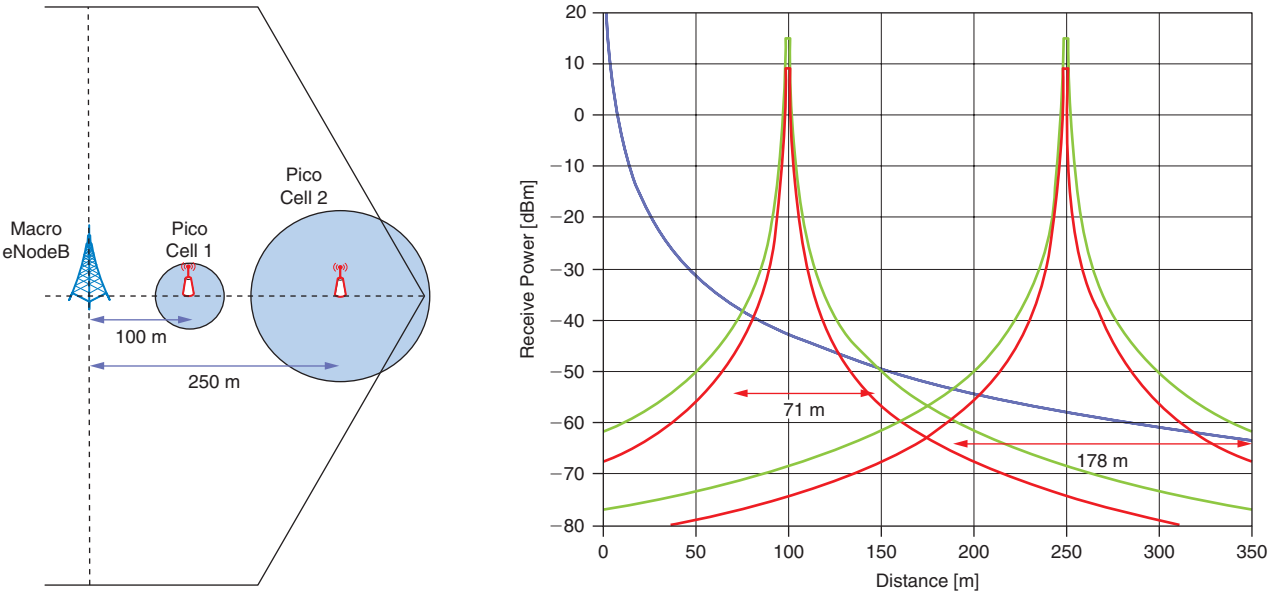


Figure 2: Macro-pico scenario with 2 pico cells (Source: Intel Corporation 2013)

The scenarios described above require operation of the mobile devices (UEs) under severe interference conditions. 3GPP Release 10 and Release 11 have introduced different methods for interference coordination. Specifically:

- Time-domain separation (eICIC/FeICIC)
- Frequency-domain separation (carrier aggregation)
- Frequency- and time-domain separation (carrier aggregation with cross-carrier scheduling / enhanced Physical Downlink Control Channel (ePDCCH) / Coordinated Multi Point (CoMP))

Figure 3 provides a schematic view of different coordination methods; all coordination methods rely on synchronization of the signal transmission of all evolved NodeBs (eNodeBs).

"The eICIC/FeICIC method has the disadvantage that a complete subframe needs to be muted for the macro cell—roughly one in every eight subframes…"

The eICIC/FeICIC method has the disadvantage that a complete subframe needs to be muted for the macro cell—roughly one in every eight subframes needs to be muted to support a reliable Hybrid Automatic Repeat Request (HARQ) process of the victim cells. The operator will lose about 15 percent of macro cell capacity. This only makes sense if at least four pico cells are deployed in a macro cell antenna sector. The eICIC/FeICIC method does not allow the operator to have a smooth transition from a homogeneous network to a heterogeneous network. It is more likely that operators will start deployment-based coordination methods as shown in Figure 3b and Figure 3c.

In Figure 3b, the operator initially deploys carrier aggregation on a homogeneous network (macro cell only) and enables cross-carrier scheduling for the secondary carrier. Now, pico cells can be installed on

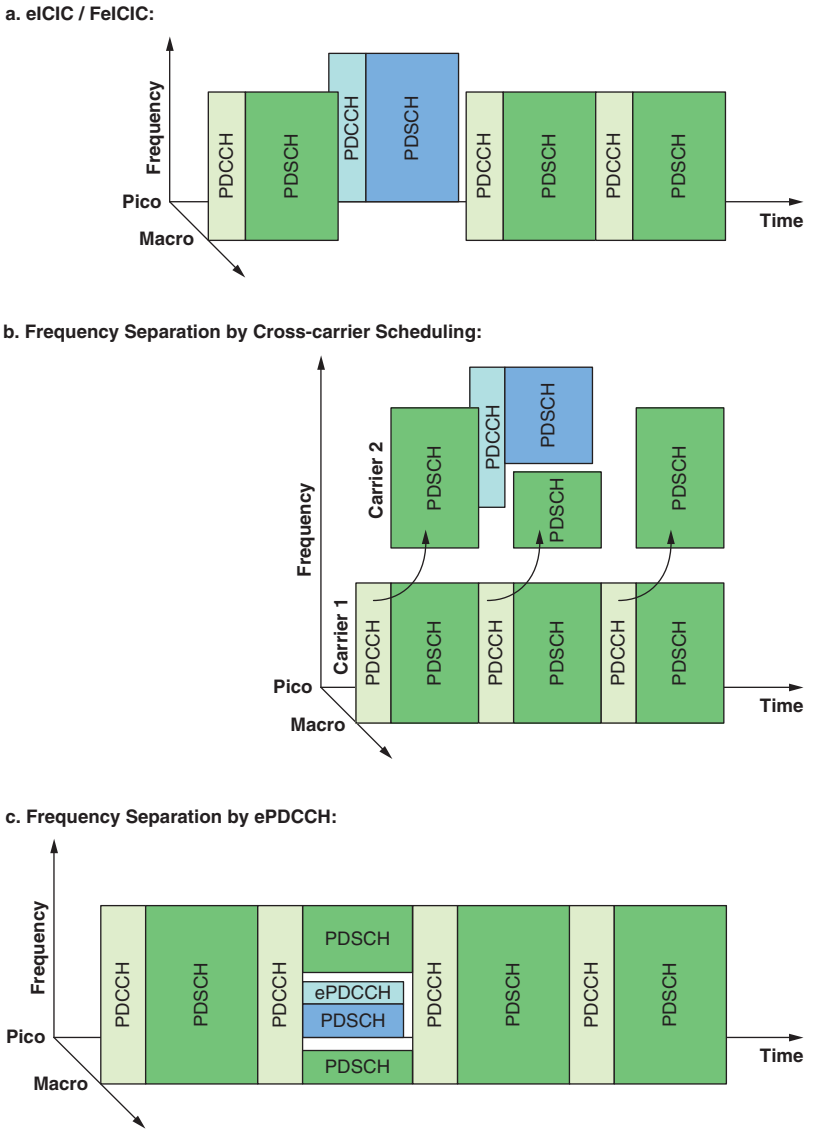


Figure 3: Interference coordination methods (Source: Intel Corporation 2013)

the secondary carrier and the bandwidth of the secondary carrier can be divided dynamically between macro and pico cells. Capacity of the macro cell is only removed as much as it is required by the installed pico cells. Figure 3c describes a similar deployment scenario for operators who only have access to single carrier spectrum. Here, the separation can be done based on the ePDCCH. Again, capacity can be split seamlessly between macro and pico cells.

In Figure 4, the set of resource elements (REs) belonging to a physical resource times subframe (with a size of one resource block in frequency direction and one subframe in time direction = RBSF) is shown for a normal cyclic prefix.

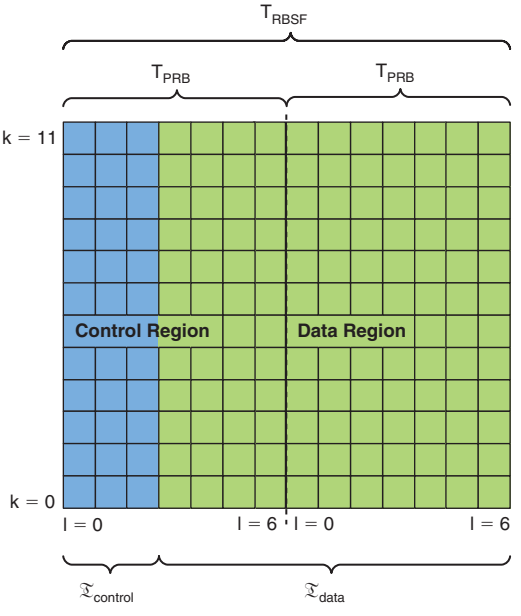


Figure 4: Physical resource block (Source: Intel Corporation 2013)

Each RBSF spans a bandwidth of 180 kHz and has a time duration of 1 ms, that is, two consecutive LTE physical resource blocks. For normal cyclic prefix, it consists of 12 × 14 resource elements; for extended cyclic prefix it consists of 12 × 10 resource elements.

Depending on the cell ID of the aggressor cell and the target cell, the common reference signals (CRSs) are either colliding or non-colliding. Figure 5 shows block diagrams of these two scenarios for one interfering cell.

In these scenarios, the interfering cell is transmitting a non-scheduled RBSF, whereas the target cell is transmitting a scheduled RBSF. At the receiver, the signal consists of target cell signal plus interfering signal plus additive white Gaussian noise (AWGN).

In the non-colliding scenario (upper part of Figure 5), the CRS of the interfering cell disturb resource elements that are used for data or control signals. Algorithms are required that perform interference mitigation for these channels. In the colliding scenario (lower part of Figure 5), the CRS of the interfering cell disturb CRS resource elements of the target cell. The CRSs are used to perform channel estimation and provide measurements on the quality of the channel. Improved channel estimation/measurement algorithms are required to handle the colliding CRS scenario. In general, more than one interfering cell might be present. Then any combination of colliding and non-colliding CRSs might occur or it can happen that multiple interferers have the same shift of the CRS.

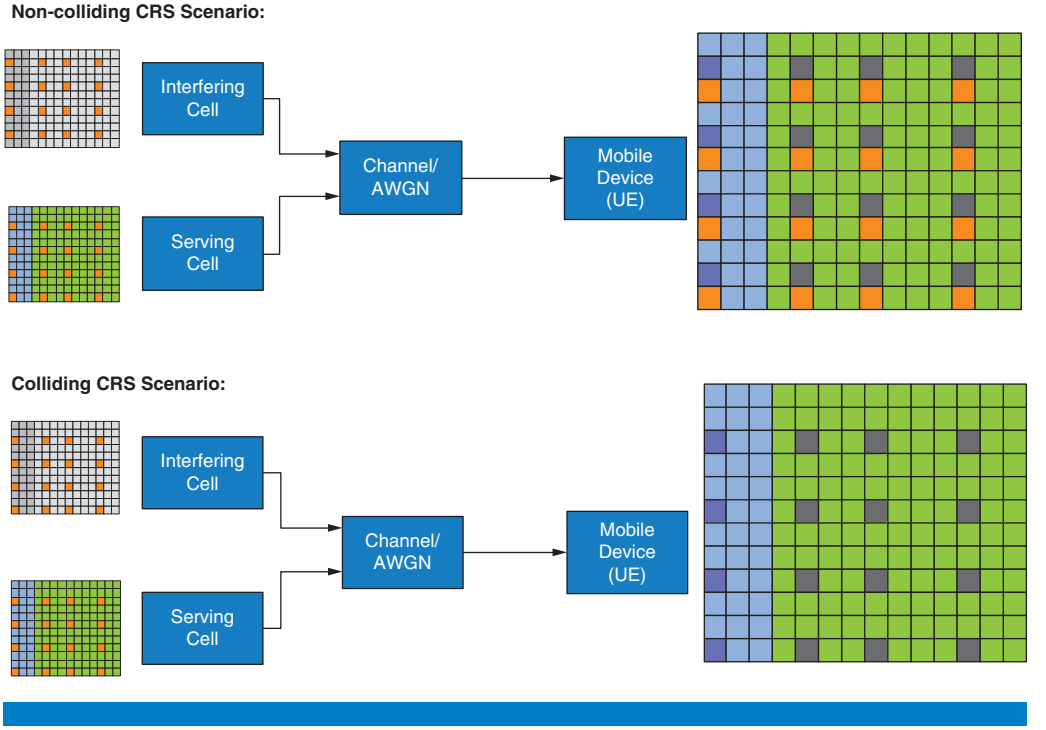


Figure 5: Colliding and non-colliding CRS scenarios (Source: Intel Corporation 2013)

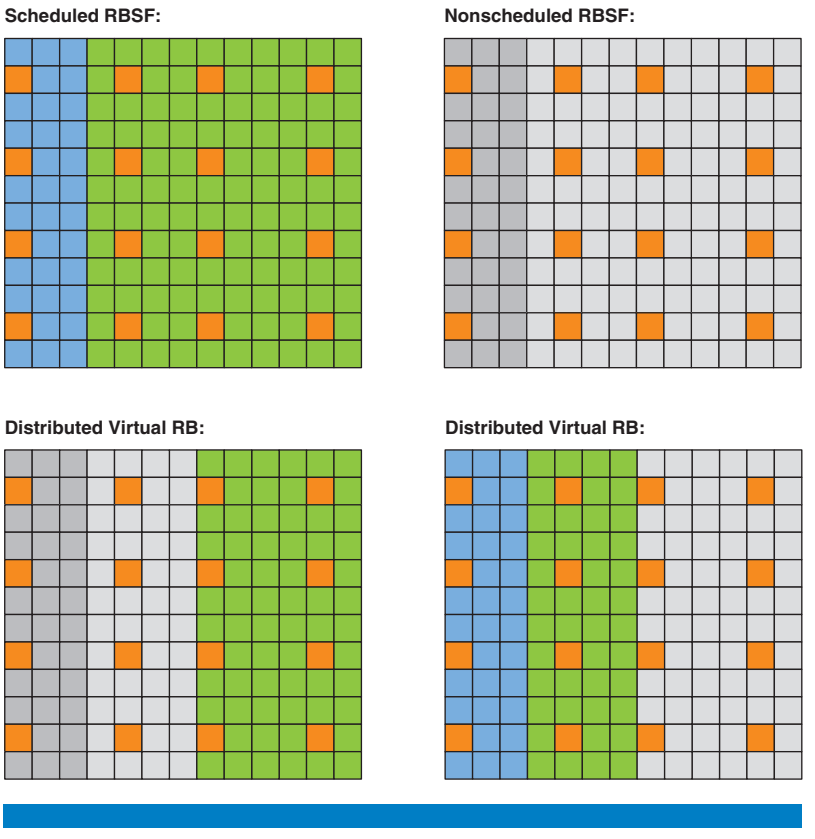


Figure 6: Scheduling options
(Source: Intel Corporation 2013)

Figure 6 provides examples of different scheduling options (normal cyclic prefix). In a scheduled subframe, all the resource elements might be transmitted. In a nonscheduled subframe, shown on the right side of Figure 6, the data and control resource elements are transmitted with power zero (marked in grey) and only the common reference signals (CRSs) are transmitted with the same power level as in a scheduled subframe. The lower part of Figure 6 shows a distributed virtual resource block assignment. In such a scenario, a slot-by-slot frequency hopping is used. In a partially loaded system, the PDSCH of an interfering cell can be present in the first slot of a subframe, in a second slot of a subframe, or the PDSCH is present in both slots of an RBSF but, for example, different precoding matrices could be used for every slot. Additionally, all combinations of scheduled/nonscheduled Physical Downlink Control Channel (PDCCH) and scheduled/nonscheduled Physical Downlink Shared Channel (PDSCH) might occur.

A subframe is called an almost blank subframe (ABS) if all resource blocks of one subframe are not scheduled (PDCCH and PDSCH). However, certain reference and control signals, such as Primary and Secondary Synchronization Signals (PSS/SSS), Physical Broadcast Channel (PBCH) and CRS, must be transmitted in an ABS because legacy devices expect the continuous reception of these signals.

"A subframe is called an almost blank subframe (ABS) if all resource blocks of one subframe are not scheduled (PDCCH and PDSCH)."

Alternatively, a Multicast Broadcast Single Frequency Network (MBSFN) ABS can be scheduled. The MBSFN-ABS has the advantage that it does not contain any CRS. However, it has the disadvantage that it cannot be scheduled in every subframe (for example, it is not possible to schedule an MBSFN-ABS in a subframe containing PSS/SSS or PBCH). Additionally, the scheduling of MBSFN-ABS cannot be changed "dynamically" depending on the traffic load of the pico cell. In an MBSFN-ABS, all resource elements are transmitted with zero power beside the RE of the first OFDM symbol, which contains PDCCH, PCFiCH, and PHICH.

For some of the CRS mitigation algorithms it is important to detect dynamically whether MBSFN subframes have been scheduled on some or all of the interfering cells because the CRS will not be present and a false subtraction might occur.

"Target and aggressor cells might use different power allocation. Depending on the algorithm used, the power allocation needs to be considered in the estimation process…"

Target and aggressor cells might use different power allocation. Depending on the algorithm used, the power allocation needs to be considered in the estimation process or resource elements with different power levels need to be omitted from the estimation.

Figure 7 provides a schematic view of a power level allocation for an RBSF of one cell. The power allocation is configured over the parameters ρ_A and ρ_B. The parameter ρ_A is UE specific; the parameter ρ_B is derived from ρ_A over a cell-specific ratio. Therefore, the UE can assume that it may receive the serving cell signal with the same power allocation over the complete bandwidth, whereas the aggressor cell might use different power allocation values for each PRB (the worst-case scenario being if virtual PRBs are used and each PRB is scheduled for a different UE).

The EPRE (energy per resource element) of the cell-specific RS is constant across the downlink system bandwidth and constant across all subframes

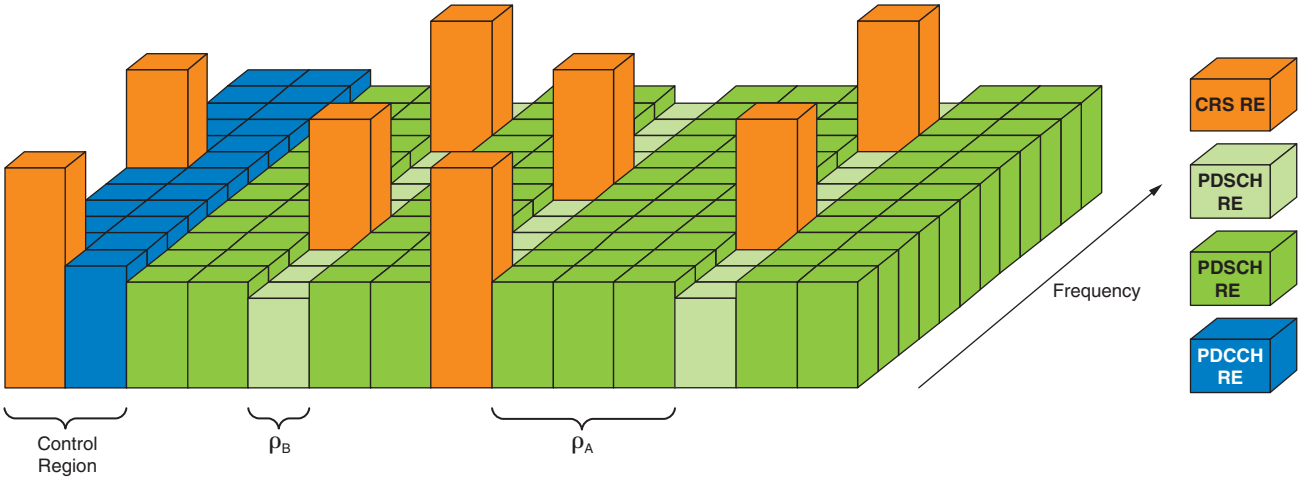


Figure 7: Different power levels in an RBSF
(Source: Intel Corporation 2013)

(until different cell-specific RS power information is received—CRS EPRE can be derived from the parameter *reference Signal Power* provided by higher layers). The parameters ρ_A and ρ_B define the ratio of PDSCH EPRE to cell-specific RS EPRE. Specifically, ρ_B is used for all symbols that contain target CRS RE, whereas ρ_A is used for all other symbols. Figure 8 shows the distribution for one and two antenna ports on the left side and for four antenna ports on the right side (normal cyclic prefix only).

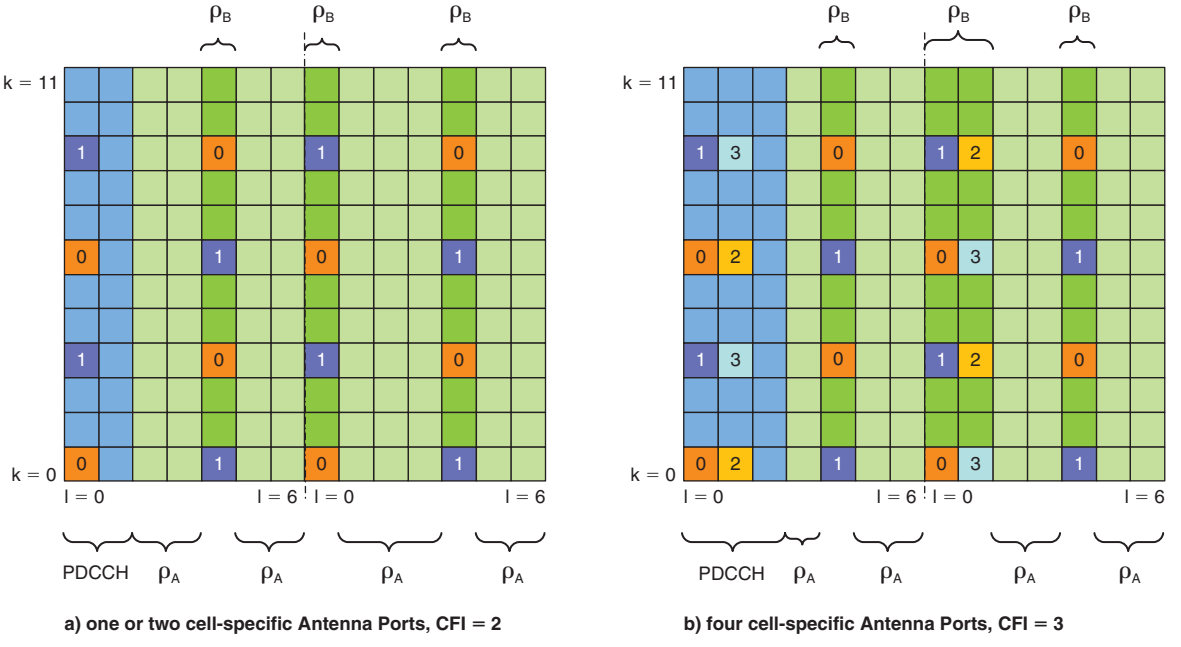


Figure 8: Distribution of power parameters (Source: Intel Corporation 2013)

In a deployment scenario, systems with different bandwidth could be mixed; for example, a macro cell uses 20 MHz, the pico cells are using only 10 MHz.

Depending on the bandwidth, the resource block boundaries between target system and aggressors are shifted by half a resource block as shown in Figure 9. In a worst-case scenario with more than one aggressor, the statistical properties of the interference can only be assumed constant over half a RBSF in frequency domain (six carriers).

"In a worst-case scenario with more than one aggressor, the statistical properties of the interference can only be assumed constant over half a RBSF in frequency domain (six carriers)."

Impact on Physical Layer Algorithms

We now provide an overview of possible algorithm enhancements that should be considered to support the eICIC/FeICIC feature discussed in the previous section. Specifically, we focus on CRS-based channel estimation, CRS interference mitigation prior to detection, and CSI feedback generation.

Channel Estimation

In the context of FeICIC, CRS-based channel estimation is faced with several new challenges.

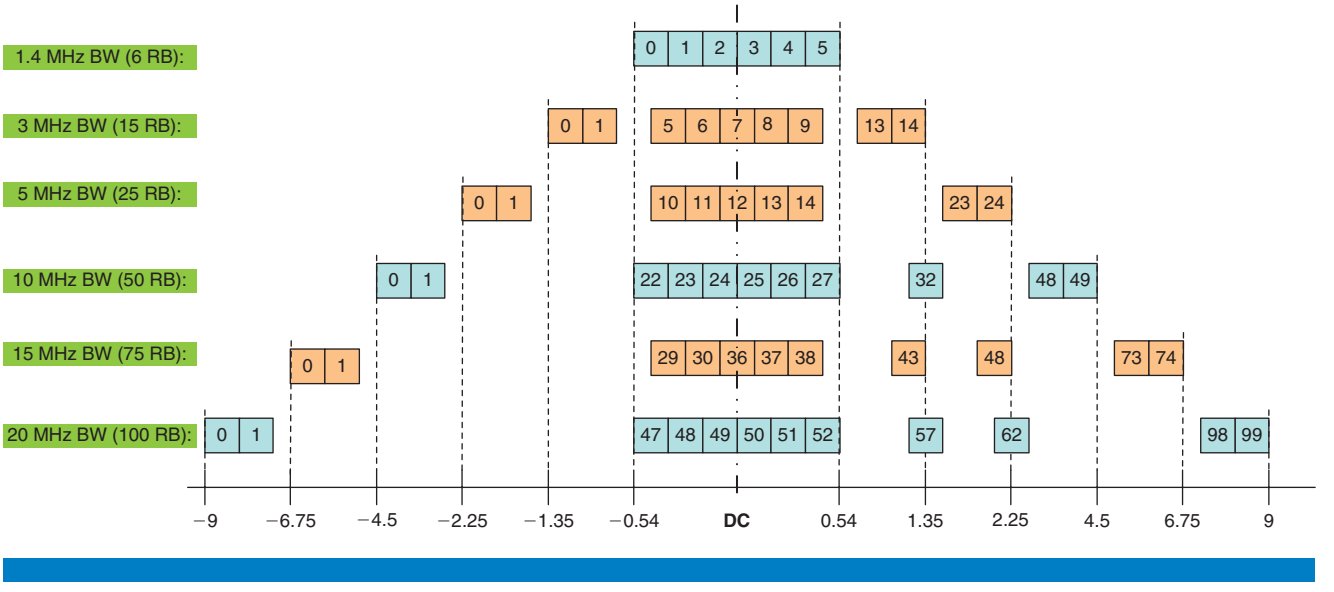


Figure 9: Resource block alignment for different system bandwidths (Source: Intel Corporation 2013)

"…the SINR may change across frequency, depending on the scheduling (per PRB) of non-colliding interference, that is, random interferer data disturbing the serving-cell CRS."

Specifically, the following tasks require some extra attention:

- *Frequency-direction filtering/interpolation of serving-cell CRS*. This filtering takes into account the channel delay spread and the signal-to-interference/noise ratio (SINR) of the CRS being filtered. Note that in a typical FeICIC scenario the SINR may change across frequency, depending on the scheduling (per PRB) of non-colliding interference, that is, random interferer data disturbing the serving-cell CRS.
- *Time-direction filtering/interpolation of (frequency-filtered) serving-cell CRS.* This filtering takes into account the channel Doppler spread and the SINR of the CRS being used. The SINR may change across time (per slot), depending on the scheduling of non-colliding interference. This is especially relevant in ABS/non-ABS scenarios.
- *Cancellation of interfering CRSs that collide with the serving-cell CRS.* These collider CRSs may be very strong and require special attention, especially for PDSCH with moderate or high SINR points of operation.
- *Ancillary tasks supporting channel estimation.* These functions comprise estimators for the SINR (now time- and possibly frequency-selective), estimators for delay and Doppler spreads and shifts, and functions for the (now dynamic) generation and/or selection of suitable channel estimation filtering coefficients.

Also, we observe that, depending on the CRS interference mitigation method applied prior to detection (see the next subsection, "CRS Interference Mitigation"), an estimate of the aggressors' channel and related parameters might be required to be able to accurately reconstruct (and, if needed, subtract) the interfering signals overlapping with the serving cell signal.

A carefully designed CRS-based channel estimator in the context of FeICIC should focus on the following basic tasks:

- colliding interference cancellation, that is, cancel or suppress colliding CRS disturbing the serving-cell CRS used in channel estimation;
- non-colliding interference mitigation, that is, minimize the effect of random data disturbing the serving-cell CRS used in channel estimation

Colliding interference cancellation must cope with the following disturbances, affecting serving-cell CRS (S-CRS) samples simultaneously:

- Colliding CRS (C-CRS): a single collider, which may be up to 9 dB stronger than the wanted signal (S-CRS). Moreover, the collider may bear considerable timing (TO) and frequency (FO) offsets relative to the wanted signal.
- Non-colliding CRS (N-CRS): one or more non-colliding random data signals, which also may add up to considerable strength.
- Background noise and interference.

Consider a set of received S-CRS samples (one- or two-dimensional) containing both colliding and non-colliding interference. In MIMO terms, this can be viewed as a number of layers (S-CRS, C-CRS, several N-CRS) stacked onto each other. Ideally, these layers should be separable by virtue of their modulation symbols, but these are in general not fully orthogonal (C-CRS) or even unknown (N-CRS). As outlined below, at least the collider(s) can be canceled or strongly mitigated.

First, the received CRSs are demodulated for all layers (S-CRS and C-CRS) to be separated (often only one collider needs to be considered, but more are also possible), thus creating two or more sets of CRS samples. These sets can be further processed by either filtering both sets first (part of channel estimation) and then cancel the colliding interference, or by canceling the collider first and then performing channel estimation filtering.

Whether or not the CRS are filtered before, collider cancellation is faced with two or more signals, which experience mutual crosstalk caused by the nonorthogonality of the individual CRS modulation sequences. The parameters of the respective crosstalk matrix model, namely, entries of the crosstalk matrix, can be derived from the known S-CRS and C-CRS modulation sequences (cell IDs known through cell search) and the filter coefficients if the CRS are filtered before collider cancellation. Note that crosstalk cancellation only hinges on the modulation sequences (and filter coefficients), that is, on deterministic parameters that are not affected by any kind of noise. This implies that FeICIC collider cancellation is independent of non-colliding interference and its mitigation.

Based on the deterministic crosstalk model, the S/C-CRS layers can finally be separated in several ways, for example, through Minimum Mean Square Error (MMSE) or Zero Forcing (ZF) filtering, at the expense of some (usually small)

"Whether or not the CRS are filtered before, collider cancellation is faced with two or more signals, which experience mutual crosstalk caused by the nonorthogonality of the individual CRS modulation sequences."

"This implies that FeICIC collider cancellation is independent of non-colliding interference and its mitigation."

noise enhancement. Because all mutual interference is thus purged from both serving-cell and all colliding CRSs, these can be further processed individually as in the non-FeICIC case, hence, baseline channel estimation (working on S-CRS) can be reused.

As a word of caution we note that the S/C-CRS modulation sequences may be collinear on some occasions, especially when the sets of CRSs are small. Then collider cancellation cannot be done for this particular set, so one has to resort to either choosing another set of CRSs (with non-collinear modulation sequences) or some other kind of cancellation gap handling, such as using the estimates from an adjacent CRS set.

After the collider has been cancelled as described above, regular channel estimation can take place. However, for the case where a payload data signal from a non-colliding CRS cell is overlapping with the serving cell CRS belonging to the channel estimator (frequency or time) filtering/interpolation window, additional steps need to be taken to mitigate the residual interference and deliver an accurate channel estimate at each RE of a given PRB (*non-colliding interference mitigation*).

To this end, we recall that the conventional channel estimation filter configuration approach often adopted in the absence of interference at CRS positions is based on a simple quantization mechanism.[7] Specifically, the channel parameters needed to configure the frequency and time filters, that is, SINR, delay spread, and Doppler spread, are estimated, quantized, and rounded up to the next value for which channel estimation filter coefficients are available in a given filter bank. This mechanism works quite well under the assumption that delay spread, Doppler spread, and SINR are constant across the time or frequency filtering/interpolation window of the channel estimator. While this is typically still true for delay and Doppler spread, the above assumption is no longer valid for the SINR in FeICIC scenarios. As a consequence, the conventional filter selection mechanism needs to be modified to account for the variable SINR levels across the two-dimensional (time and frequency) channel estimation window. To avoid the case where CRSs belonging to PRBs with too large an interference level and thus too low an SINR are included in the time or frequency filtering/interpolation window, several options are available. Among these, one could:

"…the conventional filter selection mechanism needs to be modified to account for the variable SINR levels across the two-dimensional (time and frequency) channel estimation window."

- dynamically modify the size of the channel estimation filters depending on the level of SINR variations measured across subframes/PRB so as to use only the serving cell CRS experiencing low interference;
- based on the estimated SINR, compute new channel estimation coefficients with interference suppression capabilities, that is, capable of mitigating the impact of the serving cell CRS experiencing higher interference.

Approximated approaches to the SINR-optimum solution above are also possible. In general, the final filter design/configuration choice will depend on the desired complexity versus estimation accuracy tradeoff.

CRS Interference Mitigation

Many different methods have been developed to perform efficient CRS interference mitigation in synchronized network scenarios shown in Figure 5. Most of the algorithms fall in one of the following groups:

- *CRS Puncturing and Scaling* – the soft bits of resource elements with CRS interference are either set to zero (puncturing) or scaled according to the interference level (scaling)
- *CRS Interference Cancellation* – the aggressor CRS signal is reconstructed and subtracted from the received resource elements. This method requires a channel estimate of the aggressor cell to perform the signal reconstruction
- *CRS Whitening (Interference Rejection Combining)* – a noise and interference covariance matrix, which reflects the structure of the CRS interference, is estimated and inverted through a Cholesky decomposition. The resulting whitening filter is applied to the resource elements that face CRS interference.

The different mitigation algorithms above can be classified in terms of implementation complexity, memory requirements, performance, and dependency on other estimates such as aggressor channel estimates, aggressor timing and frequency offset estimates, and aggressor Doppler estimates.

"The different mitigation algorithms above can be classified in terms of implementation complexity, memory requirements, performance, and dependency on other estimates such as aggressor channel estimates, aggressor timing and frequency offset estimates, and aggressor Doppler estimates."

For example, CRS puncturing/scaling has the lowest complexity but also relatively poor performance. CRS interference cancellation, however, ensures very good interference mitigation when the aggressors' parameters are available or can be accurately estimated, which, on its turn, increases the complexity of the receiver processing. Finally, CRS whitening can also achieve good performance without requiring accurate knowledge of the interference (but only its second order statistics).

In a typical product implementation, a combination of these methods might be employed depending on:

- the interference scenario we are facing, for example, the number of colliding or non-colliding aggressors and related powers; the downlink physical channel being decoded/demodulated (broadcast, control or shared channel) and the related performance requirements;
- the side information on the aggressors' signal parameters available at the receiver through explicit estimation or through network signaling and the accuracy of this information;
- the desired performance versus complexity tradeoff.

In Figures 10 and 11 we provide some throughput results showing the overall impact of the presence of different type of aggressors (collider only, non-collider only, both collider and non-collider) for LTE transmission mode 2 and 3. For these figures, the CRS whitening technique has been used to mitigate interference.

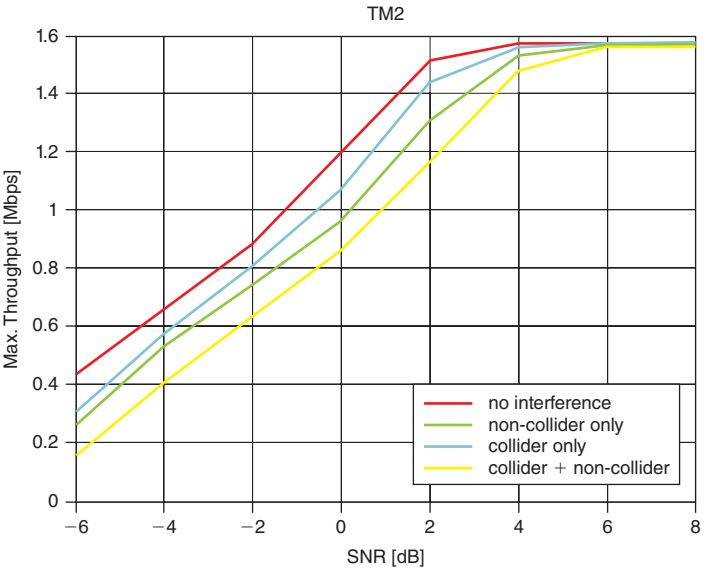


Figure 10: Throughput performance in the presence of interference for transmission mode 2. (Source: Intel Corporation 2013)

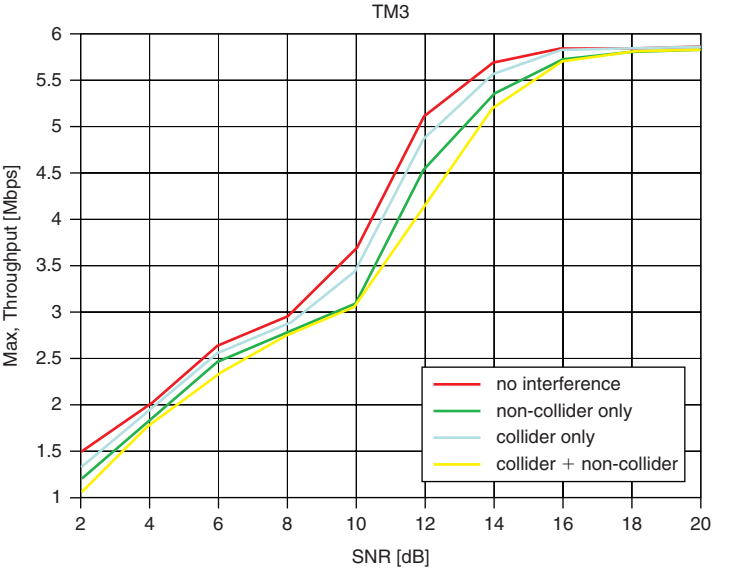


Figure 11: Throughput performance in the presence of interference for transmission mode 3. (Source: Intel Corporation 2013)

"…with the introduction of eICIC/ FeICIC, UEs can no longer assume homogeneous interference level throughout a given RB."

CSI Feedback Generation

As mentioned in the previous paragraphs, with the introduction of eICIC/ FeICIC, UEs can no longer assume homogeneous interference level throughout a given RB. The presence of ABSs implies that different REs in a subframe can experience significantly different level of interference. To

maintain accurate feedback in such conditions, a UE needs to consider two major aspects:

- *Groups of REs with distinct interference levels*. A UE can determine potential interference patterns depending on the estimate of the number of interfering cells, the number of Tx antenna ports of the interferers, and their cell ID shifts. From this, groups of REs with distinct levels of interference can be constructed. One method to counter the presence of ABSs is to estimate channel quality metrics for each of these groups and then combine them.
- *Codeblocks with distinct interference levels.* The number of codeblocks within a given transport block increases with larger allocation size and higher MCS. With the presence of ABSs, different codeblocks can experience significantly different level of interference, since CRSs are located only at certain symbols. For instance, codeblocks occupying CRS-carrying symbols might see a high level of interference while other codeblocks may face no interference at all. The former codeblocks dominate the transport block reception performance, and a single codeblock CRC error leads to a transport block error. To accurately reflect the impact of this, a UE needs to forecast the expected number of codeblocks as well as the modulation and coding scheme (MCS), and to estimate the channel quality metrics separately for each group of codeblocks with significantly distinct level of interference.

Advanced DL MIMO in LTE-A

In this section we provide an overview of the DL MIMO feature in the context of LTE-A and discuss the impact of this feature on the design of efficient baseband algorithms.

Feature Overview

In the LTE Release 8 and 9, MIMO[8][9] is mainly supported through three techniques: transmit diversity (TD), spatial multiplexing (SM), and beamforming (BF), with both open-loop (OL) and closed-loop (CL) operation being possible for SM. LTE TD is based on space-frequency block coding (SFBC) of neighbouring OFDM subcarriers. The goal of TD is to maximize diversity by transmitting dependent user bits over multiple antennas in the hope of receiving independently faded replicas of the signal, which can then be properly combined at the receiver to achieve a better signal-to-noise ratio.

When knowledge of the radio channel impulse response is available at the transmitter (such as via proper feedback signalling mechanisms or exploiting channel reciprocity), BF can be applied. BF attempts to maximize the signal-to-interference-plus-noise ratio (SINR) by individually and appropriately weighting the transmit signal at each transmit antennas. In LTE Release 8 and 9, single- and dual-layer beamforming for TDD have been standardized. Note that in case of BF the number of transmit antennas needs not be known to the UE.

In contrast to TD and BF, SM exploits the spatial domain to increase the peak data rate by transmitting parallel independent data streams over multiple antennas.

In LTE SM the vector of transmit signals for the multiple antennas is multiplied by a precoding matrix, which must be carefully selected to match to the instantaneous radio channel conditions. Precoding in LTE is codebook based, that is, the applicable PM is restricted to a predefined finite set, the codebook. In closed loop operations, the UE chooses the best PM to optimize its reception and signals back the corresponding index (precoding matrix indicator, PMI) to the eNB. Additionally, the rank of the channel (rank indicator, RI) is fed back to the eNB. The rank indicates the maximum number of independent spatial data streams that can be transmitted over that specific MIMO channel. The codebook for 4x4 MIMO is based on Householder reflections[1] and its elements have a nested property, that is, columns of a lower-rank PM are a subset of those of a higher-rank PM, which simplifies the PMI selection mechanism. The codebook for the 2-antenna case is designed more intuitively without any special structure.

The above-mentioned feedback on the quality and properties of the downlink channel (PMI and RI, together with channel quality indicator, CQI) for LTE Release 8 and Release 9 is based on the common reference signals (CRSs), which are also used for regular channel estimation and demodulation.

"To this end new reference signals have been defined and a new PDSCH transmission mode 9 (TM9) has been standardized, featuring data transmission from eight antenna ports and compressed feedback mechanisms for signalling preferred precoding."

In order to support the higher data rates of LTE-Advanced, new MIMO techniques have been introduced in LTE Release 10.

To this end new reference signals have been defined and a new PDSCH transmission mode 9 (TM9) has been standardized, featuring data transmission from eight antenna ports and compressed feedback mechanisms for signalling preferred precoding.

To enable downlink SU-MIMO transmission with up to eight layers, new precoded UE-specific demodulation reference signals (DM-RS) have been defined. A typical pattern for DMRS is shown in Figure 12 for both normal and extended CP.

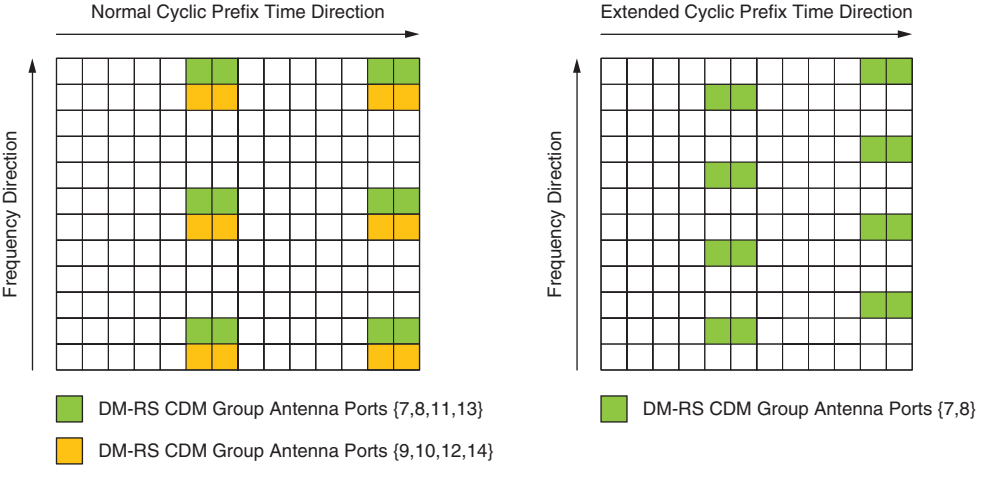


Figure 12: DMRS symbol pattern
(Source: Intel Corporation 2013)

In order to ensure the orthogonality of the DMRS among each other and with already existing CRSs, a combination of frequency multiplexing (FDM) and code division multiplexing (CDM) is used. Specifically, DMRSs are transmitted on antenna port 7–14.[1] Antenna ports (AP) 11 and 13 use the same locations as AP 7 and 8, while AP 12 and 14 use the same locations as AP 9 and 10. For extended Cyclic Prefix (eCP) only two antenna ports are used: AP7 and AP8. Although several APs use the same resource element simultaneously (AP 7, 8, 11, 13, and AP 9, 10, 12, 14 mutually), they are modulated with Orthogonal Cover Codes (OCCs) in the time domain. The OCC allows almost perfect separation of the DMRS for slowly fading channels. With at most four transmit antennas, the OCCs are pairwise orthogonal for the first pair (symbols 5 and 6) and second pair (symbols 12 and 13) of DMRS, which allows a moderately high Doppler spread. However with eight transmit antennas and 5–8 layers, the OCCs are only orthogonal for the full length of four OFDM-symbols, which requires a much lower Doppler velocity.

The DMRS are transmitted only in the RBs where demodulation takes place, thus significantly reducing the overhead compared to the CRSs, which are transmitted across the whole bandwidth. Also, DMRS undergo the same precoding as the data symbols, which has a twofold advantage:

- the precoding matrix applied at the eNodeB does not need to be signaled to the UE, thus reducing the network signaling overhead;
- the base station can implement multilayer beamforming schemes in a manner that is transparent to the UE. This is also known as non-codebook–based precoding.

Finally, we note that in some UE configurations, the precoder can be assumed to be the same across all RBs with a single precoding resource block group. This aspect can be exploited to improve the accuracy of channel estimation.

"The DMRS are transmitted only in the RBs where demodulation takes place, thus significantly reducing the overhead compared to the CRSs, which are transmitted across the whole bandwidth."

LTE Release 10 also introduces channel state information reference signal (CSI-RS) for obtaining channel state information for up to eight antenna ports and thus assist the eNodeB in the choice of the precoder.[1] The use of CSI-RS enables multilayer beamforming also for the case of FDD, while LTE Release 8 dual-layer beamforming (also known as transmission mode 8) is effective in the case of TDD, where the channel reciprocity can be exploited to select the beamforming weights.

"LTE Release 10 also introduces channel state information reference signal (CSI-RS) for obtaining channel state information for up to eight antenna ports and thus assist the eNodeB in the choice of the precoder."

A typical CSI-RS pattern is shown in Figure 13. In contrast to the DMRS, CSI-RS are unprecoded and defined over the entire system bandwidth; they are orthogonally multiplexed and sparse to limit overhead.

For the case of eight CSI-RS ports, a new codebook and related enhanced feedback have been defined.

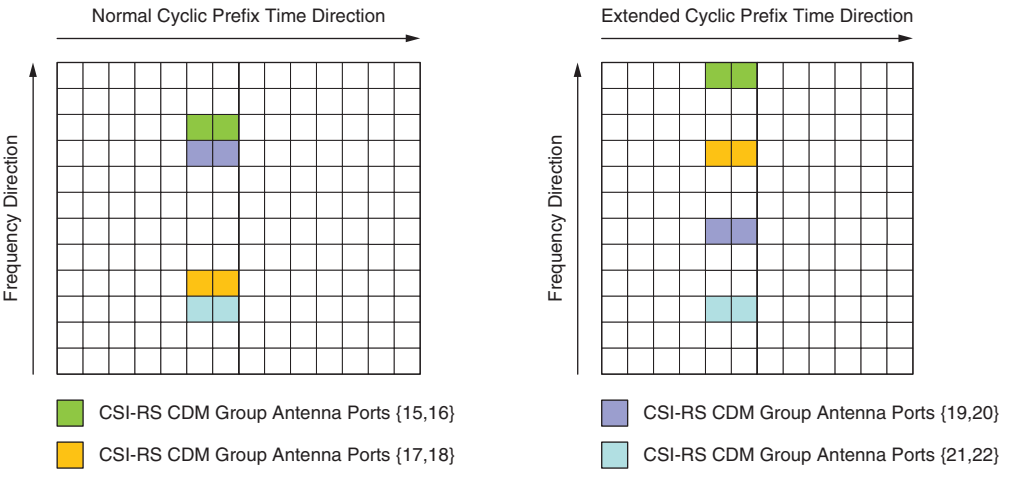


Figure 13: CSI-RS symbol pattern
(Source: Intel Corporation 2013)

The 8-Tx codebook is defined for each candidate number of layers (or rank) v with a first PMI value of $i_1^{(v)}$ and a second PMI value of $i_2^{(v)}$

This decomposition of PMI into two values arises from the two-stage nature of the 8-Tx precoding matrix $\boldsymbol{W}$. Specifically, $\boldsymbol{W}$ can be viewed as a matrix multiplication of two sub-precoding matrices $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$, that is, $\boldsymbol{W} = \boldsymbol{W}_1 \cdot \boldsymbol{W}_2$, where $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ belong to different, smaller codebooks. The precoder $\boldsymbol{W}_1$ represents channel correlation properties due to the cluster of beams from a co-polarized uniform linear subarray (ULA) of eNB Tx antennas, and $\boldsymbol{W}_2$ represents one particular beam among a given cluster of beams, as well as instantaneous effects of co-phased beams among cross-polarized (uncorrelated) pairs of eNB Tx antennas.

This peculiar structure of the precoder reflects the following observation. The channel correlation properties due to ULA beams, in general, exhibit spatial coherence, spectral coherence, and temporal coherence. Spatial coherence, in heuristic terms, means that the channel characteristics corresponding to two spatially adjacent beams are typically similar, and gradually change across beams. To put it more precisely, the mutual information (MI) per beam does not fluctuate drastically across adjacent beams, and the optimal beam's immediate neighbors likely result in high MI with little throughput degradation compared to the optimal choice. Macro-cell downlinks typically exhibit substantial spatial coherence, where an eNB faces few nearby scatterers (narrow angular spread of departure). Spectral coherence means that the channel properties stay relatively constant over a contiguous spectrum. In the context of LTE-A, it is assumed that the channel characteristics due to ULA beams exhibit spectral coherence over the whole band. This is why there is only wideband report defined for $\boldsymbol{W}_1$. Temporal coherence means that the channel

properties are slowly varying, that is, they remain relatively constant over a period of time. In LTE, it is assumed that the channel characteristics due to ULA beams exhibit temporal coherence over multiple CSI feedback reports. Thus periodic reporting supports a mechanism for updating $\boldsymbol{W}_1$ once per multiple $\boldsymbol{W}_2$ updates.

Impact on Physical Layer Algorithms

We now provide an overview of the possible algorithm enhancements enabling the advanced downlink MIMO feature discussed in the previous section. Specifically, we focus on DMRS-based channel estimation, MIMO detection in the spatial multiplexing mode, and CSI feedback generation.

Channel Estimation

The advanced DL MIMO feature requires DMRS-based channel estimation (in addition to regular CRS-based channel estimation), which is faced with the following tasks and challenges:

- *Frequency-direction filtering of DMRS samples from PRB groups of size 1, 2, or 3 PRB.* This filtering needs to take into account the—in general not-equally-spaced—DMRS positions, channel delay spread, and the SINR. Depending on interferer scheduling, the SINR may change per PRB, although not as drastically as in FeICIC.
- *Time-direction filtering of (frequency-filtered) DMRS samples of one subframe* – longer filtering is not appropriate since the precoding may change at SF boundaries.
- *Layer separation*. This important new function separates up to eight MIMO layers. These eight layers are "encoded" into two DMRS groups separated in frequency (group 1 for antenna ports 7, 8, 11, 13, group 2 for AP 9, 10, 12, 13) and a Walsh cover code with four codewords in time. The latter are separable by simple sign-and-add operations, or equivalently by first demodulating each layer (including Walsh code) and then performing time-direction channel estimation filtering, thus canceling cross-layer interference. This second option merges layer separation with time-direction channel estimation.
- *Interpolation between frequency/time-filtered DMRS* for all RE and layers for which a channel estimate is to be generated.
- *Ancillary tasks supporting DMRS-based channel estimation*.

  These functions comprise estimators for the SINR (preferably per PRB), estimators for delay spreads and shifts, and functions for the generation and/or selection of suitable channel estimation filtering coefficients.

"Layer separation. This important new function separates up to eight MIMO layers."

Detection

The new categories 6, 7, and 8 introduced by 3GPP Release 10 require, in the spatial multiplexing mode, MIMO detection for up to $N = 4$ or even up to $N = 8$ M-QAM modulated spatial layers, where M can be 4, 16, or 64. This significantly increases the MIMO detection complexity compared to the conventional Release 8 case of $N = 2$.

Per resource element, the optimum brute-force maximum likelihood MIMO detection algorithm would compute a metric for each of the M^N modulation symbol vector hypothesis, and select the hypothesis corresponding to the best metric. It is clear that for large M, N, this algorithm is by far too complex to be implemented. Furthermore, the MIMO detection problem was shown to be NP complete, and thus no algorithm with asymptotically better worst-case complexity than the brute-force approach is known in the literature.

One way to deal with this situation is to use suboptimum detectors like linear pre-filtering, followed by QAM-detection per spatial layer, treating the other layers as interference. However, in many practical scenarios, especially for high SNR, there is a large performance gap versus the optimum detection.

"A viable alternative is then to resort to the class of the so-called tree-search–based detectors that compute the optimum solution with variable complexity depending on the condition of the MIMO system equation under consideration."

A viable alternative is then to resort to the class of the so-called tree-search–based detectors that compute the optimum solution with variable complexity depending on the condition of the MIMO system equation under consideration.

A tree-search–based detector is preceded by a linear pre-filter (typically a QR decomposition or a permuted Cholesky decomposition) that sufficiently reduces the MIMO system of equations to an equivalent $N \cdot N$ system and exposes the tree-structure of the problem. The actual tree-search tries to construct the modulation symbol vector hypothesis with good metrics first (for example, mean square error or Euclidian metric) in order to find the optimum solution as soon as possible.

Thanks to link adaptation, the MIMO equations to be processed are typically relatively well conditioned. As a consequence, the tree-search–based detectors are able to find the optimum solution with affordable complexity in most cases, while still providing suboptimum solutions otherwise.

CSI Feedback Generation

One of the major challenges in implementing efficient CSI feedback generation for LTE-A comes from the size of the 8-Tx codebook, which is substantially larger than those of the 4-Tx or 2-Tx codebooks. Thus, it is of practical interest to devise a precoding matrix search method with a reduced search set and whose associated performance is comparable to that of full search.

"…it is of practical interest to devise a precoding matrix search method with a reduced search set and whose associated performance is comparable to that of full search."

The 8-Tx codebook defines 256, 256, 64, 32, 4, 4, 4, and 1 precoding matrices, for rank candidates $v = 1, 2, 3, \ldots, 8$, respectively. Thus, for instance, a UE that supports up to two layers faces a precoder search set size of 512 matrices, and a UE that supports up to four layers had to contend with a search set size of 608. To put this in perspective, the UE supporting up to four-layer transmission from a 4-Tx eNB only faces a maximum search set size of 64 matrices.

Two of the potential algorithms that can reduce the precoder search set while maintaining an acceptable level of accuracy are summarized below. In particular, these methods exploit spatial and temporal coherence properties of a typical cellular channel as well as the two-stage nature of the 8-Tx precoding matrices.

- *Spatial sampling.* Spatial coherence can be exploited in a multistage search format, at each stage of which only a sampled subset of search space is looked at. A two-stage format naturally arises from the two-stage structure of $\boldsymbol{W}$. Specifically, we first search through $\boldsymbol{W}_1$ with a fixed $\boldsymbol{W}_2$, then search over $\boldsymbol{W}_2$ given the $\boldsymbol{W}_1$ selected at the first stage. Fixing $\boldsymbol{W}_2$ at the first stage is equivalent to taking one sample beam per cluster and fixing the co-phasing factor, hence the name "spatial sampling."

