Large Scale Integrating Project

EXALTED

Expanding LTE for Devices

FP7 Contract Number: 258512

WP3 – LTE-M System

D3.1
First Report on LTE-M Algorithms and Procedures

Contractual Date of Delivery to the CEC:	31 August 2011
Actual Date of Delivery to the CEC:	31 August 2011
Responsible Beneficiary:	ALUD
Contributing Beneficiaries:	VGSL, EYU, ALUD, TKS, CEA, UNIS, CTTC, TUD, UPRC
Estimated Person Months:	50
Security:	Public
Nature	Report
Version:	2.0

Document Information

Document ID:	EXALTED_WP3_D3.1_v2.0.doc
Version Date:	31 January 2012
Total Number of Pages:	84
Abstract	This report summarises EXALTED proposals for LTE-M algorithms and concepts and maps them in the LTE system architecture.
Keywords	EXALTED, M2M communications, LTE-M, Release 11, PHY, MAC, RRC, cost efficiency, energy efficiency, spectrum efficiency, scalability, signalling reduction

Authors

Name	Organisation	Email
Stephan Saur (Editor)	ALUD	Stephan.Saur@alcatel-lucent.com
Srdjan Krco	EYU	srdjan.krco@ericsson.com
Nemanja Ognjanovic	TKS	nemanjao@telekom.rs
Gorica Nikolic	TKS	gorican@telekom.rs
Bojana Jakovljevic	TKS	bojanaja@telekom.rs
Bertrand Devillers	CTTC	bertrand.devillers@cttc.es
Christian Ibars	CTTC	christian.ibars@cttc.cat
Giorgio Corbellini	CEA	giorgio.corbellini@cea.fr
Emilio Calvanese Strinati	CEA	emilio.calvanese-strinati@cea.fr
Walter Nitzold	TUD	walter.nitzold@ifn.et.tu-dresden.de
Prakash Bhat	VGSL	prakash.bhat@vodafone.com
Angeliki Alexiou	UPRC	alexiou@unipi.gr
Athanasios Lioumpas	UPRC	lioumpas@unipi.gr
Petros Bithas	UPRC	pbithas@unipi.gr
Georgina Abou-Elkheir	UPRC	georgina@unipi.gr
Serdar Vural	UNIS	s.vural@surrey.ac.uk
Parisa Cheraghi	UNIS	p.cheraghi@surrey.ac.uk
Chuyi Qian	UNIS	c.qian@surrey.ac.uk

Approvals

	Name	Organisation	Date	Visa
Internal Reviewer 1	Srdjan Krco	EYU	26/08/2011	OK
Internal Reviewer 2	Trevor Gill	VGSL	26/08/2011	OK
Internal Reviewer 3				
Technical Manager	Pirabakaran Navaratnam	UNIS	26/08/2011	OK
Project Manager	Djelal Raouf	SC	26/08/2011	OK

Executive Summary

An extended radio access network with respect to Machine-to-Machine (M2M) communications also referred to as Long Term Evolution for Machines (LTE-M), is an integral part of the EXALTED system architecture. This report presents the current status of the research on LTE-M algorithms and concepts after almost one project year.

A detailed study of the functionality of LTE Releases 8, 9 and 10 discloses their inability to serve a multitude of small data rate devices in an energy-, spectrum- and cost-efficient way. On the other hand, Long Term Evolution (LTE) is being deployed and major changes in its overall system architecture are unlikely to be accepted by the 3rd Generation Partnership Project (3GPP). Therefore, EXALTED will propose improvements that can be easily integrated in the existing system and maintain backward compatibility to previous Releases. In this sense, EXALTED partners intend to contribute to the Release 11 Work Item “System Improvements to Machine Type Communications (SIMTC)” and a Study Item proposed by Vodafone dealing with Low Cost MTC User Equipments (UE) based on LTE.

After this top-down view of LTE-M, the following sections of the report present the basic principles of the individual proposals together with a description, which of the technical requirements in [1] are addressed. A preliminary performance assessment is attached where available. Also, each proposal is complemented with an outlook to the next steps in the EXALTED research activities. We distinguish between proposals that need a modification of the LTE specification in sections and those that are of a more generic nature and can thus be applied almost standard independently.

Improvements of the LTE-M Physical (PHY)-Layer primarily aim at the support of a big number of devices in one cell, a simplification of signal processing in the device in order to reduce energy consumption and cost, as well as a reduction of the Peak-to-Average Power Ratio (PAPR). For this, we present two alternative waveforms for the radio access, namely Generalized Frequency Division Multiplexing (GFDM) and Wideband Code Division Multiple Access (W-CDMA)-like waveforms, both providing good spectral characteristics. Other requirements for the LTE-M PHY are improvement of the link robustness and range extension. This can be achieved with the application of special antenna techniques. We propose an efficient channel feedback scheme for Multi-User Multiple Input Multiple Output (MU-MIMO) and a technique decreasing the effort for channel estimation. In this way, it is possible to gain energy savings and greater energy efficiency of the M2M system.

As specified by 3GPP, Layer 2 is divided into three sublayers: Media Access Control sublayer (MAC), Radio Link Control sublayer (RLC) and Packet Data Convergence Protocol (PDCP) sublayer. In this report we focus on the MAC-sublayer. We present a method called Slotted Access that can effectively prevent Radio Access Network (RAN) overload while achieving a high random access success rate, even if a multitude of devices compete for Random Access Channel (RACH) resources. The idea is to allow the random access procedure of a device only in the frame where it is paged. Moreover, the exchange of scheduling information between eNodeB and UE can be significantly reduced by the so called persistent scheduling configuration. Uplink scheduling can be further improved with respect to M2M