  The search set size reduction can be made more aggressive by selecting the rank based on the outcome of the first round, thereby constraining the search set size in the second round. This spatial sampling method is expected to give little performance degradation compared to the full search, in high channel correlation and high spatial coherence. It is ineffective in low channel correlation and low spatial coherence. However, the throughput variation due to different PMI is small in low correlation channels, hence the corresponding performance degradation is expected to be graceful and limited.

"Spatial coherence can be exploited in a multistage search format, at each stage of which only a sampled subset of search space is looked at."

- *Time tracking*. Temporal coherence can be exploited by limiting the search set for the current search based on the outcome of the previous search. Specifically, given a previous decision on

$$\boldsymbol{W}_1 = \overline{\boldsymbol{W}}_1$$

  we limit the search set of beam clusters to the immediate neighbors of

$$\overline{\boldsymbol{W}}_1$$

  We note that this method has to be an add-on feature activated only when certain conditions are met, since it requires a previous decision. This decision by default is made for each PMI report requested by eNB, but can be triggered more often by UE. The performance with this time tracking method is expected to be comparable to that of the full search, with a sufficiently close spacing of PMI reports and in slow fading channels. This method becomes ineffective when the spacing of PMI decisions is too large compared to the channel coherence time of the correlation characteristics. A safety mechanism, which adjusts the search set size (number of neighbors) based on decision period and channel coherence time, can mitigate this shortcoming. We note that the principle of time tracking can equivalently be applied in frequency domain by exploiting spectral coherence, such that the search set of beam clusters are based on decisions for the neighboring spectral bins.

"Temporal coherence can be exploited by limiting the search set for the current search based on the outcome of the previous search."

Final Considerations

Out of the many features that have been defined to fulfill the LTE-A downlink requirements, our contribution has mainly focused on eICIC/FeICIC techniques and on advanced multiple antenna schemes. We discussed how most of the existing baseband algorithms need to be extended, enhanced, and optimized for the newly defined LTE-A framework, in order to adequately support the features above and at the same time ensure backward compatibility

with the previous standard releases.

The next challenges are to combine the existing features with those being further developed in LTE Releases 11 and 12 and to optimize algorithms and related receiver architectures to achieve the best possible performance while at the same time meeting other key requirements such as reduced power consumption and die area.

References

[1] 3rd Generation Partnership Project (3GPP). TS 36.211 v10.4.0 (2011–12): "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation."

[2] 3rd Generation Partnership Project (3GPP). TS 36.212 v10.4.0 (2011–12): "Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and Channel Coding."

[3] 3rd Generation Partnership Project (3GPP). TS 36.213 v10.4.0 (2011–12): "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures."

[4] 3rd Generation Partnership Project (3GPP). TR 36.913 v10.0.0 (2011–03): "Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN)."

[5] Damnjanovic, A., J. Montojo, Yongbin Wei, Tingfang Ji, Tao Luo, M. Vajapeyam, Taesang Yoo, Osok Song, and D. Malladi, "A survey on 3GPP heterogeneous Networks"; *IEEE Wireless Communications*, Vol.18, No.3; June 2011.

[6] Damnjanovic, A., J. Montojo, Joonyoung Cho, Hyoungju Ji, Jin Yang, and Pingping Zong, "UE's Role in LTE Advanced Heterogeneous Networks," *IEEE Communications Magazine*, February 2012.

[7] Hoeher, P, S. Kaiser, and P. Robertson, "Pilot-symbol-aided channel estimation in time and frequency," Proc. ICASSP'97, Vol. 3, pp. 1845-848, Munich, April 1997.

[8] Tse, D. and P. Viswanath, *Fundamentals of Wireless Communications* (Cambridge: Cambridge University Press, 2005).

[9] E. Biglieri, *MIMO Wireless Communications* (Cambridge: Cambridge University Press, 2007).

Author Biographies

Cecilia Carbonelli (cecilia.carbonelli@intel.com) has been a senior specialist in the Wireless System Engineering Group of Intel Mobile Communications

since 2011. She received the Laurea degree (cum laude) in Telecommunications Engineering and PhD in Information Engineering from the University of Pisa, Italy. From 2005 to 2006, she was a post-doctoral research associate at the University of Southern California. During 2006 she held a visiting appointment at Qualcomm Inc. and joined Infineon Technologies AG. Her research interests include the area of communication theory and signal processing with special emphasis on the design of physical layer algorithms.

Axel Clausen (axel.clausen@intel.com) is a member of the Wireless System Engineering group. The focus of his current work is the development of algorithms and architectures for the physical layer of LTE-Advanced (Release 10 & 11) with emphasis on concepts for interference mitigation. In the past, he has been a system architect and project lead for multiple broadband access technologies, including IEEE 802.11n, VDSL2 vectoring, and DVB cable modem.

Franz Eder (franz.eder@intel.com) joined Intel in 2011 and is working in the wireless system engineering team on baseband algorithms for LTE Advanced. His focus is on MIMO detection and channel decoding algorithms taking into account architectural aspects to ensure efficient implementations.

Stefan A. Fechtel (stefan.fechtel@intel.com) has been with with Intel Mobile Communications since 2011, where he is currently focusing on LTE-Advanced baseband development. He received his PhD from Aachen University of Technology in 1993. In 1997 he joined Siemens Semiconductors / Infineon Technologies, working in the areas of DVB-T/H, 2G, WLAN, and LTE.

Yeong-Sun (Yongsun) **Hwang** (yeong-sun.hwang@intel.com) is a staff wireless system engineer at Intel Mobile Communications. He works on design, implementation, and verification of signal processing algorithms for wireless communication applications, with emphasis on AGC, RF impairment compensation, synchronization, channel parameter estimation, and channel quality estimation for OFDMA-based modems.

MU-MIMO SYSTEM CONCEPTS AND IMPLEMENTATION ASPECTS

Contributors

Biljana Badic
Product Engineering Group,
Intel Corporation

Rajarajan Balraj
Product Engineering Group,
Intel Corporation

Ingmar Groh
Product Engineering Group,
Intel Corporation

Peter Fazekas
Department of Networked Systems and Services, Budapest University of Technology and Economics

Multiuser Multiple-Input and Multiple-Output (MU-MIMO) systems have become promising in the context of achieving high data rates envisioned for 4G (4th Generation) systems and since LTE (Long Term Evolution) Release 8 MU-MIMO has been one of the key topics in 3GPP and research communities. However, due to the significant limitation in MU-MIMO systems introduced through 3GPP standardization, MU-MIMO's practical application is still facing serious challenges. With heterogeneous networks and small cells becoming the reality, importance of MU-MIMO has increased and 3GPP is continuously improving MU-MIMO capabilities in order to achieve theoretically established multiuser gains.

This article discusses some of the key practical MU-MIMO drawbacks in the baseband design, their development through the LTE evolution from Release 8 to Release 12, and potential ways to address those shortcomings. System aspects such as accurate feedback link, importance of MU-MIMO channel models, and antenna configuration, as well as receiver design aspects such as channel estimation and detection procedures, are discussed in this article and evaluated in various MU-MIMO practical scenarios.

Introduction

Among many features in the LTE systems, the multiuser multiple-input-multiple-output (MU-MIMO) scheme has been identified as one of the key enablers for achieving high spectral efficiency and it was already introduced in LTE (Long Term Evolution) Release 8 but in a very primitive form.[1] A crucial limitation for Release 8 MU-MIMO transmission is that there is no specific MU-MIMO signaling and dedicated MU-MIMO precoding. Therefore, MU-MIMO gain could not be fully exploited and as such MU-MIMO is not a part of current commercial deployments of LTE systems. LTE Release 9 and LTE-Advanced Release 10 enhanced the operations of the MU-MIMO system. UE-specific, demodulation reference signals (DMRS), channel state information (CSI) RS, dual-layer support, and dynamical switching between single- and multiuser transmission are the key MU-MIMO enhancements. Furthermore, a dual codebook structure is adopted for the 8-transmit antenna configuration, where one codebook represents wideband and long-term channel properties, while the other represents frequency-selective and/or short-term channel properties.

"…the accuracy of channel state information (CSI) available at both the receiver and transmitter plays an important role in the design of MU-MIMO systems."

Since MU-MIMO is a closed-loop transmission, the accuracy of channel state information (CSI) available at both the receiver and transmitter plays an important role in the design of MU-MIMO systems. While it is convenient for

theoretical investigations to assume that perfect CSI is available, in real-world scenarios, a receiver needs to estimate CSI via a pilot symbol disturbed with noise and report this CSI to the transmitter with finite granularity. Since the channel is time-varying due to the Doppler spread, it is most likely that CSI at the transmitter will be outdated. This difference will create mismatches of the channel estimated and as result the MU-MIMO performance will degrade. Furthermore, CSI reporting in LTE systems is optimized for single-user transmission because it targets maximizing the throughput of that particular UE. Due to this limitation the system throughput may not be maximized in multiuser transmission.

"Predicting the CSI is a challenging and resource-consuming task in practical cellular systems and obtained information is always unavoidably imperfect."

Predicting the CSI is a challenging and resource-consuming task in practical cellular systems and obtained information is always unavoidably imperfect. Therefore, in order to design systems that can meet minimum 4G requirements and deliver expected high data rates, it is critical to understand factors impacting the accuracy of feedback link in MU-MIMO systems. This article presents the analysis of what we believe are the three most important relevant practical factors in MU-MIMO systems, namely 1) MU-MIMO channel modeling and antenna configurations, 2) pilot-based channel estimation, and 3) link adaption. The article starts with system aspects of MU-MIMO systems and demonstrates the importance of antenna configuration in MU-MIMO systems. Then the channel estimation procedure in MU-MIMO practical systems is explained in detail, with focus given to channel estimation based on user-specific reference signals. Finally, the article discusses practical issues of link adaptation and evaluates the performance under different antenna configurations.

System Model

Both single- and multilayer MU-MIMO transmissions in LTE-Advanced frequency division duplex (FDD) systems with multi-cell downlink scenario are considered in this article. It is assumed that all eNodeBs are equipped with N_T transmit antennas and the target UEs are equipped with N_R receive antennas. The received subcarrier-specific signal vector at the target UE can be represented as

$$\underline{r} = \underline{G}_{\mathrm{I}}\underline{d}_{\mathrm{T}} + \underline{G}_{\mathrm{I}}\underline{d}_{\mathrm{I}} + \sum_{i=1}^{N_{\mathrm{I-Cell}}} \underline{G}_{\mathrm{C},i}\,\underline{d}_{\mathrm{C},i} + \underline{n}$$

$$= \sum_{i=1}^{N_{\mathrm{L}}} \underline{g}_{\mathrm{T},i} d_{\mathrm{T}} + \sum_{i=1}^{N_{\mathrm{I}}} \underline{g}_{\mathrm{I},i} d_{\mathrm{I},i} + \sum_{i=1}^{N_{\mathrm{I-Cell}}} \underline{G}_{\mathrm{C},i}\,\underline{d}_{\mathrm{C},i} + \underline{n}, \qquad (1)$$

where $\underline{G}_{\mathrm{T}}$ and $\underline{G}_{\mathrm{I}}$ are the effective and intra-cell interference channel matrices between the serving eNodeB and the target UE with physical channel matrix H_{T} and the corresponding applied precoding matrices P_{T} and P_{I} respectively. Both P_{T} and P_{I} are generated at eNodeB based on the UE feedback channel information. $G_{\mathrm{C,i}}$ is the channel matrix for the inter-cell interference from the ith interference cell. N_{L}, N_{I}, and $N_{\mathrm{I\text{-}Cell}}$ are the total number of layers to the target UE, layers to the intra-cell interference UE, and the total

number of interference cells, respectively. The signal vectors for the target UE and interferences are represented as $\underline{d}_1$, $\underline{d}_I$, and $\underline{d}_{C,i}$ and they are mutually independent with unit transmission energy per symbol. The AWGN vector is denoted by $\underline{n}$.

MU-MIMO System Aspects

Despite theoretically shown advantages MU-MIMO has been slow in getting adapted in 4G systems. The main reasons for this may be attributed to the lack of dedicated symbols to support multiuser MIMO channel estimation and signal detection, the lack of an accurate multiuser MIMO channel model, and even regional differences in antenna configuration. This section discusses key characteristics of channel model and antenna polarization and their importance for multiuser MIMO systems.

MU-MIMO Channel Model

"…proper correlation modeling is critical for MU-MIMO performance assessment."

As shown by Duplicy et al.[2], proper correlation modeling is critical for MU-MIMO performance assessment. Considering the downlink, there are multiple transmit antennas at the base station site and there is a certain number of users that are served by base station in parallel, using the same physical resource blocks of the time-frequency channel. All these users are equipped with single or multiple antennas. Each user can be characterized by a unique instantaneous channel matrix. The entries of this matrix describe the channel state between a given antenna at the base station and an antenna of that user device, namely $H_T(m,b)$ for a particular UE at discrete time instant m and on subcarrier b contains complex entries $h_{t,r}(m,b)$ that describe the channel effect between transmit antenna t and receive antenna r for that user. Note that $H_T(m,b)$ is the extended notion for the channel matrix H_T applied in the context of (1); in particular in (1) the variability of the channel in time and frequency dimensions was omitted for the sake of less complicated notations.

There are numerous models available to describe the MU-MIMO channel as shown by Duplicy et al.[2] and references therein.

The most interesting question in modeling the MU-MIMO channel is the correlation among channel matrices of different users, because low correlation allows spatially multiplexed transmission to multiple customers and the handling of interference among the transmissions to different users. The optimum transmission and reception would require all the complex entries of all channel matrices to be known instantaneously and for all subcarriers. The availability of this complete channel state information would allow optimum precoding and scheduling at the transmitter side and optimized reception at the receiver side.

"…spatial correlation of channels of different UEs determines the capacity achievable by the applied MU-MIMO scheme."

In practice, spatial correlation of channels of different UEs determines the capacity achievable by the applied MU-MIMO scheme. Low correlation allows the achievement of highest sum rate capacity (assuming proper precoding), while high correlation results in unresolvable intra-cell interference and hence low performance. Accordingly, the MU-MIMO scheduler should try

to allocate the same PRBs (Physical Resource Blocks) to UEs with the least possible spatial correlation over those resources.

Theoretical parametric models of MIMO channels (for example 3GPP SCM Spatial Correlation Model [TR25.996]) are usually based on modeling scatterers or clusters of scatterers around receivers and calculating sums of reflected planar waves. Basically the same methodology is often applied for MU-MIMO channel modeling. Based on these models nonparametric models are also derived, where transmitter-and receiver-side correlations can be estimated and explicitly used.

Based on modeling and available measurement results, some conclusions can be drawn (which are mainly intuitive as well): MU-MIMO channels are often modeled as being independent for different users; however, in scenarios where UEs are close to each other, this assumption is not valid and transmit correlation can be significant. Similarly, if propagation is such that the line of sight component for different users is dominant, correlation between their channels will be very high. Antenna configuration and polarizations as well as polarization errors after reflections also affect correlations, depending on the models applied.

"MU-MIMO channels are often modeled as being independent for different users…"

Currently the standard does not have a reporting mechanism that would enable the exact estimation of correlation between the channels of UE pairs. Hence eNodeB should chose UE pairing for MU transmission based on PMI (Precoding Matrix Index), RI (Rank Indicator), and CQI (Channel Quality Indicator) reporting. One traditional approach is to select those UEs for a given PRB that reported good CQI values. In case of MU-MIMO transmission this could lead to impairment because similar CQI values may also indicate high LoS (Line of Sight) factor or proximity of UEs.

Antenna Configuration

Current LTE networks offer different antenna deployment and configuration settings that is mainly region specific. For example, uniform linear array (ULA) is mainly deployed in China and dual polarization with cross-polarization discrimination (XPD) is mainly deployed in Europe. However, the choice of antenna configuration and polarization plays an important role in design of MU-MIMO systems. Unlike SU-MIMO, the best performance with MU-MIMO is achieved with spatially correlated antennas (if the precoding vectors are orthogonal) requiring only slowly varying phase slope precoding to steer beams to different users. Furthermore, the amount of separation between antennas will have different impacts on the potential MU-MIMO gains. Realizing these gains puts conflicting demands on the antenna separation and different choices of antenna separation will result in different system performance profiles.

"…the choice of antenna configuration and polarization plays an important role in design of MU-MIMO systems."

"…the best performance with MU-MIMO is achieved with spatially correlated antennas requiring only slowly varying phase slope precoding to steer beams to different users."

Currently, LTE-Advanced defines up to 8x8 MU-MIMO systems with a maximum of four UEs scheduled for transmission and maximum of two layers allocated per UE but no more than four layers in total. With single-pole ULA configuration, the antenna ports are highly correlated and hence the probability of rank-2 transmission per UE is low. Theoretically each such single-pole array can be used to form a beam and direct it to a particular UE

since with such correlated antenna ports, the eNodeB can spatially separate different users based on either long term statistics like AoA (Angle of Arrival) or azimuth properties of the precoding matrices. For example in a single-pole 8-transmit antenna system up to eight UEs can be scheduled simultaneously. Scheduling more than four UEs can result in complicated scheduling and hence single-layer MU-MIMO with up to four UEs would be more common.

With dual polarized deployments the eNodeB can use each polarization array to direct a single- or dual-layered data stream to a particular UE. Scheduling more than two UEs simultaneously would cause high inter-user interference and hence either single- or dual-layered MU-MIMO transmission with two UEs would be more common and beneficial in such systems.[3]

MU-MIMO Receiver Aspects

"The main issue of signal detection for MU-MIMO systems is the presence of different interference sources, including co-layer intra-cell and inter-cell interferences."

The main issue of signal detection for MU-MIMO systems is the presence of different interference sources, including co-layer (from dual-layer transmission), intra-cell (from MU-MIMO transmission), and inter-cell interferences. Similar to the study in Release 8 MU-MIMO transmission[1], interference-aware receivers are required for both single- and multi-layer MU-MIMO to eliminate the residual spatial interference and to improve data detections. Typical receivers that can be utilized for MU-MIMO detection are interference rejection combiner (IRC), minimum mean square error (MMSE), or fast maximum likelihood (Fast-ML) receivers.[4] Since there is no explicit signaling in Release 10 LTE-Advanced systems to indicate whether single-layer or dual-layer MU-MIMO transmission occurs, the challenge for MU-MIMO is to construct the knowledge of interference and to eliminate them by the interference-aware receivers. Conventional methods of constructing interference knowledge are to detect the interference covariance matrix, which can be used by IRC, MMSE, and Fast-ML receivers. Interference covariance matrix estimation schemes have been investigated by Thiele et al.[5] Different procedures of the covariance estimation, namely estimations based on reference signal (RS) based and user data, have been summarized in 3GPP R4-115213[6] and the references therein. Ohwatari et al.[7] have presented the performance of using different estimations in LTE systems in a multi-cell interference scenario. Interference estimation for multilayer multiuser multiple-input and multiple-output (ML-MU-MIMO) transmission for LTE-Advanced has been investigated by Bai et al.[4]

Interference Aware Receivers

"…a crucial component in design of MU-MIMO systems is a detector to eliminate the residual spatial interference in order to avoid the detection error floor."

The main issue of signal detections in MU-MIMO is the presence of different interference sources, including co-layer (from dual-layer transmission), intra-cell (from MU-MIMO transmission) and inter-cell interferences, which cannot be handled by a conventional, interference-unaware receiver. Therefore, a crucial component in design of MU-MIMO systems is a detector to eliminate the residual spatial interference in order to avoid the detection error floor. The simplest interference-aware receiver is the interference rejection combiner (IRC) that utilizes the knowledge on the covariance matrix of the total interference plus noise and pre-whitens the received signals being followed by a

maximum ratio combiner (MRC).[8] The ideal IRC receiver in multilayer MU-MIMO transmission in Equation 1 is presented by:

$$\underline{M} = \underline{G}_{\mathrm{T}}^{\mathrm{H}} \underline{R}_{\mathrm{II}}^{-1},$$

with

$$\underline{R}_{\mathrm{II}} = \underline{G}_{\mathrm{I}} \underline{G}_{\mathrm{I}}^{\mathrm{H}} + \sum_{i=1}^{N_{\mathrm{i-Cell}}} \underline{G}_{\mathrm{C},i} \underline{G}_{\mathrm{C},i}^{\mathrm{H}} + N_0 \mathrm{I}.$$

The IRC maximizes the post signal to noise plus interference ratio (SINR) and outperforms the classic MRC as demonstrated in Figure 1. As can be seen from the results, both types of receivers show similar performance for dual XPD polarization, but for ULA configuration the IRC outperforms the MRC receiver significantly. Obviously, the IRC receiver is a simple solution to overcome the limitation due to different antenna polarizations in MU-MIMO systems.

"The IRC maximizes the post signal to noise plus interference ratio (SINR) and outperforms the classic MRC…"

"…IRC receiver is a simple solution to overcome the limitation due to different antenna polarizations in MU-MIMO systems."

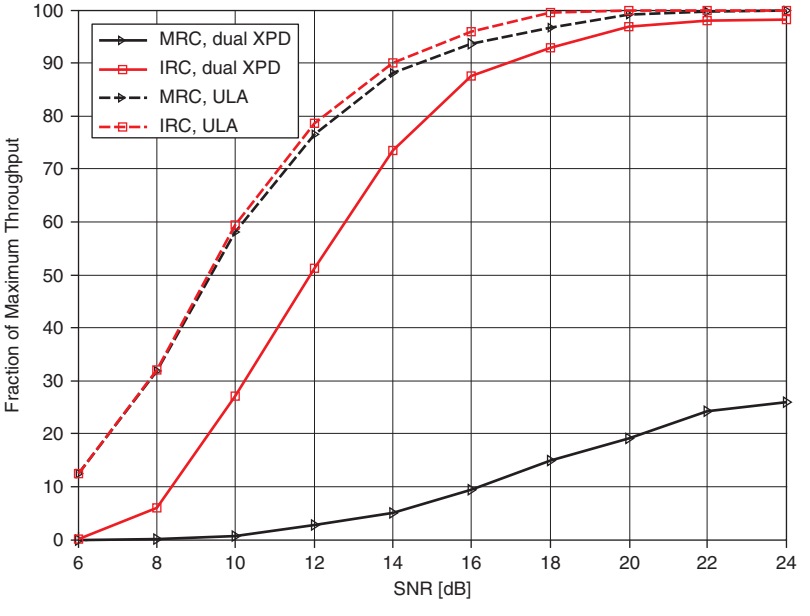


Figure 1: Receiver performance in different antenna polarizations in highly correlated channels for 4x2 MU-MIMO (four users), 10 MHz bandwidth and 64 QAM, C/R = 1/2.
(Source: Intel Corporation, 2013)

Interference-aware receivers (referred to as *advanced receivers*) have been discussed widely in 3GPP Release 10 as well as in research communities. The main focus of the advanced receiver study in Release 12 is to enhance current advanced receivers with improved mitigation of intra- and inter-cell interference and user experience. Furthermore, advanced receiver enhancements for heterogeneous deployments using even larger cell-selection offsets have become an important study item in LTE Release 12.[8]

"…noise covariance estimation in MU-MIMO systems can be either based on cell reference signals (CRS) or UE reference signals (DMRS)."

"…the user-specific RS-based covariance estimation scheme performs significantly better and provides more complete and accurate covariance than the CRS scheme."

Channel Estimation in MU-MIMO

As discussed by Bai et al.[4], noise covariance estimation in MU-MIMO systems can be either based on cell reference signals (CRS) or UE reference signals (DMRS). DMRS are orthogonal for layers through an orthogonal code but only quasi-orthogonal between users through the scrambling identity. The key idea of CRS-based estimation is to use the CRS subcarriers to estimate the covariance matrix of inter-cell interference and noise. Due to the spatially orthogonal CRS design, the estimation of the covariance matrix including intra-cell multiuser interference is difficult in this scheme and additional precoding matrix estimation is required to obtain the real interference channels. In contrast to the CRS-based covariance matrix estimation, the DMRS-based covariance matrix estimation includes an intra-cell MU interference source together with inter-cell interference and noise. This estimation scheme also avoids the mismatch of the target channel matrix and residual cross-covariance coefficients between target signals and interference signals.

Figure 2 illustrates the performance difference between CRS- and DMRS-based channel estimation applying the IRC receiver. The results are shown for MU-MIMO transmission with two interference layers in a low spatial correlation scenario and for three different MCS, QPSK with 1/3 code rate, 16-QAM with 1/2 code rate and 64-QAM with 1/2 code rate. From the simulation results it can be seen that the CRS covariance estimation scheme faces strong performance degradation and error floor. With 16-QAM and 64-QAM MCS, no more than 30 percent throughput can be achieved. In contrast, the user-specific RS-based covariance estimation scheme performs significantly better and provides more complete and accurate covariance than the CRS scheme.

DMRS-based reference estimation is explained with more details in the next section.

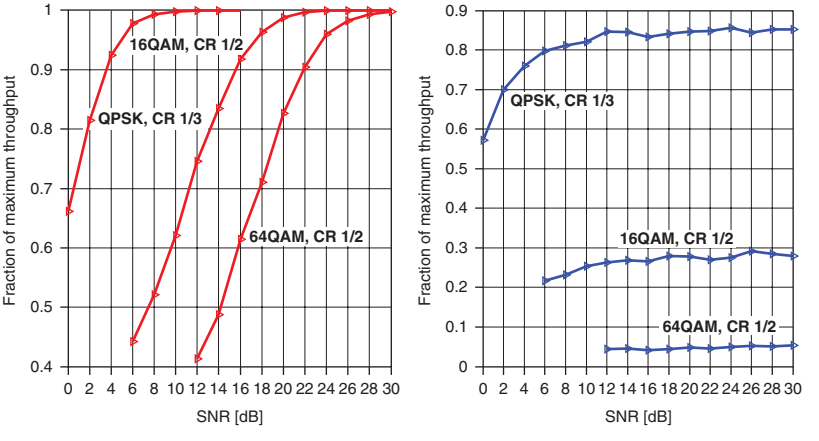


Figure 2: IRC performance for DMRS-based (left) and CRS-based (right) channel estimation in dual-layer 8x4 MU-MIMO with two interfering UEs, I/N = 10 dB, EPA5.
(Source: Intel Corporation, 2013)

UE Reference Signal–based Channel Estimation

Channel estimation for MU-MIMO transmission in LTE (Release 8 and 9) and LTE Advanced (Release 10 and 11) utilizes UE-specific reference symbols for the physical downlink shared channel (PDSCH), where the reference symbols corresponding to each of the two or four layers are distinguished from each other by a sign pattern. This is required for the separation of different transmission layers: for normal cyclic prefix, this is performed in the time interpolation (see the reference symbol pattern in Figure 3) since the same number of reference symbols with the plus and with the minus sign, that is, an even number of reference symbols, is needed. Therefore, for extended cyclic prefix, channel estimation can already be implemented in the frequency filtering (see the reference symbol pattern in Figure 4).

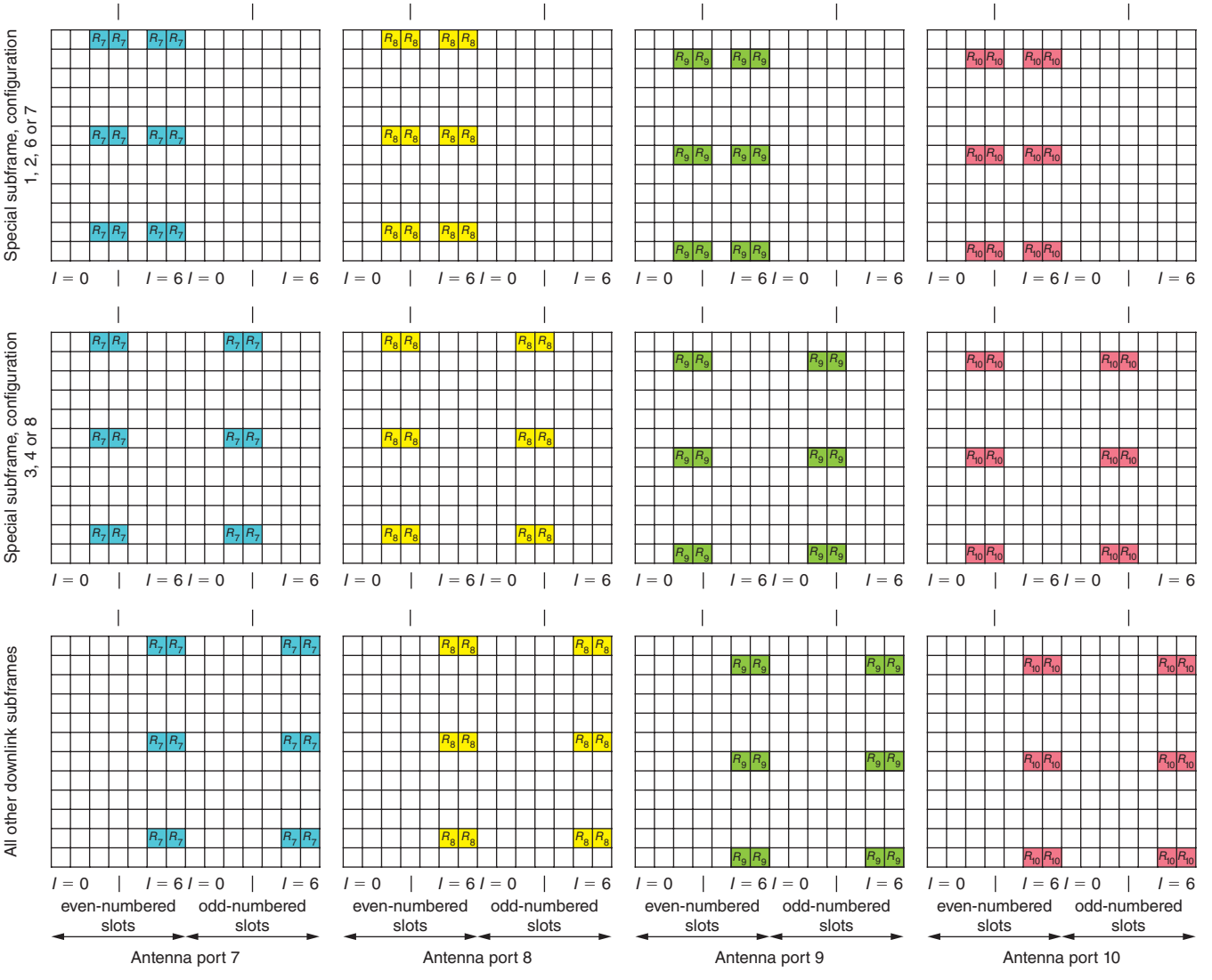


Figure 3: Mapping of UE-specific reference symbols, antenna ports 7, 8, 9, and 10 (normal cyclic prefix). (Source: 3GPP TS 136 211, v.10.7.0, 2013)

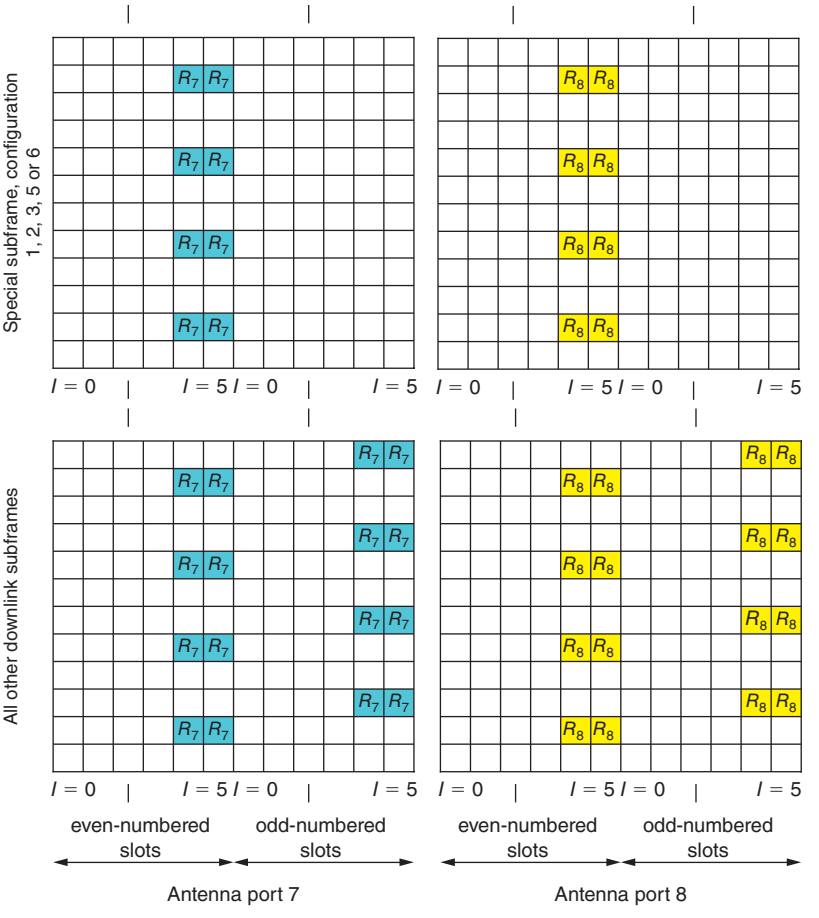


Figure 4: Mapping of UE-specific reference signals, antenna ports 7 and 8 (extended cyclic prefix).
(Source: 3GPP TS 136 211, v.10.7.0, 2013)

In general, there are four stages in the channel estimation. In the first stage, the reference symbols of each layer are demodulated, so the sign patterns of the antenna ports are observed at the output of the reference symbol demodulator. For normal cyclic prefix, the second stage of the channel estimation consists of a single phase frequency filter, which yields the frequency-filtered reference symbols at its output. The separation of the different layers is done in the third stage of the channel estimator, the time interpolation, as four OFDM symbols with reference symbols on the carriers with the frequency index $k = 1$, $k = 6$, and $k = 11$ are available. In this case the Wiener filter in time direction corresponds to an autocorrelation matrix, which equals the sum of the autocorrelation matrices corresponding to all layers. In the final stage of the channel estimation, the frequency interpolation yields the channel estimates for the carriers without time-filtered reference symbols.

For extended cyclic prefix (Figure 4), the second stage of the channel estimation consists of a two-phase frequency filter. This is required since the reference symbols of the even and of the odd slots are shifted by one subcarrier. Because four subcarriers with reference symbols per resource block are available, the

separation of the two layers is done already by the frequency filtering. The Wiener filter in frequency direction corresponds to an autocorrelation vector and an autocorrelation matrix, which equals the sum of the autocorrelation matrices corresponding to antenna port 7 and antenna port 8. The third and the fourth steps of the channel estimator are the time interpolation, which yields for any subcarrier with reference symbols the channel estimates for all 12 OFDM symbols and the frequency interpolation, which yields the channel estimates for the subcarriers without reference symbols.

Link Adaption Challenges in MU-MIMO

In order to keep low feedback overhead, 3GPP has agreed to use the same codebook optimized for single-user MIMO (SU-MIMO) in MU-MIMO.[1] Therefore, the selection of PMI and CQI for MU-MIMO is limited to SU-MIMO feedback. This is a significant drawback of MU-MIMO systems as such a suboptimal PMI selection and CQI reporting limit the gain of multiuser transmission. Failing to accurately estimate MU-MIMO CQI can lead to the wrong packet scheduling and link adaption and thus overall performance degradation in MU-MIMO. Bai et al.[9] investigate and refer to various MU-MIMO CQI prediction schemes for LTE systems with different releases.

"…the selection of PMI and CQI for MU-MIMO is limited to SU-MIMO feedback."

"Failing to accurately estimate MU-MIMO CQI can lead to the wrong packet scheduling and link adaption and thus overall performance degradation in MU-MIMO."

In this section we demonstrate the impact of different antenna polarization on MU-MIMO link adaption. Link adaptation is the process of selecting modulation and coding scheme (MCS) in a wireless link to maximize throughput while meeting reliability constraints like Block Error Rate or Packet/Frame Error Rate. In LTE and LTE-Advanced systems, MSC is selected based on feedback between UE and eNodeB, that is, CQI/PMI/RI feedback. The problem in MU-MIMO systems is that the link adaption is optimized for SU-MIMO, which creates a mismatch between the estimated MU-MIMO CQI and the true channel CQI leading to significant degradation of the system performance. Furthermore, the mismatch in the estimated MU-MIMO CQI could lead to a wrong MU-MIMO pairing decision as well as incorrect assignment of the MCS.[2][9] CQI adaption schemes optimized for MU-MIMO have been widely investigated in 3GPP as well as in research communities. Badic et al.[12] present a comprehensive overview.

As explained in the earlier section "Antenna Configuration" and demonstrated in Figure 1, antenna configuration has significant impact on MU-MIMO performance. Therefore, this section evaluates impact of different antenna configurations on link adaption in a Figure 1: Receiver performance in different antenna polarizations MU-MIMO system with four UEs scheduled simultaneously. Single-layer MU-MIMO has been evaluated but similar observations have been made for higher layer MU-MIMO as well. The IRC receiver with DMRS-based noise covariance estimation has been applied for 10 MHz system bandwidth. The results have been obtained for wideband feedback (PMI/CQI) and users are paired by Orthogonal Precoding Vector Pairing (OPVP), that is, the UE selects the best PMI corresponding to the precoding vector from the codebook by maximizing the received signal power.[7] MU-MIMO CQI prediction based on rank adaption[10] is applied in this evaluation.

As shown in Figure 5, MU-MIMO throughput in ULA outperforms dual-polarized antenna systems since the residual inter-user interference is less in ULA array systems. Additionally inter-user interference in a correlated ULA-arrays system is more predictable and hence MU-CQI prediction from a given SU-CQI is more accurate. This results in better link adaptation and hence better throughput.

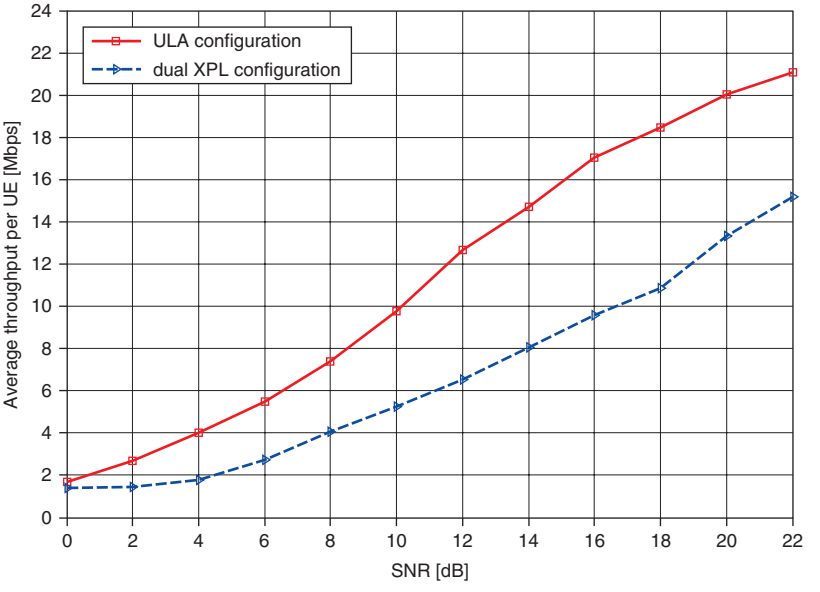


Figure 5: 4x2 MU-MIMO (four users) link adaption performance for various antenna configurations in high correlated channels and 10 MHz system bandwidth.
(Source: Intel Corporation, 2013)

LTE Release 12 pays special attention to improving CSI feedback for MU-MIMO in realistic scenarios. Based on feedback from network operators the following CSI feedback enhancements are planned[13]:

- Enhanced PMI feedback for 4Tx with cross-polarized arrays. The enhancement could be possibly similar to two staged Release 10 8Tx codebooks or finer granularity codebooks
- CQI-related enhancements for MU-MIMO for better link adaptation

Conclusion

Considerations of real-world effects are presented in the area of channel modeling and related channel estimation and reporting schemes and their drawbacks in 3GPP standardization. Results are shown to prove that the use of UE-specific reference signals for channel estimation provides significant gains over cell-specific reference signals. Practical questions of receiver algorithms and link adaptation are also considered in this article. It has been

shown that advanced interference-aware receivers show significant potential in overcoming the burdens of cross-polarized antenna configurations, which is the typical deployment scenario in Europe. As the residual inter-user interference is smaller and more predictable in a ULA array system, MU-MIMO link adaption based on standard defined low overhead SU-MIMO feedback becomes more accurate and reliable while increasing MU-MIMO throughput.

Acknowledgments

The authors wish to gratefully acknowledge the support of their colleagues at Intel Mobile Communications and at the Department of Communication Technologies of the University of Duisburg-Essen.

References

[1] 3GPP 36.201, "Evolved Universal Terrestrial Radio Access (E-UTRA); LTE physical layer; General description," 3GPP, Sophia Antipolis, Technical Specification 36.201 v8.3.0, March 2009.

[2] Duplicy, J., B. Badic, R. Balraj, R. Ghaffar, P. Horváth, F. Kaltenberger, R. Knopp, I. Z. Kovács, H. T. Nguyen, D. Tandur, and G. Vivier, "MU-MIMO in LTE Systems," *EURASIP Journal on Wireless Communications and Networking,* vol. 2011, Article ID 496763, 13 pages, 2011. doi:10.1155/2011/496763.

[3] Lim, C., T. Yoo, B. Clerck, B. Lee, and B. Shim, "Recent Trend of Multiuser MIMO in LTE-Advanced," *IEEE Comm. Magazine*, vol. 51, no. 3, pp. 127–135, March 2013.

[4] Bai, Z., B. Badic, S. Iwelski, R. Balraj, T. Scholand, G. Bruck, and P. Jung, "Interference Estimation for Multi-Layer MU-MIMO Transmission in LTE-Advanced Systems," IEEE PIMRC 2012, Sidney, Australia, Sep. 2012.

[5] Thiele, L., M. Schellmann, T. Wirth, and V. Jungnickel, "On the value of synchronous downlink MIMO-OFDMA systems with linear equalizers," in *Wireless Communication Systems*. 2008. ISWCS '08. IEEE International Symposium on, Oct. 2008, pp. 428–432.

[6] 3GPP R4-115213 "Reference receiver structure for interference mitigation on enhanced performance requirement for LTE UE," NTT DOCOMO, Approval R4–115213, Oct. 2011.

[7] Ohwatari, Y., N. Miki, T. Asai, T. Abe, and H. Taoka, "Performance of advanced receiver employing interference rejection combining to suppress inter-cell interference in

LTE-Advanced downlink," in Vehicular Technology Conference (VTC Fall), 2011 IEEE, Sep. 2011, pp. 1–7.

[8] http://www.3gpp.org/Release-12

[9] Bai, Z., B. Badic, S. Iwelski, T. Scholand, R. Balraj, G. H. Bruck, and P. Jung, "On the receiver performance in MU-MIMO transmission in LTE," in *Proceedings of the Seventh International Conference on Wireless and Mobile Communications*, ICWMC 2011, Jun. 2011, pp. 128–133.

[10] Badic, B., R. Balraj, T. Scholand, Z. Bai, and S. Iwelski, "Impact of Feedback and User Pairing Schemes on Receiver Performance in MU-MIMO LTE Systems," IEEE Wireless Communications and Networking Conference (WCNC 2012), Paris, France, April 2012

[11] 3GPP Technical Specification Group Radio Access Network: Evolved Universal Terrestrial Radio Access (E UTRA); "Physical Layer Procedures (Rel-8)" 3GPP TS36.213 V8.7.0, 2009.

[12] Badic, B., R. Balraj, T. Scholand, Z. Bai, and S. Iwelski, "Analysis of CQI prediction for MU-MIMO in LTE systems," IEEE Vehicular Technology Conference (VTC2012-Spring), Yokohama, Japan 2012.

[13] 3GPP Technical Specification Group Radio Access Network: "Evolved Universal Terrestrial Radio Access (E-UTRA); Downlink Multiple Input Multiple Output (MIMO) enhancement for LTE-Advanced," TR 36.871 2013.

Author Biographies

Biljana Badic is a system engineer in Intel's Product Engineering 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

Ingmar Groh is a system engineer in Intel's Product Engineering Group, has been with Intel since August 2011, and has participated in development of 4G wireless modems with emphasis on synchronization, channel estimation, and equalization. He is also actively involved in Intel research activities on 4G and beyond systems. Email: ingmar.groh@intel.com

Rajarajan Balraj is a system engineer in Intel's Product Engineering Group and has been with Intel since 2010. He has made key contributions to various interference cancellation and mitigation algorithms on Intel's 3G and 4G

wireless modems. Rajarajan Balraj has filed over 25 patents related to LTE/LTE-A and UMTS. Email: rajarajan.balraj@intel.com

Peter Fazekas is a researcher at Budapest University of Technology and Economics (BME), Department of Networked Systems and Services. He is with BME since 2001 and has been involved in industrial, national, and EU research activities in various topics of 4G and Beyond 4G wireless interfaces. His lecturing activities cover mobile systems, network planning, and electronics. Email: fazekasp@hit.bme.hu

IMPLEMENTATION CHALLENGES FACING USER EQUIPMENT IN COORDINATED LTE NETWORKS

Contributors

Michael Hosemann
Intel Mobile Communications

Daejung Yoon
Intel Mobile Communications

Biljana Badic
Product Engineering Group,
Intel Corporation

Rajarajan Balraj
Product Engineering Group,
Intel Corporation

Constantinos B. Papadias
Athens Information Technology

Communications systems are often defined by their peak data rate. Such peak rates are rarely available to the individual user. Recently, user experience at the cell edge has come into focus. In order to break the system limitation due to the co-channel interference in cellular networks and to improve the cell-edge user experience, Coordinated Multipoint (CoMP) techniques have been developed and standardized for LTE Advanced systems. Recent research efforts in CoMP networks yielded both satisfactory system efficiency improvements while highlighting the significant gains in the order of high double-digit percentages for cell-edge user throughput. As CoMP operates mostly in a closed-loop transmission fashion, the development of CoMP raises new challenges in the User Equipment (UE) implementation. The aim of this article is to demonstrate benefits of CoMP techniques and to discuss their practical implementation challenges in current 4G networks. Different CoMP scenarios and their challenges for a UE are explained in detail. Furthermore, the article describes potential CoMP solutions that might be supported in beyond LTE-Advanced networks. Performance benefits are evaluated by Monte Carlo simulations in different CoMP scenarios.

Introduction

One of the key technologies introduced in LTE Release 11 to improve coverage, cell-edge user throughput and spectral efficiency is Coordinated Multipoint (CoMP) transmission. The basic concept behind CoMP is the cooperative multiple-input and multiple-output (MIMO) transmission, where geographically distributed transmitters and receivers work jointly in order to enhance the received signal quality and to decrease the received spatial interference. If the UE in the cell-edge region operates in the CoMP mode, it may be able to receive signals from multiple cell sites and its transmission may be received at multiple cell sites regardless. By coordinating the signaling transmitted from the multiple cell sites, the downlink (DL) performance can be improved significantly. A wide variety of coordination schemes has been described. For LTE Release 11 four different scenarios for deploying CoMP techniques have been considered by the standardization committees.[1] For Release 12 and beyond more advanced techniques are currently considered.[2]

"Coordination between geographically separated cell sites can improve cell-edge performance significantly."

In this article we first explain the motivation for CoMP and then describe coordination schemes and deployment scenarios as they will be used in LTE Release 11 in the next two sections, respectively. We then present challenges for UEs in Release 11 in the section "Design Challenges in LTE Release 11 CoMP Networks" before considering further ideas and their challenges for Release 12 and beyond in the section "Further Developments of CoMP: Release 12 and Beyond.

Motivation for Coordinated Networks

Demand for mobile broadband communications will increase and diversify dramatically in the following 5 to 10 years. Already, the market success of smartphones and broadband-ready portable devices has generated exponential mobile data traffic growth, and new applications are continuously raising expectations for higher data rates and increased quality of service in the existing mobile networks. These large capacity demands can be met only by highly efficient, optimized, and coordinated network infrastructures. Therefore, 3GPP as well as other research communities started to look into the potential benefits of coordinated multipoint transmission and later on 3GPP introduced CoMP in LTE specifications for Release 11 as one of the key transmission techniques. The main motivation for CoMP is to improve cell-edge user throughput and the average system throughput. This has been demonstrated in literature.

Coordination Schemes and Deployment Scenarios

Various coordination schemes and deployment scenarios for coordination schemes have been described in literature and by the 3GPP standardization bodies. An overview is given in Figure 1.

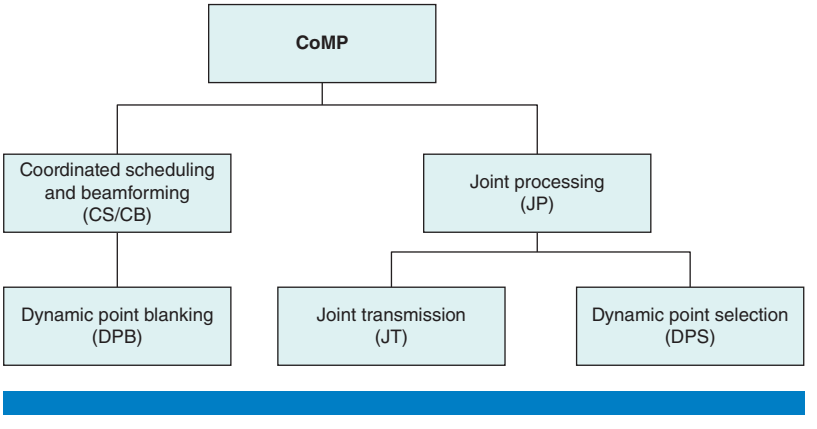


Figure 1: Concept diagram of CoMP schemes (Source: Intel Corporation, 2013)

Coordination schemes are classified into schemes performing

- coordinated scheduling and beamforming
- joint processing

Coordinated scheduling (CS) and coordinated beamforming (CB) are applied between cells to reduce the interference caused to other cells. CS only turns off transmission at an interfering cell to create a more favorable signal-to-interference ratio (SIR) for the transmission of another cell to a UE. In the context of LTE Release 11 this is also called Dynamic Point Blanking (DPB).

Coordinated beamforming goes further by exploiting the spatial locations of UEs in their cells. Transmissions in the cells are scheduled simultaneously and beamforming is applied for each UE in its respective cell. Each cell selects a beamforming pattern so that the signal at its scheduled UE is maximized

"Coordinated scheduling (CS) and coordinated beamforming (CB) are applied between cells to reduce the interference caused to other cells. CS only turns off transmission at an interfering cell to create a more favorable signal-to-interference ratio (SIR)…"

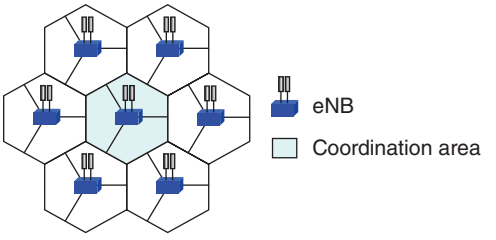


Figure 2: Scenario 1 – Homogeneous network with intra-site CoMP (Source: 3GPP TR 36.819, 2013)

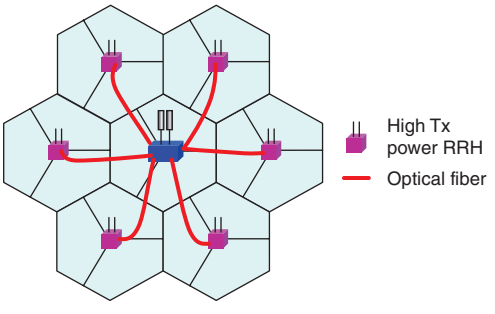


Figure 3: Scenario 2 – Homogeneous network with high TX power RRHs (Source: 3GPP TR 36.819, 2013)

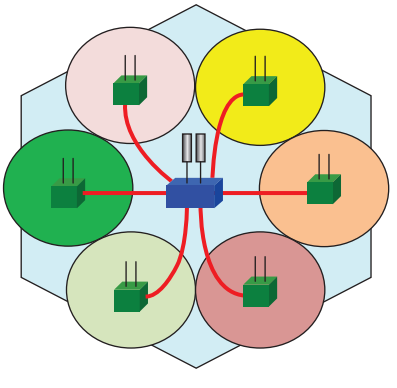


Figure 4: Scenario 3 – Heterogeneous network with low power RRHs within a macro cell (cells have the different cell IDs) (Source: 3GPP TR 36.819, 2013)

while minimizing the interference to the UE scheduled at the same time in the neighboring cell. Selecting the beamforming patterns requires frequent measurements by the UE to select the best pattern.

In Joint Processing schemes the signals are transmitted from multiple transmission points at the same time and have to be combined in the UE. A somewhat simplified scheme is dynamic point selection. Here the signals are transmitted from one of multiple transmission points. For each subframe, the transmission point is selected a new depending on feedback from the UE and scheduling requirements. This can also be combined with Dynamic Point Blanking. In this case one TP transmits while the others are blanked.

The JT scheme can be further divided into coherent-JT and non-coherent JT. For coherent-JT, a CoMP UE is required to measure specific Inter-point (inter CSI-RS) phase feedback or to report aggregated Precoding Matrix Indicator (PMI) information. Coherent-JT is not included in Release 11 but 3GPP RAN1 has introduced phase information feedback in Release 12. This will be discussed the section "Further Developments of CoMP: Release 12 and Beyond."

The 3GPP CoMP specification[3] stipulates four scenarios concerning cell deployments. The following scenarios were selected for the evaluation of DL CoMP:

- *Scenario 1:* Homogeneous network with intra-site CoMP illustrated in Figure 2.
- *Scenario 2:* Homogeneous network with high TX power remote radio heads (RRHs) illustrated in Figure 3.
- *Scenario 3:* Heterogeneous network with low power RRHs within the macrocell coverage where the transmission/reception points created by the RRHs have different cell IDs as the macro cell, illustrated in Figure 4.
- *Scenario 4:* Heterogeneous network with low power RRHs within the macrocell coverage where the transmission/reception points created by the RRHs have the same cell IDs as the macro cell, illustrated in Figure 5.

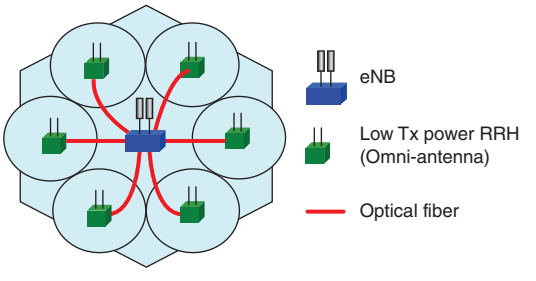


Figure 5: Scenario 4 – Heterogeneous network with low power RRHs within a macrocell (cells have the same cell IDs) (Source: 3GPP TR 36.819, 2013)

The RRHs in scenarios 2 through 4 can be thought of as small cells that are controlled (in a varying degree) by a macro cell. The possible cognitive spectrum usage of some of these small cells (when they don't belong to the same operator as the macro cell), opens up further challenges regarding their coordination—see, for example, the recently initiated project ADEL (EU FP7 project #619647).

"Another important factor of CoMP network is backhaul support between cells and or transmission points."

Another important factor of CoMP network is backhaul support between cells and or transmission points. There are two basic types of backhaul support linking RRHs and a base station eNB.

- TPs in a CoMP network can be connected via fiber-optic cable allowing for a nearly perfectly synchronized network (for example, using the CPRI or OBSAI protocols).
- TPs in a CoMP network can be connected via microwave or cable links introducing latency and time and frequency offsets in an asynchronous network.

Even though practical connection types have been considered for all deployment scenarios, rather generous frequency offsets between TPs have been specified in test cases.

For test cases up to 200 Hz frequency offset and 2 microseconds timing offset between signals arriving from different TPs have been specified. It should be noted that these frequency offsets are specified even for tightly synchronized TPs. For an assumed carrier frequency of 1 GHz the offsets are about the same as could be expected from two unsynchronized small cells. The benefits of tight coupling through fiber-optic links still need to be explored for CoMP networks.

System Model

Figure 6 shows a typical CoMP scenario where Dynamic Point Switching is employed. We have chosen this scenario because it best illustrates the challenges that CoMP poses for the UE.

One transmission point (TP1) always transmits the Physical Downlink Control Channel (PDCCH) and the associated Cell-Specific Reference Symbols. The transmission of the Physical Downlink Shared Channel (PDSCH) can be done from up to three other transmission points. With the PDSCH, the TP will also transmit the UE-specific reference symbols (DM-RS) and reference symbol for calculating channel state information (CSI-RS). While DM-RS are transmitted in every subframe, CSI-RS are transmitted at rather sparsely. Transmission intervals of up to 80 subframes can be selected by the network. While this might seem very infrequent it should be considered that CoMP is meant to improve performance in rather locally constrained, low-speed environments. For test cases, Doppler shifts of no more than 5 Hz have been specified.

"Differences in reference frequency and Doppler shift will lead to frequency offset differences. Differences in propagation delay will cause differences in time of arrival."

In addition, differences in propagation need to be considered. Differences in propagation delay can lead to Doppler shift which will add to frequency offset differences and propagation delay will cause differences in time of arrival.

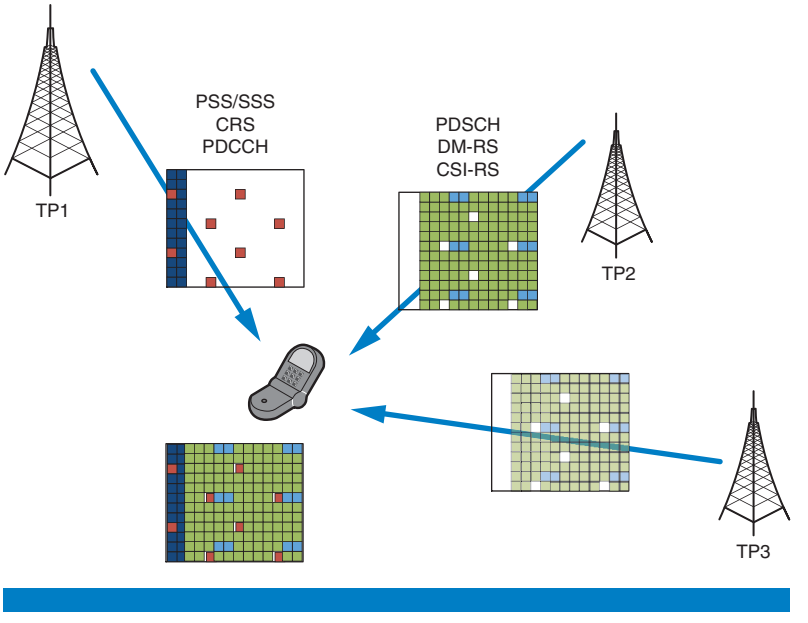


Figure 6: Combining of channels from multiple transmission points (Source: Intel Corporation, 2013)

Design Challenges in LTE Release 11 CoMP Networks

In this section we are outlining some of the design challenges posed by the CoMP schemes in LTE Release 11 networks. We first consider downlink and uplink processing before also discussing challenges posed by the combination of CoMP with other features of LTE Release 11.

"…the UE needs to be able to combine signals that have been transmitted from various transmission points and demodulate them successfully."

Downlink CoMP

As described in the previous sections, the UE needs to be able to combine signals that have been transmitted from various transmission points and demodulate them successfully. The main design challenge for the UE is the combination of resource elements that have been received with different frequency and timing offsets and channel parameters. Three main steps need performing to achieve this:

- The individual frequency and timing offsets and parameters need to be estimated utilizing the appropriate reference resources.
- The individual frequency and timing offsets need to be corrected based on the estimates.
- Channel characteristics for the channel from each transmission point first need estimating and have to be applied to the corresponding physical channels subsequently.

All tasks pose their individual challenges: estimation for the PDSCH needs to be based largely on DM-RS. These can only be used in the allocated bandwidth of the PDSCH. If this is small, very few reference symbols will be available and in low SINR conditions estimation accuracy and hence demodulation

performance will suffer. CSI-RS can be used for supplementing the estimation resources but is available only infrequently as outlined before. This creates additional control complexity in the receiver.

Furthermore, the UE needs to know which reference resources can be used for estimation. This depends on the deployment scenarios, in particular the cell IDs. Furthermore, the combinations of antenna ports used for transmitting PDSCH and reference symbols are signaled via Radio Resource Control (RRC) messages. Separate combinations for each transmission point can be preconfigured when the UE is set up. For each subframe, the configuration used is signaled as part of Downlink Control Information (DCI). Hence, when considering estimation results from multiple subframes, the UE needs to do this on a per-configuration basis. While we do not expect this to require a significant amount of additional memory, it does increase complexity and verification effort. We will further elaborate on the estimation scenarios in the next section.

Frequency Offset Estimation

An issue of a CoMP UE is frequency offset recovery, which obviously affects demodulation performance of a receiver. In an asynchronous network, all TPs in a CoMP set can have different frequency offsets. An UE must be able to estimate them and recover the PDSCH signal from any TPs in the CoMP set.

A CoMP UE can utilize CRS for frequency offset estimation thanks to the CRS and CSIRS quasi-colocation assumption. However, CRS mapping and signaling are made differently depending on CoMP scenarios; estimation algorithms must be designed fitting in each scenario. Let's assume a CoMP set with two TPs as shown in Figure 7. A behavior-B UE needs to achieve synchronization to both TPs simultaneously and perform offset recovery. PDSCH can be transmitted by dynamically switching between the two TPs. A CoMP network can also by synchronized through a high-performing backhaul, or it can make an asynchronous network. A CoMP UE takes into account both cases.

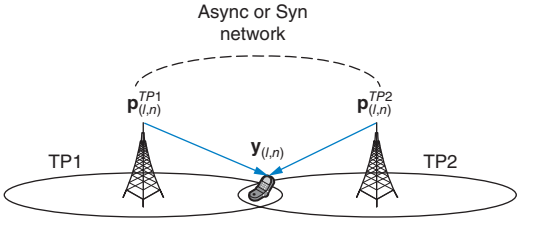


Figure 7: CoMP network model of frequency synchronization
(Source: Intel Corporation, 2013)

First, we discuss a frequency offset model of CoMP scenarios. A baseline offset estimator of a CoMP UE is performed using CRS. We need to understand CRS configurations for estimator designs. The received signal from each TP can be expressed as

$$\mathbf{y}_{(l,n)}^{TP1} = \left(\mathbf{p}_{(l,n)}^{TP1} * \gamma^{TP1}\mathbf{h}_{(l,n)}^{TP1} + \mathbf{n}_{(l,n)}^{TP1}\right) \cdot expf\left(\varepsilon_{cfo}^{TP1}\right) \quad (1)$$

$$\mathbf{y}_{(l,n)}^{TP2} = \left(\mathbf{p}_{(l,n)}^{TP2} * \gamma^{TP2}\mathbf{h}_{(l,n)}^{TP2} + \mathbf{n}_{(l,n)}^{TP2}\right) \cdot expf\left(\varepsilon_{cfo}^{TP2}\right) \quad (2)$$

where y^{iTP} is a time-domain received signal through a link between ith TP and an UE, p is CRS pilot signal in time-domain, h is a time-domain multipath channel profile, and n indicates noise; $\gamma$is a channel propagation attenuation, and ε_{cfo}^{itp}

is a frequency offset difference given from iTP TPs. Subscript (l,n) is a time index, l is an OFDM symbol index and n is a sample index. We define

$$expf(\varepsilon) = \exp(j2\pi\varepsilon(n + l(N + N_g))/N)$$

where N is a size of FFT and N_g is a size of guard interval.
As shown in the deployment scenarios earlier, CoMP sets can consist of cells with different IDs where the CRS are either colliding or non-colliding and CoMP set with the same cell ID. This is always a colliding case. Let's consider a received signal of CoMP DPS/DPB regarding CRS:

1. Colliding CRS with different cell ID:

$$\mathbf{y}_{(l,n)} = \mathbf{y}^{TP1}_{(l,n)} + \mathbf{y}^{TP2}_{(l,n)}$$

2. Non-colliding CRS with different cell ID:

$$\mathbf{y}_{(l,n)} = \mathbf{y}^{TP_{PDSCH}}_{(l,n)}, \mathbf{TP}_{PDSCH} \in \{TP1, TP2\}$$

3. Colliding CRS with same cell ID:

$$\mathbf{y}_{(l,n)} = (\mathbf{p}_{(l,n)} * (\gamma^{TP1}\mathbf{h}^{TP1}_{(l,n)} \cdot expf(\varepsilon^{TP1}_{cfo}) + \gamma^{TP2}\mathbf{h}^{TP2}_{(l,n)} \cdot expf(\varepsilon^{TP2}_{cfo}))) + \mathrm{n}_{(l,n)}$$

Case 1 indicates that a received signal from colliding CRS is simply a sum of two TPs CRS signals. Because $p^{TP1} \neq p^{TP2}$, an UE can still demodulate y^{TP1}, y^{TP2} respectively and detect the frequency offset. Case 2 non-colliding has a CRS signal as it is or contaminated by interference, and a UE has no problem in performing separate frequency offset tracking. Because case 3 has $p^{TP1}=p^{TP2}$, it has a received signal mixed up two frequency offsets. A UE has no method to detect each frequency offset. Therefore, case 3 is only available in a synchronous network scenario.

Timing Offset Estimation

Different timing offsets among TPs are problematic to a CoMP UE. While frequency synchronization can be controlled by network backhaul, timing offsets cannot be controlled effectively due to different channel propagation delays. A UE at cell edge inevitably experiences timing offset difference due to different cell sizes.

In a heterogeneous network configuration, PDCCH is transmitted from eNB, and PDSCH can be transmitted by DPS or JT. Therefore a CoMP UE must achieve timing synchronization for both TPs to receive PDCCH and PDSCH. As illustrated in Figure 8, channel propagation paths in a heterogeneous network have different inherent propagation delays ($\tau^{TP1} \neq \tau^{TP2}$). Due to different timing offsets, a UE requires a special algorithm to capture OFDM symbol boundary. An eNB transmits CRS and PDCCH, and PDSCH can be switched between an eNB and an RRH. If setting a FfT-window reference point at a PDCCH symbol, the PDSCH symbols may arrive in early timing (negative timing offset in PDSCH), or the PDSCH symbols may arrive later (positive timing offset in PDSCH) as shown in Figure 9.

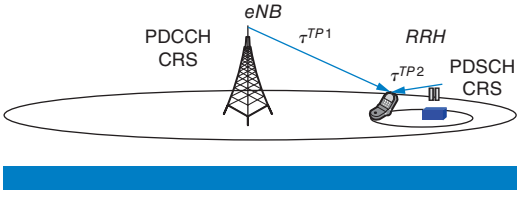


Figure 8: CoMP network model for timing synchronization
(Source: Intel Corporation, 2013)

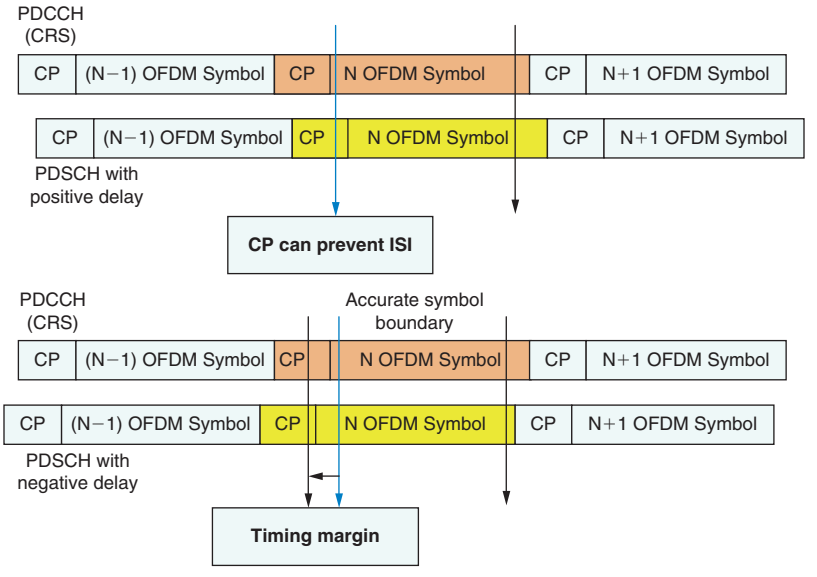


Figure 9: PDSCH timing alignment with positive timing offset (up) and negative timing offset (down)
(Source: Intel Corporation, 2013)

Depending on positive or negative timing offsets in PDSCH, the timing offset needs adjusting differently. Figure 9 shows the OFDM symbol alignments of received signals from HetNet CoMP network. UE demodulation performance can be critically degraded when the OFDM symbol misalignment causes inter-symbol interference (ISI) in a received signal.

In the case of PDSCH misalignment with positive offset, ISI can be prevented by CP region. In other words, CP is regarded as a kind of buffer covering the misalignment. As long as the positive timing offset is smaller than CP length, ISI does not occur. Therefore, the positive offset case is a trivial aspect of UE design. However, the negative timing offset may easily cause ISI. A solution to the negative timing offset problem can be approached by setting a timing margin. The FfT window position needs choosing very carefully to prevent timing misalignment of PDSCH transmission and thus ISI.

Frequency and Timing Offset Correction

We investigated the throughput performance of an UE supporting CoMP-DPS/DPB. A CoMP UE is required to establish and maintain synchronization to multiple TPs since PDSCH transmission can be dynamically switched among the TPs. The UE synchronization procedures are illustrated in Figure 10. When a UE enters a network, first, its synchronization process is locked to a serving cell. Assume that TP1 is the serving cell from which a UE receives CoMP-related signals. Then an eNB can configure PDSCH transmission from TP2 and PDSCH-mute at TP1 (no interference from TP1). In this scenario, the UE is supposed to make synchronization to TP1 through pre-FfT compensation because it is a serving cell, and signals from other TPs are recovered through post-FfT

chains. In simulation, PDSCH is assumed to be transmitted from TP2, which can be a pico cell or a local RRH. While a CoMP UE makes synchronization to TP1 through pre-FfT compensation chains, timing and frequency offsets of TP2 are compensated through a post-FfT chain as shown in Figure 10.

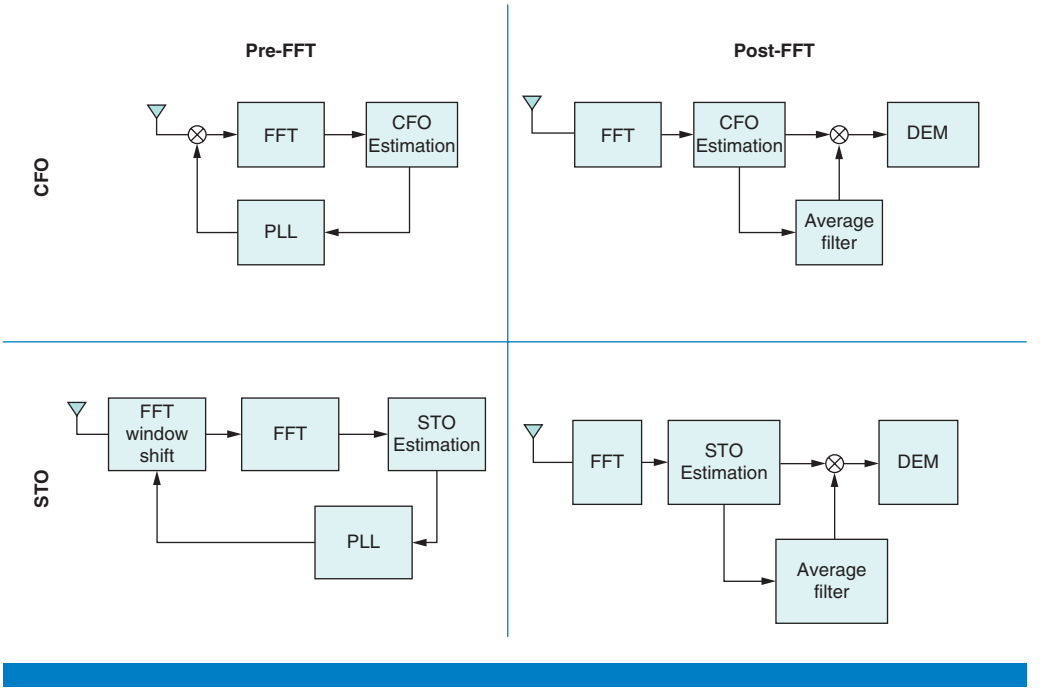


Figure 10: Synchronization concept block diagram (Source: Intel Corporation, 2013)

Figure 11 shows that the sampling timing offset (STO) affects PDSCH receptions. The simulation sets colliding CRSs in scenario 3, and the transmitter utilizes 16QAM with a code rate of 0.53 in open-loop channel state reporting. A low correlation ETU-5-Hz channel is used. Although a CoMP-UE tries to compensate for it through post-FfT chain, it is not able to completely remove ISI due to negative timing offset as discussed in the previous section. Simulation results show that ISI with a timing offset of −2 μs causes performance degradation of about SNR = 1.5 dB in the scenario. Therefore an eNB needs to make delicate transmitter timing alignment to prevent significant negative timing offset at cell-edge UEs. Note that the ETU-5-Hz channel has non-line-of-sight (NLOS) type of multipath profiles. The first tap with significant energy appears at 0.2 μs in the ETU profile, which also causes additional phase rotation

Figure 12 shows that impacts of carrier frequency offsets (CFO) from TPs. Basically, a CoMP network is assumed to be deployed with high-performing backhaul. In terms of frequency synchronization, an optical fiber cable can be an option to achieve near-perfect frequency synchronization among CoMP TPs. However, if a network considers acceptable tolerance on frequency offsets, the small amount of frequency offset can be compensated through post-FfT chain without performance loss. Depending on frequency differences, PDSCH BLERs are shown like Figure 12. Frequency offset larger than 400 Hz seems very critical to our test case.

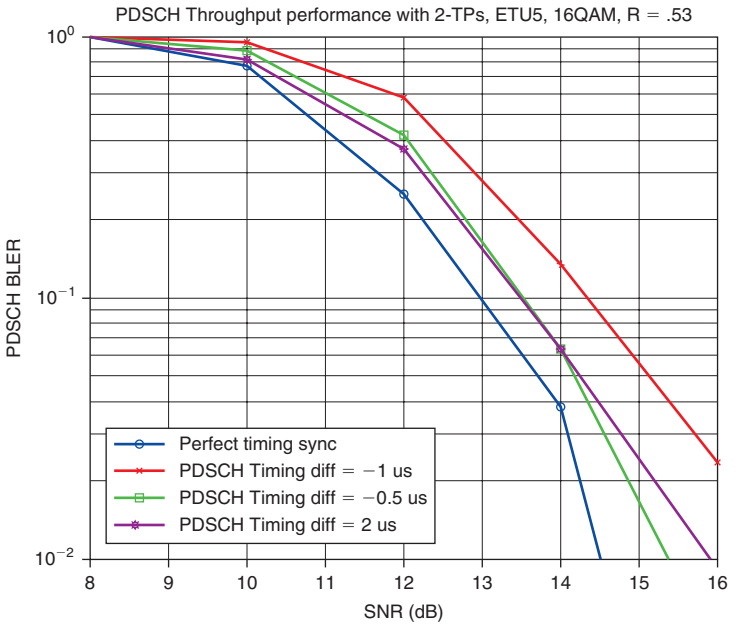


Figure 11: CoMP-UE BLER performance with timing offsets of PDSCH transmission
(Source: Intel Corporation, 2013)

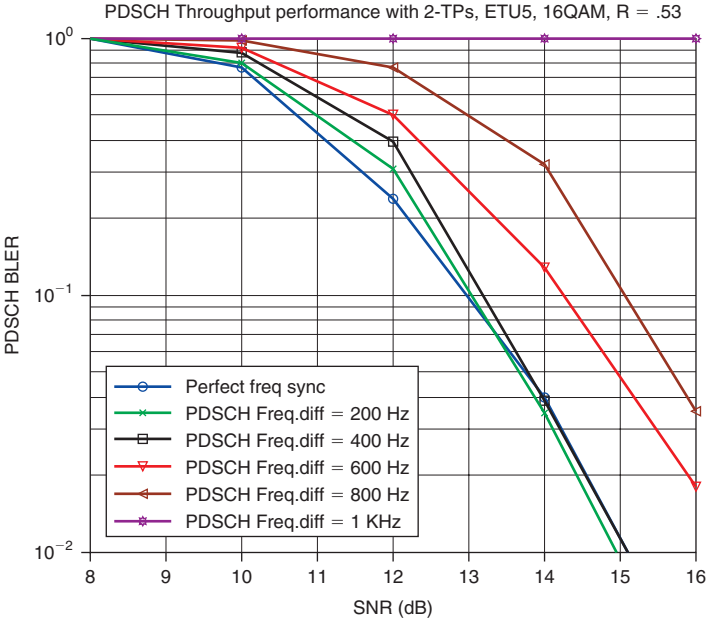


Figure 12: CoMP-UE BLER performance with carrier frequency offsets of PDSCH transmission
(Source: Intel Corporation, 2013)

Automatic Gain Control (AGC)

Since the distance between two TPs and a UE can vary considerably, propagation losses for each path will vary as well. CoMP is intended to cope with differences in received signal strength from two TPs of up to 9 dB. These changes occur within a subframe, that is, from one symbol to the next.

There are two possibilities to cope with that:

1. The AGC needs to be extremely agile so that it can adjust gain levels within the first few samples of the cyclic prefix. This approach would create a lot of spurious responses and we do not think it to be feasible.
2. On each subframe, estimate signal strength for each transmission point separately. When adjusting the gain for receiving the next subframe, select a gain that provides not only sufficient signal-to-quantization-noise ratio for the weakest TP and but also still avoids compression for the strongest TP.

Channel State Information Feedback (CSI-FB)

The eNodeB requires information about the channels from each transmission point to the UE and the interference situation between them. In order to provide this information the UE performs CSI-FB calculations. As this is the topic of another article by some of our colleagues in this journal ("Feedback Generation for Cell-Edge Transmission in Unsynchronized Coordinated Networks"), we will just mention it for completeness but not elaborate any further.

Uplink CoMP

Uplink CoMP does not pose any particular challenges to the signal processing in the UE. All combining is done in the base stations and is outside the scope of this article.

Combinations of Features

"Combinations of features cause further implementation challenges."

It should be mentioned that CoMP is just one new feature to be deployed in Release 11. Other features include Further enhanced Inter-Cell Interference Coordination (FeICIC), enhanced PDCCH (ePDCCH) and Carrier Aggregation (CA). A UE will have to support various combinations of them. This creates more implementation challenges. As an example, channel state information feedback needs to be calculated for more channels and carriers, leading to higher computational loads. Also, certain operating modes of CoMP serve very similar purposes to other techniques. An example is dynamic point blanking in CoMP and the Almost Blank Subframe (ABS) in FeICIC. Both are introduced in LTE Release 11 and stop an interfering cell from transmitting while sending data to a certain UE in an interference-prone location. We expect that usage of the one or other technique will be a network operator's choice. The UE, however, will have to support both techniques.

Further Developments of CoMP: Release 12 and Beyond

As discussed previously, LTE Release 11 introduced basic functionalities for CoMP transmissions. Since CoMP technologies are mainly designed to reduce inter-cell interference, uplink CoMP is to a large extent not related to the

specification but it is an implementation choice. However, a big effort has been made to specify downlink CoMP in Release 11, especially to specify channel-state information (CSI) to be fed back from the UE to eNodeB. For this purpose, a multipoint CSI feedback framework has been introduced in Release 11.[1]

3GPP is currently working on LTE evolution towards Release 12. Further enhancements of CoMP transmission (Evolution of CoMP, eCoMP), including enhancements with respect to channel-state information feedback and, more importantly, enhancing practical applicability of CoMP transmission with relaxed backhaul requirements, are under discussion.[2]

The following section describes some of possible extensions CoMP that might be supported in LTE Release 12 and beyond.

Implicit Feedback Generation Based on PMI Selection

As previously discussed, CoMP schemes are used with closed-loop transmission modes and thus one of the most critical implementation challenges in CoMP systems is the feedback generation. Current 3GPP specification does not support alignment between transmission points when sending the feedback, that is, the choice of precoder is independent between the transmission points. This is rather a significant limitation in CoMP networks, because the effect of inter-transmission-point interference cannot be reduced efficiently.

"…the most critical implementation challenges in CoMP systems is the feedback generation."

As shown in Figure 13, by aligning the transmit signal between the all transmission points, up to 3dB performance improvement over the nonaligned scenario can be achieved due to the increased coordination. Detailed discussion on aligned CoMP transmission can be found in Bai et al.[3] and references therein.

Feedback Generation Improvements Based on Phase Adjustments between Transmission Points

Phase feedback adjustments have been discussed in Release 11 together with PMI feedback. However to keep complexity of CoMP at the minimum, 3GPPP decided to move CSI feedback based on phase adjustments to Release 12 discussions. Here, we outline the performance benefits of phase adjustments between transmission points in different CoMP scenarios. Phase feedback was first introduced in 3GPP R1-083546[4], where it was recommended to introduce to UE feedback additional phase factors for coherent combining of beams from different transmission points.

The main aim of phase feedback is to provide additional degrees of freedom for precoding signals (increase the number of available precoders) and to align channels from different TPs in a coherent manner so that more performance improvements can be achieved through techniques such as coherent reception.

"The main aim of phase feedback is to provide additional degrees of freedom for precoding signals and to align channels from different TPs in a coherent manner…"

As shown in Figure 14, adjusting the phase between the transmission points can lead up to 4 dB depending on realizations.

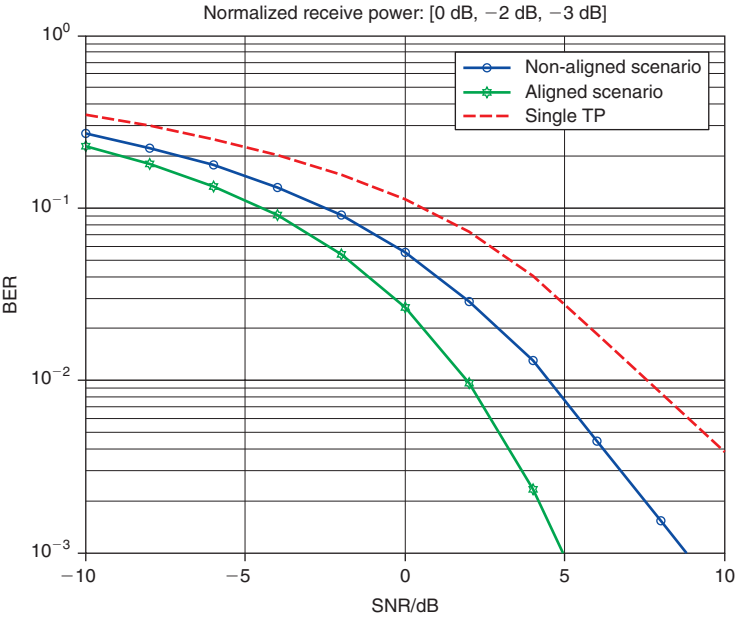


Figure 13: Impact of aligned transmission in 10 MHz 4x4 MIMO CoMP, 16QAM, SCM-B SIR = 10 dB, IRC receiver
(Source: Intel Corporation, 2013)

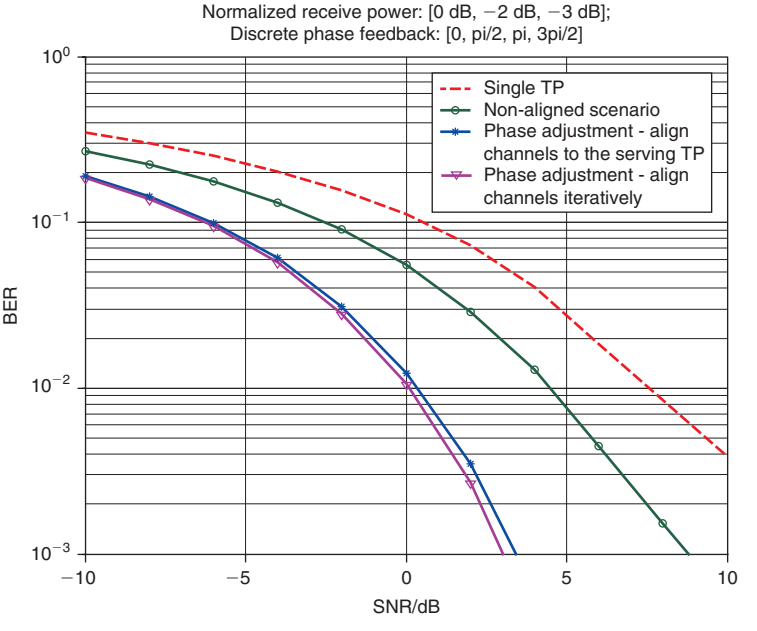


Figure 14: Phase adjustments in 10 MHz 4x4 MIMO CoMP, 16QAM, SCM-B, SIR = 10dB, IRC receiver.
(Source: Intel Corporation, 2013)

Conclusion

In this article challenges and solutions for UEs implementing the coordinated network features in the upcoming LTE Releases 11 and 12 have been discussed. For Release 11 the transmissions from multiple transmission points require the UEs to track estimation results for each transmission point. We have showed that correcting for some of these can be done in the frequency domain with little impact from inter-carrier and inter-symbol interference. This allows for reducing the receiver's computational complexity versus a classical approach with one receiver for each transmission point. Furthermore, this article demonstrates possible improvements of CoMP performance in Release 12 and beyond. By aligning the feedback between the transmission points or by enhancing the CSI feedback in CoMP, we have shown that a gain of up to 5 dB can be achieved compared to conventional Release 11 CoMP networks.

The work on evolution of CoMP continues in 3GPP. Enhancements with respect to channel-state information feedback from the terminals and the practical applicability of CoMP solutions with relaxed backhaul requirements are among the key topics under discussion.

Acknowledgments

The authors wish to gratefully acknowledge the support of their colleagues at Intel Mobile Communications and at the Department of Communication Technologies of the University of Duisburg-Essen.

References

[1] 3GPPTSG-RANWG1 "Coordinated multi-point operation for LTE physical layer aspects, Sophia Antipolis, Technical Report 36.819 v1.0.0 Dec. 2011.

[2] 3GPPTSG-RANWG1, "Coordinated Multi-point operation for LTE physical layer aspects," 3GPP, Sophia Antipolis, Technical Report 36.819 v11.2.0, Sep. 2013.

[3] Bai, Z., B. Badic, R. Balraj, S. Iwelski, T. Scholand, G. Bruck, and P. Jung, "Near Optimal Implicit Feedback Generation for Coordinated Multipoint Transmission" *Communications Letters*, IEEE, 2013.

[4] 3GPP R1-083546, "Per-cell precoding methods for downlink joint processing CoMP," 3GPP TSG RAN WG1 Meeting #54bis, ETRI 2008.

Author Biographies

Michael Hosemann has been a senior staff engineer with Intel Mobile Communications for two years. He gained his Diplomingenieur and Doctorate degrees from Technische Universität Dresden, Germany, and has since worked in various companies in Germany and England as system architect on wireless

and medical systems. Amongst other topics Michael is currently working on synchronization issues for multipoint communications systems. Michael is a member of the IEEE and can be reached at Michael.Hosemann@ieee.org.

Daejung Yoon is a system engineer with Intel Mobile Communication. He received his PhD in electrical engineering from the University of Minnesota at Twin Cities in 2010. From 2009 to 2010, he worked as a senior engineer at Samsung Information Systems America, in San Jose, California. His research interests are in areas of multiple point transmission features of LTA-A systems and advanced signal processing for estimation and interference rejection algorithms. Email: daejung.yoon@intel.com.

Biljana Badic is a system engineer in Intel Product Engineering 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 in Austria and has published over 50 scientific articles and filed over 10 4G patents. Email: biljana.badic@intel.com.

Rajarajan Balraj is a system engineer in Intel Product Engineering Group and has been with Intel since 2010. He has made key contributions to various interference cancellation and mitigation algorithms on Intel's 3G and 4G wireless modems. Rajarajan Balraj has filed over 25 patents related to LTE/LTE-A and UMTS. Email: rajarajan.balraj@intel.com.

Constantinos B. Papadias is a professor at Athens Information Technology (AIT) in Greece. He is also the head of AIT's Broadband Wireless and Sensor Networks (B-WiSE) Research Group. Dr. Papadias is a Distinguished Lecturer of the IEEE Communications Society for 2012–2013 and an IEEE Fellow. Email: cpap@ait.gr.

FEEDBACK GENERATION FOR CELL-EDGE TRANSMISSION IN UNSYNCHRONIZED COORDINATED NETWORKS

Contributors

Zijian Bai
Department of Communication Technologies, University of Duisburg-Essen, Germany

Stanislaus Iwelski
Department of Communication Technologies, University of Duisburg-Essen, Germany

Biljana Badic
Product Engineering Group, Intel Corporation

Guido Bruck
Department of Communication Technologies, University of Duisburg-Essen, Germany

Peter Jung
Department of Communication Technologies, University of Duisburg-Essen, Germany

Multi-point transmission in coordinated networks enables reliable coverage and high data rates for cell-edge users in 4G Long Term Evolution (LTE) systems. To provide theoretical proven multi-point coordination gain in the practical deployed networks, accurate channel state information (CSI) feedback is required. The focus of this article is on CSI feedback generation, that is, on precoding matrix index (PMI) selection, supporting coherent reception for the users located at the cell-edge of unsynchronized coordinated LTE networks. The received signal at user equipment and the theoretical lower bound of the joint transmission are analyzed. The impact of timing offset on feedback generation in coordinated networks is evaluated as well. With analytical and numerical results, the study in this article shows that aligned transmissions from multiple points via coordinated PMI selection provides near maximum performance for the cell-edge users in unsynchronized coordinated LTE networks even with the presence of high timing offset between the transmission points.

Introduction

The 3GPP continues to add more capabilities to Long Term Evolution (LTE) standards to enable cellular communications of even higher data rates.[1] One of the key features introduced since Release 11[2] is coordinated multipoint transmission (CoMP) based on a cooperative multiple-input and multiple-output (MIMO) technique between distributed base stations (eNodeBs), also referred to as transmission points (TPs). As the transmission to cell-edge users is more often encountered in modern cellular deployment with a heterogeneous structure (network deployments using a mix of macro, pico, femto, and relay base stations), the aim of coordinated TPs is to reduce interference and increase capacity for the cell-edge users. As a result, the coverage of high data rates of one cell is extended, and the corresponding user experience is significantly improved.

In LTE Release 11[2], four major CoMP scenarios are defined for the downlink (DL) signal transmission: Scenario 1 and Scenario 2 cover homogenous networks, while Scenario 3 and Scenario 4 are related to the heterogeneous network (HetNet). Generally, CoMP transmissions in Scenario 3 and Scenario 4 are considered as the most practical developments. CoMP Scenario 3 describes CoMP transmission between different eNodeBs (such as macro eNodeB with femto or pico eNodeB); whereas CoMP Scenario 4 represents CoMP transmission between macro eNodeB and its remote radio heads (RRHs). DL CoMP transmission schemes in coordinated networks are

mainly categorized into joint processing (JP) and coordinated scheduling/ Beamforming (CS/CB). Furthermore, JP includes an important subcategory of joint transmission (JT), in which simultaneous transmissions from multiple TPs to the target cell-edge users are applied with focuses on signal strength enhancement.[2] The CS/CB scheme aims at the interference elimination or avoidance, while JT/DPS operates in the way of turning the interference to be the useful signals. All the cooperative schemes in coordinated networks operate mostly in the closed-loop transmission mode[2], that is, all rely on DL channel state information (CSI) at TPs. Because implementation of the explicit CSI feedback is less feasible due to a high overhead[2], in this article the implicit CSI feedback for the JT scheme in coordinated networks is taken as the basis of the investigation to enable the promised advantages of CoMP transmission.

"…the cooperative schemes in coordinated networks operate mostly in the closed-loop transmission mode, that is, all rely on DL channel state information…"

"The main challenge for the cell-edge users to support CoMP trannsmission in coordinated networks and to gather JT and coherent reception gain is the proper imoplicit CSI feedback generation,…"

The main challenge for the cell-edge users to support CoMP transmission in coordinated networks and to gather JT and coherent reception gain is the proper implicit CSI feedback generation, that is, precoding matrix index (PMI) feedback.[1] Comprehensive studies of suitable PMI selection for the CoMP transmission have been carried out recently. ETRI[3] introduces per-cell PMI selection for JT in coordinated networks, which is a straightforward extension of the single point transmission case. Conventional PMI generation schemes[6] can be employed in the case of ETRI[3] to optimize transmission per TP individually. The coordination-oriented PMI generation based on concatenated cell codebook and large scale fading factors of different TPs are intensively studied by Peng et al.[4] Su et al.[5] propose a layered feedback strategy at the system level to achieve the optimal CB/CS. The physical spacing between TPs has been identified by HiSilicon Huawei[7] to have significant impact on the PMI selection. Kotzch et al.[8] propose interference suppression and cancellation techniques to overcome the interference in an unsynchronized network. Liu et al.[9] reconstruct the synchronization between TPs by adjusting the transmit signal at each TP.

In comparison with these previous works, this article focuses on the signal processing at cell-edge users and the evaluation of the link level CoMP transmission quality. The aim of the study is to elaborate the coordination-aware PMI selection approaches of Bai et al.[11][12] and obtain an insight into the feasibility of applying these approaches to enable expected CoMP gain for cell-edge users in unsynchronized coordinated networks via performance evaluation. Link level simulations are carried out to facilitate numerical results and to illustrate the promised benefits achieved by the coordination-aware PMI selection with the criterion of channel alignment in unsynchronized coordinated networks.

In what follows, complex baseband mathematical notation is used for the description of signal processing. Vectors/matrices are denoted with lower- or uppercase characters in boldface, while $\Re(\bullet)$ denotes the real part of a complex value. Furthermore, $(\bullet)^T$ and $(\bullet)^H$ are the vector/matrix transpose and Hermitian transpose operations, while $E(\bullet)$ gives the expected value of given random variables.

Signal Transmission and Reception in Coordinated Networks

To facilitate the design and evaluation of the coordinated CoMP transmission in synchronized and unsynchronized HetNet, the DL signal processing is represented in this section. The focuses are given on the system function (received signal) in mathematical expression and the optimal receiver filter structure, which captures the coordination gain and recovers the transmitted signals.

System Model

A typical layout of a coordinated HetNet with three TPs is depicted in Figure 1, where the distance between the target cell-edge user equipment (UE) and the mth TP is given as s_m, $m = 1, 2, 3$. The target UE is equipped with N_R receive antennas. Signals from different TPs carrying the same user data, $\mathbf{e}_m \in \mathbb{C}^{N_R \times 1}$, propagate to the target UE via independent paths.

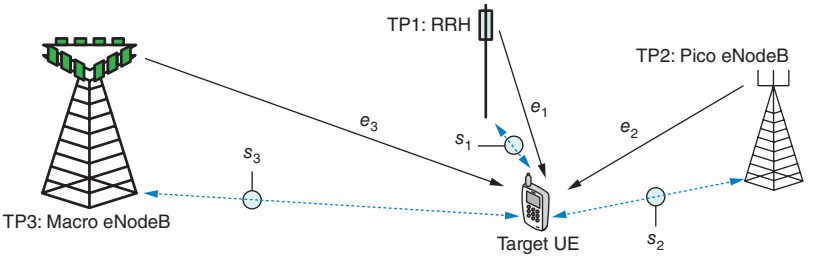


Figure 1: CoMP joint transmission in unsynchronized coordinated HetNets (Source: Department of Communication Technologies at the University of Duisburg-Essen, 2013)

For the CoMP JT in Figure 1, the subcarrier-specific MIMO system is assumed. Without loss of generality, M TPs are considered, in which the mth TP is mounted with $N_{T,m}$ transmit antennas. The system function of the receive signals can be represented (3GPP[2][10]) by

$$\boldsymbol{r} = \sum_{m=1}^{M} \boldsymbol{H}_m \boldsymbol{P}_m \boldsymbol{d}_m + \underbrace{\boldsymbol{n}_I + \boldsymbol{n}}_{=\boldsymbol{n}_T}, \quad (1)$$

where $\boldsymbol{H}_m \in \mathbb{C}^{N_R \times N_{T,m}}$ is the channel matrix from the mth TP to the target UE, $\boldsymbol{P}_m \in \mathbb{C}^{N_{T,m} \times N_{L,m}}$ denotes the applied precoding matrix at the mth TP. Furthermore, $\boldsymbol{d}_m \in \mathbb{C}^{N_{L,m} \times 1}$ is the $N_{L,m}$ layer transmit data vector out of modulation constellation at the mth TP, with $\mathrm{E}(\boldsymbol{d}_m \boldsymbol{d}_m^H) = E_d \boldsymbol{I}_{N_{L,m}}$. In addition, $\boldsymbol{r} \in \mathbb{C}^{N_R \times 1}$ is the received signal vector, $\boldsymbol{n}_I \in \mathbb{C}^{N_R \times 1}$ is the co-channel interference vector with zero mean, and the covariance matrix of $\mathrm{E}(\boldsymbol{n}_I \boldsymbol{n}_I^H) = \boldsymbol{R}_{\boldsymbol{n}_I \boldsymbol{n}_I}$, $\boldsymbol{n} \in \mathbb{C}^{N_R \times 1}$ is the additive white Gaussian noise (AWGN) vector with $\boldsymbol{n} \sim CN(\boldsymbol{0}, N_0 \boldsymbol{I}_{N_R})$. Channel coefficients, transmit data, co-channel interference, and noise samples are assumed to be mutually uncorrelated.

Since the signal to the target cell-edge UE undergoes a long propagation path, the received signal strength can drop dramatically. In addition, co-channel signals from neighbor cells are as strong as the useful signal resulting in significant interference. As a result, the target UE located at the border of macro cell suffers from a small signal-to-interference-and-noise ratio and needs power and diversity gain in a coordinated network to combat this loss. Hence the single layer joint transmission (JT) transmission is applied in this article, that is, different TPs send the same data to the target UE, which means $N_{\mathrm{L},m} = 1$, $m = 1,\ldots, M$ and $\boldsymbol{d}_m = d$ at all TPs. Without loss of generality, it is further assumed that $N_{\mathrm{T},m} = N_{\mathrm{T}}$, $m = 1,\ldots, M$. Consequently, the system function (Equation 1) in a particular subcarrier with the index k can be rewritten as

"…the target UE located at the border of macro cell suffers from a small signal/to-interference-and-noise ratio and needs power and diversity gain in coordinated network to combat this loss."

$$\boldsymbol{r}[k] = \sum_{m=1}^{M} \boldsymbol{H}_m[k]\, \boldsymbol{p}_m[k]\, d\,[k] + \underbrace{\boldsymbol{n}_{\mathrm{I}}[k] + \boldsymbol{n}[k]}_{= n_{\mathrm{T}}[k]}\,, \tag{2}$$

where $\boldsymbol{p}_m \in \mathbb{C}^{N_{\mathrm{T}}\times 1}$ degrades to be vectors.

Receiver Structure

To detect the transmitted signal d in JT scheme under co-channel interference (CCI) $\boldsymbol{n}_{\mathrm{I}}[k]$, the interference rejection combining (IRC) processing scheme is known as the adequate scheme.[13] This scheme utilizes the knowledge of $\boldsymbol{R}_{\boldsymbol{n}_{\mathrm{I}}\boldsymbol{n}_{\mathrm{I}}}$ and pre-whitens the receive signals with a whitening filter (WF) followed by a matched filter (MF). Following the definition of interference plus noise term $\boldsymbol{n}_{\mathrm{T}}$ in Equation 2 and the corresponding covariance matrix of $\boldsymbol{R}_{\boldsymbol{n}_{\mathrm{T}}\boldsymbol{n}_{\mathrm{T}}} = \boldsymbol{R}_{\boldsymbol{n}_{\mathrm{I}}\boldsymbol{n}_{\mathrm{I}}} + N_0 \boldsymbol{I} = \boldsymbol{R}^{\frac{1}{2}}_{\boldsymbol{n}_{\mathrm{T}}\boldsymbol{n}_{\mathrm{T}}} \boldsymbol{R}^{\frac{1}{2}\mathrm{H}}_{\boldsymbol{n}_{\mathrm{T}}\boldsymbol{n}_{\mathrm{T}}}$, the IRC scheme for the system (Equation 2) equals

$$\begin{aligned}\boldsymbol{m}^{\mathrm{T}}_{\mathrm{IRC}}[k] &= \underbrace{\left(\boldsymbol{R}^{-\frac{1}{2}}_{\boldsymbol{n}_{\mathrm{T}}\boldsymbol{n}_{\mathrm{T}}}[k] \sum_{m=1}^{M} \boldsymbol{H}_m\,[k]\, \boldsymbol{p}_m\,[k]\right)^{\mathrm{H}}}_{\mathrm{MF}} \underbrace{\boldsymbol{R}^{-\frac{1}{2}}_{\boldsymbol{n}_{\mathrm{T}}\boldsymbol{n}_{\mathrm{T}}}[k]}_{\mathrm{WF}} \\ &= \left(\sum_{m=1}^{M} \boldsymbol{H}_m\,[k]\, \boldsymbol{p}_m\,[k]\right)^{\mathrm{H}} \boldsymbol{R}^{-1}_{\boldsymbol{n}_{\mathrm{T}}\boldsymbol{n}_{\mathrm{T}}}[k].\end{aligned}$$

Applying the IRC scheme on the receive signals, the filtered received signal of Equation 2 equals

$$\begin{aligned}\tilde{d}[k] &= \boldsymbol{m}^{\mathrm{T}}_{\mathrm{IRC}}\,[k]\boldsymbol{r}\,[k] \\ &= \boldsymbol{m}^{\mathrm{T}}_{\mathrm{IRC}}\,[k] \sum_{m=1}^{M} \boldsymbol{H}_m\,[k]\, \boldsymbol{p}_m\,[k] d\,[k] + \boldsymbol{m}^{\mathrm{T}}_{\mathrm{IRC}}\,[k]\boldsymbol{n}_{\mathrm{T}}\,[k],\end{aligned} \tag{3}$$

with the associated post-processing signal to noise ratio (SNR) of

$$\gamma\,[k] = \frac{\mathrm{E}\left(\left|\boldsymbol{m}^{\mathrm{T}}_{\mathrm{IRC}}\,[k] \sum_{m=1}^{M} \boldsymbol{H}_m[k]\, \boldsymbol{p}_m\,[k] d\,[k]\right|^2\right)}{\mathrm{E}\left(\left|\boldsymbol{m}^{\mathrm{T}}_{\mathrm{IRC}}\,[k]\, \boldsymbol{n}_{\mathrm{T}}\,[k]\right|^2\right)} \tag{4}$$

$$= E_{\mathrm{d}} \boldsymbol{m}_{\mathrm{IRC}}^{\mathrm{T}}[k] \left(\sum_{m=1}^{M} \boldsymbol{H}_m[k]\, \boldsymbol{p}_m[k] \right) \left(\sum_{m=1}^{M} \boldsymbol{H}_m[k]\, \boldsymbol{p}_m[k] \right)^{\mathrm{H}} \boldsymbol{m}_{\mathrm{IRC}}^{*}[k]$$

$$\boldsymbol{m}_{\mathrm{IRC}}^{\mathrm{T}}[k] \boldsymbol{R}_{\boldsymbol{n}_{\mathrm{T}}\boldsymbol{n}_{\mathrm{T}}}[k]\, \boldsymbol{m}_{\mathrm{IRC}}^{*}[k]$$

$$= E_{\mathrm{d}} \left(\sum_{m=1}^{M} \boldsymbol{H}_m[k]\, \boldsymbol{p}_m[k] \right)^{\mathrm{H}} \boldsymbol{R}_{\boldsymbol{n}_{\mathrm{T}}\boldsymbol{n}_{\mathrm{T}}}^{-1}[k] \left(\sum_{m=1}^{M} \boldsymbol{H}_m[k]\, \boldsymbol{p}_m[k] \right).$$

According to the concept of IRC[13], $\gamma[k]$ is maximized under the given $\boldsymbol{H}_m[k]\, \boldsymbol{p}_m[k]$.

Precoding Selection to Achieve Maximum Coordination Gain

"The feedback parameter PMI provides the preferred precoder (quantized beamforming) out of a predefined codebook and guides the particular TP to inquire DL CSI and align the transmit signals to the target UE."

As discussed previously, the key challenges for the target cell-edge UE to support the JT scheme in coordinated networks and to obtain the maximum coordination advantage is CSI feedback, realized by PMI feedback generation in LTE systems.[3] The feedback parameter PMI provides the preferred precoder (quantized beamforming) out of a predefined codebook and guides the particular TP to inquire DL CSI and align the transmit signals to the target UE. Based on the review of the precoder selection in single point transmission, the focuses of this selection are the uncoordinated and coordinated PMI selection aiming at maximum CoMP gain in synchronized HetNet.

Precoding Selection in One-Point Transmission

A brief review of PMI selection in single-point transmission is given in this part. The state-of-the-art solution uses the maximum post-SNR (MaxSNR) criterion. , being an adequate one to find the optimal PMI as studied by Bai et al.[6] Applying this solution for a particular TP in the considered HetNet, the precoding vector $\boldsymbol{p}_m$ (and the associated index as PMI) for the mth TP can be selected by

$$\begin{aligned} \boldsymbol{p}_m &= \arg\max_{\boldsymbol{q}_i \in \wp} \sum_{k=1}^{N_{\mathrm{SUBC}}} \gamma_{m,i}[k] \\ &= \arg\max_{\boldsymbol{q}_i \in \wp} \boldsymbol{q}_i^{\mathrm{H}} \left(\sum_{k=1}^{N_{\mathrm{SUBC}}} \underbrace{\boldsymbol{H}_m^{\mathrm{H}}[k] \boldsymbol{R}_{\boldsymbol{n}_{\mathrm{I}}\boldsymbol{n}_{\mathrm{T}}}^{-1}[k] \boldsymbol{H}_m[k]}_{\boldsymbol{R}_{m,m}[k]} \right) \boldsymbol{q}_i \cdot \end{aligned} \tag{5}$$

In Equation 5, $\wp = \{\boldsymbol{q}_{1,\ldots,}\boldsymbol{q}_Q\}$ represents the codebook in LTE Release 11 systems[1] with Q precoding vectors and $\gamma_{m,i}[k]$ is the post-SNR for the mth TP in the kth subcarrier with precoding vector of $\boldsymbol{q}_i$. Finally, $\boldsymbol{R}_{m,m}[k]$ denotes the transmit covariance matrix with WF in the kth subcarrier. In Equation 5, the PMI is selected based on the criterion of MaxSNR on the averaged post-SNR over N_{SUBC} subcarriers, in which the same precoder is applied.[1]

Based on the baseline of Equation 5, different PMI selection approaches are discussed by Bai et al.[11][12] under two types of the LTE coordinated networks: the single frequency network (SFN) and multiple single point network (MSPN). SFN presents the infrastructure configuration when all TPs perform

exactly the same processing for the transmit data, such as the macro eNodeB with its RRH.[2] In contrast, MSPN allows TPs to construct DL signals on its own demand under coordinated scheduling, such as heterogeneous eNodeBs in coordinated networks.[2] While some of the PMI selection approaches in these two networks are straightforward extensions from single TP transmission, others are coordination-aware and more advanced with the criterion of TP channel alignment. For the purpose of convenient discussion in this article, the PMI selection approaches for synchronized HetNets discussed by Bai et al.[11][12] are summarized in following subsections.

"While some of the PMI selection approaches in these two networks are straightforward extensions from single TP transmission, others are coordination-aware and more advanced with the criterion of TP channel alignment."

Coordination-Unaware PMI Selection

When the SFN infrastructure is present, the same precoder shall be applied at all TPs. Accordingly, the system function of Equation 2 turns into

$$\boldsymbol{r}[k] = \sum_{m=1}^{M} \boldsymbol{H}_m[k]\,\boldsymbol{p}_s\,d[k] + \boldsymbol{n}_{\mathrm{T}}[k]. \tag{6}$$

Three different approaches for $\boldsymbol{p}_s$ selection in SFN are examined by Bai et al.[11]

The first approach, denoted by SFN-SingleTP, utilizes the PMI selection (Equation 5) based on the channel of the serving TP and ignores the other TPs during the feedback. Such a situation occurs when the target UE is connected by the macro eNodeB, where the other RRH are transparent to it. With $m = 1$ the serving macro eNodeB, the PMI selection is carried out by

$$\boldsymbol{p}_s = \arg\max_{\boldsymbol{q}_i \in \wp} \boldsymbol{q}_i^{\mathrm{H}} \left(\sum_{k=1}^{N_{\mathrm{SUBC}}} \boldsymbol{R}_{1,1}[k] \right) \boldsymbol{q}_i, \tag{7}$$

The second approach, denoted by SFN-AdaptiveTP, applies the single TP-based PMI selection. However, the target UE is now aware of the presence of all TPs. In the case of being close to a certain TP (for example, RRH), the target UE adapts the PMI selection for this TP to obtain the power gain given by the small transmission distance. With $\boldsymbol{H}_m[k]$ being estimated for individual TP channel and $\|\boldsymbol{A}\|_{\mathrm{F}}$ being the Frobenius norm of the matrix $\boldsymbol{A}$,

$$m_s = \arg\max_{m=1,\ldots,M} \left(\sum_{k=1}^{N_{\mathrm{SUBC}}} \left\| \boldsymbol{H}_m[k] \right\|_{\mathrm{F}}^2 \right). \tag{8}$$

The preferred TP and the associated PMI can therefore be selected as

$$\boldsymbol{p}_s = \arg\max_{\boldsymbol{q}_i \in \wp} \boldsymbol{q}_i^{\mathrm{H}} \left(\sum_{k=1}^{N_{\mathrm{SUBC}}} \boldsymbol{R}_{m_s m_s}[k] \right) \boldsymbol{q}_i. \tag{9}$$

In the third approach, denoted by SFN-VirtualTP, the system given in Equation 6 is taken as a virtual system with the equivalent channel matrix $\boldsymbol{H}_{\mathrm{E}}[k] = \sum_{m=1}^{M} \boldsymbol{H}_m[k]$ and the corresponding precoding vector selection is given as

$$\begin{aligned}\boldsymbol{p}_s &= \arg\max_{\boldsymbol{q}_i \in \wp} \boldsymbol{q}_i^{\mathrm{H}} \left(\sum_{k=1}^{N_{\mathrm{SUBC}}} \boldsymbol{H}_{\mathrm{E}}^{\mathrm{H}}[k] \boldsymbol{R}_{\boldsymbol{n}_{\mathrm{I}}\boldsymbol{n}_{\mathrm{T}}}^{-1}[k]\ \boldsymbol{H}_{\mathrm{E}}\ [k] \right) \boldsymbol{q}_i \\ &= \arg\max_{\boldsymbol{q}_i \in \wp} \boldsymbol{q}_i^{\mathrm{H}} \left(\sum_{k=1}^{N_{\mathrm{SUBC}}} \sum_{m=1}^{M} \boldsymbol{H}_m^{\mathrm{H}}[k] \boldsymbol{R}_{\boldsymbol{n}_{\mathrm{I}}\boldsymbol{n}_{\mathrm{T}}}^{-1}[k] \sum_{m=1}^{M} \boldsymbol{H}_m\ [k] \right) \boldsymbol{q}_i .\end{aligned} \tag{10}$$

As discussed by Bai et al.[11], the MSPN infrastructure allows different precoders being applied at different TPs yielding additional freedom of the channel alignment and the potential SNR enhancement by the coherent reception of multiple signals. Therefore, the most straightforward way of generating the PMI feedback for each TP feedback is by applying Equation 5 with $m = 1,\ldots, M,$ and it is denoted as MSPN-LocalSNR in the following. In this approach all selected PMIs, that is $\boldsymbol{p}_m$, $m = 1,\ldots, M$, are reported to the network.

Coordination-Aware PMI Selection

Although the approaches in the previous subsection captured a certain level of beamforming and power gain in coordinated networks via CoMP JT, the global post-SNR in Equation 4 is not maximized with respect to $\boldsymbol{q}_{\mathrm{i}}$ since the cross-multiplication terms, that is, coordinated JT gain of Equation 4, are missing. Hence, the total achievable performance improvement is significantly limited as shown in the following section, "Performance Evaluation."

"To achieve the maximum performance gain for the target cell-edge UE in coordinated networks, coordionated PMI feedback is required."

To achieve the maximum performance gain for the target cell-edge UE in coordinated networks, coordinated PMI feedback is required. The optimal solution of generating this feedback, denoted by MSPN-Optimal, is to maximize the global post-SNR by a full-blown search over all possible hypotheses of $[\boldsymbol{p}_1,\ldots, \boldsymbol{p}_{\mathrm{M}}]$[11], which is represented by

$$\begin{aligned}&[\boldsymbol{p}_1,\ldots, \boldsymbol{p}_M] \\ &= \arg\max_{[\boldsymbol{q}_{i_1},\ldots, \boldsymbol{q}_{i_M}] \in \wp^{M\times 1}} \sum_{k=1}^{N_{\mathrm{SUBC}}} \left(\left(\left(\sum_{m=1}^{M} \boldsymbol{q}_{i_m}^{\mathrm{H}} \boldsymbol{H}_m^{\mathrm{H}}[k] \right) \boldsymbol{R}_{\boldsymbol{n}_{\mathrm{I}}\boldsymbol{n}_{\mathrm{T}}}^{-1}[k] \left(\sum_{m=1}^{M} \boldsymbol{H}_m[k]\ \boldsymbol{q}_{i_m} \right) \right) \right) \\ &= \arg\max_{[\boldsymbol{q}_{i_1},\ldots, \boldsymbol{q}_{i_M}] \in \wp^{M\times 1}} \left(\begin{array}{l} \sum_{m=1}^{M} \boldsymbol{q}_{i_m}^{\mathrm{H}} \sum_{k=1}^{N_{\mathrm{SUBC}}} \boldsymbol{R}_{m,m}[k] \boldsymbol{q}_{i_m} \\ + \sum_{\substack{m,n \in [1,M] \\ m \neq \mathrm{n}}} \boldsymbol{q}_{i_m}^{\mathrm{H}} \sum_{k=1}^{N_{\mathrm{SUBC}}} \underbrace{\boldsymbol{H}_m^{\mathrm{H}}[k]\ \boldsymbol{R}_{\boldsymbol{n}_{\mathrm{I}}\boldsymbol{n}_{\mathrm{T}}}^{-1}[k]\ \boldsymbol{H}_n[k] \boldsymbol{q}_{i_m}}_{= \boldsymbol{r}_{m,\mathrm{n}}[k]} \end{array} \right)\end{aligned} \tag{11}$$

"The computational complexity of the full-blown search is high and grows exponentially when the number of transmission points increases and as such it is practically not implementable."

where $\boldsymbol{q}_{i_m}$ is the i_mth precoder in $\wp$ and the candidate precoding vector for the mth TP with $i_m \in [1, \mathrm{Q}], \forall\ m$. Furthermore, $\boldsymbol{r}_{m,\mathrm{n}}[k]$ denotes the cross-TP covariance matrix with WF in the kth subcarrier. The computational complexity of the full-blown search is high and grows exponentially when

the number of transmission points increases and as such it is practically not implementable.

To exploit the JT gain with low implementation complexity, a coordination-aware channel-alignment–based PMI generation scheme, denoted by MSPN-Alignment is proposed by Bai et al.[12] to produce the coordinated PMI feedback. By expanding the sum of the effective channels in the parentheses, Equation 11 can be rewritten as

$$
\begin{aligned}
&[\boldsymbol{p}_1,\ldots,\boldsymbol{p}_M] \\
&= \arg\max_{[\boldsymbol{q}_{i_1},\ldots,\boldsymbol{q}_{i_M}] \in \wp^{M\times 1}} \sum_{k=1}^{N_{\mathrm{SUBC}}} \left(\sum_{m=1}^{M} \sum_{n=1}^{M} \boldsymbol{q}_{i_m}^{\mathrm{H}} \boldsymbol{H}_m^{\mathrm{H}}[k] \boldsymbol{R}_{\boldsymbol{n}_{\mathrm{I}}\boldsymbol{n}_{\mathrm{I}}}^{-1}[k] \boldsymbol{H}_n[k]\, \boldsymbol{q}_{i_n} \right) \\
&= \arg\max_{[\boldsymbol{q}_{i_1},\ldots,\boldsymbol{q}_{i_M}] \in \wp^{M\times 1}} \sum_{m=1}^{M} \sum_{n=1}^{M} \boldsymbol{q}_{i_m}^{\mathrm{H}} \left(\sum_{k=1}^{N_{\mathrm{SUBC}}} \boldsymbol{H}_m^{\mathrm{H}}[k] \boldsymbol{R}_{\boldsymbol{n}_{\mathrm{I}}\boldsymbol{n}_{\mathrm{I}}}^{-1}[k] \boldsymbol{H}_n[k] \right) \boldsymbol{q}_{i_n} \\
&= \arg\max_{[\boldsymbol{q}_{i_1},\ldots,\boldsymbol{q}_{i_M}] \in \wp^{M\times 1}} \left(\begin{array}{l} \displaystyle\sum_{m=1}^{M} \boldsymbol{q}_{i_m}^{\mathrm{H}} \sum_{k=1}^{N_{\mathrm{SUBC}}} \boldsymbol{R}_{m,m}[k]\, \boldsymbol{q}_{i_m} \\ \quad + 2\Re \left(\displaystyle\sum_{M=1}^{M-1} \sum_{n=m+1}^{M} \boldsymbol{q}_{i_m}^{\mathrm{H}} \sum_{k=1}^{N_{\mathrm{SUBC}}} r_{m,\mathrm{n}}[k]\, \boldsymbol{q}_{i_n} \right) \end{array} \right) \qquad (12) \\
&= \arg\max_{[\boldsymbol{q}_{i_1},\ldots,\boldsymbol{q}_{i_M}] \in \wp^{M\times 1}} \sum_{m=1}^{M} \underbrace{\left(\begin{array}{l} \boldsymbol{q}_{i_m}^{\mathrm{H}} \displaystyle\sum_{\mathrm{k}=1}^{N_{\mathrm{SUBC}}} \boldsymbol{R}_{m,m}[k]\, \boldsymbol{q}_{i_m} \\ \quad + 2\Re \left(\displaystyle\sum_{n=1}^{m-1} \boldsymbol{q}_{i_m}^{\mathrm{H}} \sum_{\mathrm{k}=1}^{N_{\mathrm{SUBC}}} \boldsymbol{r}_{m,\mathrm{n}}[k]\, \boldsymbol{q}_{i_n} \right) \end{array} \right)}_{=\gamma_{x,m}}.
\end{aligned}
$$

For a particular m, the term of $\gamma_{\mathrm{cross},m}$ in Equation 12 represents the mth TP local post-SNR in addition with the cross-multiplication components between mth and all nth TP, $n < m$. The key concept of Equation 12 is to break down the global SNR metric in Equation 11 into the individual terms $\gamma_{x,m}$, $m = 1,\ldots, M$, which is more advanced than the term of $\gamma_{m,i}[k]$ in Equation 5 by taking the coordination gain into account. Hence, MSPN-Alignment performs close to the optimal solution MSPN-Optimal as shown in the "Performance Evaluation" section. By focusing on $\gamma_{\mathrm{cross},m}$, the representation of Equation 12 results in the low-cost individual PMI selection for the mth TP as

$$
\boldsymbol{P}_m = \arg\max_{\boldsymbol{q}_{i_m} \in \wp} \gamma_{x,m} \qquad (13)
$$

Hence, $\boldsymbol{p}_m$ is obtained with individual metric optimization in an iterative fashion. As elaborated by Bai et al.[12], a significant implementation complexity gain is achieved by MSPN-Alignment in contrast to MSPN-Optimal. If the

"…96.5 percent of the computational effort is reduced by the iterative channel-alignment based PMI selection method."

calculation of $\boldsymbol{q}_i^{\mathrm{H}} \sum_{k=1}^{N_{\mathrm{SUBC}}} \boldsymbol{R}_{m,m}[k]\boldsymbol{q}_i$ in Equation 5 for a particular "i" has the complexity of O (1), then MSPN-Optimal has the complexity of O (Q^M) due to Q^M hypotheses of $[\boldsymbol{p}_1,\ldots,\boldsymbol{p}_M]$ to be tested in Equation 11. The coordination-aware feedback approach of MSPN-Alignment in Equation 13 requires $\boldsymbol{R}_{m,m}[k]$ and $\boldsymbol{r}_{m,n}[k]$ as given in Equation 12 yielding computational efforts of O (mQ) for the mth TP. With M TPs, it comes to $\mathrm{O}\left(\sum_{m=1}^{M} mQ\right) = \mathrm{O}\left(\left(M^2+M\right)/2Q\right)$. With the system configuration of $N_{\mathrm{T}} = 4$, $Q = 16$ and $M = 3$[1], the implementation complexity of MSPN-Alignment and MSPN-Optimal reads O (144) and O (4096), respectively. Obviously 96.5 percent of the computational effort is reduced by the iterative channel-alignment–based PMI selection method.

Performance Evaluation

The simulated DL performance of CoMP JT in a synchronized coordinated network with 3 TPs is depicted in Figure 2 with different PMI selection approaches. The LTE CoMP transmission in Release 11[1] is used as the baseline and is summarized in Table 1.

Parameter	Setting
Bandwidth / Sampling rate	10 MHz / 15.36 MHz
FfT size / Active subcarriers	1024 / 600
MIMO Configuration	4×4
Transmission mode	Closed-loop single layer to the target UE with 16 QAM and PMI selection per resource block, i.e. $N_{\mathrm{SUBC}} = 12$
Channel model	Spatial channel model, SCM-B[15]
Channel estimation	Estimator with mean square error of 0.01 for coefficients in channel matrix
Receiver type	IRC
Inter-cell co-channel interference	S/I = 10 dB (Signal to Interference Ratio)

Table 1: Downlink LTE parameters for the transmissions in coordinated HetNets
(Source: Department of Communication Technologies at the University of Duisburg-Essen, 2013)

In the scenario of synchronized networks, the location of the target UE is given by $s_{\mathrm{m}} = \mathrm{s}$, $m = 1,\ldots,M$ in Figure 1. Hence, the profile of $P_{\mathrm{W}} = [0, 0, 0]$ dB can be expected, where $P_{\mathrm{W}}(m)$ denotes the normalized average received signal power from the mth TP at the UE. At the required bit error rate (BER)

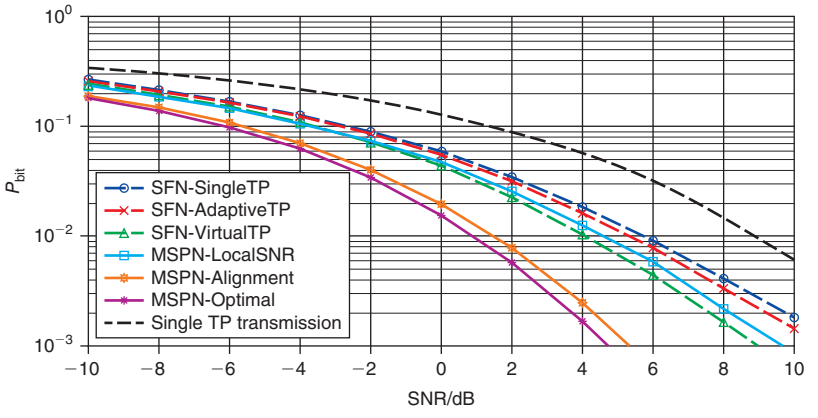


Figure 2: Performance comparison between PMI selection approaches in synchronized coordinated HetNets with the balanced received signal power profile P_W = [0, 0, 0] dB
(Source: Department of Communication Technologies at the University of Duisburg-Essen, 2013)

$P_{bit} = 10^{-2}$, SFN-SingleTP in Figure 2 only reaches the power gain of 3.2 dB from other TPs in contrast to the single TP transmission. SFN-AdaptiveTP achieves the additional dynamic point selection diversity of 0.4 dB in comparison with SFN-SingleTP, while SFN-VirtualTP reaches the power and beamforming gain jointly and outperforms single TP transmission with 4.8 dB. Regarding JT transmission in MSPN, MSPN-LocalSNR achieves non-coherent suboptimal beamforming gain and MSPN-Optimal performs the best by capturing the maximum JT benefits in coordinated networks via coordinated PMI feedback. The low complex coordination-aware channel alignment approach of MSPN-Alignment outperforms the MSPN-LocalSNR scheme with 3.2 dB at $P_{bit} = 10^{-2}$ and performs only 0.5 dB close to MSPN-Optimal. It is worth mentioning that SFN-VirtualTP achieves better performance than MSPN-LocalSNR. While the link beamforming gain by optimal PMI selection per TP is reached, the superposition of these beams is however not coherent with MSPN-LocalSNR. On the other hand, SFN-VirtualTP does maintain the power and beamforming again jointly via $\boldsymbol{H}_E[k]$ and as such it outperforms MSPN-LocalSNR.

"While the link beamforming gain by optimal PMI selection per TP is reached, the superposition of these beams is however not coherent with MSPN-LocalSNR."

CoMP Implementation in Unsynchronized Networks

In the investigation in the previous section, receive signals are assumed to be synchronized at the target UE. However, as we concentrate on the CoMP JT scheme to the target cell-edge UE, the transmission is commonly impacted by the timing offset (TO) between the receive signals. Such TOs inherit from the unsynchronized network due to two possible reasons: the propagation delays from different TPs are different, and the TPs do not

"…presence of TO in single point transmission network has been identified to have significant impact on DL transmission with distributed antenna ports."

transmit signals at the same time exactly. Regarding the layout in Figure 1, unsynchronized coordinated network can occur with $s_1 < s_2 < s_3$. In this case the propagation delays of the three signals, $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$, are different even if all TPs are synchronized over cellular network backhaul, yielding TO between these signals. Furthermore, such TO may indicate the difference of signal arrival time between heterogeneous eNodeBs or RRHs / distributed antenna ports[7] under the consideration of Scenarios 3 and 4 in coordinated networks. For the sake of the general representation, the signals from the first transmit antenna of the first TP, $\boldsymbol{e}_{1,1}$ is taken as the reference signals and $\tau_{m,t}$ are the relative TOs between $\boldsymbol{e}_{m,t}$ and $\boldsymbol{e}_{1,1}$, $m = 1,\ldots, M$, $t = 1,\ldots, N_T$. Certainly, it holds $\tau_{1,1} = 0$. The presence of TO in single point transmission network[14] has been identified to have significant impact on DL transmission with distributed antenna ports. The focuses of this section are given on the evaluation of the impact of TO on the coordinated PMI feedback and the transmissions to the target cell-edge UE in the unsynchronized coordinated networks.

Implementation Challenges

To facilitate the analysis of coordinated transmission with TO, the system functions in the section "System Model" is first extended in the unsynchronized HetNet. With the presence of $\tau_{m,t}$ in the receive signals $\boldsymbol{e}_{m,t}$ in the time domain (regardless of noise and interference components), the corresponding receive signal in the kth subcarrier $\boldsymbol{r}_{m,t}[k] \in \mathbb{C}^{N_R \times 1}$ in the frequency domain can be represented by

$$\begin{aligned}\boldsymbol{r}_{m,t}[k] &= \frac{1}{\sqrt{N_F}} \sum_{l=0}^{N_F-1} \boldsymbol{e}_{m,t}[l + N_c - \tau_{m,t} f_s] \exp\left(-j2\pi \frac{kl}{N_F}\right) \\ &= \frac{1}{N_F} \sum_{l=0}^{N_F-1} \sum_{k'=0}^{N_F-1} \left(\begin{array}{l} \boldsymbol{h}_{m,t}[k]\, p_{m,t} d[k'] \\ \times \exp\left(j2\pi \frac{k'(1-\tau_{m,t} f_s)}{N_F}\right) \exp\left(-j2\pi \frac{kl}{N_F}\right) \end{array} \right) \\ &= \boldsymbol{h}_{m,t}[k]\, p_{m,t}\, d[k] \exp(j\varphi_{m,t}[k]),\end{aligned} \tag{14}$$

where $\varphi_{m,t}[k] = -2\pi k\, \tau_{m,t} f_s \,/\, N_F$ represents the subcarrier specific phase rotation caused by the TO parameter $\tau_{m,t}$. Furthermore, N_F, N_C and f_s denote the FFT size, the cyclic prefix (CP) length and the sampling frequency of the system, respectively. In addition, $\boldsymbol{h}_{m,t}[k] \in \mathbb{C}^{N_R \times 1}$ gives the channel coefficient vector from the tth transmit antenna of the mth TP, while $p_{m,t}$ equals the precoder coefficient for the corresponding channel vector. According to Equation 14, system function of CoMP transmission in Equation 2 can be extended for the unsynchronized coordinated network such as

$$\begin{aligned}\boldsymbol{r}[k] &= \sum_{m=1}^{M} \sum_{t=1}^{N_T} \boldsymbol{r}_{m,t}[k] + \boldsymbol{n}_T[k] \\ &= \sum_{m=1}^{M} \underbrace{\boldsymbol{H}_m[k]\, \boldsymbol{\Phi}_m[k]}_{=\underline{\boldsymbol{H}}_m[k]} \boldsymbol{p}_m\, d[k] + \boldsymbol{n}_T[k]\end{aligned} \tag{15}$$

with the diagonal phase rotation matrix of

$$\boldsymbol{\Phi}_m[k] = \begin{pmatrix} \exp(\mathrm{j}\varphi_{m,1}[k]) & 0 & \cdots & 0 \\ 0 & \exp(\mathrm{j}\varphi_{m,1}[k]) & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \exp(\mathrm{j}\varphi_{m,N_\mathrm{T}}[k]) \end{pmatrix} \tag{16}$$

Obviously, Equation 15 differs from Equation 2 by the additional phase rotation matrix of $\boldsymbol{\Phi}_m$[k]. However, the systematic representation of receive CoMP JT signal in unsynchronized coordinated networks does not change since $\boldsymbol{\Phi}_m$[k] belongs to the part of channel matrix $\boldsymbol{H}_m[k]$. Hence, the last step in Equation 15 equals Equation 2 by replacing $\boldsymbol{H}_m[k]$ with $\underline{\boldsymbol{H}}_m[k]$. Consequently, the IRC scheme for the system Equation 15 equals

$$\boldsymbol{m}_{\mathrm{IRC}}^{\mathrm{T}} = \sum_{m=1}^{M} \boldsymbol{p}_m^H[k]\, \underline{\boldsymbol{H}}_m^H[k]\, \boldsymbol{R}_{\boldsymbol{n}_\mathrm{T}\boldsymbol{n}_\mathrm{T}}^{-1}[k],$$

yielding the averaged post-SNR over N_{SUBC} subcarriers of

$$\gamma = \sum_{k=1}^{N_{\mathrm{SUBC}}} E_\mathrm{d} \sum_{m=1}^{M} \boldsymbol{p}_m^{\mathrm{H}}[k]\, \underline{\boldsymbol{H}}_m^H[k]\, \boldsymbol{R}_{\boldsymbol{n}_\mathrm{T}\boldsymbol{n}_\mathrm{T}}^{-1}[k] \sum_{m=1}^{M} \underline{\boldsymbol{H}}_m[k]\, \boldsymbol{p}_m[k]. \tag{17}$$

PMI Selection under TO

For the PMI selection in unsynchronized coordinated networks, the transmit covariance matrix

$$\underline{\boldsymbol{R}}_{m,\,m}[k] = \underline{\boldsymbol{H}}_m^H[k]\, \boldsymbol{R}_{\boldsymbol{n}_\mathrm{T}\boldsymbol{n}_\mathrm{T}}^{-1}[k]\, \underline{\boldsymbol{H}}_m[k] = \boldsymbol{\Phi}_m^{\mathrm{H}}[k]\, \boldsymbol{R}_{m,\,m}[k]\, \boldsymbol{\Phi}_m[k]. \tag{18}$$

plays an important role. Furthermore, the modified cross-TP covariance matrix

$$\underline{\boldsymbol{r}}_{m,n}[k] = \underline{\boldsymbol{H}}_m^{\mathrm{H}}[k]\, \boldsymbol{R}_{\boldsymbol{n}_\mathrm{T}\boldsymbol{n}_\mathrm{T}}^{-1}[k]\, \underline{\boldsymbol{H}}_n[k] = \boldsymbol{\Phi}_m^{\mathrm{H}}[k]\boldsymbol{r}_{m,n}[k]\boldsymbol{\Phi}_n[k]. \tag{19}$$

impacts PMI selection as well. Obviously, $\boldsymbol{\Phi}_m[k] = \exp(\mathrm{j}\varphi_{m,1}[k])\boldsymbol{I}$ if all TOs are only dependent of the index m, that is $\tau_{k,m,1} = \tau_{k,m,t}$, $t = 2,\ldots, N_\mathrm{T}$, $\forall m$.

Considering the PMI selection for a single TP transmission with unsynchronized and distributed antenna ports, the TP specific PMI selection in Equation 5 turns into

$$\boldsymbol{p}_m = \arg\max_{\boldsymbol{q}_i \in \wp} \boldsymbol{q}_i^{\mathrm{H}} \left(\sum_{k=1}^{N_{\mathrm{SUBC}}} \underline{\boldsymbol{R}}_{m,m}[k] \right) \boldsymbol{q}_i, \tag{20}$$

according to the transmit covariance matrix in Equation 18. Hence, if there is no phase difference between antenna ports, Equation 20 equals the PMI selection in synchronized networks in Equation 5 and the selection results will not change. With the deployment of unsynchronized coordinated HetNets in SFN, SFN-SingleTP follows Equation 20 to select the PMI for the serving TP. SFN-AdaptiveTP does the same, but uses Equation 20 to select the PMI for the TP having the strongest DL channel. Similar to the single point transmission, no influence of $\tau_{k,m,t}$, $m = 1,\ldots,M$ on the PMI selection with SFN-Single TP and SFN-Adaptive TP exists if $\tau_{k,m,1} = \tau_{k,m,t}$, $t = 2,\ldots, N_\mathrm{T}$, $\forall m$. In SFN-VirtualTP the equivalent channel matrix $\underline{\boldsymbol{H}}_\mathrm{E}[k]$ is evaluated with the

phase rotation matrix $\boldsymbol{\Phi}_{\mathrm{m}}[k]$, such as $\underline{\boldsymbol{H}}_{\mathrm{E}}[k] = \sum_{m=1}^{M} \underline{\boldsymbol{H}}_{m}[k]$. Hence, the PMI selection (Equation 10) is changed to

$$\boldsymbol{p}_s = \arg\max_{\boldsymbol{q}_i \in \wp} \boldsymbol{q}_i^{\mathrm{H}} \sum_{\mathrm{k}=1}^{N_{\mathrm{SUBC}}} \left(\sum_{m=1}^{M} \underline{\boldsymbol{R}}_{m,m}[k] + \sum_{\substack{m,n \in [1,M] \\ m \neq n}} \underline{\boldsymbol{r}}_{m,n}[k] \right) \boldsymbol{q}_i,$$

according to Equation 18 and Equation 19. It is clear that the PMI selection based on SFN-VirtualTP will be affected by the different TOs between antenna ports in $\underline{\boldsymbol{R}}_{m,m}[k]$ and between TPs in $\underline{\boldsymbol{r}}_{m,n}[k]$.

The approach of MSPN-LocalSNR applies Equation 20 to select PMI for each TP. Hence, TOs between antenna ports result in impacts on the TP-specific PMI selection. If $\tau_{k,m,1} = \tau_{k,m,t}$, $\mathrm{t} = 2,\ldots,N_{\mathrm{T}}$, $\forall m$, the same PMI will be selected by MSPN-LocalSNR in unsynchronized as well as synchronized networks under the same instantaneous CSI, even if $\tau_{k,m,1} \neq \tau_{k,m',1}$, $m \neq m'$. Under the consideration of all TOs as the part of CSI, the coordinated PMI selection in Equation 11 can be updated with Equation 18 and Equation 19 resulting in

$$[\boldsymbol{p}_1,\ldots,\boldsymbol{p}_M] = \arg\max_{[\boldsymbol{q}_{i_1},\ldots \boldsymbol{q}_{i_M}] \in \wp^{M\times 1}} \left(\begin{array}{l} \sum_{m=1}^{M} \boldsymbol{q}_{i_m}^{\mathrm{H}} \sum_{k=1}^{N_{\mathrm{SUBC}}} \underline{\boldsymbol{R}}_{m,m}[k]\, \boldsymbol{q}_{i_m} \\ \quad + \sum_{\substack{m,n \in [1,M] \\ m \neq n}} \boldsymbol{q}_{i_m}^{\mathrm{H}} \sum_{k=1}^{N_{\mathrm{SUBC}}} \underline{r}_{m,n}[k]\, \boldsymbol{q}_{i_n} \end{array} \right) \tag{21}$$

for MSPN-Optimal. In the case of applying the low complex PMI selection of MSPN-Alignment in the unsynchronized coordinated network, the coordinated PMI selection based on iterative channel alignment in Equation 12 turns into

$$[\boldsymbol{p}_1,\ldots,\boldsymbol{p}_M]$$

$$= \arg\max_{[\boldsymbol{q}_{i_1},\ldots,\boldsymbol{q}_{i_M}] \in \wp^{M\times 1}} \sum_{m=1}^{M} \left(\underbrace{\boldsymbol{q}_{i_m}^{\mathrm{H}} \sum_{\mathrm{k}=1}^{N_{\mathrm{SUBC}}} \underline{\boldsymbol{R}}_{m,m}[k]\, \boldsymbol{q}_{i_m} + 2\Re\left(\sum_{n=1}^{m-1} \boldsymbol{q}_{i_m}^{\mathrm{H}} \sum_{k=1}^{N_{\mathrm{SUBC}}} \underline{r}_{m,n}[k]\, \boldsymbol{q}_{i_m} \right)}_{=\underline{\gamma}_{x,m}} \right) \tag{22}$$

$$= \arg\max_{[\boldsymbol{q}_{i_1},\ldots,\boldsymbol{q}_{i_M}] \in \wp^{M\times 1}} \sum_{m=1}^{M} \left(\begin{array}{l} \sum_{k=1}^{N_{\mathrm{SUBC}}} \left(\boldsymbol{q}_{i_m}^{\mathrm{H}} \boldsymbol{\Phi}_m^{\mathrm{H}}[k] \right) \boldsymbol{R}_{m,m}[k] \left(\boldsymbol{\Phi}_m[k]\, \boldsymbol{q}_{i_m} \right) \\ + 2\Re \left(\sum_{n=1}^{m-1} \sum_{k=1}^{N_{\mathrm{SUBC}}} \left(\boldsymbol{q}_{i_m}^{\mathrm{H}} \boldsymbol{\Phi}_n^{\mathrm{H}}[k] \right) r_{m,n}[k] \left(\boldsymbol{\Phi}_m[k]\, \boldsymbol{q}_{i_m} \right) \right) \end{array} \right)$$

based on which, the final PMI selection in Equation 13 is changed to be

$$\boldsymbol{p}_m = \arg\max_{\boldsymbol{q}_{i_m} \in \wp} \underline{\gamma}_{x,m} \ . \tag{23}$$

As it can be seen from the previous analysis, the coordinated PMI feedback generated by MSNP-Optimal and MSPN-Alignment are more affected by the TO through $\underline{\boldsymbol{R}}_{m,m}[k]$ and $\boldsymbol{r}_{m,\mathrm{n}}[k]$. However, all the updated representation of PMI selection approaches for the unsynchronized coordinated networks are the same as for the synchronized one, except that the per-TP-channel matrices

$\underline{\boldsymbol{H}}_m[k]$ are replaced by $\underline{\boldsymbol{H}}_m[k]$. Since $\boldsymbol{\Phi}_m$[k] is dependent on subcarrier index k, coefficients in $\underline{\boldsymbol{H}}_m$[k] must be obtained in the same subcarrier. With the proper receiver design, the phase rotation matrix $\boldsymbol{\Phi}_m$[k] must be covered directly in channel estimation process with $\underline{\boldsymbol{H}}_m[k]$.

Furthermore, the second step in Equation 22 indicates that the phase rotation matrix $\boldsymbol{\Phi}_m$[k] changes the phases of predefined coefficients of the precoders. Hence, the actual applied equivalent precoder per TP differs from the selected precoder (via PMI). Such difference is considered within the computation of covariance matrices of $\sum_{k=1}^{N_{\mathrm{SUBC}}} \underline{\boldsymbol{R}}_{m,m}[k]$ and $\sum_{k=1}^{N_{\mathrm{SUBC}}} \boldsymbol{r}_{m,\mathrm{n}}[k]$ so that the applied coordinated precoder matches to the desired one, that is, the one maximizing post-SNR in Equation 17.

Performance Evaluation

The performance of COMP JT in an unsynchronized coordinated network is depicted in Figure 3. This scenario is realistic for the transmission to the target cell-edge UE, which sees different distances to different TPs and faces TO of $[\tau_{k,1,1}, \tau_{k,2,1}, \tau_{k,3,1}] = [0.0, 1.0, 2.0]$ μs and an imbalanced received signal power profiler $P_{\mathrm{W}} = [0, -3, -6]$ dB. With quasi-collocated antenna ports at a single TP[16], it holds $\tau_{k,m,1} = \tau_{k,m,t}$, $t = 2,\ldots,N_{\mathrm{T}}$, $\forall m$. Comparing with Figure 2, all PMI approaches degrade in the unsynchronized and power-imbalanced networks in Figure 3. However, the main factor resulting in such degradation is the imbalanced receive signal power. In comparison to the 3.2 dB JT gain in Figure 2, both SFN-SingleTP and SFN-AdaptiveTP converge together and are only 0.9 dB better than the single TP transmission due to the loss of received signal power from TP2 and TP3. Similarly, SFN-VirtualTP and MSPN-LocalSNR perform nearly the same and provide about 1.8 dB JT gain. The benefits observed in SFN-VirtualTP are no longer significant in contrast to MSPN-LocalSNR due to the power imbalance and the phase rotation matrix. Different from coordination-unaware PMI selection approaches discussed in the section "Coordination-Unaware PMI Selection," the MSPN-Optimal approach still provides significant JT gain of 4.7 dB and shows the lower bound in unsynchronized networks. The MSPN-Alignment in Figure 3 keeps near optimal performance, that is, only 0.3 dB away from MSPN-Optimal. Obviously, taking the phase rotation matrix $\boldsymbol{\Phi}_m$[k] into account in each iterative selection stage in Equation 22 enables a robust and low complex approach of MSPN-Alignment selecting coordinated PMIs and aligning channels from different TPs coherently.

Continuing the study of the impact of TO on JT to the target cell-edge UE, the scenario in Figure 3 is further extended with additional TO between distributed antenna ports at each TP, such as $\tau_{k,m,t} = (\tau_{k,m,1} + 0.2\,(t-1))\mu\mathrm{s}$, $t = 1,\ldots,4$. The simulation results are depicted in Figure 4. With the presence of TO $[\tau_k,1,1, \ldots, \tau_k,1,4]$ between non-collocated antenna ports of the serving TP, single TP transmission is 0.3 dB worse in Figure 4 than in Figure 3. Such a loss is caused by the subcarrier dependent phase rotation matrix $\boldsymbol{\Phi}_m$ [k]. Within N_{SUBC} of 12 subcarriers, the complex-valued channel coefficient

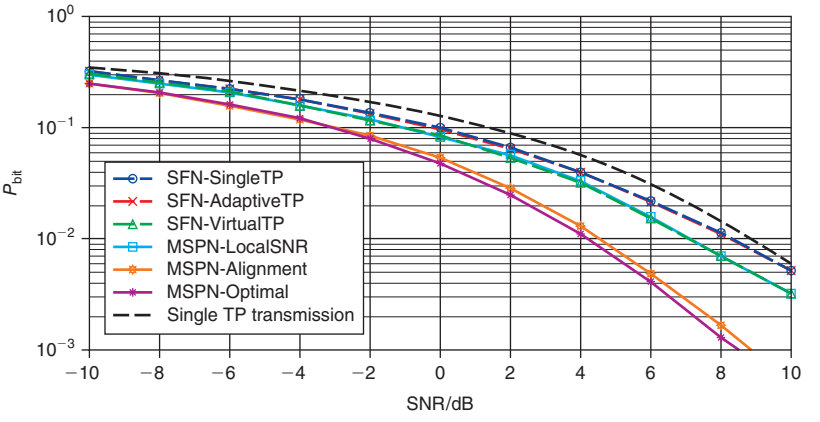


Figure 3: Performance comparison between PMI selection approaches in unsynchronized coordinated HetNets with the imbalanced received signal power profile $P_W = [-6, -3, 0]$ dB and the receive signal TO per TP (Source: Department of Communication Technologies at the University of Duisburg-Essen, 2013)

related to the last transmit antenna port has a phase change of 40 degrees due to TO. Hence the selected PMI for a single TP in Equation 20 is just a compromised result in one resource block including significant different coefficients. The same loss of 0.3 dB between results in Figure 4 and Figure 3 can be observed for selection approaches based on serving TP channel and dynamic selected TP channel, that is, SFN-SingleTP and SFN-AdaptiveTP, respectively. However, the equivalent sum channel $\boldsymbol{H}_E$ [k] based approach

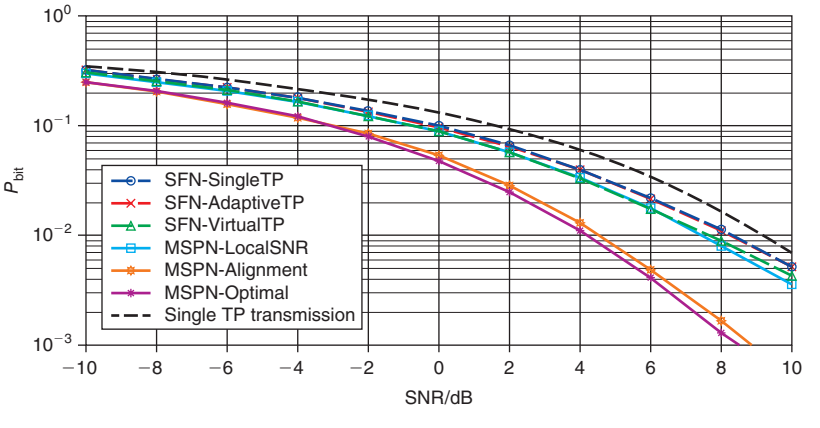


Figure 4: Performance comparison between PMI selection approaches in unsynchronized coordinated HetNets with the imbalanced received signal power profile $P_W = [-6, -3, 0]$ dB and the receive signal TO per antenna port (Source: Department of Communication Technologies at the University of Duisburg-Essen, 2013)

(SFN-VirtualTP) and per-TP-individual-selection approach (MSPN-LocalSNR) face the loss of about 0.5 dB since the TOs occur in every TP channel. While the MSPN-Optimal using full-blown search gives coordinated PMIs and achieves unchanged lower bound in Figure 4 as in Figure 3, a cause of concern is the fact that MSPN-Alignment keeps also nearly unchanged, that is, only 0.3 dB away from MSPN-Optimal. The consideration of the cross-TP covariance matrix handles the TO between TPs and between antenna ports in MSPN-Alignment in a simple and proper way so that the desired coordinated PMIs yielding major JT gain can be generated. Hence, the impact of the TO is negligible at the SNR range, at which JT gain can be mainly captured by the channel alignment between unsynchronized TP links rather than the TO alignment.

"…the impact of the TO is negligible at the SNR range, at which JT gain can be mainly captured by the channel alignment between unsynchronized TP links rather than the TO alignment."

Conclusion

In this article, system performance of CoMP JT in coordinated networks in beyond LTE Release 11 systems has been intensively studied. The main focus of the investigation is given on the transmission performance analysis with PMI selection approaches for the cell-edge users in synchronized as well as in unsynchronized coordinated networks. In both networks, the approaches generating coordinated PMIs by aligning the channels between transmission points outperform conventional CoMP scenario without such alignment by at least 3 dB. Furthermore, the systematic functions of PMI selection approaches remain unchanged, when the phase rotation effects $\Phi_m[k]$ caused by TO in unsynchronized networks is taken into account during channel estimation. In particular, reliable channel alignments and coordinated precoders can be achieved via the low complex approach of aligning PMI selection iteratively. Consequently, near maximum performance improvements can be guaranteed by this coordination-aware approach to the cell-edge users at a low implementation cost, even with the presence of high timing offsets between transmission points.

Acknowledgements

The authors would like to thank Erfan Majeed for his valuable comments to this article and wish to gratefully acknowledge the support of their colleagues at Intel Mobile Communications and at the Department of Communication Technologies of the University of Duisburg-Essen.

References

[1] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception," 3GPP Tech. spec. 36.101 v11.1.0, Sophia Antipolis, July 13, 2012.

[2] 3GPP, "Coordinated multi-point operation for LTE physical layer aspects," 3GPP Tech. rep. 36.819 v11.1.0, Sophia Antipolis, Dec. 2011.

[3] ETRI, "Per-cell precoding methods for downlink joint processing CoMP," 3GPP Tech. rep. R1-083546, Prague, Czech Rep., Sept. 29, 2008.

[4] Peng, B., W. Chen, Y. Zhang, and M. Lei, "A joint scheduling and beamforming method based on layered limited feedback," in *Proceedings of PIRMC 2012*, Sydney, NSW, Sept. 9–12, 2012, vol. 1, pp. 553–558.

[5] Su, D., X. Hou, and C. Yang, "Quantization based on per-cell codebook in cooperative multi-cell systems," in *Proceedings of WCNC 2011*, Cancun, Quintana Roo, Mar. 28–31, 2011, vol. 1, pp. 1753–1758.

[6] Bai, Z., C. Spiegel, G. H. Bruck, P. Jung, M. Horvat, J. Berkmann, C. Drewes, and B. Gunzelmann, "Closed loop transmission with precoding selection in LTE/LTE-advanced system," in *Proceedings of ISABEL 2010*, Rome, Italy, Nov. 7–10, 2010, invited paper.

[7] HiSilicon Huawei, "Discussion on the impact of time misalignment in real life transmission," 3GPP, Tech. rep. R1-112910, Zhuhai, China, Oct. 10-14, 2011.

[8] Kotzsch, V., W. Rave, and G. Fettweis, "Interference cancellation and suppression in asynchronous cooperating base station systems," in *Proceedings of WSA 2012*, Dresden, Germany, Mar. 7–8, 2012, vol. 1, pp. 78–85.

[9] Liu, Y., Y. Li, D. Li, and H. Zhang, "Space-time coding for time and frequency asynchronous CoMP transmissions," in *Proceedings of WCNC 2013*, Shanghai, China, Apr. 7–10, 2013, vol. 1, pp. 2632–2637.

[10] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation," 3GPP Tech. spec. 36.211 v10.2.0, Sophia Antipolis, Dec. 2011.

[11] Bai, Z., B. Badic, S. Iwelski, R. Balraj, T. Scholand, G. Bruck, and P. Jung, "Evaluation of Implicit Feedback in Coordinated Multipoint Transmission Beyond LTE-Advanced," in *Proceedings VTC 2013-Spring*, Dresden, Germany, June 2–5, 2013.

[12] Bai, Z., B. Badic, R. Balraj, S. Iwelski, T. Scholand, G. Bruck, and P. Jung, "Near Optimal Implicit Feedback Generation for Coordinated Multipoint Transmission," in *IEEE Commun. Lett.* PP.99 (Sept. 2013), pp. 1–4.

[13] Duplicy, J., B. Badic, R. Balraj, R. Ghaffar, P. Horvath, F. Kaltenberger, R. Knopp, I. Z. Kovács, H. T. Nguyen, D. Tandur, and G. Vivier, "MU-MIMO in LTE systems," in *EURASIP Journal*

on Wireless Communications and Networking 2011, vol. 2011, pp. 1–13, 2011.

[14] HiSilicon Huawei, "Discussion on geographically separated antennas," 3GPP Tech. rep. R4-124283, Qingdao, China, Aug. 13–17, 2012.

[15] 3GPP, "Physical layer aspect for evolved Universal Terrestrial Radio Access (UTRA)," 3GPP Tech. rep. 25.814 v7.1.0, Sophia Antipolis, Sept. 2006.

[16] Intel Corporation, "Remaining issues of antenna ports quasi co-location definition," 3GPP Tech. rep. R1-123425, Qingdao, China, Aug. 13–17, 2012.

Author Biographies

Zijian Bai is a PhD student and a member of the scientific staff at the department of Communication Technologies at the University of Duisburg-Essen. His current research interests are wireless system modeling, MIMO radio channel modeling, advanced interference-aware signal transmission/detection, signal sensing in cognitive radio and network optimization for the delivery of high bit-rate services in cellular systems. Email: Zijian.bai@kommunikationstechnik.org.

Stanislaus Iwelski is member of the scientific staff at the department of Communication Technologies at the University of Duisburg-Essen. His research interests are in signal processing in heterogeneous networks, focusing on interference-aware receivers and coordinated multipoint transmission. Email:stanislaus.iwelski@kommunikationstechnik.org.

Biljana Badic is a system engineer in Product Engineering Group, has been with Intel since 2010, and has participated in development of 4G wireless modems with the particular focus on interference cancellation 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. Email: biljana.badic@intel.com.

Guido Bruck received the Dr.-Ing. (PhD equivalent) with the focus on source image coding from the Gerhard-Mercator-University Duisburg. He joined the department of Communication Technologies in June 2000 and since then has worked in the field of software-defined radio and on adaptation of source coding methods to mobile communication systems. He is the senior manager at the department of Communication Technologies and in charge of several industrial projects and subprojects. Email: guido.bruck@kommunikationstechnik.org.

Peter Jung is a chaired professor for the department of Communication Technologies at the University of Duisburg-Essen. He received the Dr.-Ing. (PhD EE equivalent) and Dr.-Ing. habil. (DSc EE equivalent), with the

focus on microelectronics and communications technology, from University of Kaiserslautern. He was senior director of concept engineering wireless baseband with Siemens AG, now Infineon Technologies AG, and has been a member of the editorial board of *IEEE Transactions on Wireless Communications* and *Springer Journal of Wireless Personal Communications*. His areas of interest include wireless communication technology, software-defined radio, and system-on-a-chip integration of communication systems. Email: peter.jung@kommunikationstechnik.org.

LTE PHYSICAL LAYER ASPECTS FOR eMBMS: CONCEPT AND PERFORMANCE EVALUATION

Contributors

Rainer Bachl
Platform Engineering Group,
Intel Corporation

Rajarajan Balraj
Platform Engineering Group,
Intel Corporation

Ingmar Groh
Platform Engineering Group,
Intel Corporation

Consumption of bandwidth hungry applications like video streaming over mobile networks is on the rise due to the success of ubiquitous devices like smart phones and tablets. In order to meet future demands, enhanced multimedia broadcast and multicast services (eMBMS) was defined in the Release 9 specification of Long Term Evolution (LTE). This article presents the physical aspects of eMBMS in a easy-to-read manner to not only enable the reader to understand the technical details but also appreciate the design decisions made for eMBMS.

Introduction

Mobile multicasting and broadcasting is becoming increasingly attractive with smartphones being omnipresent. With the limited amount of spectrum available for mobile communications and with the ever-increasing demand for mobile services, spectral efficiency of mobile broadcasting and smooth integration with existing mobile unicast communication standards has become a focus for the mobile communications industry.

Traditionally, cellular systems have been intended for data transmission for a single user employing a dedicated point-to-point transmission, whereby it is possible to optimize transmission parameters for each user independently. In contrast, broadcast transmissions are point-to-multipoint transmissions, where adaptation of the transmission parameters is not possible in general and the transmission has to be designed for a worst-case channel scenario. The worst-case channel scenarios encountered in point-to-point transmissions are usually at the cell edge, where the interference from neighboring cells is strongest. In broadcast networks and more recently also in mobile multicast, the usage of a single frequency network (SFN) exploits over-the-air combining of the signals from multiple transmitters and thus also avoids interference at the cell edges of a mobile network by combining the signals from multiples cells.

"…the usage of a single frequency network (SFN) exploits over-the-air combining of the signals from multiple transmitters and thus also avoids interference at the cell edges of a mobile network by combining the signals from multiples cells."

The 3GPP standard for LTE includes the enhanced multimedia broadcast and multicast service (eMBMS), which has evolved from earlier standards such as 3G MBMS. In this article, we outline details of the eMBMS physical layer and demonstrate the advantages of the design decisions made for eMBMS. In particular, we introduce physical layer transmission format of eMBMS in the next section, and the channel scenarios and the benefits arising from the usage of single frequency networks are discussed in the subsequent section.

Physical Layer for eMBMS

In this section, we describe channels employed for transmitting eMBMS, as well as the frame and subframe structure used in the transmission. Single frequency networks (SFN) are introduced, and it is outline how a receiver can adapt to SFN channel scenarios.

Channels Employed for eMBMS

As illustrated in Figure 1, a set of logical channels is defined for eMBMS that correspond to certain MBMS services and consist of the multicast traffic channel (MTCH) carrying the data and a multicast control channel (MCCH) providing the necessary control information to receive MBMS services, including subframe allocation and modulation and coding scheme.

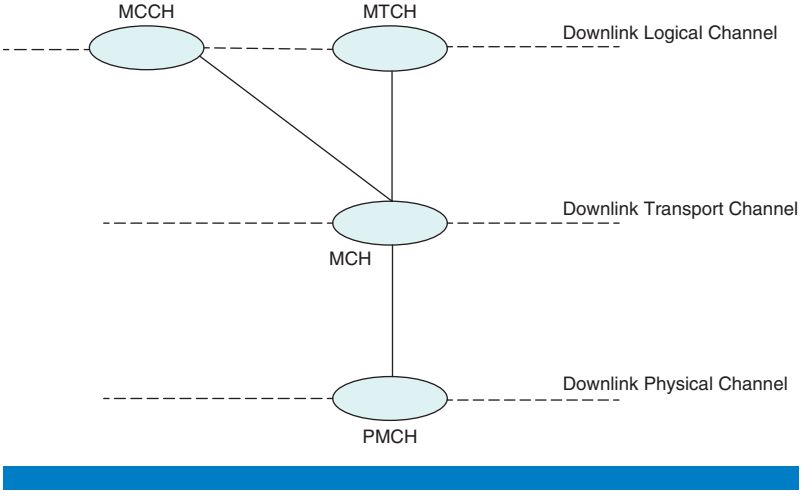


Figure 1: Logical channel mapping (Source: Intel Corporation 2013)

Several MTCH and MCCH are used to support different services concurrently and the MTCH and MCCH are then multiplexed into the downlink transport channel called the multicast channel (MCH). Finally, MCH is then transmitted via the physical multicast channel (PMCH). In the following sections, we focus on the physical layer transmission of eMBMS with PMCH.

"Several MTCH and MCCH are used to support different services concurrently and the MTCH and MCCH are then multiplexed into the downlink transport channel called the multicast channel (MCH)."

Frame Structure for eMBMS

To adapt the data rates for eMBMS, the LTE frame structure allows for flexible time-multiplexing of subframes for unicast with the physical downlink shared channel (PDSCH) and multicast transmission using PMCH, which is illustrated in Figure 2. In particular, up to 6 subframes of 1 ms within a radio frame consisting of 10 subframes can be defined to be subframes for PMCH transmission. Other subframes that cannot be used for PMCH transmission contain the broadcast channel (PBCH) or synchronization signals (PSS, SSS), or are special subframes or uplink subframes for TDD duplex mode.

"…subframes that cannot be used for PMCH transmission contain the broadcast channel (PBCH) or synchronization signals (PSS, SSS), or are special subframes or uplink subframes for TDD duplex mode."

To further adapt the data rates for eMBMS, a pattern of 1 or 4 radio frames with PMCH transmissions can defined and repeated every {1, 2, 4, 8, 16, 32}

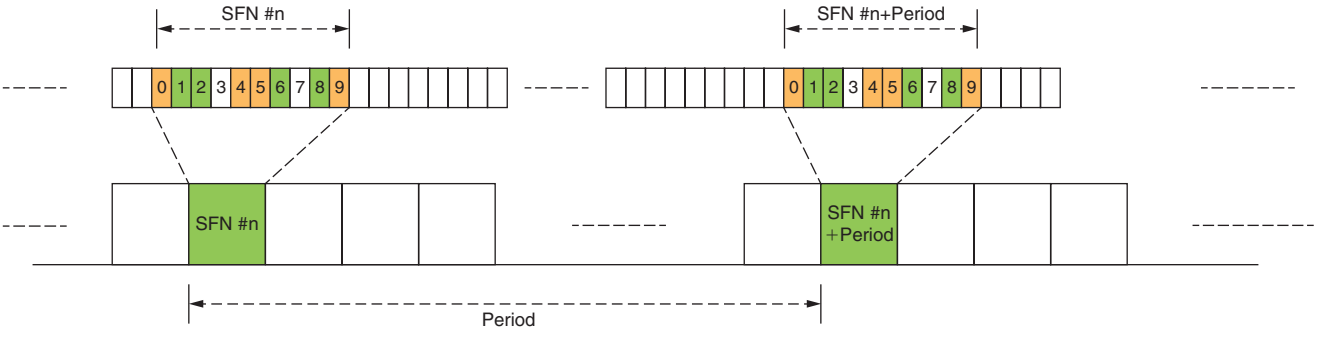


Figure 2: Example of PMCH pattern length of 1 radio frame repeating with a configurable period (Source: Intel Corporation 2013)

radio frames. Therefore, the subframe ratio available for MBMS is in the range from 1/320 to 192/320.

Single Frequency Networks for eMBMS

Like other broadcast technologies such as ATSC M/H, DVB-T, DVB-H, and DVB-T2, eMBMS is specified to use a single frequency network (SFN) configuration where multiple eNodeBs are transmitting the same eMBMS contents synchronously. In particular, the eNodeBs transmit identical signals for PMCH, which are then combined over-the-air and received by the UE. The eNodeBs transmitting the same PMCH signal make up a multicast/broadcast single frequency network (MBSFN) area and the PMCH signal is transmitted as part of MBSFN subframes. Note that overlap between different MBSFN areas is supported, and one eNodeB may belong to up to eight different MBSFN areas, which can be organized to be venue specific, region specific or even nationwide.

"The eNodeBs transmitting the same PMCH signal make up a multicast/broadcast single frequency network (MBSFN) area and the PMCH signal is transmitted as part of MBSFN subframes."

The eNodeBs in the same MBSFN area need to be synchronized, which usually also implies that the entire network is a synchronized network. Although these eNodeBs are synchronized, their signals arrive at the UE with certain propagation delays. Some examples for the resulting power delay profile

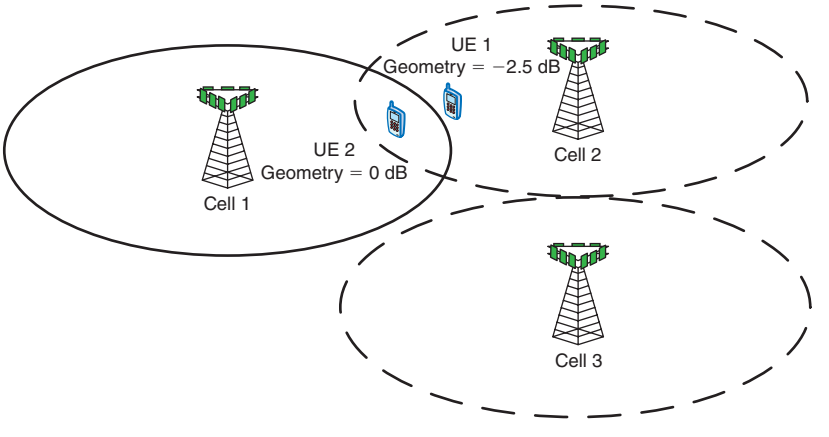


Figure 3: UE geometries with respect to unicast serving cell 1 (Source: Intel Corporation 2013)

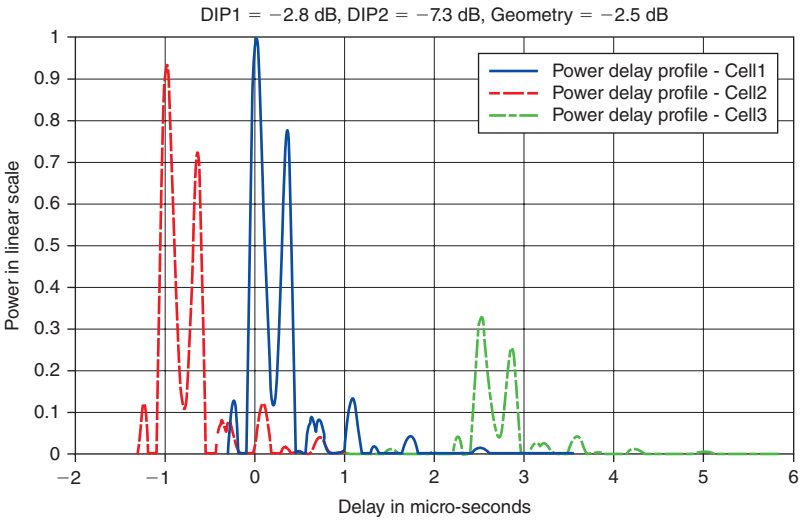


Figure 4: Example for power delay profile perceived at UE with geometry of –2.5dB
(Source: Intel Corporation 2013)

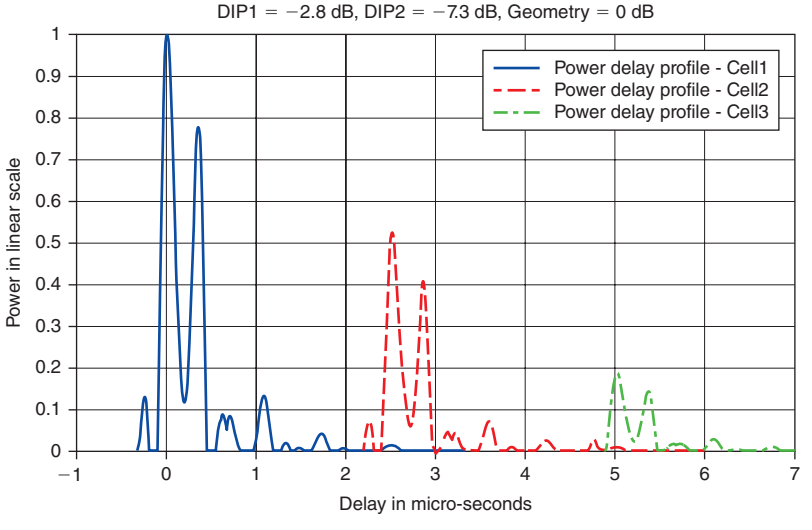


Figure 5: Example for power delay profile perceived at UE with geometry of 0dB
(Source: Intel Corporation 2013)

are depicted in Figures 4 and 5 for the UE geometries shown in Figure 3, where the dominant interferer proportion DIP1 and DIP2 correspond to the interference power in unicast scenarios from cell 2 and cell 3.

While the signals from cell 2 and cell 3 cause interference in a unicast transmission, these signals combine constructively in a MBSFN scenario provided that the length of the power delay profile is shorter than the length of the cyclic prefix in orthogonal frequency division multiplexing (OFDM), which is the same condition as in unicast transmissions for avoiding inter-

symbol interference (ISI). Note that all cells in a particular MBSFN area also transmit the MBSFN subframe using the same scrambling code.

"Since the composite power delay profile of the signal from multiple cells is longer than the power delay profile from a single cell, an extended cyclic prefix length has been chosen for the MBSFN region of a subframe…"

Since the composite power delay profile of the signal from multiple cells is longer than the power delay profile from a single cell, an extended cyclic prefix length has been chosen for the MBSFN region of a subframe, regardless of the choice of the cyclic prefix length for subframes used in unicast transmissions. The length of the extended cyclic prefix for MBSFN subframes is 16.7 μs, which covers a propagation delay corresponding to a distance of 5 km.

MBSFN Subframe Structure

Any MBSFN subframe also includes support for unicast transmissions with PCFICH, PDCCH and PHICH, which are multiplexed within the non-MBSFN region of the subframe (Figure 6). In particular, PCFICH conveys the number of PDCCH symbols, PDCCH conveys downlink control information (DCI) for uplink grants and power control, while PHICH conveys acknowledgement (ACK or NACK) for the hybrid automatic repeat request (HARQ) of corresponding to subframes where the user has been previously scheduled for transmissions in the uplink. When PMCH is transmitted in a MBSFN subframe, PDCCH does not include any DCI for downlink grants and, hence, the number of OFDM symbols carrying PDCCH can be smaller than the ones in subframes used for unicast transmission. Therefore the number of OFDM symbols supporting unicast in MBSFN subframes has been limited in the LTE specifications to a maximum of 2 OFDM symbols.

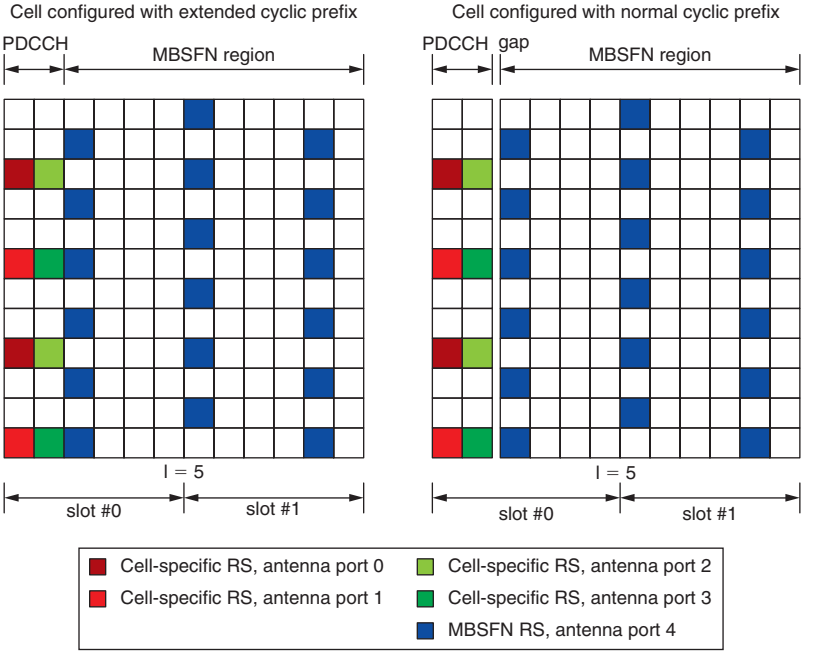


Figure 6: Example for MBSFN subframe structure for a cell configured with 4 cell-specific antenna ports and either extended or normal cyclic prefix
(Source: Intel Corporation 2013)

The non-MBSFN region of a MBSFN subframe is transmitted using the cyclic prefix format for which the particular cell is configured, which can be either normal or extended cyclic prefix, while the MBSFN region is always transmitted using the extended cyclic prefix format. In case the cell is configured with normal cyclic prefix, there is a change of the cyclic prefix format between non-MBSFN region and MBSFN region, which also implies a transmission gap between the PDCCH symbols and the MBSFN region as depicted in Figure 6.

"The non-MBSFN region of a MBSFN subframe is transmitted using the cyclic prefix format for which the particular cell is configured, which can be either normal or extended cyclic prefix, while the MBSFN region is always transmitted using the extended cyclic prefix format."

The time-frequency grid of MBSFN subframes is shown in Figure 6 for one resource block consisting of 12 subcarriers for the case that the cell is configured with 4 cell-specific antenna ports. In case the cell is configured with 4 cell-specific antenna ports, there are always two PDCCH symbols transmitted as shown in Figure 6, while in case of less than 4 cell-specific antenna ports the reference symbols for antenna port 2 and 3 are not present and the number of PDCCH symbols may be reduced to one.

Since the MBSFN channel corresponds to the superposition of the channels from all cells in the MBSFN area, the MBSFN channel is different from the unicast channel and channel estimation for the MBSFN channel is supported with dedicated MBSFN reference symbols as depicted in Figure 6. There are dedicated MBSFN reference symbols corresponding only to one antenna port because eMBMS is supported only for single layer transmission. Since the power delay profile of the MBSFN channel spans more time than the one for the unicast channel, the MBSFN channel is also more frequency selective as compared to the unicast channel. To account for the larger frequency selectivity of MBSFN channels, the MBSFN reference symbols spacing in frequency direction is smaller than the cell-specific reference symbol spacing (see Figure 6), thus making possible channel estimation with relatively small mean squared error in the UE receiver.

"To account for the larger frequency selectivity of MBSFN channels, the MBSFN reference symbols spacing in frequency direction is smaller than the cell-specific reference symbol spacing…"

With a single layer transmission, PMCH consists of only one transport block that can be decoded per subframe. The modulation and coding scheme is fixed per MBSFN area and selected usually either to allow reception even for the worst-case channel conditions within the particular MBSFN area or at least to provide coverage for a large percentage of users of the MBSFN area.

Receiver Adaptation to MBSFN Channel Characteristics

The characteristics of unicast channels and MBSFN channels are substantially different and the receiver needs to adapt to the particular propagation scenario encountered. As discussed in the previous section, the delay spread for MBSFN channels is larger than the one for unicast channels, which prompts the usage of different channel estimation filters based on MBSFN reference signals and cell-specific reference signals.

"The characteristics of unicast channels and MBSFN channels are substantially different and the receiver needs to adapt to the particular propagation scenario encountered."

Since MBSFN channels correspond to the superposition of the channels from multiple cells, the received signal power for MBSFN channels could be significantly higher than the received signal power for unicast channels. While eNodeB transmit power is often adapted for the particular channel in unicast

transmissions, the transmit power of the eNodeB for multicast transmissions is usually chosen to be constant. However, the power of cell-specific reference signals and MBSFN reference signals can also be set independently, which is a further reason to choose a dedicated and MBSFN-area–specific gain control for multicast reception.

The UE receiver compensates frequency offset due to errors in the eNodeB carrier frequency and the Doppler shift caused by the velocity of the UE with respect to the serving eNodeB. For reception with MBSFN channels, however, the frequency offset will be less pronounced since relative velocities with respect to multiple eNodeBs participating in the MBSFN transmission need to be considered and there is not only a single Doppler shift creating some distinct frequency offset.

The receiver symbol timing is usually chosen such that inter-symbol interference (ISI) is minimized. For unicast transmissions, increasing the propagation delay with increased distance to the serving eNodeB implies that there is also a larger receiver symbol timing offset. For multicast transmissions, however, the symbol timing in the receiver should be chosen to minimize ISI for the MBSFN channel. Hence, the receiver symbol timing needs to be chosen specifically for the MBSFN area considered, and there is no longer a simple relationship between symbol timing and propagation delay with respect to a single transmission point (TP).

MBSFN Channel Scenarios and Performance

MBSFN channels in synchronous networks do not suffer from interference generated by neighboring cells, provided (1) the neighboring cells participate in the transmission of the same MBSFN subframe and therefore are part of the same MBSFN area, and (2) the channel power delay profile for all cells are within the same cyclic prefix length. The following sections present simulation results for post-equalization SINR and spectral efficiency when employing MBSFN subframes.

Post-Equalization SINR in Case of Dominant Interferer Proportions

The benefit from employing MBSFN channels can be evaluated by comparing the post-equalization SINR for MBSFN channels with the one from unicast channels. In Figures 7 and 8, we compare the post-equalization SINR for MBSFN channels with the one for unicast channels whereby the UE has two receiver antennas. The receiver employs maximum ratio combining (MRC) for MBSFN subframes, and it employs MRC or interference rejection combining (IRC) for the unicast channel, whereby the latter is equivalent to using a MMSE receiver.[1] The scenario simulated consists of two dominant interferer proportions (DIP)[2], respectively, that have been chosen to be represent typical interferer profiles for unicast transmissions.[2][3] A homogeneous network with full load and inter-site

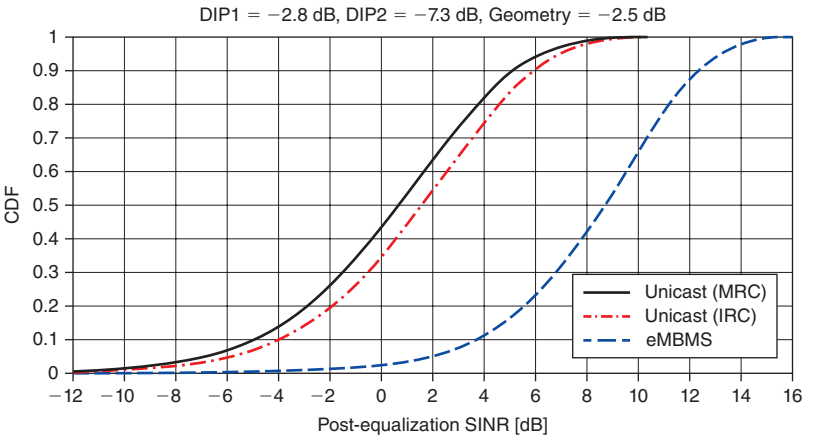


Figure 7: Post-equalization SINR for MBSFN and unicast at geometry of –2.5 dB
(Source: Intel Corporation 2013)

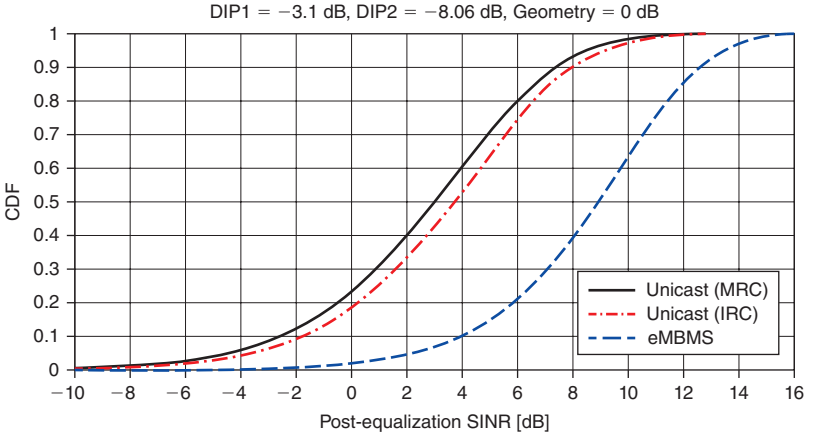


Figure 8: Post-equalization SINR for MBSFN and unicast at geometry of 0 dB
(Source: Intel Corporation 2013)

distance of 500 m was used in the simulations, a carrier frequency of 2 GHz and bandwidth of 10 MHz was selected; handover margin was 3 dB.

As can be seen from Figures 7 and 8, the largest gain of MBSFN transmissions over unicast transmissions is given for the low geometries, which can be as high as 8 dB in post-equalization SINR in this scenario. MBSFN transmissions also increase the SINR at higher geometries closer to the cell center, but there is no additional benefit for eMBMS transmissions since the modulation and coding scheme need to be chosen for reception with the worst-case channel conditions considered.

"…the largest gain of MBSFN transmissions over unicast transmissions is given for the low geometries,…"

Post-Equalization SINR for Multicast vs. Unicast in Handover Scenario

We consider a simple handover scenario with a network consisting of two eNodeBs, whereby the UE is moving from transmission point 1 (TP1) towards transmission point 2 (TP2). The scenario is depicted in Figure 9 for a homogeneous network with TP1, TP2 having equal power of 46 dBm, a distance of 1800 m, and handover margin of 3 dB, while Figure 10 shows a

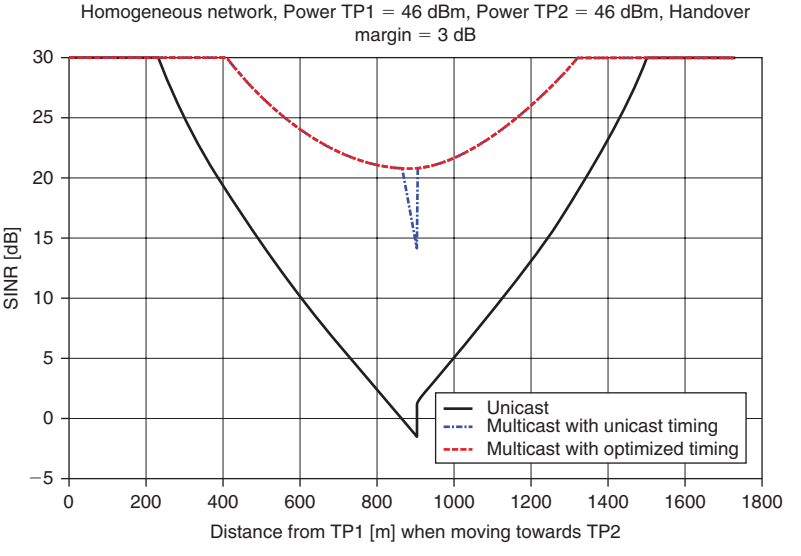


Figure 9: Post-equalization SINR for unicast and MBSFN channels in a homogeneous network scenario
(Source: Intel Corporation 2013)

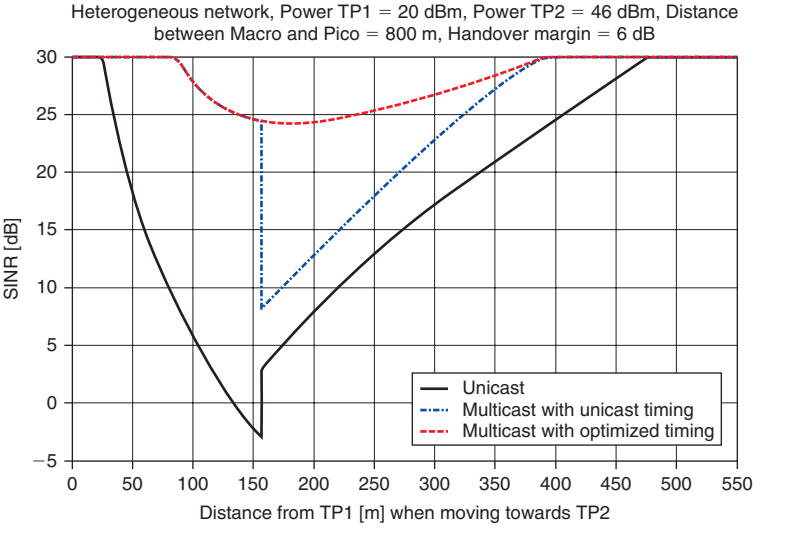


Figure 10: Post-equalization SINR for unicast and MBSFN channels in a heterogeneous network scenario
(Source: Intel Corporation 2013)

unicast handover scenario for a heterogeneous network with a 20-dBm pico cell and a 46-dBm macro cell spaced by 800 m and 6 dB handover margin. Note that support for heterogeneous network scenarios is one of the key enhancements in LTE Release 12.

Figures 9 and 10 depict the post-equalization SINR versus the position of the UE with respect to TP1. In the cell center of TP1 and TP2, the SINR is merely limited by transmitter and receiver imperfections. When the UE moves from TP1 towards TP2, however, the interference from TP2 becomes significant in unicast reception and decreases the SINR to values below 0dB. When the SINR for reception from TP2 is larger than the SINR for reception from TP1 plus handover margin, handover from TP1 to TP2 takes place, which corresponds to the discontinuity in the unicast SINR curves. With the handover from TP1 to TP2, the UE also adjusts the symbol timing to minimize ISI for reception from TP2 instead of TP1. When the UE moves further towards the cell center of TP2, the interference from TP1 decreases and SINR is again increased. For a homogeneous network scenario depicted in Figure 9, the unicast SINR is symmetrical with respect to the cell edge, except for the area close to the cell edge itself where the handover occurs. For the heterogeneous network scenario depicted in Figure 10, the different power setting between pico and macro cell implies that the handover point is much closer to the pico cell, that is, the pico cell covers a much smaller area as compared to the macro cell. However, the handover margin is usually chosen to be larger in a heterogeneous network as compared to the one for homogeneous networks to extend the cell range of the pico cell.

The post-equalization SINR for unicast is compared with the one for MBSFN reception of multicast. When moving from TP1 towards TP2, the SINR decreases as the signal is weakened by the path attenuation and the noise floor is no longer negligible as compared to the signal power. However, the signal from TP2 is summed with the signal from TP1 rather than causing any interference and therefore the SINR for MBSFN reception remains relatively high.

In Figures 9 and 10, we distinguish the cases of MBSFN reception (1) with symbol timing aligned with the symbol timing that was optimized for the unicast channel and (2) with symbol timing optimized for the MBSFN channel itself.

Symbol timing derived from the unicast channel yields large losses in homogeneous networks at the cell edge where suboptimum symbol timing yields large ISI (see Figure 9). The SINR degradation becomes even more severe for unicast channel based symbol timing in heterogeneous networks, where differences of the propagation delays at the handover point also cause larger ISI. Note that the modulation and coding scheme for multicast transmissions over MBSFN areas is selected to allow reception even for the worst-case channel conditions or to provide at least coverage for a large percentage of users of the MBSFN area. Hence, SINR losses from ISI due to imperfect symbol timing, as shown with the blue curves in Figures 9 and 10, are highly undesirable.

"Symbol timing derived from the unicast channel yields large losses in homogeneous networks at the cell edge where suboptimum symbol timing yields large ISI…"

"Multicast reception over MBSFN channels with optimized symbol timing, however, can maintain high SINR over the entire network…"

Multicast reception over MBSFN channels with optimized symbol timing, however, can maintain high SINR over the entire network (refer to the red curves in Figures 9 and 10) and can therefore also achieve relatively high spectral efficiency.

Spectral Efficiency

Spectral efficiency for MBSFN subframes has been evaluated by multiple companies that participated in the 3GPP standardization for eMBMS.[4] However, the spectral efficiency has been evaluated for MBSFN carriers with 7.5 kHz subcarrier spacing, which has not been standardized by 3GPP for a time-multiplex among multicast and unicast transmissions. In a scenario with MBSFN carriers with 7.5 kHz subcarrier spacing, the cyclic prefix length is doubled as compared to the cyclic prefix length for 15 kHz subcarrier spacing. Hence, the results from TR25.912[4] are not representative for eMBMS spectral efficiency in a time-multiplex with unicast transmissions in channel scenarios with large delay spread. Moreover, at most 6 out of 10 subframes can be used as MBSFN subframes as outlined earlier in the section "Frame Structure for eMBMS."

A realistic evaluation of spectral efficiency for eMBMS can be found by Alexiou et al.[6], where the subframe structure corresponds to MBSFN subframes as defined in LTE for a time-multiplex between unicast and multicast. The evaluation by Alexiou et al.[6] is done for an inter-site distance of 1732 m at a carrier frequency of 2 GHz and bandwidths of 1.4 MHz and 5 MHz. Three different approaches for transmitting eMBMS have been evaluated. In a first approach, eMBMS is transmitted using a modulation and coding scheme that ensures that even users with the lowest SINR can receive the MBSFN service. In a second approach, the modulation and coding scheme is selected to ensure that the maximum spectral efficiency over all users is achieved. In a third approach, a predefined setting for a target modulation and coding scheme is used. In all but the first approach, a spectral efficiency above 1 bit per second per hertz and per user can be achieved for MBSFN subframes. Note that a service provisioning for the user with the worst case channel conditions obviously comes at the expense for large spectral efficiency losses, which discards the first approach for practical networks. Optimization of the spectral efficiency over all users is also not practical as it would require exhaustive feedback from the users, but the second approach serves as an upper bound of the spectral efficiency over all users that can be achieved in the network. Finally, the third approach demonstrates that the high spectral efficiency can be achieved in practice with a predetermined modulation and coding scheme that depends only on the network topology rather than the individual users' channel conditions.

Conclusions

Multicast transmissions for eMBMS have been specified for LTE as a time-multiplex with unicast transmissions. The transmission format and the receiver structure are very similar among multicast and unicast transmissions, which allows for an easy and economical implementation of eMBMS within a LTE network. The usage of SFN yields a high spectral efficiency for multicast

transmissions, which is also attractive to operators that need to compete with other broadcasting services.

References

[1] Bai, Zijian, Biljana Badic, Stanislaus Iwelski, Tobias Scholand, Rajarajan Balraj, Guido Bruck, and Peter Jung, "On the equivalence of MMSE and IRC in MU-MIMO systems," *IEEE Comms. Letters*, pp. 1288–1290, December 2011.

[2] TS36.829, "Enhanced performance requirements for LTE user equipment (UE) (Release 11)," 3GPP, 2012.

[3] TS25.963, "Feasibility study on interference cancellation for UTRA FDD user equipment (UE)," V.11.0.0, 3GPP, 2012.

[4] TR25.912, "Feasibility study for evolved universal terrestrial radio access (UTRA) and universal terrestrial radio access network (UTRAN) (Release 11)," 3GPP, September 2012.

[5] TS36.211, "Evolved universal terrestrial radio access (E-UTRA), physical channels and modulation (Release 10)," 3GPP, June 2012.

[6] Alexiou, Antonios, Christos Bouras, Vasileios Kokkinos, Andreas Papazois, and George Tsichritzis, "Spectral efficiency performance of MBSFN-enabled LTE networks," *IEEE 6th International Conference on Wireless and Mobile Computing Networking and Communications*, pp. 361–371, 2010.

Author Biographies

Rainer Bachl joined Intel in 2012, working as a systems engineer on algorithms development for the LTE and LTE-A receiver design in Platform Engineering Group. Rainer has been working in communications industry R&D for 18 years on physical layer receiver design for various standards including ISDN, ADSL, UMTS, LTE, and LTE Advanced. Rainer graduated from Munich Technical University in 1989. He received a PhD from the National University of Singapore in 1995. He can be contacted at rainer.w.bachl@intel.com.

Rajarajan Balraj is a system engineer in the Platform Engineering Group and has been with Intel since 2010. He has made key contributions to various interference cancellation and mitigation algorithms on Intel's 3G and 4G wireless modems. Rajarajan Balraj has filed over 25 patents related to LTE/LTE-A and UMTS and can be contacted at rajarajan.balraj@intel.com

Ingmar Groh is a system engineer in the Platform Engineering Group, has been with Intel since August 2011, and has participated in development of 4G wireless modems with emphasis on synchronization, channel estimation, and equalization. He is also actively involved in Intel research activities on 4G and beyond systems. He can be contacted at ingmar.groh@intel.com.

INTERFERENCE-AWARE RECEIVER DESIGN FOR MU-MIMO IN LTE: REAL-TIME PERFORMANCE MEASUREMENTS

Contributors

Sebastian Wagner
Intel Mobile Communications

Florian Kaltenberger
EURECOM

Multiuser (MU) MIMO is a promising technique to significantly increase the cell capacity in LTE systems. However, users scheduled for MU-MIMO may still experience strong MU interference if the channel state information at the base station is outdated or in small cells with a limited number of users available. To tackle the MU interference, an interference-aware (IA) receiver design is employed. Unlike the interference-unaware (IU) receiver, the IA receiver exploits the information about the interfering data stream in the decoding process, resulting in a significant performance gain while maintaining a moderate complexity. We evaluate the performance of both receivers in terms of throughput through real-time measurements carried out with the OpenAirInterface, an open-source hardware/software development platform created by EURECOM. The measurement results show that the IA receiver achieves significantly higher data rates compared to the IU receiver if the user has multiple receive antennas.

Introduction

"…MU-MIMO can significantly increase cell throughput compared to SU-MIMO…"

It is well known that multiuser (MU) multiple-input multiple-output (MIMO) transmission can significantly increase the cell throughput compared to single-user (SU) MIMO transmission due to MU diversity. Therefore, MU-MIMO is already implemented in the 3GPP long-term evolution (LTE) standard Release 8[1], where it is referred to as transmission mode (TM) 5. However, since TM5 only supports two co-scheduled user equipment (UEs) with a single data stream each, the MU-MIMO mode has been extended in TM8 and TM9 in LTE Release 9 and 10, respectively, by introducing UE-specific (precoded) reference signals (RS).[2][3] In TM8 and TM9 the base station (referred to as eNodeB [eNB] in context of LTE) can schedule up to four users with a single data stream where both precoding technique and number of co-scheduled users are entirely transparent to the UE.

"…The throughput at the UEs greatly depends on the amount of interference from co-scheduled users"

In MU-MIMO, the throughput at the UEs greatly depends on the amount of interference from co-scheduled users. This MU interference can be managed at the eNB through efficient precoding or at the UE via interference cancellation. If the precoding is effective, there will not be any significant MU interference at the UEs and thus no need to cancel that residual interference. However, in TM5 where the precoding is based on a very limited set of possible precoding vectors, efficient precoding can only be achieved if the number of users in the cell is large. The same holds true for non-codebook–based precoding schemes as enabled in TM8 and TM9, unless very accurate channel state information is available at the eNB, which in turn is very difficult to obtain.

Consequently, the precoding is likely to be incapable of efficiently mitigating the MU interference at the UEs especially in small cells with a very limited number of users. Therefore, it is of paramount importance that the UEs are able to effectively mitigate the residual MU interference by exploiting its structure.

"…It is of paramount importance that the UEs are able to effectively mitigate the residual MU interference…"

To achieve effective interference mitigation at the UE, different receiver designs have been proposed in the literature. Ghaffar and Knopp[4] propose an optimal simplified interference-aware (IA) receiver based on the maximum likelihood (ML) criterion. However, the optimal IA receiver requires knowledge of the interfering symbol constellation, which is unavailable to the UE. Therefore, Ghaffar and Knopp[4] propose to use a fixed constellation and show that the performance degradation of the sub-optimal IA receiver is acceptable. Bae et al.[5] try to overcome this disadvantage of the IA receiver by implementing an interference modulation estimator prior to the IA receiver. It is shown through simulations that this joint receiver outperforms the IA receiver with fixed interfering constellation especially when the interference power is high. The question of the performance-complexity tradeoff of different receivers is addressed by Bai et al.[6], where a linear receiver termed interference rejection combiner (IRC) is applied to MU-MIMO. Based on simulation results under various channel conditions, the authors conclude that the IRC achieves the best performance-complexity tradeoff. However, it is important to note that the simulation results of Bai et al.[6] assume an infinite number of potential users with the 4 transmit-antenna codebook suggesting that the MU interference is rather low. Under higher interference levels, the simulation results of Bae et al.[5] show a significant performance loss of the IRC compared to the IA receiver. However, in practice, the amount of MU interference is highly dependent on the algorithms (scheduling, precoding, and so on) implemented at the eNB and it is therefore difficult to identify a "typical" MU-MIMO scenario.

In this article, we focus on TM5 applied to small-cell scenarios with *few* users in the cell and the two transmit-antenna codebook resulting in *high* residual MU interference at the UEs. The focus on TM5 is further motivated by the fact that at lower bandwidths (5 MHz and lower), the number of possible UE-specific downlink control information (DCI) in the physical downlink control channel (PDCCH) is limited and it is very likely that the eNB is unable to co-schedule more than two UEs. We implemented the IA receiver proposed by Ghaffar and Knopp[4] on the OpenAirInterface real-time platform[7] and evaluate its performance through measurements under realistic channel conditions.

"We implemented the IA receiver proposed by Ghaffar and Knopp on the OpenAirInterface real-time platform…"

The remainder of the article is organized as follows. The next section, "System Model," introduces the system model and briefly reviews the IA receiver design. In the section "Simulation Results," we carry out simulations to evaluate the IA receiver performance under false assumptions on the interfering symbol constellation. The section "Real-Time Measurements" describes the real-time

measurement setup and presents our results. Finally, we summarize our results in the "Conclusion" section.

Notation: In the following, boldface lowercase and uppercase characters denote vectors and matrices, respectively. The operators $(\cdot)^H$ and $\|\cdot\|$ denote conjugate transpose and norm, respectively. The $N \times N$ identity matrix is denoted $\mathbf{I}_N$, z^R and z^I are the real and imaginary part of $z \in \mathbb{C}$, respectively. The imaginary unit is denoted **i**. A random vector $\mathbf{x} \sim \mathbb{NC}(\mathbf{m}, \mathbf{\Theta})$ is complex Gaussian distributed with mean vector **m** and covariance matrix **Θ.**

System Model

Consider a system with an n_t- antenna eNB and K scheduled UEs, each endowed with n_r receive antennas. We assume that the eNB transmits a *single* data stream s_k to UE k,($k = 1,2, \ldots, K$)and applies a linear precoding technique. Under narrow-band transmission, the received signal $\boldsymbol{y}_k \in \mathbb{C}^{n_r}$ of user k takes the form

$$\boldsymbol{y}_k = \underbrace{\mathrm{H}_k \mathbf{g}_k s_k}_{\text{useful signal}} + \underbrace{\mathrm{H}_k \sum_{j=1, j\neq k,}^{K} \mathbf{g}_j s_j}_{\text{MU interference}} + \underbrace{\mathbf{n}_k}_{\text{noise}}$$

where $\mathbf{H}_k = [\mathbf{h}_{k1}, \mathbf{h}_{k2}, \ldots, \mathbf{h}_{kn_r}]^H \in \mathbb{C}^{n_r \times n_t}$ is the channel from the eNB to UE k, $\mathbf{G}_k = [\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_K]^{n_t \times K}$ is the concatenated precoding matrix and $\mathbf{n}_k \sim \mathbb{NC}(0, \sigma^2 \mathbf{I}_{n_r})$ is the noise vector. Defining the *effective* channels of user k as $\bar{\mathbf{h}}_i \triangleq \mathbf{H}_k \mathbf{g}_i$, ($i = 1, 2, \ldots, K$), the received signal reads

$$\boldsymbol{y}_k = \bar{\mathbf{h}}_k s_k + \sum_{j=1, j\neq k,}^{K} \bar{\mathbf{h}}_j s_j + \mathbf{n}_k.$$

"The key challenge in MU-MIMO is to minimize the MU interference"

The key challenge in MU-MIMO is to minimize the MU interference. This interference can be mitigated at the eNB by computing an appropriate precoder **G** or the interference can be accounted for in the receiver by exploiting its potential structure. It is well known that efficient interference mitigation at the eNB requires precise downlink channel knowledge, which can only be acquired through extensive user feedback. On the other hand, interference management at the receiver necessitates an estimate of the effective channels $\bar{\mathbf{h}}_i$ as well as the interfering symbol alphabet A_j, $s_j \in A_j (j \neq k)$. The LTE standard defines three possible symbol alphabets $A_k \in \{Q_4, Q_{16}, Q_{64}\}$, that is, QPSK, 16QAM, and 64QAM. In the following sections, we discuss the LTE MU-MIMO mode in more detail.

MU-MIMO in LTE Release 8

LTE Release 8[1] defines MU-MIMO in TM5 where two UEs can be scheduled simultaneously each receiving a *single* data stream. The UEs are aware of a

co-scheduled user through the downlink power offset value signaled in the DCI. Moreover, LTE Release 8 adopts a codebook-based precoding scheme as a compromise between performance and feedback overhead. The codebook Γ for $n_t = 2$ is defined as

$$\Gamma = \frac{1}{\sqrt{2}} \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ i \end{pmatrix}, \begin{pmatrix} 1 \\ -i \end{pmatrix} \right\}$$

and $\mathbf{g}_k \in \Gamma$. Since this codebook only offers a very limited choice of precoding vectors, there will remain a significant amount of MU interference at the UEs, especially in small cells with few users. Therefore, it is crucial to implement an IA receiver that exploits the knowledge about the MU interference, as opposed to an IU receiver that treats the interference as noise.

"…It is crucial to implement an IA receiver that exploits the knowledge about the MU interference, as opposed to an IU receiver that treats the interference as noise."

MU-MIMO in LTE Release 9 and Beyond

The LTE Releases 9 and beyond allow for MU-MIMO transmission with UE-specific reference symbols (RS) defined as TM8 (Release 9) and TM9 (Release 10). Transmission modes 8 and 9 enable the scheduling of up to four users with a single data stream or up to two users with two data streams each. The UE-specific RS are precoded the same way as the data, thus leaving the actual precoding open to implementation and entirely transparent to the UEs. Hence, the users are completely oblivious to whether the eNB applied a linear precoding technique like zero-forcing (ZF) or regularized ZF[8][9] or even a nonlinear technique such as Tomlinson-Harashima precoding. However, note that such non-codebook-based approaches require accurate downlink channel estimate at the eNB, which can only be obtained via quantized codebook-based feedback in FDD systems. Furthermore, with UE-specific RS the UE does not know if there are any co-scheduled users or if it is operating in SU-MIMO mode. If the precoding is working well, there will not be any significant MU interference and thus an IA receiver will not improve the performance compared to an IU receiver. The UE can monitor the power of the interfering streams by estimating the effective channel $\bar{\mathbf{h}}_j$ since the UE-specific RS among the potentially co-scheduled users are quasi-orthogonal. Subsequently, if the interference power $\|\bar{\mathbf{h}}_j\|^2$ exceeds some given threshold, the UE will cancel this interference with an IA receiver.

Recently 3GPP created a new study item for LTE Release 12 on Network-Assisted Interference Cancellation and Suppression (NAICS)[10] to study the potential advantages of providing additional information to the UE in order to support its interference cancellation abilities. In the context of MU-MIMO, such information could include the interfering modulation order and the number of interferers (co-scheduled users). If the UE receiver is capable of decoding and subtracting the interfering data stream, then information about the interfering Modulation and Coding Scheme (MCS) and resource allocation is required. This could be achieved by providing the UE with the Radio Network Temporary Identifier (RNTI) of the interfering users to decode the interfering DCIs and subsequently the data for successive interference cancellation.

"The measurements presented in this article investigate how valuable this interfering modulation order is under different propagation environments."

In our context of TM5 and the IA receiver, the only additional information that is required, is the interfering modulation order (that is, QPSK, 16QAM, or 64QAM). The measurements presented in this article investigate how valuable this interfering modulation order is under different propagation environments.

Interference-Aware Receiver

The IA receiver design has been proposed by Ghaffar and Knopp[4] and exploits the potentially available information about the MU interference, that is, the interfering effective channels $\overline{\mathbf{h}}_j$ ($j \neq k$) and the interfering symbol alphabet A_j. In the following, we briefly review the principle of the IA receiver.

As discussed in the previous section, each user has access to the effective channels $\overline{\mathbf{h}}_j$ either through cell-specific RS and the a-priori known codebook as in LTE Release 8, or through UE-specific RS as in LTE Release 9 and beyond. Concerning the interfering symbol alphabet A_j , this information is not readily available to the UEs. The symbol alphabets A_j could be estimated from the statistics of the received signal as shown by Bae et al.[5], but this approach is computationally complex. Ghaffar and Knopp[11] always chose $A_j = Q_{16}$, independent of A_k, which is reasonable if $A_j = A_k$. However, if $A_j \neq A_k$, we observe through simulations in the subsequent section, that assuming identical alphabets, that is, $A_j = A_k$ performs very well even if the true interfering constellation is different.

To compute the log-likelihood ratios (LLRs) Λ, required as an input to the channel decoder, we apply the classical ML criterion with subsequent Max-log approximation. Without loss of generality, we focus on UE k and hence drop the index k. The minimum distance λ reads

$$\lambda = \max_{s_i \in A_i} \left\{ -\left\| \mathbf{y} - \sum_{i=1}^{K} \overline{\mathbf{h}}_i s_i \right\|^2 \right\}.$$

For $n_t = K = 2$, omitting the common term $\|\mathbf{y}\|^2$ and separating into real and imaginary parts we obtain

$$\lambda = \max_{\substack{s_1 \in A_1 \\ s_2 \in A_2}} \left\{ -\|\overline{\mathbf{h}}_1\|^2 |s_1|^2 - \|\overline{\mathbf{h}}_2\|^2 |s_2|^2 + 2[\bar{y}_1^R s_1^R + \bar{y}_1^I s_1^I] + 2|\eta_1| |s_2^R| + 2|\eta_2| |s_2^I| \right\} \quad (1)$$

with

$$\eta_1 = \rho_{12}^R s_1^R + \rho_{12}^I s_1^I - \bar{y}_2^R$$
$$\eta_2 = \rho_{12}^R s_1^I - \rho_{12}^I s_1^R - \bar{y}_2^I$$

where we defined the matched filter outputs $\bar{y}_1 \triangleq \overline{\mathbf{h}}_1^H \mathbf{y}$, $\bar{y}_2 \triangleq \overline{\mathbf{h}}_2^H \mathbf{y}$, and the correlation coefficient $\rho_{12} \triangleq \overline{\mathbf{h}}_1^H \overline{\mathbf{h}}_2$. Note that in Equation 1 we do not require the sign of the interfering symbol $s_2$since Equation 1 is maximized if s_2^R and s_2^I have the opposite signs of η_1 and η_2, respectively. Moreover, the search space for the ML detection can be reduced by one complex dimension since both η_1 and η_2 are *independent* of s_2 and hence by equating the derivative of the expression in braces in Equation 1 to zero the optimal values of $|s_2^R|$ and $|s_2^I|$ are directly given as

$$|s_2^R|^* = \frac{\|\overline{\mathbf{h}}_2\|^2}{|\eta_1|} \; and \; |s_2^I|^* = \frac{\|\overline{\mathbf{h}}_2\|^2}{|\eta_2|}.$$

"…The search space for the ML detection can be reduced by one complex dimension…"

For detailed expressions of the LLRs under various symbol alphabets the reader is referred to Ghaffar and Knopp.[4]

The above IA receiver is able to cancel a *single* interferer. An extension to multi-interference cancellation is not straightforward since the optimal interference amplitude can only be directly computed for one interfering symbol. To cancel more than one interferer requires a full ML detection, which quickly increases the complexity of the receiver. But as previously mentioned, the UE can monitor the strength of the interfering users $\|\overline{\mathbf{h}}_j\|^2$ and decide to cancel the strongest interferer if beneficial.

"An extension to multi-interference cancellation is not straightforward…"

Precoder Selection

User k selects the optimal precoding vector $\mathbf{g}_k^*$ that maximizes his desired *effective* channel magnitude $\|\mathbf{H}_k\mathbf{g}_k\|$, that is,

$$\mathbf{g}_k^* = \arg\max_{\mathbf{g}\in\Gamma}\{\|\mathbf{H}_k\mathbf{g}\|\}$$

and sends the corresponding precoding matrix indicator (PMI) to the eNB. In the test configuration presented in this article, we always assume that two users with orthogonal precoding vectors are scheduled for transmission. Moreover, the above maximization is carried out over the average channel per sub-band as opposed to wideband PMI described by Bai et al.[6]

Simulation Results

Before carrying out real-time measurements, we do link-level simulations to identify the performance loss incurred by a false assumption on the interfering symbol alphabet A_j . Given various modulation and coding schemes (MCS), we measure the Block-Error Rate (BLER) for the SCM-C channel model [12] and average over 10,000 independent channel realizations. The system parameters are given in Table 1.

Carrier Frequency	1907.6 MHz
System Bandwith	5 MHz
Number of Transmit Antennas at eNB	2
TDD Configuration	3
DL Transmit Subframe	7
UL Transmit Subframe	3
RB Allocation	8191 (all 25 RBs)
Number of PDCCH symbols	1

Table 1: System Configuration Parameters
(Source: EURECOM, 2013)

Figure 1, Figure 2, and Figure 3 show the simulation results for QPSK, 16QAM, and 64QAM interference, respectively. From these figures it can be observed that if the desired user has a QPSK alphabet, then the assumption on

the interfering constellation has little impact on the IA receiver performance. Even the IU receiver performs almost as well as the IA receiver. If the desired constellation is 16QAM or 64QAM, we observe that the performance loss is more significant, especially if the interfering modulation order is high but a low modulation order is assumed. From these simulations we conclude that

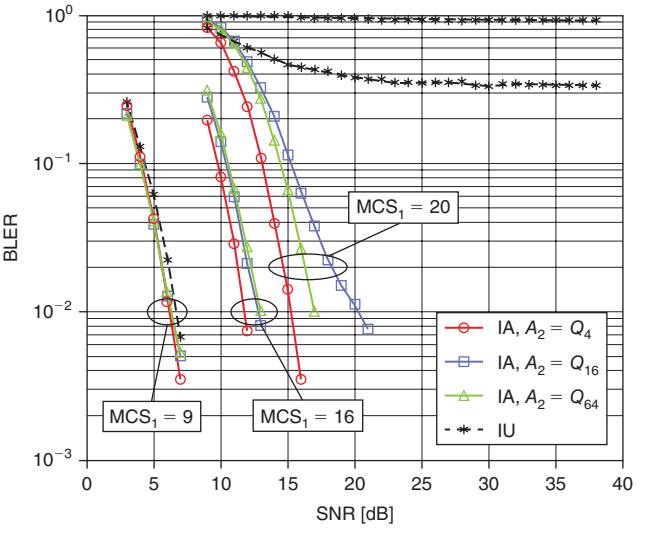


Figure 1: QPSK interference, BLER vs. SNR, MCS_1 = {9,16,20}, MCS_2 = 9, n_r = 2, SCM-C, no HARQ, 10,000 channel realization
(Source: EURECOM, 2013)

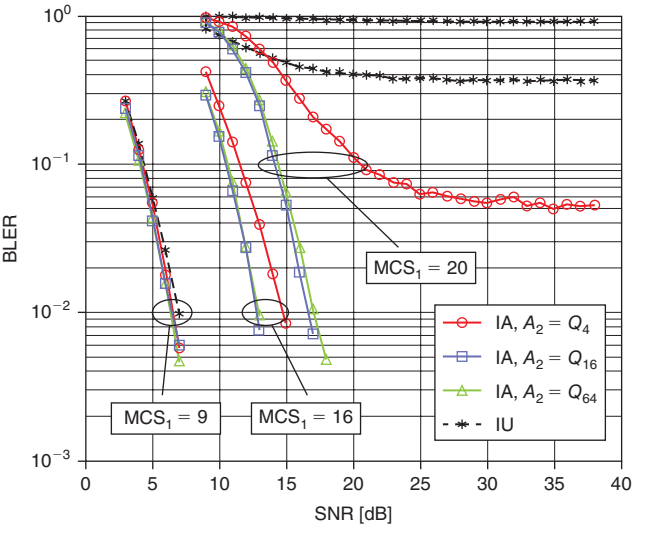


Figure 2: 16QAM interference, BLER vs. SNR, MCS_1 = {9,16,20}, MCS_2 = 16, n_r = 2, SCM-C, no HARQ, 10,000 channel realization
(Source: EURECOM, 2013)

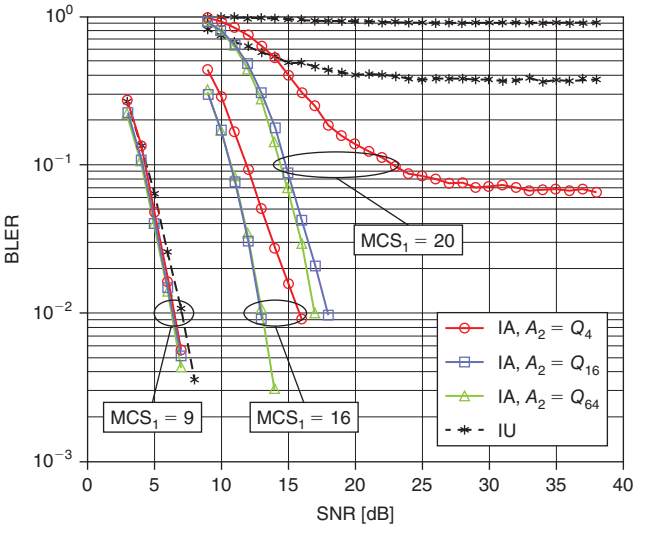


Figure 3: 64QAM interference, BLER vs. SNR, MCS_1 = {9,16,20}, MCS_2 = 20, n_r = 2, SCM-C, no HARQ, 10,000 channel realization
(Source: EURECOM, 2013)

choosing the symbol alphabet for the interfering stream identical to the desired stream, that is, $A_j = A_k$, is robust and results only in a small performance loss. Even under a false assumption on A_j the IA receiver always outperforms the IU receiver significantly.

We are now interested in the computational complexity of the IU and IA receiver and choose to measure the processing time of each receiver on an Intel® Xeon™ CPU E5-2690 dual-core processor clocked at 3 GHz. It should be noted that the implementation makes heavy use of the Streaming SIMD Extension (SSE) 4 instruction set. The results are presented in Figure 4, where we assume that $A_1 = A_2$ and also plot the processing time of the Turbo decoder for comparison. From Figure 4, we observe that the processing time of the IU receiver increases only slightly with the modulation order, whereas the IA receiver complexity increases significantly. For 64QAM the processing time of the IA receiver is almost 5 times that of the Turbo decoder and 12 times the amount of the IU receiver. Although the IA receiver greatly increases the computational complexity at high modulation orders, it is precisely in that region where the subsequent measurements show a tremendous performance gain over the IU receiver. Note that the real-time implementation of the LLR computation uses multiple threads to meet the real-time requirements.

"Choosing the symbol alphabet for the interfering stream identical to the desired stream is robust and results only in a small performance loss"

"…The implementation makes heavy use of the Streaming SIMD Extension (SSE) 4 instruction set"

Real-Time Measurements

In this section we describe the real-time measurement setup and assumptions, the equipment, and the different measurement scenarios. The throughput is measured for both IU and IA receivers in TM5.

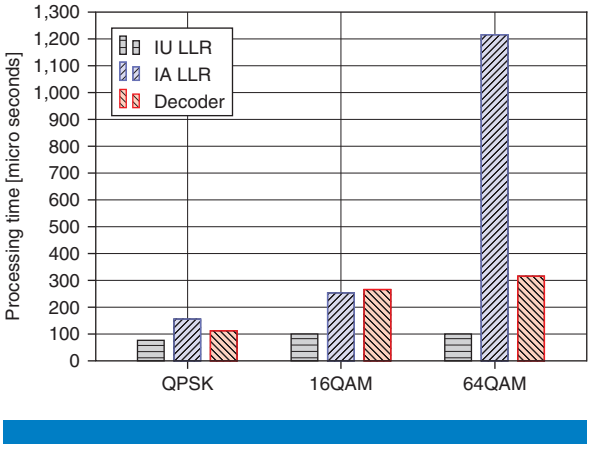


Figure 4: Processing time per subframe of the IA and IU receiver with $n_r = 2$
(Source: EURECOM, 2013)

Setup

For the test setup, we configured Time-Division Duplex (TDD) mode on LTE band 33 (1900–1920 MHz). The eNB and UE have two antennas, whereas the UE uses one antenna for transmission and one or two antennas for reception. The important system configuration parameters are summarized in Table 1.

In subframe (SF) 3, the UEs transmit their measured PMIs on the PUSCH, which is subsequently used in SF 7 by the eNB to precode the signals of both users. In our test setup, only SF 7 carries downlink data.

"The IA receiver should only be applied to those resources with interference…"

Note that in TM5, the data for the interfering UE (user 2) always occupies exactly the same time-frequency resources as the data for user 1, because the downlink (DL) power offset parameter, signaled in the DCI and indicating the presence of another user, is valid for the entire subframe. In TM8 and TM9, the interference could only be present partially within a codeword. The IA receiver should only be applied to those resources with interference; otherwise, in absence of interference, the IA receiver will perform worse than the IU receiver. As previously mentioned, the presence of an interfering user could be determined by monitoring the interference power and applying a suitable threshold to decide if interference cancellation is used.

In the measurements we use two IA receivers. One IA receiver assumes that the interference modulation order is the same as the desired modulation order. This receiver is simply termed "IA". The other receiver is assumed to obtain the correct interfering modulation order through network-aided (NA) signaling and is referred to as "NA-IA" receiver.

Assumptions

In the measurements, we consider the scenario where only two users are available for TM5. The eNB always uses the PMI reported by user 1 (the

desired user) and assigns orthogonal PMIs to user 2 regardless of the PMI report of user 2. This scheduling scheme is optimal for user 1 but suboptimal for user 2 and from a cell capacity point of view. However, since we focus on the throughput of user 1, this scheduling scheme is adequate. Note that we have only two transmit antennas at the eNB and hence use a *lower* resolution codebook than Bai et al.[6] Thus, the interference experienced by user 1 is relatively high. This assumption is realistic in small cells where the number of users is likely to be small and orthogonal PMIs might not be available, which will result in higher MU interference.

Concerning the PMI feedback, we make several assumptions. First, we ensure that the uplink (UL) is always error-free by transmitting with sufficient power. This is necessary to avoid errors in the PMI that would impair our receiver performance measurements. Secondly, we implement sub-band PMI measurements similarly to TM6, which is not foreseen in LTE Release 8 but later in Release 9 and beyond. However, this has no impact on the relative performance of the IA and IU receiver.

Since the PMI is measured in SF 2 and applied in SF 7, the channel is supposed to be approximately constant during 5 SFs or equivalently 5 ms, which was the case during the measurements.

The LTE modem ran without protocol stack (no Hybrid ARQ) and UL and DL resources were statically configured. Note that, although we disabled the higher layers for this measurements, a similar MU-MIMO setup has been successfully demonstrated with complete protocol stack during the SAMURAI project.[13][14]

"…a similar MU-MIMO setup has been successfully demonstrated with complete protocol stack during the SAMURAI project"

During the measurements, the receiver type is changed per frame, and MCS_1 is random and uniformly distributed between 0 and 27. To make MCS_2 available to the NA-IA receiver without explicit signaling, it is coupled to the system frame number (SFN) as $MCS_2 = SFN \bmod 28$. Although MCS_2 is not truly random, no significant change in performance compared to a random MCS_2 has been observed. Moreover, each of the subsequent results was obtained by measuring over a time period of about 2 minutes.

Equipment

The measurements are carried out with the EURECOM experimental OpenAirInterface (OAI) platform. The OAI implements a software-defined radio of the 3GPP LTE Release 8.6 standard, which runs on common x86 Linux machines. To ensure real-time operation, we utilize the real-time application interface (RTAI). Furthermore, the real-time signals are transmitted via the PCI Express interface to the EURECOM Express MIMO 2 board (see Figure 5), where the base-band signal is modulated and transmitted via an additional RF front-end as depicted in Figure 6. The Express MIMO 2 board is able to receive and transmit on four channels independently and for a wide range of frequencies.

"The OAI implements a software-defined radio of the 3GPP LTE Release 8.6 standard, which runs on common x86 Linux machines"

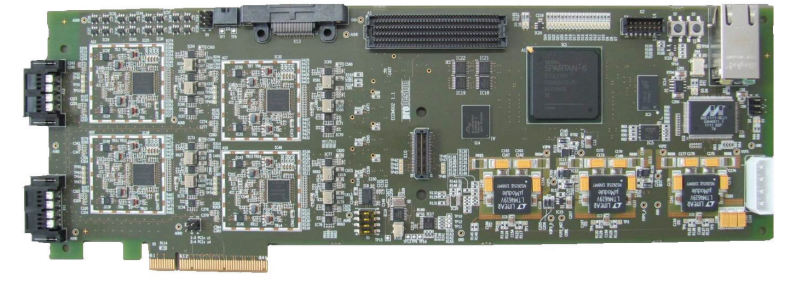


Figure 5: Express MIMO 2 board
(Source: EURECOM, 2013)

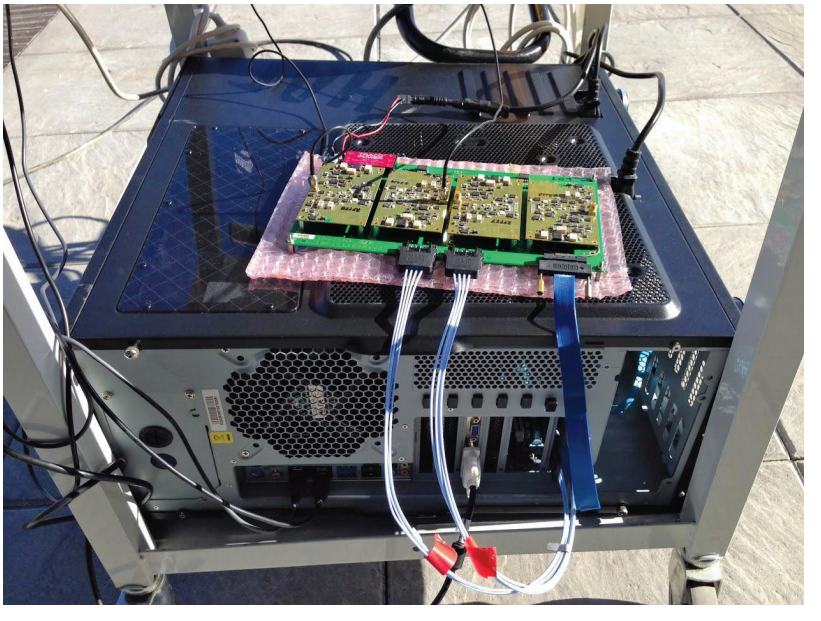


Figure 6: User equipment with RF board
(Source: EURECOM, 2013)

Scenarios

We consider three different scenarios:

1. Indoor scenario (UE is inside building, eNB is outside) to measure throughput for different number of receive antennas
2. Outdoor scenario with a strong line-of-sight (LOS) channel
3. Outdoor scenario with non-LOS (NLOS) channel conditions

In all scenarios the UE is moved at low speeds to avoid a strong Doppler effect but to allow for an averaging of the performance over sufficiently different channel realizations.

Measurement Results

The throughput measurements for all three scenarios are presented in the following sections.

Indoor Scenario with Variable Number of Receive Antennas

The measurement setup consists of an eNB situated outside on top of the EURECOM building and the UE placed inside the building.

Figure 7 and Figure 8 show the average throughput of IU, IA, and NA-IA receivers with one or two receive antennas, respectively, for different values of MCS_1. From these figures we observe that, for QPSK (that is, MCS 0,1,…,9),

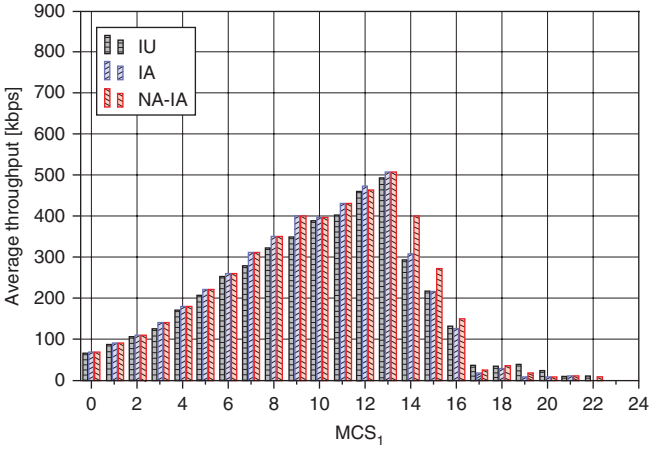


Figure 7: MCS_1 vs. average throughput with $n_r = 1$, indoor scenario
(Source: EURECOM, 2013)

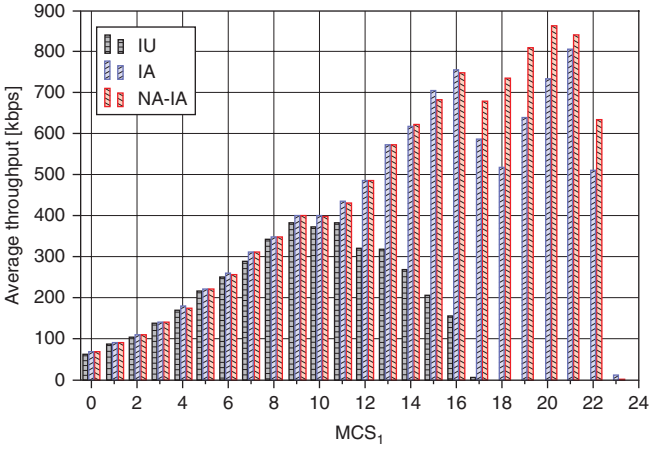


Figure 8: MCS_1 vs. average throughput with $n_r = 2$, indoor scenario
(Source: EURECOM, 2013)

all receivers achieve about the same throughput irrespective of the number of receive antennas.

From Figure 7, we observe that the IA receiver does not offer a significant throughput increase if only a single receive antenna is available. However, the NA-IA receiver can achieve moderately higher throughput for higher MCS.

"With two receive antennas the performance of the IA and NA-IA receivers improve drastically…"

With two receive antennas the performance of the IA and NA-IA receivers improve drastically (see Figure 8), whereas the IU receiver shows a small performance loss, which may be explained by the varying channel conditions. The NA-IA receiver outperforms the IA receiver for $MCS_1 > 16$, that is, for 64QAM.

We conclude that an IA receiver can significantly improve the performance of the UE in TM5, especially if an additional receive antenna is present to allow for effective interference mitigation. Furthermore, the NA-IA receiver can improve the performance if 64QAM is used.

Outdoor Scenario with Strong Line-of-Sight Component

Figure 9 shows the measurement environment with the UE in the foreground and the eNB on the roof in the background. During the measurement we move the UE slowly in one direction and back multiple times.

Figure 9: Outdoor scenario with strong LOS channel
(Source: EURECOM, 2013)

Figure 10 depicts the measurement results for all three receivers. It can be observed that both IA and NA-IA receivers achieve maximum throughput for $MCS_1 = 16$ and the IU receiver at $MCS_1 = 14$. Although the maximum throughput of IA and NA-IA receivers are almost identical, the NA-IA receiver achieves a significantly higher throughput for $MCS_1 > 16$, that is, for 64QAM.

"…The NA-IA receiver achieves a significantly higher throughput for 64QAM"

Outdoor Scenario without Line-of-Sight Component

Figure 11 shows the NLOS environment, where the UE was slowly moved straight until the bridge.

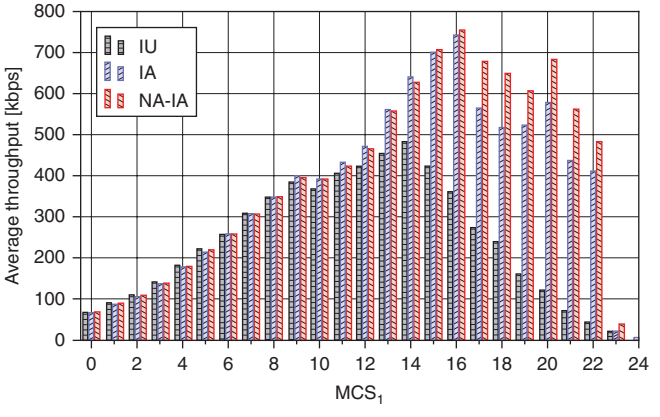


Figure 10: MCS_1 vs. average throughput with $n_r = 2$, outdoor scenario with strong LOS channel
(Source: EURECOM, 2013)

Figure 11: NLOS environment with eNB on the roof of right building
(Source: EURECOM, 2013)

The throughput results for a NLOS channel are presented in Figure 12. It can be seen that the difference in maximum throughput between IA and NA-IA receivers is negligible and almost identical compared to the results in the LOS environment. Moreover, as in the LOS channel, the NA-IA receiver significantly outperforms the IA receiver for 64QAM modulation. These results suggest that the *relative* performance of IA and NA-IA receivers is robust to the propagation environment, that is, their throughput difference is similar in LOS and NLOS channels.

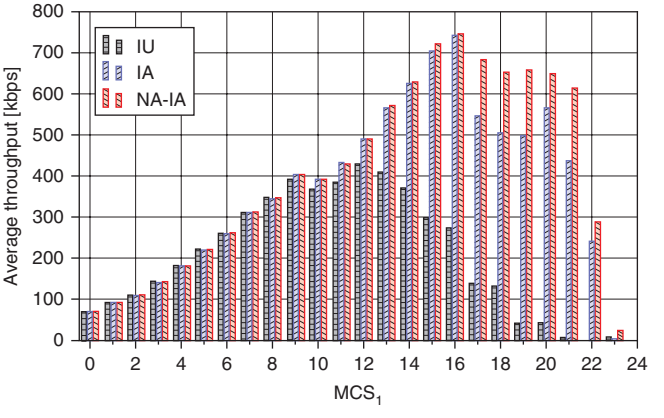


Figure 12: MCS_1 vs. average throughput with $n_r = 2$, outdoor scenario with non-LOS channel (Source: EURECOM, 2013)

Regarding the IU receiver, we observe significant performance degradation in NLOS compared to LOS channels, especially for $MCS_1 > 12$. For instance, the maximum throughput in LOS of 481 kbps decreased to 428 kbps in the NLOS channel. Even more drastic is the loss at higher order modulations, for example, for $MCS_1 = 17$ the throughput decreased from 273 kbps to 137 kbps. We conclude that the IU receiver benefits significantly from LOS environments, especially at higher order modulations. This is in line with the findings of Bai et al.[6], who showed that channel correlation is beneficial for MU-MIMO. The LOS component of the channel increases the channel correlation and hence renders the precoding more effective, resulting in lower MU interference.

"The IU receiver benefits significantly from LOS environments, especially at higher order modulations"

Conclusion

This article evaluated the potential performance improvements of IA receiver designs over an IU receiver in TM5 through real-time field measurements in LOS and NLOS propagation environments.

In the case of a single receive antenna, the measurements indicate that the IA receiver offers almost no advantage compared to the IU receiver.

However, for both single- and dual-receiver antennas, the measurements revealed that the NA-IA receiver significantly outperforms the IA receiver

for higher order modulations, such as 64QAM. This result suggests that the signaling of the interfering modulation order can greatly improve performance in the case where 64QAM is applied. For lower order modulations the simplified IA receiver without knowledge of the interfering modulation order performs equally well as the NA-IA receiver.

Moreover, the measurements indicate that the IU receiver benefits significantly from LOS channels compared to the IA receivers, especially at higher order modulations. In the case of QPSK, even the IU receiver achieves the same throughput as the IA receivers.

We conclude that a UE with IA receiver can greatly increase both cell and user throughput, especially with additional network assistance.

"…A UE with IA receiver can greatly increase both cell and user throughput, especially with additional network assistance"

References

[1] 3rd Generation Partnership Project, "Physical Channels and Modulation," 3GPP TS 36.211 V8.6.0, 2009.

[2] Sesia, S., I. Toufik, and M. Baker, *LTE, The UMTS Long Term Evolution: From Theory to Practice,* 2 ed., Wiley & Sons, 2011.

[3] Duplicy, J. B. Badic, R. Balraj, R. Ghaffar, P. Horvath, F. Kaltenberger, R. Knopp, I. Z. Kovács, H. T. Nguyen, D. Tandur, G. Vivier, "MU-MIMO in LTE systems," *EURASIP Journal on Wireless Communications and Networking*, 2011.

[4] Ghaffar, R. and R. Knopp, "Interference-Aware Receiver Structure for Multi-user MIMO and LTE," *EURASIP Journal on Wireless Communications and Networking*, no. 1, pp. 1-17, 2011.

[5] Bae, J. H., S. Kim, J. Lee, and I. Kang, "Advanced Downlink MU-MIMO Receiver for 3GPP LTE-A," in *IEEE International Conference on Communications (ICC)*, 2012.

[6] Bai, Z., S. Iwelski, G. Bruck, P. Jung, B. Badic, T. Scholand, and R. Balraj, "Receiver Performance-Complexity Tradeoff in LTE MU-MIMO Transmission," in *Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, 2011.

[7] EURECOM, "OpenAirInterface Hardware/Software Development Platform," Sep 2013. [Online]. Available: http://www.openairinterface.org.

[8] Peel, C. B., B. M. Hochwald, and A. L. Swindlehurst, "A Vector-perturbation Technique for Near-Capacity Multiantenna Multiuser Communication - Part I: Channel Inversion and Regularization," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195-202, 2005.

[9] Wagner, S., R. Couillet, M. Debbah, and D. T. Slock, "Large System Analysis of Linear Precoding in Correlated MISO Broadcast Channels under Limited Feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4509-4537, 2012.

[10] 3rd Generation Partnership Project, "Network-Assisted Interference Cancellation and Suppression for LTE (Release 12)," 3GPP TR 36.866, 2013.

[11] Ghaffar, R. and R. Knopp, "Interference Sensitivity for Multiuser MIMO in LTE," in *IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2011.

[12] 3rd Generation Partnership Project, "Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations," TR 25.996, 2009.

[13] Badic, B., A. F. Cattoni, M. Dieudonne, J. Duplicy, P. Fazekas, F. Kaltenberger, I. Z. Kovács, and G. Vivier, "Advances in Carrier Aggregation and Multi-User MIMO for LTE-Advanced: Outcomes from SAMURAI project," http://www.ict-samurai.eu/content/publications, 2012.

[14] Cattoni, A. F., H. T. Nguyen, J. Duplicy, D. Tandur, B. Badic, R. Balraj, F. Kaltenberger, I. Latif, A. Bhamri, and G. Vivier, "Multi-user MIMO and Carrier Aggregation in 4G systems: The SAMURAI Approach," in *IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, 2012.

Acknowledgements

This work was partially funded by the European FP7 Network of Excellence Newcom#.

Author Biographies

Sebastian Wagner is an LTE firmware engineer at Intel Mobile Communications since 2013. He received his MS in Electrical Engineering from the Technische Universität Dresden in 2008 and the PhD from Télécom ParisTech (EURECOM) in 2011. He joined EURECOM as a post-doctoral fellow in 2012 where he was involved in the LTE Modem development for the OpenAirInterface platform. His research interests include signal processing, MIMO communication systems, receiver design and the application of random matrix theory to wireless communications. He can be contacted at sebastian.wagner@intel.com.

Florian Kaltenberger is an Assistant Professor in Mobile Communications at EURECOM and part of the team managing the real-time open-source test platform OpenAirInterface.org for experimentation in LTE systems. He

received his MS and his PhD both in Technical Mathematics from the Vienna University of Technology in 2002 and 2007, respectively. His research interests include signal processing for wireless communications, MIMO communication systems, receiver design and implementation, MIMO channel modeling and simulation, and hardware implementation issues. He can be contacted at florian.kaltenberger@eurecom.fr.

AN EMPIRICAL LTE SMARTPHONE POWER MODEL WITH A VIEW TO ENERGY EFFICIENCY EVOLUTION

Contributors

Mads Lauridsen
Aalborg University, Denmark

Laurent Noël
Vidéotron, Quebec, Canada

Troels B. Sørensen
Aalborg University, Denmark

Preben Mogensen
Aalborg University, Denmark

"The smartphone power consumption model includes the main power consumers in the cellular subsystem as a function of receive and transmit power and data rate…"

"…the combination of high data rates and long sleep periods is the optimal combination from a user equipment energy-saving perspective."

Smartphone users struggle with short battery life, and this affects their device satisfaction level and usage of the network. To evaluate how chipset manufacturers and mobile network operators can improve the battery life, we propose a Long Term Evolution (LTE) smartphone power model. The idea is to provide a model that makes it possible to evaluate the effect of different terminal and network settings to the overall user equipment energy consumption. It is primarily intended as an instrument for the network engineers in deciding on optimal network settings, but could also be beneficial for chipset manufacturers to identify main power consumers when taking actual operating characteristics into account. The smartphone power consumption model includes the main power consumers in the cellular subsystem as a function of receive and transmit power and data rate, and is fitted to empirical power consumption measurements made on state-of-the-art LTE smartphones. Discontinuous Reception (DRX) sleep mode is also modeled, because it is one of the most effective methods to improve smartphone battery life.

Energy efficiency has generally improved with each Radio Access Technology (RAT) generation, and to see this evolution, we compare the energy efficiency of the latest LTE devices with devices based on Enhanced Data rates for GSM Evolution (EDGE), High Speed Packet Access (HSPA), and Wi-Fi*. With further generations of RAT systems we expect further improvements. To this end, we discuss the new LTE features, Carrier Aggregation (CA) and Enhanced Physical Downlink Control Channel (EPDCCH), from an energy consumption perspective.

Not surprisingly, the conclusion is that having the cellular subsystem ON, and in addition, transmit powers above 10 dBm, have the largest effect on UE power consumption, and that the combination of high data rates and long sleep periods is the optimal combination from a user equipment energy-saving perspective.

Introduction

The battery life of smartphones has become shorter as smartphones have become more advanced, both due to slow battery capacity evolution, but also due to bigger displays, faster and more processor cores, and more complex Radio Access Technologies (RATs).

The power consumed due to use of various RATs depends on the hardware and software within the device, and in addition on the RAT network setup. To analyze and minimize the power consumption caused by suboptimal

network setup, the responsible network engineers require a model that describes the smartphone power consumption as a function of relevant parameters.

In recent literature the smartphone power consumption has been studied either by running a meter application on the phone[1][2] or by using a dummy battery[3][4][5], which logs the current drain. The latter option seems to be the best because it does not introduce any additional signal processing and hardware routines in the smartphone. In some articles, the authors[4][2] have connected the smartphone to a live, commercial network, while others have performed conducted tests using a base station emulator in a laboratory.[3][5] The emulator setup is preferable because it provides the full control and logging of all relevant network parameters such as resource allocation and power levels, while also being a realistic "live" connection.

"The emulator setup is preferable because it provides the full control and logging of all relevant network parameters…"

Few public measurements of LTE smartphones are available, and most of the literature unfortunately only reports power consumption for a limited number of parameters. One article[4] provides the power consumption as a function of data rates, but with no information about the transmit (Tx) and receive (Rx) power levels, while another[5] only reports power consumption as a function of Tx power. Therefore we decided to provide a new model, which includes the most relevant network parameters, that is Tx and Rx power levels and data rates. Our first LTE power model[6] was based on commercial Universal Serial Bus (USB) dongles, which were not optimized for low power consumption, but the model did not include DRX and cell bandwidth. Therefore we presented an updated model[3] where the power consumption of three different LTE smartphones, commercially available in fall 2012, was examined. Comparing our dongle and smartphone measurements, it is clear that the cellular subsystems develop fast and that the power consumption improves with each generation. Therefore it is of interest to examine how it has evolved with the launch of the latest LTE chipsets.

"…it is clear that the cellular subsystems develop fast and that the power consumption improves with each generation"

In this article we present our recent measurements on LTE smartphones and compare with the previous generations.[3][6] We also discuss the observed energy efficiency (EE) improvement and compare it with other wireless RATs. Finally we discuss how the LTE power consumption can be lowered in the future by use of micro sleep and Carrier Aggregation.

The article is organized as follows: in the next section we introduce our smartphone power consumption model, and in the following section, "Experimental Assessment," we present the measurement campaign we have carried out to assign meaningful values to the model. Then we define energy efficiency (EE) and provide a comparison of EE in wireless RATs in the section "Energy Efficiency Evolution," and in connection with this we discuss micro sleep and carrier aggregation as future power optimization possibilities in the section "Energy Efficiency Improvements." In the last section we present our conclusions.

Smartphone Power Consumption Model

In this section the smartphone power consumption model, originally developed for the dongle measurements[6] but also applicable here, is presented.

The model covers the LTE cellular subsystem and the overall power consumption is defined as:

$$P_{cellular} = m_{con} \times P_{con} + m_{idle} \times P_{idle} + m_{DRX} \times P_{DRX} \text{ [W]} \quad (1)$$

where *m* is a binary variable describing whether the UE is in RRC_connected (con), RRC_idle (idle), or DRX mode. The associated *P* value describes the power consumption in the given mode as a function of mode specific parameters.

"The power consumption model of RRC_connected mode is divided into Transmit and Receive Base Band and Radio Frequency functional blocks…"

"…the power consumption can be measured independently of the other blocks' contributions by varying the block-specific parameter."

The power consumption model of RRC_connected mode is divided into Tx and Rx Base Band (BB) and Radio Frequency (RF) functional blocks, which each define the power consumption as a function of either Tx or Rx power levels (*S*) and data rates (*R*). The model, sketched in Figure 1, was divided into those blocks[6] because they each have a distinct parameter, for example, transmit power S_{Tx} in the Tx RF, which primarily affects the power consumption of that block. Therefore the power consumption can be measured independently of the other blocks' contributions by varying the block-specific parameter. Our empirical measurements[3][6] have consolidated this division.

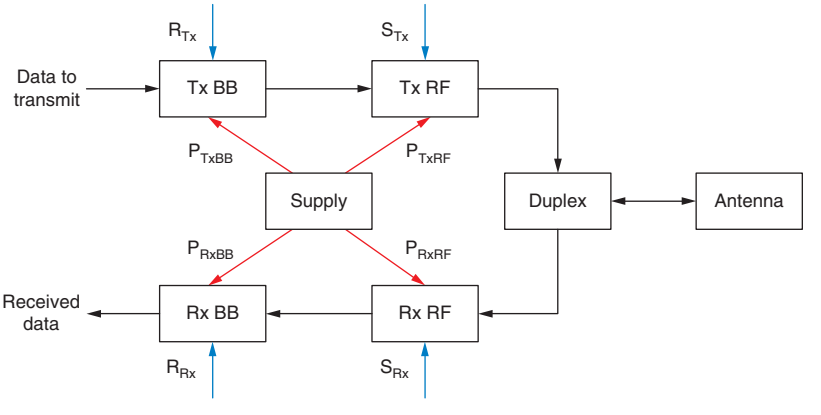


Figure 1: LTE smartphone cellular subsystem power model (Source: Mads Lauridsen, Aalborg University, 2013)

The RRC_connected mode power consumption is:

$$P_{con} = P_{on} + m_{Rx} \times (P_{Rx} + P_{RxBB}(R_{Rx}) + P_{RxRF}(S_{Rx})) + m_{Tx} \times (P_{Tx} + P_{TxBB}(R_{Tx}) + P_{TxRF}(S_{Tx})) \text{ [W]} \quad (2)$$

The constants P_{on}, P_{Rx}, and P_{Tx} describe the power consumed when the cellular subsystem is ON, the receiver is actively receiving, and the transmitter is actively transmitting, respectively.

In RRC_idle mode the UE is mainly in a low-power sleep mode. It wakes up periodically to see whether there is an incoming paging message from the network. The period is defined by the network-controlled paging cycle t_pc. This behavior resembles the DRX power consumption, and therefore the DRX model, which is presented in the next section, is used to calculate RRC_idle mode power consumption P_idle of Equation 1. This is however an approximation because the number of tasks required in RRC_idle is far less compared to RRC_connected.

"Sleep modes are one of modern RATs' most important methods to achieve high energy efficiency."

DRX Power Consumption Model

Sleep modes are one of modern RATs' most important methods to achieve high EE. The Connected Mode DRX sleep mode is standardized in LTE, and has also been included in recent versions of 3G. The idea is that the UE is scheduled periodically by the network, hence it knows when to be active and when it can sleep. The LTE DRX allows for periods of 10–2560 ms, so the period can be well adjusted to the data type. Furthermore the network can specify how long the UE must remain ON during each period, known as the On Duration t_onD, and whether it must remain active for a certain period after successfully decoding data. The UE power consumption as a function of time, when using DRX, could therefore be expected to look as sketched in Figure 2.

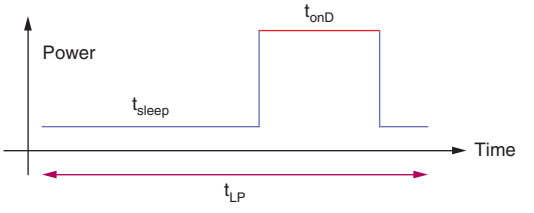


Figure 2: Ideal LTE DRX behavior (Source: Mads Lauridsen, Aalborg University, 2013)

"…the UE is scheduled periodically by the network, hence it knows when to be active and when it can sleep. The LTE DRX allows for periods of 10–2560 ms…"

There are however multiple tasks that prevent the phone from performing as in Figure 2. They include but are not limited to[7]:

- The use of different clocks. In deep sleep mode the UE typically uses a low-power low-precision 32 kHz crystal to keep track of the System Frame Number (SFN), whereas it needs to power on a high-precision clock to achieve a proper phase reference for all clocks used when the cellular subsystem is ON.
- The wakeup phase. To enable demodulation, the UE obviously needs a phase lock of the BB Phase Locked Loop (PLL) synthesizer, but also a stable RF subsystem. The latter entails phase-locked RF PLL, stable Automatic Gain Control (AGC), programming of channel filters, and possibly a calibration of certain components.
- The synchronization phase. This requires demodulation of LTE's primary and secondary synchronization signals, which are sent every 5 ms, and possibly also decoding the Physical Broadcast Channel to get the SFN and other basic information. While this is being performed, channel estimation is also carried out.
- Power-down phase. In this phase the UE does not need to perform decoding, calibration, or any other time-consuming tasks, but powering down the components also takes time, and therefore the phase is included.

Due to the aforementioned tasks, the LTE DRX UE power consumption is as illustrated in Figure 3.

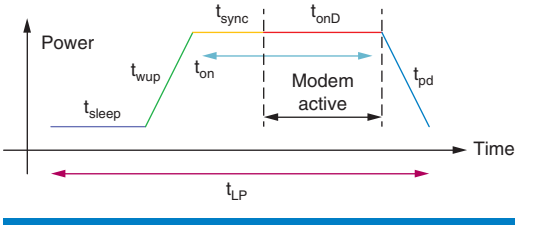


Figure 3: Realistic LTE DRX behavior (Source: Mads Lauridsen, Aalborg University, 2013)

Comparing Figure 3 with Figure 2, you can see that the standardized t_{onD} remains the same while the sleep time t_{sleep} is shortened due to the introduction of the wakeup (t_{wup}), synchronization phase (t_{sync}), and power-down phase (t_{pd}), all of which are functions of the DRX period t_{LP}, because it is the deciding

factor for which sleep power level is used. The shorter sleep time means that the average power consumption is increased.

The sleep time is calculated as

$$t_{sleep}(t_{LP},t_{onD}) = t_{LP} - (t_{wup}(t_{LP}) + t_{sync}(t_{LP}) + t_{pd}(t_{LP}) + t_{onD})[s] \quad (3)$$

Combining it with the energy consumed during the wakeup, synchronization, and power-down phases $E_{wup/pd+sync}(t_{LP})$, the average DRX power consumption, excluding the power consumed in the On Duration, is

$$P_{DRX}(t_{LP}, t_{onD}) = (t_{sleep}(t_{LP},t_{onD}) \times P_{sleep} + E_{wup/pd+sync}(t_{LP}))/(t_{LP} - t_{onD})[W] \quad (4)$$

Combining this value with the power consumption and length of the On Duration, the total power and energy consumption of a DRX period can be calculated and applied in system level simulations.

"The DRX model uses average power, and therefore the results cannot be used for Transmission Time Interval simulations, but only system-level simulations with a longer time perspective"

The DRX model uses average power, and therefore the results cannot be used for Transmission Time Interval (TTI) simulations, but only system-level simulations with a longer time perspective. Note however that P_{con} is applicable on the TTI level.

Experimental Assessment

Each of the proposed model's functional blocks depend on one specific parameter, and in this section it is described how the functions are derived using experimental measurements.

The assumption is that a given block's function can be assessed experimentally by varying the function-specific parameter, such as the R_{Tx} of the Tx BB, while keeping the other parameters S_{Tx}, R_{Rx}, and S_{Rx} constant and at a level where they will influence the measurement the least.

The parameters are varied by adjusting the Modulation and Coding Scheme (MCS), number of Physical Resource Blocks (PRBs), and Rx and Tx powers *S*. For example the receive data rate R_{Rx} can be varied by adjusting the Downlink (DL) MCS and the number of DL PRBs.

A least one test case (TC) is then designed for each of the model's four functions (see Table 1), and to enable a comparison with our previous work[3] the same TCs are used. The varied parameter is shown in brackets. We have previously[3] applied the TCs in 10, 15, and 20 MHz cell bandwidth. The measurements showed a very linear relationship with bandwidth and therefore the TCs are only performed in 20 MHz cell bandwidth in this study.

The TCs in Table 1 are furthermore designed such that a common point exists. The point uses DL MCS 0, DL PRB 3/4, Uplink (UL) MCS 5/6, UL PRB 100, and constant powers. In addition TC 2 and 4 have an initial test point using 0 PRBs in either DL or UL. By comparing these three points the cellular subsystem ON power P_{on}, and the power consumption of having active reception P_{Rx} and transmission P_{Tx} can be determined.

Test Case		Downlink parameters			Uplink parameters		
		MCS	PRB	S_{Rx}	MCS	PRB	S_{Tx}
Rx BB	1	[0,28]	100	−25	6	100	−40
	2	0	[0,100]	−25	6	100	−40
Rx RF	3	0	100	[−25,−90]	6	100	−40
Tx BB	4	0	3	−25	6	[0,100]	−40
	5	0	3	−25	[0,23]	100	−40
Tx RF	6	0	3	−25	6	100	[−40,23]

Table 1: Measurement parameters. Tests are made for cell bandwidth of 20 MHz. In DL both 1 and 2 code words (CWs) are tested.
(Source: Lauridsen, Mads et al.[3], 2013)

Measurement Setup

In this study, measurements on two LTE Release 8 category 3 smartphones were performed to obtain updated and realistic values for the smartphone power model. The main characteristics of the Device Under Test (DUT) are listed in Table 2. They are both touchscreen phones running the Android* OS, and are connected to LTE band 4 with carrier frequency 2145 MHz (DL UARFCN 2300).

"In this study, measurements on two LTE Release 8 category 3 smartphones were performed to obtain updated and realistic values for the smartphone power model."

	UE1	UE2
OS	Android 4.0.4	Android 4.1.2
Launch date	June 2012	April 2013
Modem & CPU	Part #A	Part #B
Modem & CPU CMOS node	28 nm	28 nm
RF transceiver	Part #C	Part #D
RF transceiver CMOS node	65 nm	65 nm
Band 4 PA	Part #E	Part #F
LTE bands	4, 17	1, 2, 4, 5, 17

Table 2: DUT main characteristics
(Source: Laurent Noël, Vidéotron, 2013)

UE2 is one generation newer than UE1, which we previously have examined[3], and therefore the UEs do not share modem and RF transceiver components as indicated in the table.

"UE2 is one generation newer than UE1 and therefore the UEs do not share modem and RF transceiver components…"

Power consumption measurements are performed under conducted test conditions, that is, the DUT is connected to an Anritsu 8820c eNodeB emulator via a pair of RF coaxial cables. A Faraday cage is used to ensure adequate DUT RF isolation from surrounding commercial LTE and HSPA+ networks. An Agilent N6705B power supply is connected to the DUT via the OEM's respective dummy batteries. Both supply voltage and current consumption are logged with microsecond time accuracy over at least 30 seconds per measurement point. Each power consumption log is then post-processed on a computer to determine the average power consumption. The accuracy of the measurement is estimated at +/− 10 mW in cell-connected mode. The setup is illustrated in Figure 4.

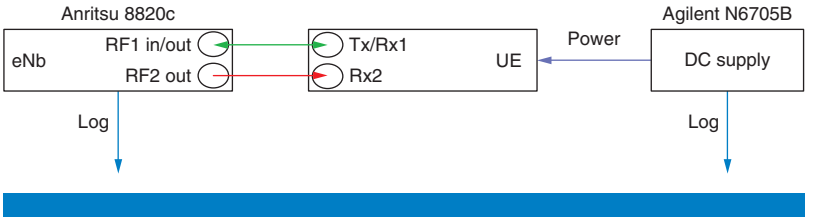


Figure 4: Measurement setup using the eNodeB emulator (Source: Mads Lauridsen, Aalborg University, 2013)

Note that the use of a dummy battery, and especially the length and diameter of the connecting wires, have a non-negligible effect on the accuracy of the DRX time measurements, which is difficult to prevent.

Uplink Characterization

The main contributor to cellular subsystem power consumption is the transmitter, and in this section the power consumption as a function of the two UL parameters in Equation 2 is reported.

"The PA's high gain mode is activated when the required transmit power exceeds a certain limit, and it entails a major increase in power consumption."

A transmitter is usually composed of a single chip RF transceiver and one external power amplifier (PA). The PA's high gain mode is activated when the required transmit power exceeds a certain limit, and it entails a major increase in power consumption. We have however previously shown[8] that transmitting with high power and high data rates may be the most energy-efficient solution, depending on the type of data and propagation scenario.

Previous measurements[3][6] on older generations of LTE UEs, including UE1 of this study, have revealed a major power consumption increase when the transmit power exceeds 10 dBm. As illustrated in Figure 5, based on TC 6, this is also the case for the new UE2. Comparing the power consumption of UE1 and UE2, it is clear that the baseline power consumption has improved considerably in UE2. For transmit powers below 0 dBm the improvement is in the order of 35 percent. On the other hand, the PA used in UE2 is not as efficient as the one used in UE1, since the power consumption gap decreases for transmit powers above 10 dBm. This means the energy savings are reduced for high transmit power, but this may not be the case for other UEs because the PA is a component, which is available in many versions and designs, and because many tradeoffs are possible when specifying PA performance. For further information refer to the discussion in Holma et al. [7] on PAs.

The blue dotted line in Figure 5 represents the model fit for the functional Tx RF block. The design of the fit and the function's values are presented in the subsection "Model Parameterization." Observe the blue dotted line is present in the following measurement results as well, and that it covers the related functional block fit.

"…the UE2 power consumption is completely independent of the UL data rate."

The 35 percent power reduction between UE1 and UE2 is also observed when examining the power consumption as a function of UL data rate as in TC 4 and 5, where the transmit power and DL parameters are kept constant. The result of TC 5 is shown in Figure 6, and it illustrates that the UE2 power consumption is completely independent of the UL data rate. In some UEs,

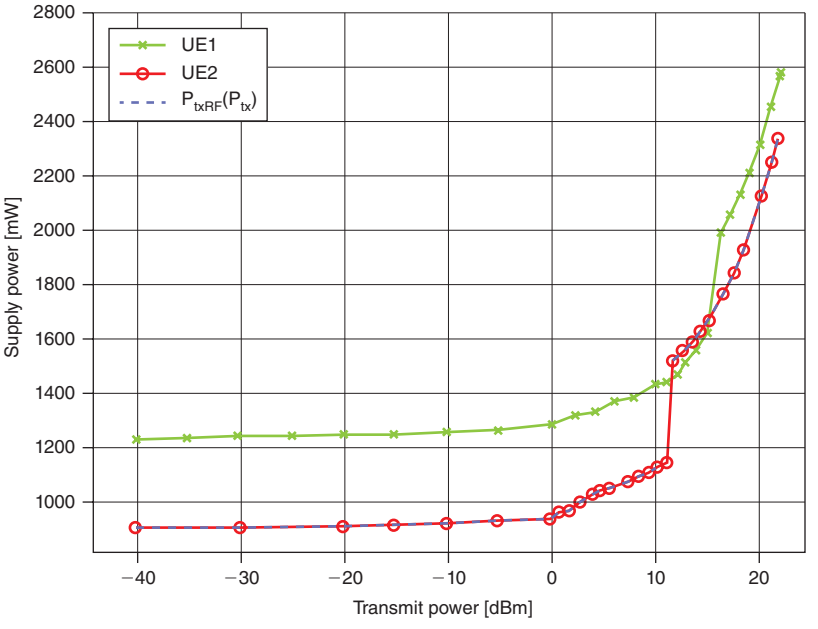


Figure 5: Supply power consumption as a function of transmit power (Source: Mads Lauridsen, Aalborg University, 2013)

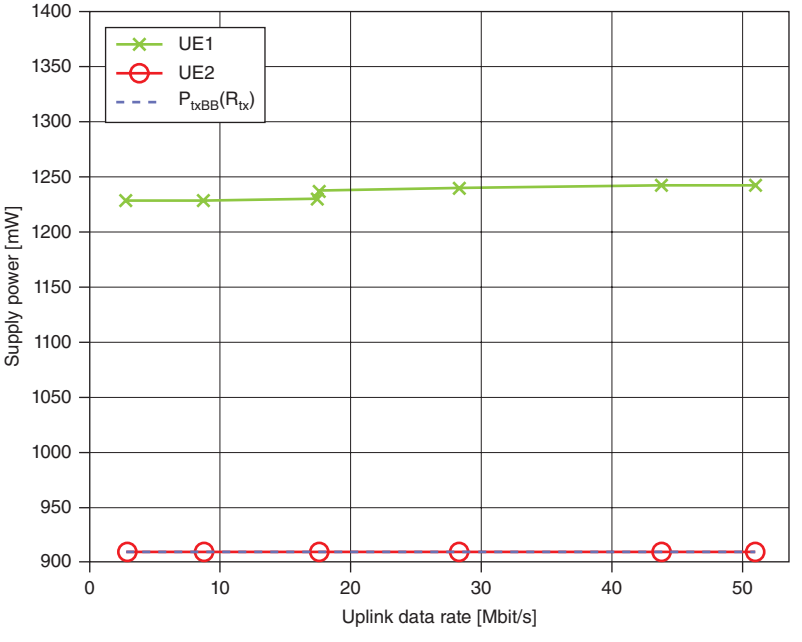


Figure 6: Supply power consumption as a function of UL data rate (Source: Mads Lauridsen, Aalborg University, 2013)

including the UE1, a step is observed when the modulation scheme is changed from QPSK to 16QAM. The reason is believed to be that certain PAs require a different bias/linearity setting to deliver the best compromise between power consumption and PA spectral emissions.

"…certain PAs require a different bias/linearity setting to deliver the best compromise between power consumption and PA spectral emissions."

Comparing the common point (UL MCS 6, UL PRB 100, Tx P -40 dBm) of Figure 5 and Figure 6, a maximum difference of 3.3 mW is observed, so there is good consistency between the measurements.

The UL parameter results illustrate how the new UE2 have improved the power consumption approximately 35 percent, but also that the choice of PA greatly affects the overall power consumption and that it can eliminate the advantage obtained by switching to a newer transceiver.

"…network designers can aim for the highest UL data rate without affecting UE power consumption, but that transmit powers above 10 dBm must be avoided when possible."

The results furthermore show that network designers can aim for the highest UL data rate without affecting UE power consumption, but that transmit powers above 10 dBm must be avoided when possible.

Downlink Characterization

As opposed to the UL, the DL of LTE Release 8 allows for use of Multiple Input Multiple Output (MIMO) antenna configuration, more specifically 2x2. Actually all UEs are required to have two Rx antennas, and therefore Rx diversity can be expected to be applied for all single-stream receptions. Furthermore spatial multiplexing, using two streams, is applicable in favorable channel conditions. This can greatly improve the DL data rate, but since the examined UEs are category 3, the DL data rate is limited to 100 Mbit/s.[9]

Figure 7 shows the power consumption as a function of received power, based on TC 3, and as expected the improvement from UE1 to UE2 is at least 30 percent. Furthermore observe that UE2 applies a different gain adjustment scheme. The insert of Figure 7 highlights how the scheme adjusts the gain of the Low Noise Amplifier in multiple steps in order to ensure a good compromise between the amplifier's linearity and power consumption.

The result of TC 1, used to examine the DL data rate's effect on power consumption, is shown in Figure 8. The Turbo decoding complexity is known to scale linearly with DL data rate[10], and this is clearly observable in the figure. The decoder power consumption does however not scale with the same proportion because increasing the data rate by a factor 10, for example from 5 to 50 Mbit/s, only increases the power consumption about 5 percent. This implies it is much more energy-efficient to run at high data rates. This is good for high data rate applications such as file transfers and high quality video streaming.
Finally it is interesting to observe that the use of two CWs only add a constant offset to the power consumption.

"…increasing the data rate by a factor 10, for example from 5 to 50 Mbit/s, only increases the power consumption about 5 percent."

Comparing the common point (DL MCS 0, DL PRB 100, Rx P–25 dBm) of Figure 7 and Figure 8, good consistency is again observed because the maximum difference is 1.4 mW.

The measurements on DL parameters showed the same 30–35 percent power consumption improvement as in UL, and that high data rates, similar to UL, results in the best EE.

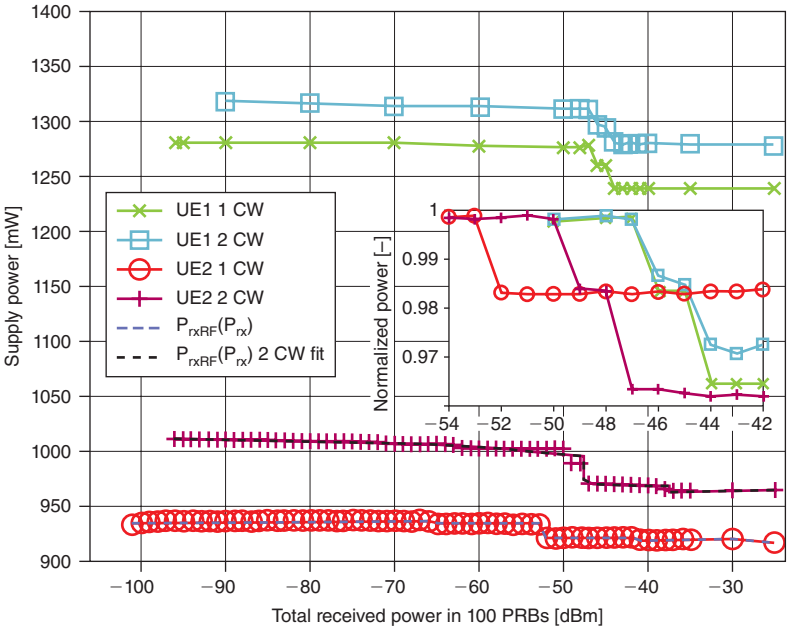


Figure 7: Supply power consumption as a function of receive power (Source: Mads Lauridsen, Aalborg University, 2013)

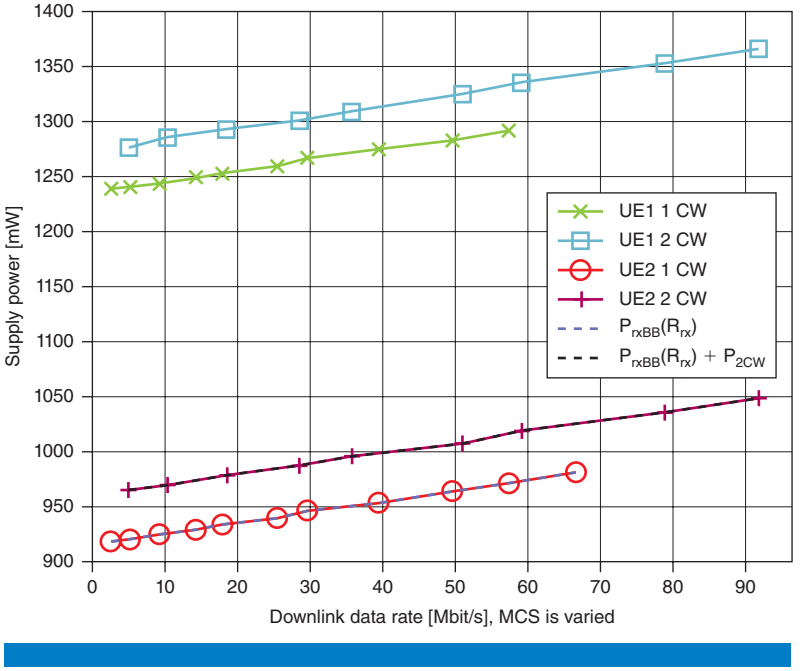


Figure 8: Supply power consumption as a function of DL data rate (Source: Mads Lauridsen, Aalborg University, 2013)

"When LP is less than 32 ms the UE does not enter a sleep mode at all and therefore DRX Short Period was not examined."

DRX Characterization

To examine DRX power consumption, the UEs were connected to the base station emulator and Connected Mode DRX was initiated. The DRX Long Period (LP) was varied from 32 ms to 256 ms, while the On Duration was set to 1 ms. When LP is less than 32 ms the UE does not enter a sleep mode at all and therefore DRX Short Period was not examined. Figure 9 shows two measurements on UE2 using DRX LP of 40 and 64 ms.

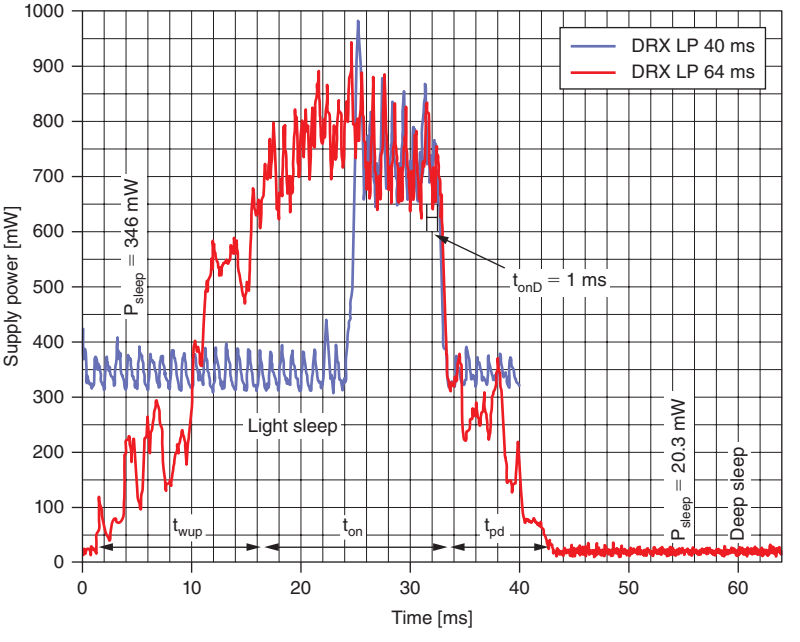


Figure 9: UE2 power consumption for DRX LPs 40 and 64 ms (Source: Mads Lauridsen, Aalborg University, 2013)

Based on the DRX measurements, values for power consumption and duration of each phase have been derived. The results are given in Table 3

Device	t_{LP}	P_{sleep}	t_{wup}	t_{pd}	t_{sync}	$E_{wup/pd+sync}$
UE1	≤ 40 ms	570 mW	6 ms	9 ms	8 ms	19.2 mJ
UE2	≤ 40 ms	346 mW	0.7 ms	0.6 ms	6.7 ms	6.45 mJ
Improvements		39%	88%	93%	16%	66%
UE1	≥ 80 ms	29 mW	26 ms	21 ms	21 ms	41.4 mJ
UE2	≥ 64 ms	20 mW	16 ms	10 ms	16 ms	19.3 mJ
Improvements		31%	38%	52%	24%	53%

Table 3: Measured DRX parameters. On Duration is 1 ms (Source: Mads Lauridsen, Aalborg University, 2013)

"The ratio between power consumed in the active and sleep mode has improved to 39.9 for deep sleep…"

The light and deep sleep power has improved 31–39 percent. This is as expected since similar improvements were noted in the previous sections. The ratio between power consumed in the active and sleep modes has however also improved from 1.8 to 2.2 and 35.6 to 39.9 for light and deep sleep respectively. This means the use of the sleep modes is even more effective.

In addition the wakeup and power-down times have also become shorter in the new UE2. In particular, the mode change times for light sleep have improved about 90 percent, which means it is much more applicable for short sleep periods. Previously it was discussed[3] how Nokia's widely used LTE DRX power model[11] does not correspond well with reality because the active-to-sleep ratio was assumed to be 50 and the transition time 1 ms, but the current results at least indicate the UEs are approaching Nokia's estimates. The sub-millisecond transition time has now been achieved for light sleep, but in this case the active to sleep ratio is far from 50. Still the conclusion remains that DRX is a key method to improve smartphones' battery life.

"…DRX is a key method to improve smartphones' battery life."

The synchronization time has also improved, but not as substantially as the aforementioned times, and the reason is the inherent limitation given by LTE's synchronization structure, where the synchronization signals only appear every 5 ms. Examining the synchronization phase after exiting deep sleep, it seems like there is room for improvement, but it must also be noted that achieving proper AGC and a valid channel estimate becomes more difficult when the UE has been sleeping for longer, because the old settings and estimates will be outdated.

"…the inherent limitation given by LTE's synchronization structure, where the synchronization signals only appear every 5 ms."

The 40 and 64 ms LPs were selected for Figure 9 because they represent the switching point where UE2 is applying either light or deep sleep. The light sleep is used when the LP is short or the On Duration is long, in either way eliminating the use of longer sleep periods. Furthermore the use of light sleep also represents the lowest energy consumption. This is illustrated in Figure 10, where the energy consumption as a function of DRX LP and sleep mode has

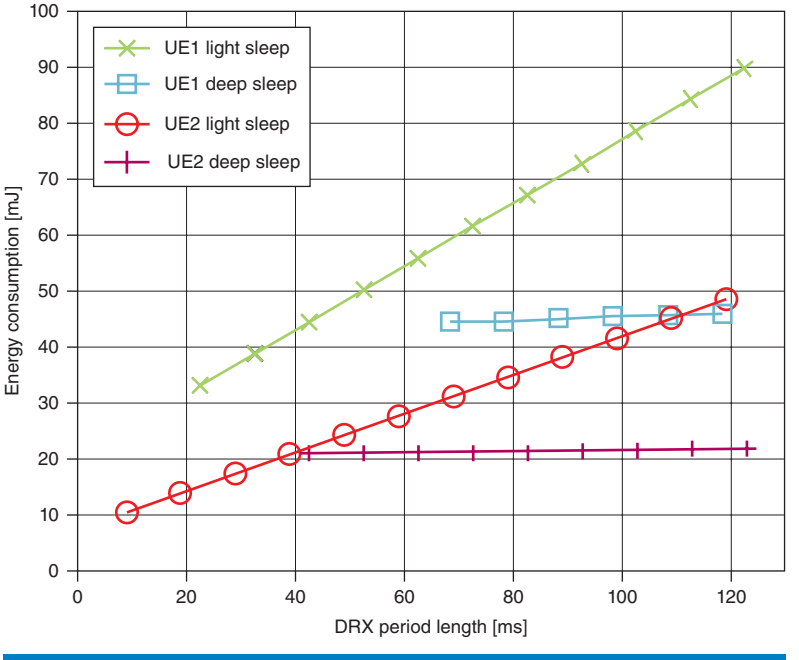


Figure 10: Energy consumption when using DRX sleep modes as a function of DRX LP length
(Source: Mads Lauridsen, Aalborg University, 2013)

been calculated using Table 3. The figure clearly illustrates that deep sleep is more energy-efficient for LP ≥ 64 ms, hence the UE sleep settings are well chosen in terms of energy consumption.

"when developing an empirical model it is important to evaluate how easy it is to update with new measurements."

Model Parameterization

When developing an empirical model it is important to evaluate how easy it is to update with new measurements. In our work more specifically the required number of measurement points per TC needed to achieve a proper fit. We previously discussed it[3], but now suggest the Tx BB is modeled using three points, since it is linear apart from the minor steps related to modulation scheme change. One point at each end of the data rate range is therefore sufficient, while a point in between is necessary as a sanity check. The Rx BB is also linear as a function of DL data rate and therefore three points should be sufficient. The Rx RF always includes UE-specific gain adjustments, which affect the power consumption, and therefore 5–6 points will be required to detect and model the steps of this piecewise linear function. The Tx RF exhibits a linear relation up to 0 dBm, thus the first part can be modeled using three points. For higher transmit powers, one measurement point per dB increase in transmit power is suggested. The reason for this accurate modeling is that the PA is the dominant power consumer.

Based on the above discussion it is clear how the functional blocks of the model in Figure 1 must be fitted to the measured data presented in the earlier subsections on uplink and downlink characterization. Each fit, representing one function in Equation 2, is based on adjusting a polynomial to the measured data by minimizing the least square error. The function of transmit power is divided into three piecewise linear sections due to its nonlinear behavior, while the function of UL data rate is a constant. The function of receive power is also divided into two sections due to the observed gain adjustment steps. The DL data rate function is a first order linear polynomial.

As mentioned earlier, the TCs in Table 1 are designed to have a common point, and the mean value of the four TCs in this point is 908 mW. By comparing this point with the 0 PRB point of TC 2 and 4, the cellular subsystem ON power, the active reception, and transmission power consumption were calculated. These values were then subtracted from the previously determined polynomials such that they can be applied in Equation 2 without contributing multiple times. The estimated polynomials are given in Table 4 and can be directly applied in Equation 2.

"…the cellular subsystem contributes to 30–50 percent of the total power consumption depending on transmit power, screen brightness, and CPU load."

For information on how the UE cellular subsystem compares with the power consumption of screen, central processing unit (CPU), and graphics processing unit (GPU), refer to our previous measurements.[3] We concluded that the cellular subsystem contributes to 30–50 percent of the total power consumption depending on transmit power, screen brightness, and CPU load.

The accuracy of the model fit is examined by comparing each of the measurement results with the model's estimated value. The relative error for each test point in each TC is illustrated in Figure 11. The maximum relative error is 3.3 percent hence a good fit has been achieved.

Part	Polynomial	Comment
P_{TxRF}	$0.78 \times S_{Tx} + 23.6$	$S_{Tx} \leq 0.2$ dBm
P_{TxRF}	$17.0 \times S_{Tx} + 45.4$	0.2 dBm $< S_{Tx} \leq 11.4$ dBm
P_{TxRF}	$5.90 \times S_{Tx}^{2} - 118 \times S_{Tx} + 1195$	11.4 dBm $< S_{Tx}$
P_{TxBB}	0.62	
P_{RxRF}	$-0.04 \times S_{Rx} + 24.8$	$S_{Tx} \leq -52.5$ dBm
P_{RxRF}	$-0.11 \times S_{Rx} + 7.86$	$S_{Tx} > -52.5$ dBm
P_{RxBB}	$0.97 \times R_{Rx} + 8.16$	
ON	853, 29.9, 25.1	Cellular subsystem, Tx, Rx active

Table 4: Polynomial fits in mW for the UE2-based model
(Source: Mads Lauridsen, Aalborg University, 2013)

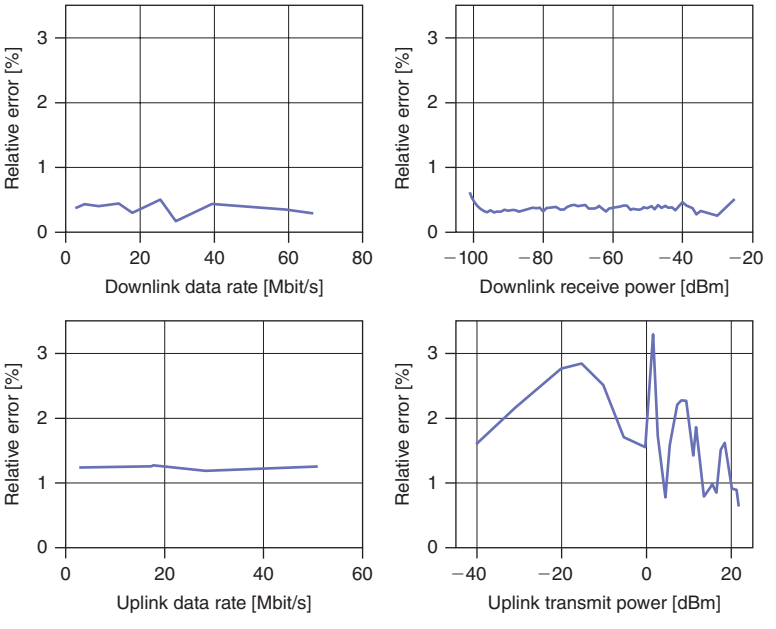


Figure 11: Relative error between estimated and measured power
(Source: Mads Lauridsen, Aalborg University, 2013)

Energy Efficiency Evolution

The measurements, presented in the previous sections, clearly showed the power consumption of LTE UEs have improved with each new chipset generation. The question is how the improvement compares with other modern RATs. We use the metric EE defined as the number of joules required to transfer one bit. In most articles the instantaneous power consumption is given as a function of data rate, but this is equal to EE:

"The measurements clearly showed the power consumption of LTE UEs have improved with each new chipset generation."

P [W] / R [bit/s] > W × s/bit = J/bit

"The downlink energy efficiency has improved with each RAT generation, as a result of improvements in CMOS node and devices in general, but also due to changes in the technologies used in the RATs …"

Our study is based on a review of power consumption measurements reported in recent literature; EDGE, HSPA, and Wi-Fi* 802.11g (Wi-Fi) has been reported for an HTC Hero by Wang and Manner[12], Friedman et al.[13] analyzed Bluetooth* (BT) 2.0 and Wi-Fi power consumption in a Samsung i900*, while Perruci et al.[14] covered BT 2.0, GPRS, HSDPA, and Wi-Fi using a Nokia N95. Xiao et al.[15] examined Wi-Fi using both a Nokia N95* and a Nexus S*. Our measurements on LTE dongles[6] and smartphones[3] are also included. In addition Texas Instruments have reported the power consumption of their standalone Bluetooth 4.0 Low Energy (BT LE)* chip.[16] Finally System-on-Chip measurements on BT LE and ZigBee are reported for UL by Siekkinen et al.[17] The latter two studies obviously differ because they only cover the RAT chip and not a fully functional phone.

The DL EE is shown in Figure 12 for the examined devices. Usually the power consumption is reported as a function of increasing data rate, and therefore Figure 12 includes the dependency on both low and high data rates.

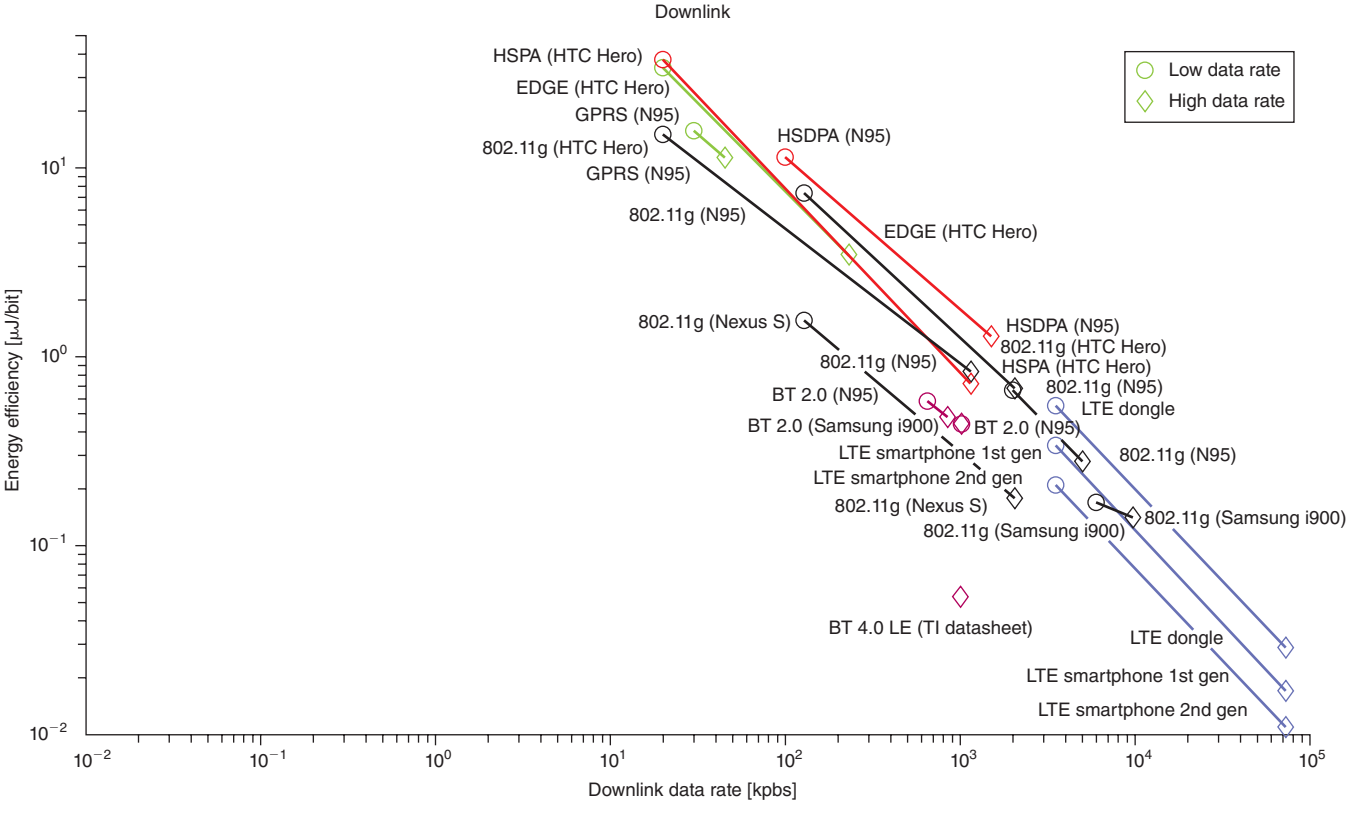


Figure 12: Downlink energy efficiency for modern RATs (Source: Mads Lauridsen, Aalborg University, 2013)

The DL EE has generally improved with each RAT generation, as a result of improvements in complementary metal–oxide–semiconductor (CMOS) node and devices in general, but also due to changes in the technologies used in the RATs, such as switching from CDMA to OFDMA.[18]

LTE achieves both the highest data rates and the best EE, while Wi-Fi is number two in both categories. One interesting observation is that the slope between the low and high data rate points is similar for all technologies.

If the target is Machine Type Communications (MTC), with low data rate, none of the systems seem optimal, because the EE rapidly decreases as the data rate is lowered. Currently the 3rd Generation Partnership Project (3GPP)[19] is running a study on MTC for LTE, which includes reducing the bandwidth and peak data rates, together with a single RF chain and lower transmit power to make LTE cost competitive and energy efficient.

"If the target is Machine Type Communications, with low data rate, none of the systems seem optimal, because the energy efficiency rapidly decreases as the data rate is lowered."

Comparing the EE for UL transmission is more complicated because the transmit power and the general range of the system plays an important role.

The result of the literature review is shown in Figure 13. As in Figure 12, low and high data rates are reported, when available, and additionally the transmit power of the device is included.

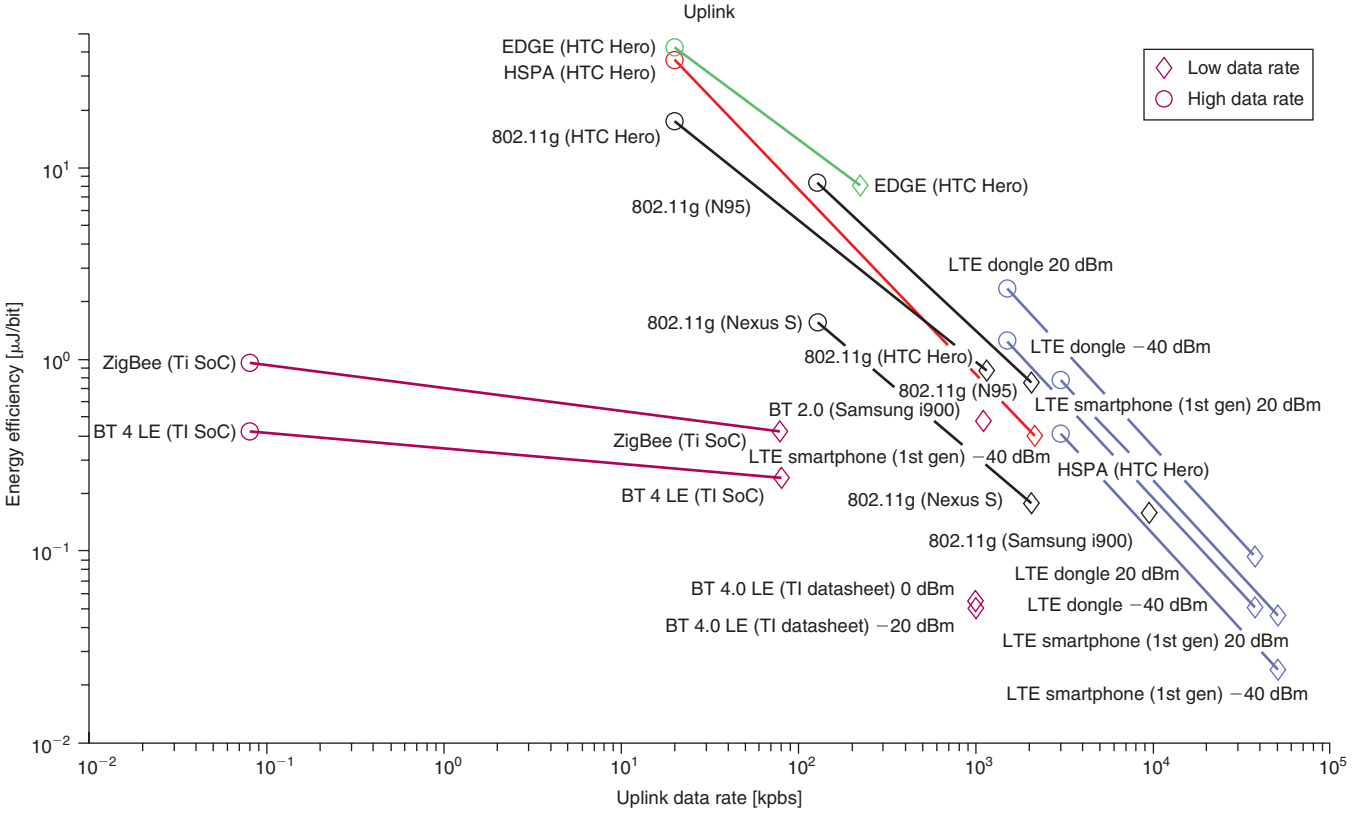


Figure 13: Uplink energy efficiency for modern RATs (Source: Mads Lauridsen, Aalborg University, 2013)

Again LTE proves to be the most energy-efficient RAT at high data rates, both for low and high transmit powers. The LE version of BT, based on TI's datasheet[16], is almost as efficient at a data rate approximately two orders of magnitude lower. This means it is very useful for MTC, but one could

have expected the BT LE to be even more efficient because it was developed specifically for low power purposes. The limiting factor is the low data rate, which prevents the EE from improving significantly. The chip implementations of BT and ZigBee[17] are a little less efficient, but at data rates below 100 kbps, no other RAT can compare with them in terms of EE. The authors[17] furthermore mention BT can be made even more efficient using another protocol stack.

It is important to note that the "communication range" of the RATs differs a lot. The mobile communication systems, such as HSPA and LTE, can have a range of several kilometers, whereas Wi-Fi and BT are limited to 10–100 m. This affects their applicability in certain MTC scenarios, and therefore the 3GPP work on LTE for MTC[19] is important.

"…LTE has proven to be the currently most energy-effective RAT for transferring data…"

To conclude, LTE has proven to be the currently most energy-effective RAT for transferring data, and based on our new measurement and the observed trend it is not expected to change. It is for further study to evaluate how tail energy[4], which covers the energy consumed after the actual data transmission is completed and is due to network and RAT dependent timeouts, affect each RAT.

Energy Efficiency Improvements

The device maturity may not be enough to guarantee user satisfaction with regards to the battery life and therefore researchers are investigating methods that do not affect the current LTE standard, but decrease the power consumption. In the following subsections we discuss the micro sleep concept and how CA may affect the battery life.

Micro Sleep

One issue in LTE is that the UE is forced to receive and buffer the Physical Downlink Data Channel (PDSCH) while it is decoding the Physical Control Format Indicator Channel (PCFICH) and Physical Downlink Control Channel (PDCCH), which carry scheduling information about PDSCH.[9] This occurs every subframe and if the UE is not scheduled it will be receiving and buffering PDSCH for no purpose.

To deal with this issue the Fast Control Channel Decoding[20] concept has been proposed. The idea is to perform a fast decoding of the control channels, stop buffering the PDSCH if the UE is not scheduled, and then power down specific RF and BB components. The UE has to wake up and receive the next subframe, meaning the sleep period is no longer than 7–9 symbols (0.47–0.60 ms) hence the label "micro sleep." The concept is illustrated in Figure 14. The cost is that the UE will not receive the Reference Signals (RS) in the latter part of the subframe. In literature[20] this has been described as an SNR loss, which was simulated to result in a throughput degradation of 1–4 percent. On the other hand potential energy savings of 5–25 percent were reported and therefore the concept was deemed valuable.

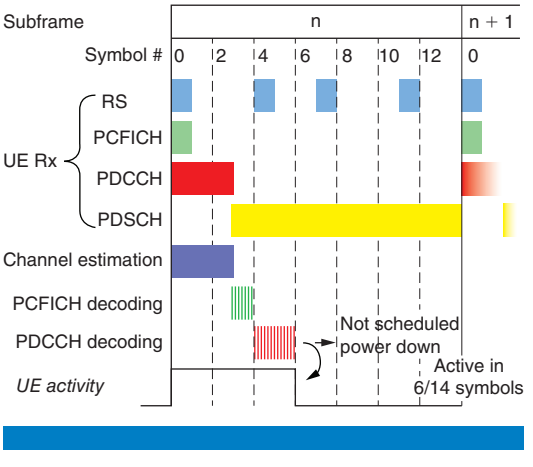


Figure 14: Micro sleep in LTE Release 8 (Source: Lauridsen et al. 2012[20])

Further advantages of the micro sleep concept include the fact that it can complement DRX, and it fits all types of traffic scenarios as opposed to DRX, which require a periodic pattern to be effective. In addition there is no increase in control message overhead as in DRX where configuration parameters are transferred. Finally the network scheduler will not be affected, because the concept is applied autonomously and individually by each UE.

"Further advantages of the micro sleep concept include the fact that it can complement DRX, and it fits all types of traffic scenarios as opposed to DRX…"

Comparing the assumptions in the micro sleep literature[20] with current smartphones' DRX capabilities, as presented in the section "DRX Characterization," it is clear that the instantaneous power consumption cannot be lowered as much as initially expected. The reason is the wakeup and power-down times caused by powering ON and OFF of UE components, which was measured to be 0.6–0.7 ms. Table 3 does however show great improvements in wakeup and power-down times from UE1 to UE2 and therefore UE manufacturers may be able to apply the micro sleep concept in future LTE generations.

Unfortunately the introduction of the Enhanced PDCCH (EPDCCH) in LTE Release 11 has precluded the use of micro sleep. The reason is that the E-PDCCH is spread across the whole subframe time-wise, as illustrated in Figure 15, in order to obtain a frequency diversity gain by only using selected resource blocks in the frequency domain.

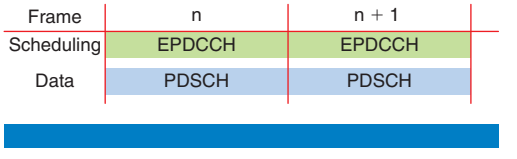


Figure 15: Control and data channel position in LTE Release 10
(Source: Lauridsen et al. 2014[18])

In a recent proposal[21] for a next generation RAT, the control and data channel position has however been reordered such that the control data is a whole frame ahead of the data as illustrated in Figure 16. This allows for efficient pipelining and micro sleep.

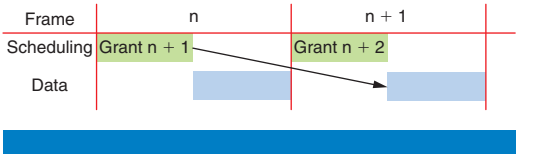


Figure 16: Control and data channel position in a 5G concept[21]
(Source: Lauridsen et al. 2014[18])

Carrier Aggregation

Carrier Aggregation is included in LTE Release 10 to improve user throughput and coverage. The standardization of CA entails a more complicated transceiver design, because the UE needs to be able to receive at least two (up to five) carriers simultaneously each up to 20 MHz wide. The additional hardware can potentially lead to increased UE power consumption, hence the search for even higher data rates may worsen the users' battery life.

"Carrier Aggregation is included in LTE Release 10 to improve user throughput and coverage, but entails a more complicated transceiver design…"

To examine this issue we proposed a narrow and a wideband UE power model[22] and calculated the energy consumption in a heterogeneous network (HetNet) scenario using macro and small cells. The narrowband model applies two RF front ends and two analog-to-digital converters (ADCs), while the wideband model applies a single RF front end and ADC, but with double bandwidth capability.

The users were set to receive a file either via single carrier LTE Release 8 or using two carriers. The narrowband CA UE was estimated to consume 20 percent more power on average as compared to the Release 8 UE, but as illustrated in the simulation results in Figure 17, the energy consumption is approximately the same for both UEs. The reason is that with CA the

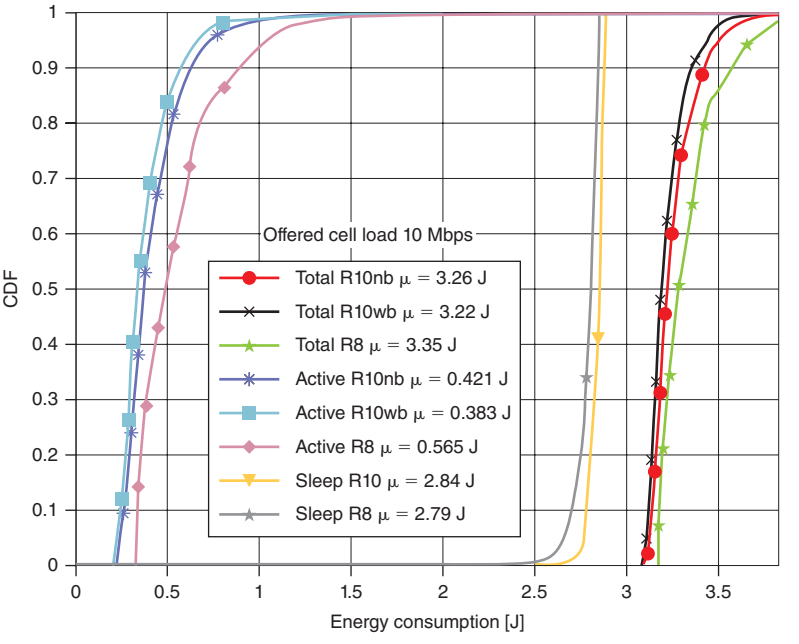


Figure 17: Carrier Aggregation energy consumption in a HetNet scenario (Source: Lauridsen et al. 2013[22])

throughput increased more than 50 percent in the simulated scenario, and therefore the UE can receive the file faster with little penalty on the power consumption of the Rx BB as illustrated in Figure 8. Then the UE can enter a low power sleep mode and achieve high EE. If the throughput does not increase at least 20 percent[22], CA will lead to decreasing EE.

"If the throughput does not increase at least 20 percent, Carrier Aggregation will lead to decreasing energy efficiency."

When the work was performed there was no knowledge about how fast the transition to sleep mode could be made, but the measurements in the section "DRX Characterization" have made it clear that shifting to DRX sleep mode takes a considerable amount of time. Therefore CA will mainly be effective for very large file transfers, where the time difference between CA and Release 8 UEs, including the transition time, is large. Otherwise the penalty on the user may be heavy due to the increased instantaneous power consumption.

As mentioned, the work[22] was based on a theoretical extension of the existing power model[6], but recently a vendor has launched a CA device[23], which could help clarify if the assumptions were correct.

Conclusion

"The model covers the cellular LTE subsystem and is based on measurements on the newest generation of LTE smartphones."

In this work an empirical smartphone power model was presented. The model covers the cellular LTE subsystem and is based on measurements on the newest generation of LTE smartphones. By comparing with our previous measurements on older LTE generations, power consumption improvements of approximately 35 percent were noted and attributed to device maturity. The LTE Discontinuous Reception feature was also examined and the results

show the deep sleep power is now as low as 1/40 of the active mode power. Furthermore the new smartphone is able to enter and exit the sleep modes at least 30 percent faster, which makes DRX more applicable in a real network and enhances the possibility for use of micro sleep.

The cellular subsystem model is intended for use in system level simulations to evaluate how specific network settings affect user equipment power consumption. The measurements show that the power consumption is dominated by the subsystem being ON, consuming about 0.9 W, and also very affected by transmit powers above 10 dBm, consuming an additional 0.6–1.5 W. The power consumption is almost independent of uplink and downlink data rates, and therefore the combination of high data rates and long sleep periods must be the target of an energy-efficient network setup.

"…the power consumption is dominated by the subsystem being ON, consuming about 0.9 W, and also very affected by transmit powers above 10 dBm, consuming an additional 0.6–1.5 W."

"…the combination of high data rates and long sleep periods must be the target of an energy-efficient network setup."

The work also surveys the energy efficiency, in terms of number of joules required to transfer one bit, of multiple radio access technologies. For high data rates, LTE is superior to older technologies such as EDGE, HSPA, and 802.11g Wi-Fi.

Finally it was evaluated that Carrier Aggregation, which is a prominent new LTE feature, will affect the energy efficiency positively if the throughput can be increased 20 percent as compared to conventional single-carrier LTE UEs.

Acknowledgements

Thanks to Frank Laigle, Vidéotron, Taehee Song, Anritsu Canada, and Janus Faaborg, Agilent Technologies Denmark for supporting this work. The work is partly funded by the Danish National Advanced Technology Foundation and the 4GMCT project.

References

[1] Murmuria, R., J. Medsger, A. Stavrou, and J. M. Voas, "Mobile Application and Device Power Usage Measurements," IEEE Sixth International Conference on Software Security and Reliability, 2012.

[2] Wang, C., F. Yan, Y. Guo, and X. Chen, "Power Estimation for Mobile Applications with Profile-Driven Battery Traces," IEEE ISLPED, 2013.

[3] Lauridsen, M., L. Noël, and P. Mogensen, "Empirical LTE Smartphone Power Model with DRX Operation for System Level Simulations," VTC, Fall 2013.

[4] Huang, J., F. Qian, A. Gerber, Z. M. Mao, and S. Sen, "A Close Examination of Performance and Power Characteristics of 4G LTE Networks," MobiSys'12, 2012.

[5] Dusza, B., C. Ide, L. Cheng, and C. Wietfeld, "An Accurate Measurement-Based Power Consumption Model for LTE Uplink Transmissions," INFOCOM 2013.

[6] Jensen, A. R., M. Lauridsen, P. Mogensen, T. B. Sørensen, and P. Jensen, "LTE UE Power Consumption Model – For System Level Energy and Performance Optimization," VTC Fall 2012.

[7] Holma, Harri, Antti Toskala, and Pablo Tapia (eds.), *HSPA+ Evolution to Release 12: Performance and Optimization* (Wiley, 2014), ISBN: 9781118503218.

[8] Lauridsen, M., A. R. Jensen, and P. Mogensen, "Reducing LTE Uplink Transmission Energy by Allocating Resources," VTC Fall 2011.

[9] Holma, Harri and Antti Toskala (eds.), *LTE for UMTS: OFDMA and SC-FDMA Based Radio Access* (Chichester, UK: Wiley, 2009) ISBN 9780470994016.

[10] Holma, Harri and Antti Toskala, *LTE for UMTS: Evolution to LTE-Advanced,* 2d Ed (Chichester, UK: Wiley, 2011), ISBN 9780470660003.

[11] Nokia, DRX Parameters in LTE – 3GPP R2-071285, 2007.

[12] Wang, L. and J. Manner, "Energy Consumption Analysis of WLAN, 2G and 3G interfaces," IEEE GreenCom 2010.

[13] Friedman, R., A. Kogan, and Y. Krivolapov, "On Power and Throughput Tradeoffs of Wi-Fi and Bluetooth in Smartphones," IEEE Infocom 2011.

[14] Perrucci, G.P. and F. Fitzek, "Measurements campaign for energy consumption on mobile phones," Technical report, Aalborg University 2009.

[15] Xiao, Y., Y. Cui, P. Savolainen, M. Siekkinen, A. Wang, L. Yang, A. Ylä-Jääski, and S. Tarkoma, "Modeling Energy Consumption of Data Transmission over Wi-Fi," *IEEE Transactions on Mobile Computing*, 2013.

[16] Texas Instruments, CC2541, 2.4-GHz Bluetooth low energy and Proprietary System-on-Chip. Datasheet, 2012.

[17] Siekkinen, M., M. Hiienkari, J. K. Nurminen, and J. Nieminen, "How Low Energy is Bluetooth Low Energy? Comparative Measurements with ZigBee/802.15.4," WCNC Workshop 2012.

[18] Lauridsen, M., G. Berardinelli, T. B. Sørensen, and P. Mogensen, "Ensuring Energy-Efficient 5G User Equipment by Technology Evolution and Reuse," Submitted for VTC Spring 2014.

[19] 3rd Generation Partnership Project, "Study on provision of low-cost Machine-Type Communications (MTC) User Equipments (UEs) based on LTE," TR 36.888 v.12.0.0, 2013.

[20] Lauridsen, M., A. R. Jensen, and P. Mogensen, "Fast Control Channel Decoding for LTE UE Power Saving," VTC Spring 2012.

[21] Mogensen, P., K. Pajukoski, B. Raaf, E. Tiirola, E. Lahetkangas, I. Z. Kovacs, G. Berardinelli, L. G. U. Garcia, H. Liang, and A. F. Cattoni, "B4G local area: High level requirements and system design," IEEE Globecom Workshops 2012.

[22] Lauridsen, M., H. Wang, and P. Mogensen, "LTE UE Energy Saving by Applying Carrier Aggregation in a HetNet Scenario," VTC Spring 2013.

[23] Qualcomm, "World's First Mobile Device with LTE Advanced Carrier Aggregation Powered by the Qualcomm Snapdragon 800 Processor," http://www.qualcomm.com/media/blog/2013/06/26/worlds-first-mobile-device-lte-advanced-carrier-aggregation-powered-qualcomm, 2012.

Author Biographies

Mads Lauridsen received his MSc EE from Aalborg University in 2009 and is currently pursuing a PhD at Aalborg University in the Radio Access Technology Section. His work focuses on user equipment energy consumption in LTE and future radio access technologies. He can be reached at ml@es.aau.dk.

Laurent Noel received a degree in mathematics and physics from the University of Montpellier II, France, in 1991, and a degree as a microelectronics engineer at the Institut des Sciences de l'Ingenieur (ISIM: now Polytech-Montpellier) in 1994. From 1994 to 2000 he worked at British Telecom (VT) Research Laboratories. From 2000 to 2010, he worked at Philips Semiconductors. He became Senior Principal System Architect at ST-Ericsson and worked on reconfigurable multimode, multiband, digital RF solutions for LTE, FDD-WCDMA, EGPRS. He currently works in the mobile phone certification team at Videotron, Montreal (Canada). He can be reached at Laurent.Noel@ videotron.com.

Troels B. Sørensen graduated in 1990 (MSc EE) from Aalborg University and received his PhD degree from the same university in 2002. He has had experience in industry (telecom operator) and academia since 1997. He holds a position as associate professor in the Radio Access Technology Section at Aalborg University, Department of Electronic Systems, where his primary involvement is in the supervision of Master and PhD students and research related to wireless communication. He can be reached at tbs@es.aau.dk.

Preben Elgaard Mogensen received his MSc EE and PhD in 1988 and 1996 from Aalborg University, Denmark. He is currently Professor at Aalborg University leading the Radio Access Technology Section. Preben Mogensen is also associated with Nokia Siemens Networks. His current research work is related to heterogeneous network deployment, cognitive radio, and beyond 4G. He can be reached at pm@es.aau.dk.

EVOLVING THE NETWORK FOR MACHINE-TO-MACHINE COMMUNICATION

Contributors

Dr. Robert Zaus
Platform Engineering Group,
Intel Corporation

Hyung-Nam Choi
Mobile and Communications Group,
Intel Corporation

Machine-to-Machine (M2M) communication is a fast-growing new business segment for cellular mobile networks. M2M communication is used by a variety of applications. Typical for many of them is a large number of terminals, each of which is generating only little user data traffic. So the challenge for the network is not so much a demand for high data rates, but rather to cope with the amount of signaling data that can suddenly be generated by the huge number of M2M devices and at the same time to provide a sufficient level of service to human users attached to the same network. In the present article we describe the various enhancements to the architecture and the protocols of 3GPP mobile networks to meet the increasing demand for M2M communication and to protect the network against signaling overload. While some of the new mechanisms were designed specifically for M2M devices, many of them will also be used to keep the signaling created by the fast-growing number of smartphones under control. In the second part of the article we discuss "device triggering" and other new features that were developed by 3GPP particularly for M2M devices.

Introduction and Overview

Machine-to-Machine (M2M) communication is a rapidly expanding new business segment for cellular mobile networks. Devices for M2M communication are omnipresent: for example logistics enterprises are using them for fleet management, road tolling, and anti-theft protection, supply companies for power, gas, and water metering. Further applications include the control of surveillance systems and maintenance of vending machines. New generations of wearable devices are going to facilitate smaller and more convenient remote health diagnostic systems.

"…the challenge for the network is not so much to satisfy the demand for high data rates, but rather to cope with the large amount of signaling data…"

For many of these applications it is typical that there is a large number of terminals, each of which is generating only little user data traffic. Furthermore, for these terminals the ratio between signaling data and user data exchanged with the network is usually higher than for a regular mobile phone. So the challenge for the network is not so much to satisfy the demand for high data rates, but rather to cope with the large amount of signaling data that can suddenly be generated by the huge number of M2M devices and at the same time to provide a sufficient level of service to human users attached to the same network.

Beginning with Release 10 (2009–2011), 3GPP reacted on the growing interest in machine-to-machine (M2M) communication and initiated work on various network enhancements and optimizations under the work item title "network improvements for machine-type communications (NIMTC)" based on the service requirements specified in 3GPP TS 22.368.[1]

Since operators had already gained first experience with the signaling load that can be generated by smartphones, the focus in Release 10 was on the introduction of new mechanisms for overload and congestion control, especially for the signaling traffic. The only entity that was added to the network architecture is an MTC application server (MTC AS) hosting the M2M applications, which can be operated by the network operator or by an independent service provider (see Figure 1).

"…the focus in Release 10 was on the introduction of new mechanisms for overload and congestion control, especially for the signaling traffic."

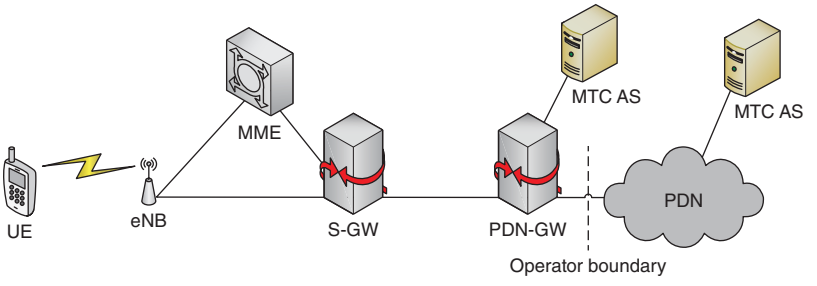


Figure 1: Network architecture for M2M communication in Release 10 (Source: Intel, 2013, based on 3GPP TS 23.401[2] and 3GPP TR 23.888, v 1.0.0[3])

In Release 11 (2010–2012), a new feature called *device triggering* was introduced. As this was based on the short message service (SMS), 3GPP developed various enhancements to this service. Furthermore, operators required better support of a subscription including only packet services ("PS only" subscription), specifically removal of the need to assign a mobile subscriber ISDN number (MSISDN).

"…operators required better support of a subscription including only packet services…, specifically removal of the need to assign a mobile subscriber ISDN number (MSISDN)."

For the current Release 12 (2012–2014), 3GPP is continuing the work on device triggering and SMS enhancements. As new topics, signaling optimizations for the transfer of small data and UE power consumption optimizations are being studied.

Overload and Congestion Control

For the control of signaling overload and congestion, 3GPP designed a set of mechanisms forming several lines of defense (see Figure 2). The various mechanisms take effect:

- Before the UE is accessing the radio access network (RAN)
- When the radio controller in the RAN (eNB, RNC, BSC) receives the first message from the UE
- When the first core network node (MME, SGSN, MSC) receives the first signaling message from the UE
- When the gateway to the external packet data network (PDN-GW, GGSN) receives a message from the UE

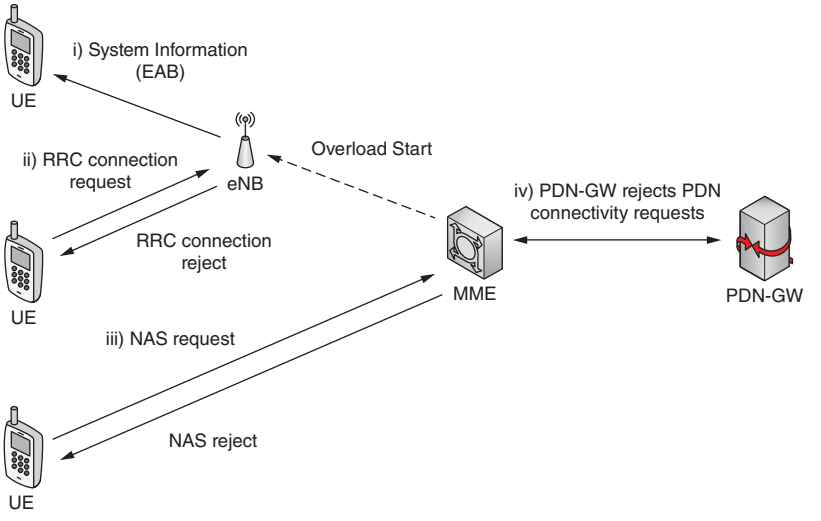


Figure 2: Overload and congestion control (Source: Intel, 2013, based on 3GPP TS 23.401[2])

Some of these mechanisms are only applicable to UEs specifically configured by their HPLMN operator, others to all UEs compliant to 3GPP Release 10 or later. For example if a UE is running only applications that are not time-critical, it can be configured for "extended access barring (EAB)" and "NAS signaling low priority." Configuration is possible either via OMA device management or on the USIM via the SIM toolkit.

As a means to restrict access to the network by a huge number of MTC devices, a UE configured for extended access barring (EAB) needs first to pass an EAB-specific check before it is allowed to proceed with a second check for common access control, which is obligatory for all UEs. For the EAB check, the UE compares its own access class, a number between 0 and 9 stored on the USIM, with the EAB information broadcast by the radio access network (see Figure 2, i). If the network indicates that the access class of the UE is barred, then the UE should not even attempt to access the network. The operator of the visited network can configure the EAB information so that it is applicable to all UEs configured for EAB or for example only to the subset of UEs configured for EAB and belonging to inbound roaming subscribers. Usually the inbound roaming subscribers constitute only a small percentage of the total number of UEs attached to a network. This is different for MTC devices that are often equipped with a USIM from a foreign country so that the device can use roaming agreements with several operators of the visited country, and consequently it can select another network, if the network to which it is currently registered fails. In Release 10, EAB was first standardized for GERAN. Support for UTRAN and E-UTRAN was added in Release 11.

"If the network indicates that the access class of the UE is barred, then the UE should not even attempt to access the network."

"…MTC devices…are often equipped with a USIM from a foreign country so that the device can use roaming agreements with several operators of the visited country,…"

For UTRAN and E-UTRAN, a UE configured for "NAS signaling low priority" will indicate the RRC establishment cause "delay tolerant" when requesting the RAN to establish a connection for a mobile originating

transaction. The establishment cause can be used by the RAN to give a lower priority to this request and, when the core network is in overload, reject the request with an "extended wait time value" in the range 1 to 1800 seconds (see Figure 2, ii). The UE is then not allowed to repeat the request before the respective time interval has elapsed. For GERAN, in order to save signaling channel capacity 3GPP developed an "implicit reject" mechanism that enables the RAN to reject access requests of several UEs with a single message.

"The establishment cause can be used by the RAN to give a lower priority to this request and, when the core network is in overload, reject the request with an "extended wait time value"…"

A UE configured for "NAS signaling low priority" will indicate this also in the first mobility management message and all session management messages sent to the core network. The first core network node (MME, SGSN, MSC) can use this indication to prioritize mobility management requests and reject the request with a back-off time in the range from 2 seconds to 3 hours 6 minutes, preventing the UE from another network access until the back-off time has elapsed (see Figure 2, iii). The criteria for rejecting the request can be for example the overall load situation or membership to a group of devices that have a subscription for a specific APN.

Both the MME/SGSN and the PDN-GW/GGSN can detect an overload situation, for example based on the rate of session management context activations for a specific access point name (APN), and reject further requests for the same APN with a back-off time in the range from 2 seconds to approximately 13 days, (see Figure 2, iii and iv).

"Both the MME/SGSN and the PDN-GW/GGSN can detect an overload situation,…, and reject further requests for the same APN with a back-off time…"

The core network can send back-off time values also in response to mobility management and session management requests without "NAS signaling low priority". That is, this back-off mechanism needs to be supported by all UEs compliant with Release 10 or later.

Subscribers belonging to a special group of users (PLMN staff, police, fire department, and so on) can use a special access class in the range 11 to 15. When accessing the network with such a special access class or for the purpose of an emergency call, the UE does not indicate "delay tolerant" in the RR/RRC establishment cause or "NAS signaling low priority" in the first message sent to the core network, and it is not subject to the overload control mechanisms described above.

Even if an application is generally delay tolerant, there can be conditions, such as theft of a device or detection of a critical health condition, where it is necessary to inform the MTC AS quickly. Since Release 11, a UE that is configured with the capability to override "NAS signaling low priority" or "EAB" can, upon request from an MTC application, ignore any back-off time values that were provided in response to requests issued with "NAS signaling low priority" and any EAB information broadcasted by the network, and send its NAS messages without "NAS signaling low priority" indication. A similar mechanism is already available in Release 10, if the device needs to establish a connection to an emergency center.

A new feature for all Release 10 UEs is an increased value range for the periodic location update timer, routing area update timer, and tracking area update

timer, although values of up to 13 days are expected to be used only for M2M devices. Furthermore, M2M devices can be configured with a longer value for the timer controlling the periodic search for higher prioritized PLMNs. Both mechanisms are intended to reduce the signaling or at least mitigate peaks in the signaling load which can occur when in a certain area a PLMN goes out-of-service or returns into service.

For a complete list of overload control mechanisms and additional features for M2M communication see 3GPP TS 23.401.[2]

Device Triggering and SMS via Packet Domain

"Device triggering provides a means for an application on the MTC application server… to send a small amount of data ("trigger payload") to the UE…"

In Release 11, further entities and interfaces were introduced to support a new feature called device triggering.[4] A 2G/3G UE can be attached to a GPRS network without having an IP address allocated. Device triggering provides a means for an application on the MTC application server (MTC AS) to send a small amount of data ("trigger payload") to the UE when it is in such a state. Inside the UE the data will be delivered to an M2M application and trigger an application specific action; for example, at a specific point in time in the future the M2M application on the UE could request the UE to get an IP address allocated by the mobile network and establish a connection with the MTC AS. Device triggering is offered as a service to the MTC AS by a new network entity called Services Capability Server (SCS), which can be operated by the network operator or by an independent service provider (see Figure 3). In principle, different mechanisms could be used to deliver the trigger payload. In Release 11 the only standardized mechanism is short message service (SMS).

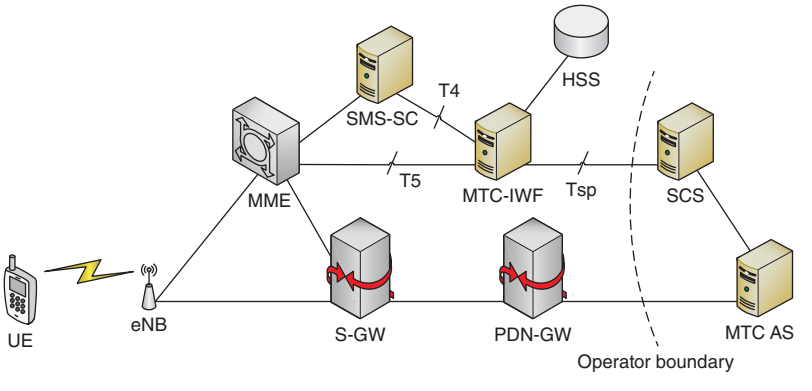


Figure 3: Network architecture for device triggering (Source: Intel, 2013, based on 3GPP TS 23.682[4])

The SCS communicates with the MTC interworking function (MTC-IWF), which authorizes any device triggering request received from an external SCS, performs a mapping from an external identifier (for example a network access identifier [NAI]) to the IMSI of the device, and initiates the signaling towards the short message service center (SMS-SC) if the trigger payload is to be delivered via SMS (trigger delivery via T4, see Figure 3).

The external identifier is needed when the device cannot be addressed via an IP address. The IMSI cannot be used for this purpose, since for privacy reasons it should not be provided to network-external third parties. On the other hand, with the increasing number of terminals, it becomes impractical to assign a telephone number (MSISDN = mobile subscriber ISDN number) to every single device, as the number range available to a network operator is limited. Since many of the protocols used in the packet domain were standardized by IETF, 3GPP adopted the network access identifier (NAI), a format defined by IETF, for the external identifier.[5] This enables operators to offer a subscription for packet services without MSISDN.

As SMS is playing such a prominent role for M2M communications, 3GPP came up with a number of enhancements to improve the usability of the service via the packet domain. These include:

- MAP (Mobile Application Part) protocol modifications to enable the delivery of short messages to devices without MSISDN number
- Signaling enhancements for 2G/3G GPRS which ensure that a device only interested in receiving packet services and SMS will actually receive the short messages via the packet domain if this is supported by all the involved network entities. So the device does not need to attach also to the CS domain just to receive short messages.
- A new option to deliver short messages via the evolved packet core ("SMS in MME"), allowing operators to avoid the deployment of IMS or of a CS domain only for this purpose
- The support of an alternative protocol solution for SMS in MME at core network internal interfaces, based on the IETF protocol Diameter and allowing operators to avoid the introduction of the ITU-T SS7 based MAP.

Since in many mobile networks, short messages are still by default delivered via the CS domain, 3GPP also created a specific subscription for packet services and SMS only ("PS and SMS only"). That is, SMS is the only service provided to such a subscriber via the CS domain.

MTC Enhancements in Release 12

For the current Release 12, 3GPP is working on mechanisms for overload control for the reference points Tsp and T4, as the signaling load generated by device triggering should not negatively affect the regular network operation.

"…the signaling load generated by device triggering should not negatively affect the regular network operation."

Further enhancements for SMS will include:

- The support of Diameter-based protocols at core network internal interfaces also for SMS via SGSN
- Protocol adaptations necessary to support SMS over IMS for subscribers without MSISDN

Other topics that were studied include signaling optimizations for the transfer of small data and UE power consumption optimizations.

"For many M2M applications, the amount of data exchanged between the M2M device and the network is relatively small, often less than 1 kB."

"Concepts under study include the introduction of a specific power saving state for the UE in which the UE remains attached to the network but stops any activity towards the network."

For many M2M applications, the amount of data exchanged between the M2M device and the network is relatively small, often less than 1 kB. Therefore, 3GPP was studying optimizations that enable the network to use network resources more efficiently and transfer small data packets with less signaling overhead. This could also be used for device triggering without SMS (trigger delivery via T5, see Figure 3).

Moreover, 3GPP was looking for options to optimize the UE power consumption to extend its battery life. Concepts under study include the introduction of a specific power saving state for the UE in which the UE remains attached to the network but stops any activity towards the network. There is also a proposal to increase the DRX cycle, that is, the interval between the points in time when the UE needs to activate its receiver, for example in order to monitor paging messages. Both mechanisms would increase the potential response time of the UE to a device triggering, but many M2M applications are considered to be tolerant to such a delay.

The results of these studies were documented by the 3GPP RAN working groups in the technical report 3GPP TR 37.869.[6] The main findings of the study are:

- On signaling optimizations for the infrequent transfer of small data, a control plane solution where small data packets are transferred piggy-backed on RRC messages could lead to noticeable performance improvements on both the radio and the S1-MME interface only in very specific use cases: less than one data exchange per minute; not more than one data packet transferred in each direction; and packet size limited to a few hundred octets. Due to these restrictions 3GPP working group SA2 decided recently not to proceed with this solution in Release 12, but to continue working only on the following proposal.
- It is also possible to optimize the signaling by reducing the number of UE state transitions. For the decision whether to keep the UE in connected mode, and in particular for the setting of some parameters like the RRC inactivity timer and the DRX timers, the eNB could benefit from assistance information provided by the core network, for example the UE mobility behavior and the traffic type/pattern.
- It is expected that the introduction of a specific power saving state or an increase of the DRX cycle have roughly the same effect on UE power consumption if the sleep times are equivalent. Furthermore, a new power-saving state or an increase of DRX cycles to values up to 10.24 seconds (for E-UTRAN) would have very limited impact on the RAN. On the other hand DRX cycle values beyond 10.24 seconds (for E-UTRAN) would have implications, for example to system information acquisition and mobility.

As part of Release 12, 3GPP working group SA2 also studied further possible enhancements like

- Monitoring of MTC devices, that is, the detection and reporting of certain device related events: for example a change in the point of attachment of the device, or loss of the association between the device and the UICC;

- Group-based features: for example group-based device triggering, allowing to address a whole group of devices with a single broadcast message; group-based policing, that is, enforcement of a QoS policy—like a certain maximum bandwidth—for a group of devices instead of a single device; or group-based charging, but work on these topics could not be completed within the time frame of Release 12.

Summary

In the next few years, 3GPP cellular networks will have to cope with a growing demand for M2M communications. Autonomous applications and human users can be quite different with regard to the traffic (user data and signaling data) they are generating and with regard to other characteristics, for example their delay tolerance when accessing the network. Therefore, 3GPP introduced various network enhancements to keep the signaling load that can be created by a huge number of M2M devices under control. Additionally, new features like device triggering in Release 11, or small data transfer and optimizations to reduce UE power consumption currently considered for Release 12, are intended to make 3GPP mobile networks even better to use for M2M communication. Ideas for further possible enhancements have already been documented by 3GPP working group SA2, and dependent on the interest of the 3GPP members (mobile operators and vendors), they could become part of one of the future releases.

"…3GPP introduced various network enhancements to keep the signaling load that can be created by a huge number of M2M devices under control."

References

[1] 3GPP TS 22.368, Service requirements for Machine-Type Communications (MTC); Stage 1.

[2] 3GPP TS 23.401, General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access.

[3] 3GPP TR 23.888, v 1.0.0; System Improvements for Machine-Type Communications; (Release 10).

[4] 3GPP TS 23.682, Architecture enhancements to facilitate communications with packet data networks and applications.

[5] IETF RFC 4282, December 2005, The Network Access Identifier.

[6] 3GPP TR 37.869, v 12.0.0; Study on Enhancements to Machine-Type Communications (MTC) and other Mobile Data Applications; Radio Access Network (RAN) aspects; (Release 12).

Author Biographies

Robert Zaus is a systems engineer in the Intel Platform Engineering Group, has been with Intel since 2012, and has been actively involved in ETSI SMG and 3GPP standardization for more than 15 years, with a focus on non-access stratum (NAS) protocols. He holds a PhD (Dr. rer. nat.) in physics from Ludwig Maximilian University of Munich, Germany. Email: robert.zaus@intel.com

Hyung-Nam Choi is a systems engineer in the Mobile and Communications Group at Intel and has over 13 years of experience in 3GPP standardization. Since he joined Intel's standards team in 2011 he has been an active contributor to protocol-related topics for 3G (UTRA, HSPA) and 4G (LTE, LTE-Advanced) in 3GPP RAN WG2. He received his Dipl.-Ing. degree in electrical engineering from Technical University of Hamburg-Harburg, Germany. Email: hyung-nam.choi@intel.com

RADIO NETWORK EVOLUTION TOWARDS LTE-ADVANCED AND BEYOND

Contributor

Sabine Roessel
Platform Engineering Group,
Intel Corporation

Michael Faerber
Mobile Communications Group,
Intel Corporation

Bernhard Raaf
Platform Engineering Group,
Intel Corporation

Prof. Dr. Josef Hausner
Platform Engineering Group,
Intel Corporation

Living up to promises, commercial LTE deployments are delivering peak data rates of 150 Mbps and will soon exceed 300 Mbps by launching LTE Release 10 features like carrier aggregation. By 2025 a media-intense lifestyle as well as future working environments will demand a thousandfold increase of mobile communication data volume and an increase of 10 to 100 times the current typical user data rates.

The present article covers two major directions of radio network evolution. Firstly, measures to increase network capacity and network performance are discussed in detail: spectral efficiency enhancements, network densification, and extended use of spectrum. Secondly, the article introduces new radio network architectures fostering traffic offloading towards additional network layers, WLAN, or towards direct user-to-user transmission.

A simple model allows forecasting for the need to deploy LTE small cells and of traffic offloading into WLAN access points at the latest from 2020 onwards.

Introduction

High-volume data transmission, high definition video streaming, massive media sharing, sharing of augmented reality scenery, social networking, online gaming, and sports and entertainment broadcasting are services and data volumes characterizing mobile communication of the coming decade.

Following Cisco's 2013 Virtual Network Index[1], cellular network operators need to prepare for exponential traffic growth. Extrapolating Cisco's 2012 to 2017 compound annual growth rate (CAGR) of 66 percent implies an almost thousandfold increase of mobile traffic demand from 2012 through 2025. Likewise, the European project METIS[2] postulates "information and sharing of data available anywhere and anytime to anyone and anything" and predicts a thousandfold increase of mobile and wireless communication data volume and an increase of 10 to 100 times the typical user data rates.

"Living up to promises, commercial LTE Release 8 deployments are delivering peak data rates of 150 Mbps. LTE Release 10 features like carrier aggregation now being launched enable peak data rates of at least 300 Mbps."

Accordingly, expectations on cellular network performance and capacity continually rise. Living up to promises, commercial LTE Release 8 deployments are delivering peak data rates of 150 Mbps. LTE Release 10 features like carrier aggregation now being launched enable peak data rates of at least 300 Mbps. Figure 1 illustrates the expected evolution of user data rates.

Matching exponential traffic growth and hence evolving cellular networks towards LTE-Advanced (4th Generation, 4G) and beyond (often called 5G) will be vital to securing cellular network operators' market shares.

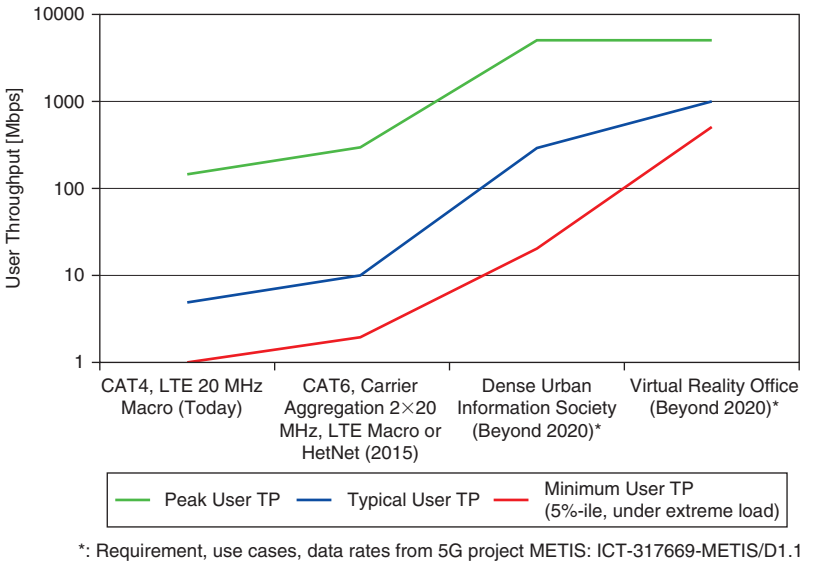


Figure 1: Evolution of user data rates
(Source: Intel Mobile Communications, 2013)

The present article covers two major directions of network evolution:

- The next section introduces measures to increase network capacity and network performance.
- The subsequent section covers network architecture evolution fostering traffic offloading.

Increasing Capacity and Performance

In order to match exponential traffic growth with network capacity, operators have to increase throughput per area in terms of bits per second per square kilometer (bps/km^2).

Network capacity or throughput per area can be instructively presented as the product of:

- Cell spectral efficiency: bits per second per hertz per cell (bps/Hz/cell),
- Density of cells: number of cells per square kilometer (cell/km^2), and the
- Amount of spectrum in hertz (Hz).

For a 10-MHz carrier, a commercial LTE Release 8 cellular network provides more than 300 Mbps/km^2 downlink (DL) capacity based on a typical cell size of 0.072 km^2 and a DL cell spectral efficiency of 2.23 bps/Hz/cell.[14]

Likewise, network performance reflects user experience and in particular user data rate in bits per second per user (bps/user). User throughput is comprised of:

- User spectral efficiency: bits per second per hertz per user (bps/Hz/user), and the
- Amount of simultaneously accessible spectrum in hertz (Hz).

"The cell-edge user data rate is the driver behind cellular network planning, while the DL peak user data rate dominates cellular network operators' marketing."

The minimum, or fifth percentile, or cell-edge user data rate is the driver behind cellular network planning. The DL peak user data rate dominates cellular network operators' marketing. While a commercial LTE Release 8 cellular network is delivering a DL peak data rate of 150 Mbps, LTE Release 10 features like carrier aggregation are about to enable DL peak data rates of at least 300 Mbps. Theoretically, the first release of LTE-Advanced, LTE Release 10, enables a DL peak data rate of 3 Gbps when aggregating five 20 MHz carriers and using 8x8 single-user multiple-input multiple-output (SU-MIMO).

Simply put, cellular network operators increase network capacity and network performance by three elements that are discussed in the subsequent subsections:

- Enhancing spectral efficiency,
- Deploying more cells, and
- Increasing the amount of (simultaneously accessible) spectrum.

Higher Spectral Efficiency

In the following sections, we will study techniques to enhance cell and user spectral efficiency for today's dominant radio access technology LTE/LTE-Advanced. We will focus on the downlink direction of an Orthogonal Frequency Division Multiplexing (OFDM) system and assume for each frequency-time resource element: N cells (or base stations or transmission points) with $N_T(n)$ transmit antennas at the nth base station, and M user devices or UEs with $N_R(m)$ receive antennas in the mth UE (see Figure 2).

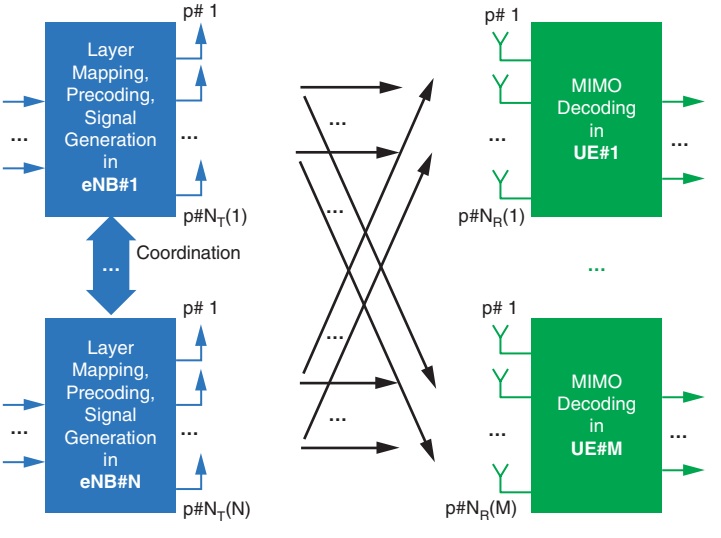


Figure 2: DL OFDM system comprising N—potentially coordinating—eNBs with $N_T(n)$ transmit antennas at the nth eNB and M UEs with $N_R(m)$ receive antennas at the mth UE (Source: Intel Mobile Communications, 2013)

Single Link Spectral Efficiency

When setting in Figure 2 the number of cells (or transmission points), UEs, transmit and receive antennas to 1 each, we obtain one single-input single-output (SISO) link for which the LTE physical layer structure with a normal Cyclic Prefix (CP) provides us 12 frequencies per Resource Block (RB) × 14 OFDM symbols per ms × 100 RBs in 20 MHz × 1,000 ms = 16,800,000 frequency-time Resource Elements (REs) in 20 MHz spectrum and 1 second.[3] If all resource elements were exploited for data communication, at the highest currently supported modulation order of 6, that is, using the 64QAM modulation scheme, and at an ideal code rate of 1, the downlink peak spectral efficiency of a single LTE Release 8 link would reach 5 bps/Hz.

"…3GPP is studying in LTE Release 12 whether or not to introduce 256QAM for small cell propagation environments."

As a method to increase downlink spectral efficiency of a single SISO link, 3GPP is studying in LTE Release 12 whether or not to introduce 256QAM for small cell propagation environments. Providing 256QAM presents an implementation challenge to both the base station transmitter as well as the UE receiver, because both transmitter Error Vector Magnitude (EVM) and receiver EVM respectively must be reduced compared to today's infrastructure and device implementations. Increasing the modulation order from 6 to 8, ideally a gain of "only" 33 percent, may not be sufficient to keep pace with future throughput demands. LTE Release 12 studies claim 10- to 30-percent spectral efficiency gain when assuming 4 percent transmitter EVM and neglecting receiver impairments.[4] Nevertheless, it can be expected that with 5G network topologies involving short-distance Massive MIMO and/or device-to-device communication, 256QAM will make an entrance.

No radio access technology can exploit 100 percent of the available frequency-time resources exclusively for data communication. Part of the resources must be spent for control and pilots (reference symbols):

- Control overhead: The LTE physical layer structure has to reserve between 7 and 21 percent of the resources for downlink control channels.
- Reference symbol overhead: Even in the case of a single SISO link, LTE Release 8 spends almost 8 percent of the downlink resources on Cell-specific Reference Symbols (CRS).

Work on the New Carrier Type (NCT) targeted at eliminating 80 percent of the CRS, marginalizing the pilot overhead, but was stopped both in LTE Release 11 and LTE Release 12 due to inconsistent performance results. Reducing CRS by 80 percent is expected to degrade channel estimation and consequently system performance. Despite the discouraging lack of progress in 3GPP, 5G research will again target an optimal tradeoff of pilots versus channel estimation and demodulation performance. Of course, this may not be achievable without fundamental changes to the physical layer structure, which then will deviate from the current LTE/LTE-Advanced physical structure.

Furthermore, transmission of Physical DL Shared Channel (PDSCH) and/or Enhanced Physical DL Control Channel (EPDCCH) in the first OFDM

symbol of a subframe or Transmission Time Interval (TTI) is expected to provide up to 6-percent cell spectral efficiency gain.[4] Wider use of the EPDCCH is expected to enable a control overhead reduction scaling with the number of scheduled users.

The LTE/ LTE-Advanced air interface operating point is at an initial Block Error Rate (BLER) of approximately 10 percent and reaches the required Block Error Rates (BLER) of typically 10^{-2} (for a VoIP call) to 10^{-6} (for Video streaming or IMS signaling) by employing a hybrid ARQ (HARQ) scheme. As an optimization, a 5G technology may accelerate the HARQ process, both by shortening the TTI and making the HARQ process more responsive.[5]

"As an optimization, a 5G technology may accelerate the HARQ process, both by shortening the TTI and making the HARQ process more responsive."

"…5G studies involve methods of dynamically reducing or eliminated the Cyclic Prefix."

With the current LTE/LTE-Advanced waveform, 7 percent of the resources are consumed by the guard period or Cyclic Prefix. Accordingly, 5G studies involve methods of dynamically reducing[5] or eliminating the Cyclic Prefix. These methods include modifying the OFDM waveform towards non-continuous OFDM[6] or Filter-Bank Multiple Carriers (FBMC)[7] including its variants like Generalized Frequency Division Multiplexing (GFMA).[7] For example, FBMC uses longer pulses of two to four symbol periods to eliminate the Cyclic Prefix. Besides the higher spectral efficiency, a key advantage of FBMC consists in its higher spectral containment: power leakage into adjacent carriers is significantly lower than in OFDM. The resulting spectrum shaping capability can be exploited both in coexistence situations occurring in small cell deployments and for Authorized/Licensed Shared Access (ASA/LSA) schemes.

MIMO Channel Spectral Efficiency

Fortunately, multiple antennas, both at the base station and the UE, add spatial multiplexing as a degree of freedom that transforms into spectral efficiency.

Figure 2 captures all techniques related to spectral efficiency enhancements from exploiting spatial multiplexing, that is, the MIMO channel. As an example, using two transmit and two receive antennas (still limiting the system to a single cell and a single UE) we obtain the 2×2 MIMO channel that ideally can be decoupled into two orthogonal sub-channels, both of which carry data independently and without mutual interference. This allows for doubling the peak data rate[8][9] compared to the SISO channel.

A single cell, single UE system could be run optimally without interference between the spatial streams if perfect channel knowledge were available such that the transmitter could perfectly precode the signal and run both streams at the respective available capacity. Channel knowledge, however, must be obtained through UE feedback. In order to minimize the amount of UE feedback, precoding options are limited to a finite and small set of codebook entries. Providing different precoding options blindly, for example by rotating through the codebook entries, is called Open-Loop MIMO (OL-MIMO). Selecting a codebook entry based on UE's Rank Indicator (RI) and Precoding Matrix Index (PMI) feedback is called Closed-Loop MIMO (CL-MIMO).

In order to increase the MIMO channel spectral efficiency LTE-Advanced and beyond networks may target at:

- Higher Order Single-User MIMO ($N_T(1) > 1$ and $N_R(1) > 1$, min ($N_T(1)$, $N_R(1)$) = number of MIMO streams):
  The LTE-Advanced (LTE Release 10) standard supports up to eight streams in the downlink direction. Eight streams run at the maximum modulation scheme, 64QAM, in an aggregated 100-MHz channel and considering the overhead of transmission mode TM9, result in a downlink peak throughput of 3 Gbps. Only very high signal-to-interference-plus-noise ratios (SINR) like in indoor hotspots allow using three to four streams by 10 to 20 percent of homogeneously distributed UEs.[10]
- High transmission rank beamforming ($N_T(1) > N_R(1) > 1$, number of MIMO streams = 1): The powers of identical data streams transmitted from different antennas are combined at the receiver's antenna therefore improving the MIMO channel's SINR.
- High transmission rank hybrid MIMO/beamforming schemes ($N_T(1) > N_R(1) > 1$, number of MIMO streams > 1): As an example, TM8-based dual-layer (8×2) beamforming has been specified in LTE Release 9.

When allowing for multiple users in a single cell, a related ingredient to increasing spectral efficiency is:

- Multi-User-MIMO ($N_T(1) \geq M_{mu}$ and $N_R(1) + \ldots + N_R(M_{mu}) \geq M_{mu}$, $\min(N_T(1), N_R(1) + \ldots + N_R(M_{mu})) \geq$ number of MIMO streams $\geq M_{mu}$), which presents an additional degree of freedom by providing MIMO streams on the same resource to M_{mu} different UEs participating in Multi-User-MIMO.

Active Antenna Systems (AAS) where the RF components are integrated into the antenna enable significant capacity gains per sector area. AAS antenna panels have rather vast dimensions, in particular an extended surface that is difficult to mount on a typical macro network's antenna mast because of wind load. AAS will likely be mounted below roof-top at walls of tall buildings and will be deployed at higher carrier frequencies where antenna structures are smaller due to smaller wavelengths.

Accordingly, major directions for evolution of MIMO techniques encompass:

- Elevation Beamforming (3D-BF), studied for LTE Release 12, which adds UE-specific vertical beam-steering to existing azimuth-only closed-loop SU-/MU-MIMO.
- Full-Dimension MIMO (FD-MIMO), also part of the LTE Release 12 studies, which exploits both the azimuth and the elevation dimensions of the multipath channel on a UE-specific basis.
- Massive MIMO ($N_T(1) >> N_R(1) + \cdots + N_R(M_{Served})$), which targets propagation environments that allow for very narrow beams. The narrow beams constitute small cells allowing for two or more spatial streams on

all frequency-time resources to all UEs in reach (M_{Served}).[11] The very small wavelengths in the 60 GHz (mmWave) spectrum allow for a very large amount of antenna elements both in horizontal and elevation dimensions, and hence for creating extremely narrow beams. This is called *pencil beamforming* and is foreseen as mmWave radio access for indoor and dense urban clutter environments.

Figure 3 summarizes the evolution of MIMO techniques in the 3GPP specifications from LTE Release 8 to LTE Release 12 and beyond.

Acronym	Transmission Scheme	Reference Symbols	Precoding Feedback	LTE	Status
2×1, 4×1 TX Div	TM2: Transmit Diversity	CRS	n/a	Rel-8	
2×2, 4×2 SU-MIMO	TM3: Open Loop Spatial Multiplexing (Rank-2 OL-MIMO)	CRS	n/a	Rel-8	
	TM4: Closed Loop Spatial Multiplexing (Rank-2 CL-MIMO)	CRS	RI and PMI	Rel-8	
4×4 SU-MIMO	TM3: Rank-4 OL-MIMO	CRS	n/a	Rel-8	Substituted by TM9 SU-MIMO
	TM4: Rank-4 CL-MIMO	CRS	RI and PMI	Rel-8	Substituted by TM9 SU-MIMO
2×1, 4×1, 4×2 MU-MIMO	TM5: CL MU-MIMO	CRS	PMI	Rel-8	Substituted by TM9 MU-MIMO
	TM6: Rank-1 CL-MIMO	CRS	PMI	Rel-8	
8×1 BF	TM7: Single-antenna port Beamforming	DM-RS	n/a	Rel-8	Substituted by TM8 and TM9 BF
8×2 BF	TM8: Dual-layer Beamforming	DM-RS	PMI, RI	Rel-9	TDD networks
4×4, 8×4, 8×8 SU-/MU-MIMO	TM9: Up-to Rank-8 CL-MIMO	DM-RS	PMI, RI based on CSI-RS	Rel-10	Substitutes Rel-8 4×4 SU-MIMO, Rel-8, Rel-9 BF
DL CoMP	TM10: DL CoMP	DM-RS	PMI, RI based on CSI-RS, CSI-IM	Rel-11	
4×4, 8×4, 8×8 SU-/MU-MIMO	Up-to Rank-8 CL-MIMO	DM-RS	New codebook and feedback	Rel-12	
3D-BF, Elevation Beamforming	UE-specific vertical beamsteering added on top of azimuth-only CL SU-/MU-MIMO		New codebook and feedback	Rel-12/13	Active Antenna System
FD-MIMO	Simultaneous UE-specific exploitation of azimuth & elevation dimensions		New codebook and feedback	Rel-12/13	Active Antenna System
N×2, N×4 CL SU-/MU-MIMO	Massive MIMO with N >= 16, N>> 16: Pencil Beamforming		New codebook and feedback	Open	Distributed Active Antenna System

Figure 3: Evolution of MIMO Techniques
(Source: Intel Mobile Communications, 2013)

White rows mark schemes that are in use. Despite the fact that the transition from two to four base station antennas is key to improve sector spectral efficiency, cellular network operators still need to complete this transition. Also, four UE (receive) antennas can hardly be expected any time soon for typical smartphones. Smartphone form factors are determined by the conflicting requirements of: being handheld, preferably in a single hand, and supporting frequencies down to 700 MHz on at least two cellular antennas.

Lines representing techniques that became obsolete due to newer, better schemes are marked grey. Green-colored lines mark techniques yet to

be deployed, while lines marked blue represent new techniques under development or yet to be developed for Release 12 and beyond.

Cell spectral efficiency and fifth-percentile user or cell-edge spectral efficiency gains vary with antenna correlation, radio propagation environment and LTE transmission mode, that is, reference symbol overhead. Figure 4 summarizes relative gains over LTE Release 8 2×2 SU-MIMO from 3GPP Release 10[12] and Release 11[13] technical reports. As shown, 4×4 MU-MIMO based on TM9 can double cell spectral efficiency compared to LTE Release 8 2×2 SU-MIMO and increase fifth-percentile user spectral efficiency by a factor of 2.5.

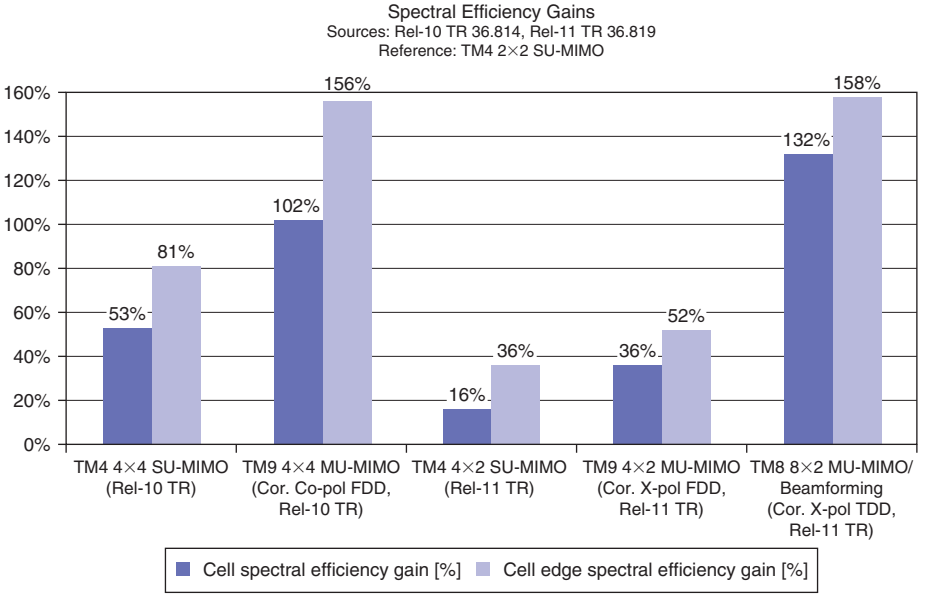


Figure 4: MIMO gains based on data in 3GPP TR 36.814[12] and 3GPP TR 36.819[13] (Source: Intel Mobile Communications, 2013)

Interference Mitigation

In a system with multiple cells (or eNBs or non-collocated transmission points) as well as with multiple UEs (full configuration in Figure 2) frequency-time resources are reused in neighbor cells and hence inter-cell interference is introduced.

Interference profiles, that is, spatially white versus colored interference, number, power, and modulation of dominant interferers, data channels or reference symbols as interference source or victim, and so on, to a great extent depend on network deployment.

While interference mitigation in LTE Release 8 is based on single cell radio resource management techniques like interference-aware scheduling, LTE Release 10 and LTE Release 11 focus on inter-cell interference coordination (ICIC) for Heterogeneous Networks (HetNet) where several low power pico eNBs transmit and receive on the same frequency as the high power macro eNB.

"Enhanced Inter-Cell Interference Coordination" (eICIC) as well as "Further enhanced ICIC" (FeICIC) have been the related LTE Release 10 and Release 11 work areas. Interference of a macro cell to UEs served by a pico cell deployed on the same frequency is reduced by the introduction of Almost Blank Subframes (ABS) in the macro cell. Furthermore, the UE is assisted by the network to improve reception such that the UE is ultimately able to cancel the interference of macro cells' Cell-specific Reference Symbols (CRS-IC) and common channels like the Physical Broadcast Channel (PBCH-IC).

Further LTE Release 11 research has focused on multi-cell radio resource management, that is, Coordinated Multi-Point (CoMP) techniques for ideal inter-eNB backhaul both in homogeneous macro networks as well as in HetNets (then called CoMP Scenario 3). DL CoMP means to jointly precode and/or coordinate a cluster of non-collocated transmission points in such a way that the wanted signal's power is maximized and the remaining interference minimized. In particular Joint Transmission CoMP therefore requires instantaneous and almost perfect channel knowledge as well as instantaneous availability of user data at all transmission points in a CoMP cluster.

"DL CoMP means to jointly precode and/or coordinate a cluster of non-collocated transmission points in such a way that the wanted signal's power is maximized and the remaining interference minimized."

For illustrating a simple DL CoMP system, we configure the system in Figure 2 with two non-collocated transmission points (or two eNBs) as a small network cluster serving two UEs. The transmission points' mutual interference at the receiving UEs can now be mitigated by eliminating inter-transmission-point interference thanks to coordination and precoding, and, in case of Joint Transmission CoMP, by allowing both transmission points to contribute to the wanted signal power.

The following classification of DL CoMP schemes can be made (see Figure 5):

- In case of (DL) Coordinated Scheduling/Beamforming (CS/CB) the UE is served by a single cell, eNB, or transmission point while the network optimizes the frequency-time resource allocation and MIMO/BF configuration of the UE, taking interfering neighbor cells into account. This optimization may also include Dynamic Point Blanking (DPB): dynamically leaving the interfering cell's frequency-time resources unscheduled.
- Dynamic Point Selection (DPS) means that the UE is served dynamically, that is, from Transmission Time Interval (TTI) to TTI, from different cells, eNBs, or transmission points.
- Finally, Joint Transmission CoMP (JT CoMP) means that the UE is simultaneously served by at least two cells, eNBs, or transmission points on the very same frequency-time resources. JT CoMP is not (explicitly) supported by LTE Release 11 because feedback for inter-transmission point phase alignment is not available.

LTE Release 11 DL CoMP performance gains (see Figure 6) are rather modest. While there is little-to-negligible cell capacity gain from LTE Release 11 DL CoMP, up to 40-percent cell-edge throughput gain can be achieved.

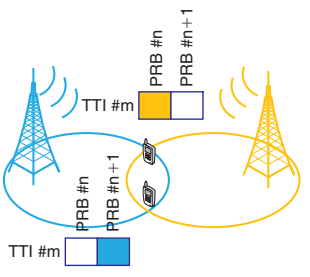
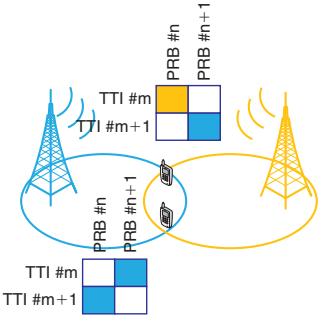
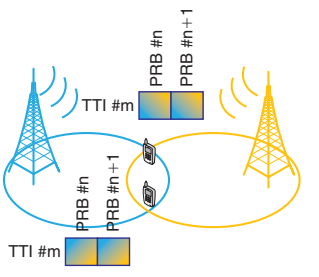
Coordinated Scheduling/ Coordinated Beamforming (CS/CB) with or without **Dynamic Point Blanking (DPB)**	**Dynamic Point Selection (DPS) with** or without **Dynamic Point Blanking (DPB)**	Single-user or **Multi-user Joint Transmission CoMP**
Availability of data: at single transmission point (TP) only	Availability of data: at all TP of CoMP Set	Availability of data: at all TP of CoMP Set
Transmission only from serving cell	Transmission is coordinated but from single TP at a time	Simultaneous transmission from multiple TPs in CoMP Set
		

Figure 5: DL CoMP schemes
(Source: Intel Mobile Communications, 2013)

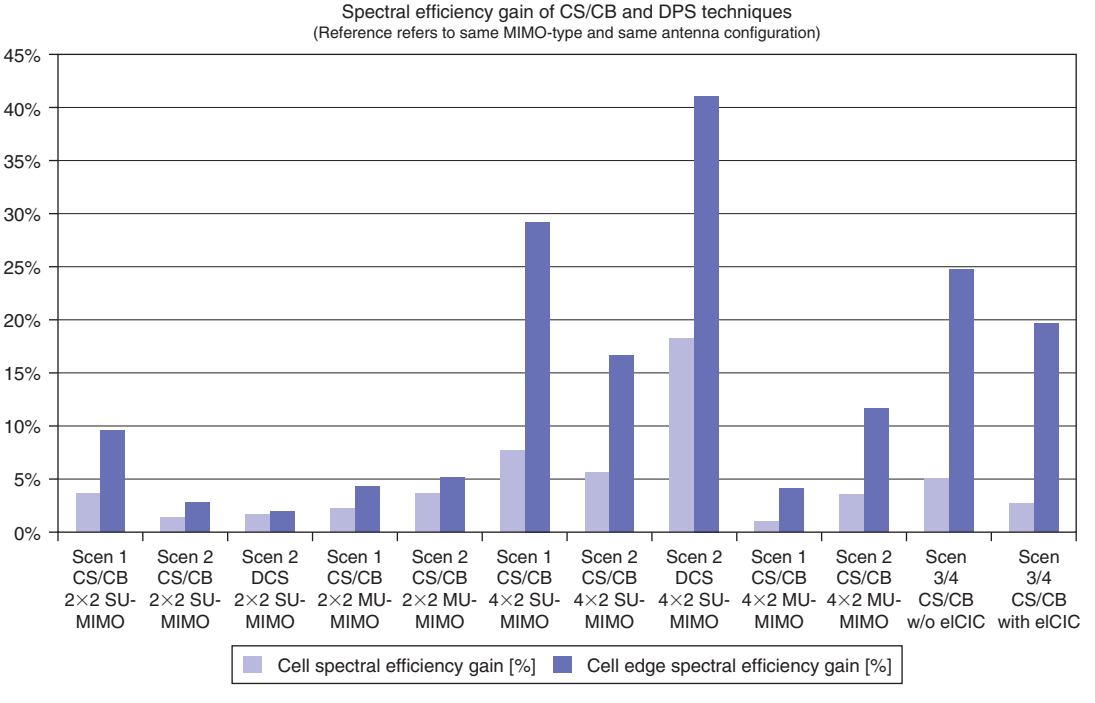


Figure 6: Throughput gains of LTE Release 11 DL CoMP techniques, based on data in TR 36.819 v.11.2.0[13]
(Source: Intel Mobile Communications, 2013)

Accordingly, the following network evolution can be foreseen:

- Getting beyond a 10-percent cell-edge performance gain is possible when upgrading the network to four antennas per sector in homogeneous networks or with two antenna base stations in CoMP Scenario 3.
- Operators will first exploit Coordinated Scheduling/ Coordinated Beamforming in macro network deployments.
- As soon as heterogeneous networks have come into wider use, Dynamic Point Selection will be rolled out independently for DL and UL service.

Multi-cell interference mitigation techniques will be further developed in LTE Release 12 and beyond and will also be extended to new techniques required to support 5G (see Figure 7).

"In general network (downlink) interference mitigation means: network coordination and network assistance."

In general network (downlink) interference mitigation means:

- Network coordination: the network shapes, reduces, or avoids interference by multi-cell or multi-transmission-point coordination, scheduling and/or beamforming, transmission point selection and/or blanking, and precoding, and/or
- Network assistance: the network signals information relevant to the UE for interference cancellation or suppression.

	Rel-8	Rel-10	Rel-11	Rel-12	Rel-13+	5G
Network	Interference-Aware • Scheduling • Power Loading • Link Adaptation • X2-based Inter-Cell Interference Coordination (ICIC)	• Enhanced ICIC w/ Almost Blank Subframes (ABS) and 6 dB Cell Range Extension (CRE)	• (Further) enhanced ICIC based on ABS and 9 dB CRE • DL CoMP Coordinated Scheduling/ Beamforming (CS/CB) w/ Dynamic Point Blanking (DPB) • DL CoMP Dynamic Point Selection (DPS) with DPB	• Network Assistance for R-ML/SL-IC Receivers • DL CoMP for non-ideal backhaul (NIB)	• Network Assistance for IC/IS Receivers (E/PDCCH) • Coordination based on New Carrier Type • Joint Transmission DL CoMP • CS/CB for massive MIMO/ extreme Beamforming	Network Assistance/ Coordination for new receiver types supporting e.g.: • Non-orthogonal Multiple Access (NOMA) • Filter-Bank Multi-Carrier (FBMC) • Full Duplex Radio (FDR)
UE	• MMSE-MRC Receiver		• CRS IC and PBCH/PSS/SSS IC Receiver • MMSE-IRC Receiver	• R-ML/SL-IC Receiver for CRS IC and PDSCH IC Receiver	• R-ML/SL-IC Receiver for CRS IC and E/PDCCH IC Receiver • Channel estimation on NCT CRS	• Receiver types supporting NOMA, FBMC, and FDR

Figure 7: Evolution of downlink interference mitigation techniques from LTE Release 8 to beyond LTE Release 12 and 5G (Source: Intel Mobile Communications, 2013)

Of course, the UE has an equally important role in interference mitigation. UE interference mitigation is a combination of the following aspects:

- Receiver Type: As part of the LTE Release 12 studies on "Network Assistance for Interference Cancellation and Suppression in LTE Receivers,"[15] 3GPP has been investigating the following receiver types: Minimum Mean Square Error Interference Rejection Combining (MMSE-IRC), reduced complexity Maximum-Likelihood (R-ML), Symbol-Level Interference Cancelling (SL-IC), and Codeword-level Interference Cancelling (CW-IC) receiver.
- Interference Cancellation (IC) capabilities ultimately comprise: cancelling interference from colliding or non-colliding CRS, common channels, data (PDSCH), other reference symbols like CSI-RS and DM-RS, or control (PDCCH).
- Blind Detection capabilities: if network assistance information on interferer parameters is not or only partially available these must be inferred by the interfered UE without (full) foreknowledge.

"UE interference mitigation is a combination of the following aspects: receiver type, interference cancellation capabilities, and blind detection capabilities."

Network assistance signaling for interference cancellation may be quite expensive. For example, the required knowledge for PDSCH interference cancellation comprises several parameters ranging from interferer's Cell ID and Control Format Indicator (CFI), to interferer's Modulation Order, Precoding Matrix Index (PMI), number of MIMO layers, and Transmission Mode (TM). Also, a sophisticated eNB scheduler may create dynamic allocation patterns changing rapidly in frequency and in time both in the serving cell as well as in the neighbor cells. Unless network coordination is involved, inter-cell interference will therefore vary rapidly, and thus network assistance signaling will have to vary rapidly as well, that is, ultimately from Resource Block (RB) to RB and from Transmission Time Interval (TTI) to TTI.

LTE Release 12 studies are comparing gains from network-assisted interference cancellation with gains from interference cancellation based on blind detection. In order to avoid network assistance signaling overhead outweighing the gains from interference cancellation, the finalized concept will leave some parameters to blind detection and will support some parameters to be signaled or coordinated by the network.

Good candidates for performance tests related to the network-assisted interference cancellation are the reduced complexity Maximum-Likeli-hood (R-ML) and the Symbol-Level Interference Cancelling (SL-IC) receivers.

The evolution of LTE/LTE-Advanced receivers towards Symbol-Level and even Codeword-Level Interference Cancellation (CW-IC) will also enable more spectrally efficient multiple access schemes. For example, an OFDM-based Non-Orthogonal Multiple Access (NOMA) scheme has been proposed as a 5G multiple access candidate[16] that targets power division multiple access. This scheme relies on the LTE/LTE-Advanced receiver's capability to perform successive interference cancellation efficiently thanks to the largely differing powers of the multiplexed data streams.

"…An OFDM-based Non-Orthogonal Multiple Access (NOMA) scheme has been proposed as a 5G multiple access candidate that targets power division multiple access and that relies on the UE receiver's capability to perform successive interference cancellation efficiently thanks to the largely differing powers of the multiplexed data streams."

More Cells

The second dimension in enhancing network capacity and network performance refers to network densification, or simply put, to providing more cells. In the following subsections, several different approaches to network densification are described.

"One method of network densification consists of sector splitting in homogeneous macro networks."

Sectorization in Macro Networks

One method of network densification consists of sector splitting in homogeneous macro networks. Sector splitting can be vertical, resulting in six sectors placed in two concentric circles around the macro base station site, or horizontal, resulting in six equivalent macro sectors just like six equal slices of a pie. Sector splitting can be implemented statically by replicating antenna equipment with fixed but narrow antenna beams, as well as dynamically by exploiting Active Antenna Systems (AAS) for vertical, horizontal, or 3D beamforming. Integrating the base station's RF elements with the antenna implies the need to (re-)define the test point for receive and transmit performance of AAS-based base stations; this is currently being studied in 3GPP.[17] Accordingly, AAS are yet to be widely deployed.

Vertical sectorization or vertical beamforming is known to achieve 65-percent capacity gain compared to a standard three-sector base station site.[18]

Site Densification in Homogeneous Networks

The second method of network densification consists of increasing the number of base station sites in an otherwise homogeneous macro network. Due to densification of base station sites in areas of high traffic demand and deployment of higher frequency bands with higher path loss, homogeneous macro network inter-site distances (ISD) have dropped from 500 meters in the first 3G deployments down to 200 meters or less in today's urban LTE deployments.

"Another method of network densification consists in augmenting a macro base station by several pico base stations or non-collocated Remote Radio Heads in areas of extreme capacity demand…"

Small Cell Deployment

The third method of network densification consists in augmenting a macro base station sector by several pico base stations or non-collocated Remote Radio Heads (RRHs) in areas of extreme capacity demand (hot spots). The resulting small cell layer is an integral part of the macro network, as pico base stations and non-collocated RRHs are subject to radio planning and operational integration like neighbor relation setup for mobility. As mobile service in the small cell layer is controlled or assisted by the macro network, this type of deployment is called coordinated or macro-assisted small cell deployment. In contrast, uncoordinated small cell deployment refers to the deployment of residential home eNodeBs or femto cells that typically do not offer seamless mobility into and from the macro network and very often can only be accessed by members of *Closed Subscriber Groups* (*CSGs*).

Since LTE Release 10, 3GPP has been studying coordinated or macro-assisted small cell deployment scenarios:

- Heterogeneous network (HetNet) or co-channel small cell deployment (see Figure 8): Several, typically four, pico base stations per macro base station sector are deployed in the same frequency channel as the umbrella macro cell. This deployment case is called a heterogeneous network (HetNet). Of course, such deployments cause extra inter-cell interference counteracting the capacity (per area) gain targeted by network densification. Accordingly, LTE Release 10 and LTE Release 11 studied different interference mitigation solutions for HetNets including Release 11 FeICIC and Release 11 DL CoMP techniques for CoMP Scenario 3.
- Dedicated frequency layer small cell deployment (see Figure 9): The small cells cover a dedicated frequency layer at rather high frequencies like 2.6 GHz or 3.5 GHz while the typical macro cell's frequency bands are at 2 GHz or below. The dedicated frequency layer small cell deployment is a key scenario of the LTE Release 12 studies on higher layer aspects of small cell enhancements (see TR 36.842[19]) as it targets at efficient traffic offloading from the macro into the small cell layer.

  A key challenge of small cell deployments is mobility robustness as due to frequent handovers the amount of radio link failures (RLFs) significantly increases. For optimizing mobility robustness, LTE Release 12 is developing UE-based solutions following the studies in TR 36.839[20], as well as dual connectivity solutions described in TR 36.842.[19]

  A particular challenge of the dedicated frequency layer small cell deployment is the need to jointly operate Time Division Duplex (TDD) and Frequency Division Duplex (FDD) spectrum as the majority of higher frequency bands are assigned a TDD frequency plan while the lower frequency bands typically adhere to the more robust FDD frequency plan. To enable macro-assisted small cell operation in a mixed duplex scenario, 3GPP is working on TDD-FDD carrier aggregation in LTE Release 12.[21]

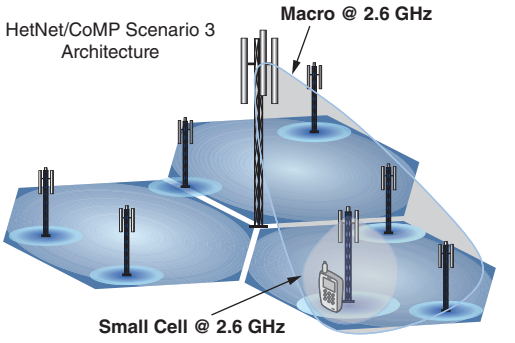


Figure 8: Coordinated (co-channel) small cell deployment also known as HetNet or CoMP Scenario 3
(Source: Intel Mobile Communications, 2013)

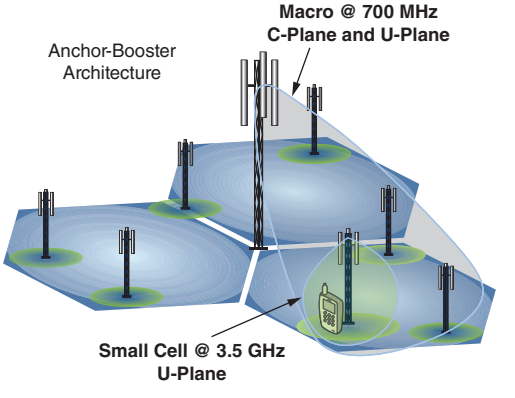


Figure 9: Dedicated frequency layer small cell deployment also known as anchor-booster or dual connectivity architecture
(Source: Intel Mobile Communications, 2013)

More Spectrum

The third dimension in enhancing network capacity refers to providing access to additional spectrum.

A major step consists of increasing the amount of accessible spectrum in cellular networks by supporting frequency bands below 700 MHz and at 3.5 GHz and above. Coordinated or macro-assisted small cell deployments for a mixed duplex scenario as well as WLAN-3GPP Radio Interoperation are means to enhance network capacity by exploiting additional spectrum.

As a next step, cellular network operators can exploit spectrum more efficiently by contracting to Authorized or Licensed Shared Access (ASA/LSA) schemes and by enabling cellular equipment to support such licensing schemes with Cognitive Radio (CR) techniques.[22]

Finally, as a strong trend in 5G research, very large spectrum blocks will be made accessible by creating a new radio access technology for the mmWave spectrum (60 GHz). While both attenuation and obstacles are severe issues for mm waves, they may be overcome by using antenna arrays with several tens (or even hundreds) of elements from multiple sites. One option is to place the antennas on top of a chip that implements an array of RF transceivers. This way beams can be steered around obstacles for example by pointing at suitable reflectors to achieve reliable communications with extremely high bandwidth.

The third dimension in improving network performance, in particular user experience and potential peak data rates, also consists of making more amount of spectrum simultaneously accessible.

"…A vast amount of carrier aggregation 3GPP RAN4 work items has been established to provide cellular network operators all over the world with guaranteed UE performance in carrier aggregation scenarios."

Starting with LTE Release 10, 3GPP has been continuously working on carrier aggregation techniques. Since then, a vast amount of carrier aggregation 3GPP RAN4 work items has been established to provide cellular network operators all over the world with guaranteed UE performance in carrier aggregation scenarios.

As of September 2013, 3GPP RAN4 lists a total of 44 frequency bands (27 FDD, 1 DL-only, and 12 TDD) as well as 103 carrier aggregation completed or ongoing work items (see Figure 10).

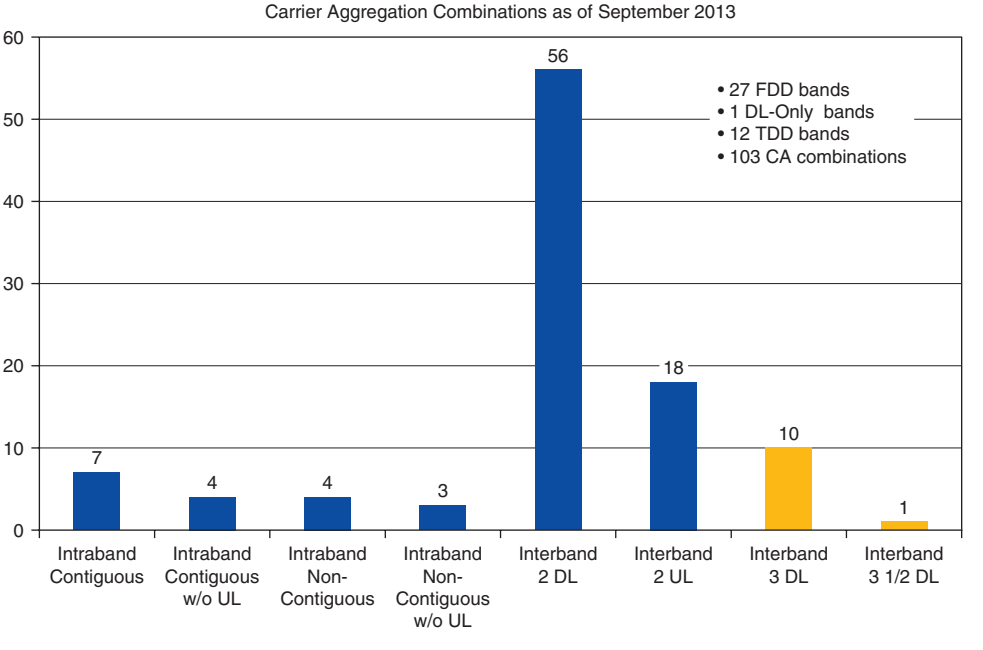


Figure 10: LTE carrier aggregation combinations as of September 2013 (Source: Intel Mobile Communications, 2013)

While LTE Release 10 has formally specified the aggregation of at maximum five 20-MHz carriers, through early 2013 only work items for aggregating two carriers had been proposed.

Recent trends involve the aggregation of three 20-MHz carriers as well as the combination of aggregating two small intra-band contiguous carriers with two further carriers on other frequency bands, one example being the combination of two Band 12 channels with a Band 2 or Band 4 carrier and a Band 30 carrier.

Complexity, cost, and size of the RF and antenna frontend components of a smartphone device will not allow for providing too many nontrivial carrier aggregation combinations. A major trend will therefore be wide-band transceivers that are capable of processing contiguous spectrum blocks of 40 MHz and more.

Finally, research on 5G networks also involves research on full duplex radio[23], which implies the application of self-interference cancellation techniques when allowing for simultaneous transmission and reception on the very same frequency-time resource.

Evolving Network Architecture

The following sections introduce new radio network architectures fostering traffic offloading towards additional network layers, WLAN, or towards direct user-to-user transmission.

Anchor-Booster or Dual Connectivity Architecture

The anchor-booster or dual connectivity architecture is the architectural concept underlying the small cell deployment scenarios.

"The anchor-booster or dual connectivity architecture is the architectural concept underlying the small cell deployment scenarios."

A macro cell eNB acts as anchor eNB, and the eNBs serving small cells are the so-called booster eNBs. Anchor-booster operation or dual connectivity means that the UE is simultaneously served by (two) different eNBs. The main tasks of the anchor eNB are mobility management and running the control plane. 3GPP is studying the following aspects of the anchor-booster architecture:[19]

- Mobility robustness
- Signaling load to the core network (CN)
- User throughput

As an example for mobility robustness improvement, the anchor-booster architecture can support radio resource control (RRC) diversity. Handover-related RRC signaling is additionally transmitted from or to a potential target cell such that a significant portion of radio link failures (RLFs) can be prevented.

Another example refers to separating the control plane and user plane: While receiving from or transmitting data to the booster eNB, the UE maintains the control-plane connection to the anchor eNB. In a dedicated frequency small cell layer with good macro coverage, RRC signaling is then safely received from the macro cell while RLFs, due to potential lack of coverage in the small cell layer, can be avoided.

Separating control plane and user plane also presents an anchor-booster architecture advantage on user throughput: a UE in the coverage of a booster cell can receive user-plane data both from the booster cell as well as from the anchor

cell. The small cell may offer channel conditions that allow for using aggressive modulation and coding schemes (MCS) and hence boosting user throughput. However, as the (backhaul) connection between booster and anchor eNBs typically is not ideal, user-plane data provided via the booster eNB should preferably be best effort. Delay-sensitive services should be provided through the anchor eNB.

Anchor-booster or dual connectivity architectures may vary in their characteristic features and differ in:

- Level of coupling between anchor and booster nodes,
- Radio access technologies, and
- Involvement of logical network elements.

As the anchor node is the manager of a sub-network and the mobility within this sub-network, the anchor node also performs radio admission and radio bearer control for traffic steering and flow control. This enables most efficient use of radio resources in booster cell enhanced macro cells, with the purpose of boosting performance: the tighter the coordination between anchor and booster nodes, the higher the gains in capacity, user throughput, and mobility robustness. New air interfaces and LTE enhancements may require new interfaces, or new messages on existing interfaces, to allow for optimized anchor-booster operation.

"Booster cells may provide access via different radio access technologies (RATs) or combine cellular and non-cellular access including LTE, WLAN, or mmWave."

Booster cells may provide access via different radio access technologies (RATs) or combine cellular and non-cellular access including LTE, WLAN, or mmWave.

No matter which radio access technology is provided by the booster cell, anchor-booster concepts will be key to cope with the exponential traffic growth that operators face in the future and hence will be a key ingredient of 5G networks.

D2D and Ad-Hoc Networks

Device-to-device communication (D2D) is the capability to enable direct communication between two devices with or without network involvement. D2D is considered as one building block of future wireless systems and as an element contributing to what is commonly called 5G.

Bluetooth* and Wi-Fi Direct* enable user-driven D2D communication. This means users proactively establish a pairing of devices for temporary data exchange. D2D can happen in different deployment scenarios:

- Standalone and self-organized
- Network-assisted and network-controlled
- Integrated in a heterogeneous network

"The present LTE Release 12 studies on "Device-to-Device Proximity Services" cover both commercial and public safety aspects…"

3GPP is working in LTE Release 12 to build the fundamentals for a D2D technology that minimizes user interaction and introduces network assistance and integration.

The present LTE Release 12 studies on "Device-to-Device Proximity Services"[25] cover both commercial and public safety aspects (see Figure 11). While commercial use is limited to the in-network-coverage case, it is a

Solutions	Use cases		In-Network		Out-of-Network
			Commercial	Public Safety	Public Safety
Discovery	LTE-Direct	Push	x	x	x[1]
		Pull	x	x	x[2]
	EPC-Level LCS-based (No RAN impact)		x		
Communication	LTE-Direct		x[1]	x	x
	Locally-Routed		x[1]	x	
	EPC-Assisted WLAN (No RAN impact)		x		

Note 1: Based on the prioritization as of RAN/SA#61 plenaries:
- D2D communication: Public Safety broadcast/groupcast in-/out-of-coverage, UE-to-network relay
- D2D discovery: LTE-direct considers in-network only.

Note 2: Pull discovery solution can be supported by LTE-Direct communication and requires no special capability for discovery.

Figure 11: 3GPP RAN focus in LTE Release 12 study on "D2D Proximity Services"[25] (Source: Intel Mobile Communications, 2013)

specific matter of public safety to also enable wireless communication in the out-of-network-coverage case for example by D2D-based ad-hoc networks. Analysis of the differing needs has also resulted in the focus on LTE direct D2D discovery for the in-network-coverage cases as well as on LTE direct broadcast/groupcast communication for the Public Safety use case both in-network-coverage and out-of-network-coverage.

According to the 3GPP working assumption, LTE direct D2D discovery takes place in the LTE Uplink spectrum using Single-Carrier Frequency Division Multiple Access (SC-FDMA) as an access scheme and is based on synchronous Time Division Multiplex (TDM) for multiplexing D2D and cellular transmissions. The proposed LTE direct D2D operation involves implementing part of the eNB receiver chain in the UE. Also, synchronization capabilities for inter-cell timing delay and asynchronous network cases must be enhanced in the UE receiver. Finally, the amount of RF components of a UE must increase to support D2D operation in addition to regular operation with a reasonable set of carrier aggregation combinations.

The D2D offloading network capacity gain directly scales with the number of successful D2D pairings. The requirements for a successful D2D connection are as follows:

- Two devices are in communication range of each other;
- The interference is sufficiently low to allow for establishing the service between the devices.

While capacity gains from D2D offloading likely are limited in reality, LTE direct D2D discovery will have the merit of catalyzing new forms of social networking as well as future broadband consumer proximity services including massively multiplayer online games (MMOG), push advertising, or venue onsite services.

"While capacity gains from D2D offloading are limited in reality, LTE direct D2D discovery will have the merit of catalyzing new forms of social networking as well as future broadband consumer proximity services…"

"The "3GPP-WLAN Radio Interworking" study targets the enabling of traffic offloading into operator-controlled WLAN networks…"

WLAN Offloading

Since Release 8, 3GPP has been working on the integration of WLAN into cellular network operation. Work has covered non-3GPP access to the Evolved Packet Core (EPC), various aspects of the Access Network Discovery and Selection Function (ANDSF), support of in-device coexistence of ISM and cellular frequency bands, and last but not least the Release 12 "3GPP-WLAN Radio Interworking"[24] study item. The study "3GPP-WLAN Radio Interworking" targets the enabling of traffic offloading into operator-controlled WLAN networks and at an optimal balance of and tradeoff between WLAN and cellular service.

The Release 12 study has proposed three different approaches to 3GPP-WLAN radio interworking. In two of the proposals, the UE is the master of selecting (or not) the WLAN access network while referring to 3GPP radio access network (3GPP RAN) assistance information like the RSRP (Received Signal Received Power) threshold or network load:

- In the first solution, the UE relies on the Access Network Discovery and Selection Function (ANDSF): while ANDSF rules are enhanced with 3GPP RAN assistance information, they always take precedence.
- In the second solution, 3GPP RAN assistance consists in providing rules that can be followed with or without ANDSF.
- In the third solution, the 3GPP radio access network has full control. WLAN measurements of the UE are scheduled by the 3GPP RAN and the 3GPP RAN controls Wi-Fi offload by dedicated traffic steering commands.

Regardless of the 3GPP-WLAN radio interworking solution finally chosen by 3GPP, WLAN offloading will be inevitable to meet future traffic demand.

Matching Network Capacity with Exponential Traffic Growth

A simple model allows for matching the traffic demand derived from Cisco's 2013 Virtual Network Index with network capacity evolution:

- It is assumed that traffic demand grows in accordance with Cisco[1] with a CAGR of 66 percent.
- A typical (European) cellular network operator will own spectrum licenses ranging from 400 MHz to 4 GHz mounting up to about 500 MHz (DL and UL) licensed spectrum in total. This allows for an increase of licensed spectrum use by a factor of 5 (10) from 2010 until 2016 (2025).
- Cell spectral efficiency is assumed to increase by a factor of 2 (2.5) from 2010 until 2016 (2020). Such an evolution requires LTE Release 12 UE receiver enhancements or eight eNB transmit antennas (see Table 1). Cell spectral efficiency enhancements beyond a factor of 2.5 will have to originate from beyond LTE Release 12 and 5G.

Configuration (eNB TX × UE RX)	Cell average spectral efficiency factor
2 × 2	1
4 × 4	2
8 × 2	2.5

Table 1: Spectral efficiency factors obtained derived from data in various studies[12][13][14]

- When supporting macro cell sectorization only, network densification can realistically increase by a factor of 4 by 2020. Beyond 2020, higher sectorization factors could be achieved with the deployment of Massive MIMO.
- LTE Release 12 small cell enhancements and WLAN offloading are supported both in the devices as well as in the LTE network from 2016 onwards.
- WLAN offloading offers the advantage of additionally exploiting large blocks of unlicensed spectrum in 2.4 and 5 GHz frequency bands.

Modeling in accordance with these assumptions, a conventional macro network might struggle to deliver the required DL capacity in 2016, at the latest in 2020. Hence, traffic offloading to LTE small cells and/or WLAN access points as described in previous sections is inevitable. In the model underlying Figure 12, network density in 2020 is approximately 20 times the network density in 2010. Furthermore, network densification through 2020 will be a combination of vertical and horizontal macro cell sectorization and deployment of LTE small cells and/or WLAN access points. Network densification beyond 2020 will be enabled by a combination of Massive MIMO and offloading.

"…A conventional macro network might struggle to deliver the required DL capacity in 2016, at the latest in 2020. Hence, traffic offloading to LTE small cells and/or WLAN access points is inevitable."

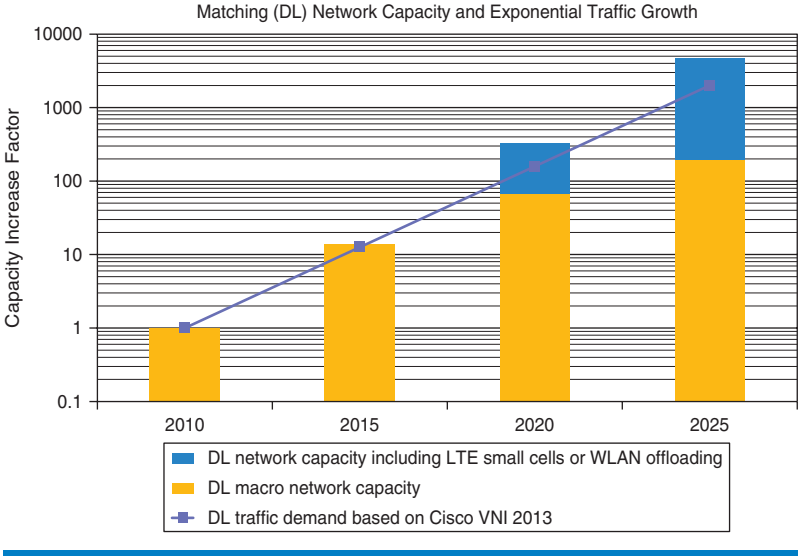


Figure 12: Evolution of DL network capacity demand vs. traffic demand based on Cisco VNI 2013
(Source: Intel Mobile Communications, 2013)

Summary and Outlook

To match future traffic demand, network capacity will have to grow in three dimensions: spectral efficiency, spectrum availability, and network density. Also, the network will have to evolve towards new architectures that foster traffic offloading. Realistic enhancements in spectral efficiency, conservative network densification based on macro cell sectorization, and exclusive use of licensed spectrum will likely not be sufficient to match traffic demand in 2020. Depending on how fast licensed spectrum can be made available, LTE small cell deployments and/or offloading to WLAN access points and unlicensed spectrum will be needed at the latest from 2020 onwards.

Besides matching future traffic demand, cellular network operators need to keep total cost of ownership and operational expenditures in particular at an acceptable level. Hence, operators seek to increase operational efficiency and network robustness by exploiting Self-Organizing Network (SON) techniques. Furthermore, operators need to target energy efficiency both in the network and for connected devices.

"Candidates of future application segments include: machine-type communication (MTC), support of wearables, device-to-device connectivity for machine-to-machine communication, the Internet of things, Public Safety, and mobile and wireless communication for a safe society."

Cellular network operators will have to enter new application segments to secure their market shares. Candidates of new application segments include: machine-type communication (MTC), support of wearables, device-to-device connectivity for machine-to-machine communication, the Internet of things, Public Safety, and mobile and wireless communication for a safe society.

Ultimately, however, the evolution of the information society and the yet-to-be-determined potential of LTE Release 12 and 5G techniques will decide the final path of radio network evolution.

Acknowledgements

The authors are grateful for feedback and valuable suggestions provided by Kenneth Stewart and Stefan Fechtel.

References

[1] Cisco Systems, Inc., Virtual Networking Index 2013, www.cisco.com.

[2] European funded project METIS: "Mobile and wireless communications Enablers for the Twenty-twenty Information Society," www.metis2020.com.

[3] 3GPP TS 36.211 V11.4.0 (2013-09): 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation, Release 11.

[4] 3GPP TR 36.872 V12.0.0: 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Small cell enhancements for E-UTRA and E-UTRAN - Physical layer aspects; September 2013.

[5] Lähetkangas, E., K. Pajukoski, G. Berardinelli, F. Tavares, E. Tiirola, I. Harjula, P. Mogensen, and B. Raaf; "On the Selection of Guard Period and Cyclic Prefix for Beyond 4G TDD Radio Access Network," European Wireless 2013 April 2013.

[6] Poston, J. D. and W. D. Horne, "Discontiguous OFDM considerations for dynamic spectrum access in idle TV channels," *Proc. First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, November 2005.

[7] Datta, R, G. Fettweis, Z. Kollar, and P. Horvath, "FBMC and GFDM Interference Cancellation Schemes for Flexible Digital Radio PHY Design," *Proc. 14th EUROMICRO Conference* (Euromicro'11), August 2011.

[8] Shannon, Claude E., "Communication in the Presence of Noise" in *Proceedings of the I.R.E.*, January 1949.

[9] Schwarz, Stefan, Michal Simko, and Markus Rupp, "On Performance Bounds for MIMO OFDM-based Wireless Communications Systems," in *2011 IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications,* Page 311.

[10] Furuskog, Johan, Karl Werner, Mathias Riback, and Bo Hagerman, "Field trials of LTE with 4 × 4 MIMO," in *Ericsson Review*, 1/2010.

[11] Jun Won Choi, Byungju Lee, Byonghyo Shim, and Insung Kang, "Low Complexity Detection and Precoding for Massive MIMO Systems," 2013 IEEE Wireless Communications and Networking Conference (WCNC), 2013.

[12] 3GPP TR 36.814 V9.0.0: 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects; March 2010.

[13] 3GPP TR 36.819 V11.2.0: 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Coordinated multi-point operation for LTE physical layer aspects; September 2013.

[14] Holma, Harri and Antti Toskala (eds), *LTE-Advanced: 3GPP Solution for IMT-Advanced* (Chichester, UK: John Wiley & Sons, 2012), ISBN: 9781119974055.

[15] 3GPP TR 36.863 V0.1.0 (2013-04): 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Network-Assisted Interference Cancellation and Suppression for LTE (Release 12), April 2014 (Release 12)

[16] Nagisa Otao, Yoshihisa Kishiyama, and Kenichi Higuchi, "Performance of Non-orthogonal Access with SIC in Cellular Downlink Using Proportional Fair-Based Resource Allocation" in Wireless Communication Systems (ISWCS), 2012 International Symposium on Digital Object Identifier: 10.1109/ISWCS.2012.6328413; Publication Year: 2012 , Page(s): 476–480

[17] 3GPP TR 37.840 V12.0.0 (2013-03): 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study of Radio Frequency (RF) and Electromagnetic Compatibility (EMC) requirements for Active Antenna Array System (AAS) base station (Release 12), March 2013.

[18] NSN White Paper: "Active Antenna Systems – A step-change in base station site performance"; product code: C401-00741-WP-201109-1-EN; http://www.scribd.com/doc/103820928/Nokia-Siemens-Networks-Active-Antenna-System-White-Paper-26-01-12.

[19] 3GPP TR 36.842 V0.3.0 (2013-08): 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Study on Small Cell Enhancements for E-UTRA and E-UTRAN – Higher layer aspects (Release 12), August 2013.

[20] 3GPP TR 36.839 V11.1.0 (2012-12): 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Mobility enhancements in heterogeneous networks (Release 11), December 2012.

[21] Nokia and NSN, "LTE TDD – FDD Joint Operation" 3GPP Work Item Description Rel-12; RP-130888.

[22] Matinmikko, Marja, Marko Palola, Harri Saarnisaari, Marjo Heikkilä, Jarmo Prokkola, Tero Kippola, Tuomo Hänninen, Markku Jokinen, and Seppo Yrjölä, "Cognitive Radio Trial Environment: First Live Authorized Shared Access-Based Spectrum-Sharing Demonstration" in *IEEE Vehicular Technology Magazine*, September 2013.

[23] Barghi, S., A. Khojastepour, and K. Sundaresan, "Characterizing the Throughput Gain of Single Cell MIMO Wireless Systems with Full Duplex Radios," 10th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 14–18 May 2012.

[24] 3GPP TR 37.834 V1.0.0 (2013-08): 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study on WLAN/3GPP Radio Interworking (Release 12).

[25] 3GPP TR 36.843 V0.1.0 (2013-10-09): 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study on Device-To-Device Proximity Services in LTE (Release 12).

Author Biographies

Sabine Roessel joined Intel Mobile Communications in December 2012 as Principal Engineer and is working on LTE and technology roadmapping. From 1986 until 2012, Sabine was first with Siemens and then with Nokia Siemens Networks (NSN). In her previous position at NSN, Sabine managed the entire NSN and Nokia LTE Research and Standardization Program. Sabine received her Dipl.-Phys. degree in theoretical physics from the Johann Wolfgang Goethe-University at Frankfurt am Main (Germany) in 1986. She is author of a textbook chapter on High-Level Synthesis and has recently filed 35 patent applications. Email: sabine.roessel@intel.com

Michael Färber joined Intel Mobile Communications Standards and Advanced Technology (SAT) in October 2012. Michael received his Dipl.-Ing. degree from the University of applied sciences Darmstadt-Dieburg (Germany) and joined Siemens in 1985. In 1990 he started in GSM standardization and followed the evolution through EDGE and UMTS to LTE standards, having roles as delegate, chairman, and vice-chairman. 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 has filed more than 100 patent applications. Email: michael.faerber@intel.com

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 the 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

Josef Hausner is leading the Wireless Systems Engineering Architecture and IP team, driving system architecture evolution and innovation across the entire mobile product portfolio. He is with Intel since the Infineon wireless acquisition in 2011. Professor Hausner holds Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from the Technical University of Munich and is a member of the board of trustees of the Fraunhofer Heinrich-Hertz Institute. Email: josef.hausner@intel.com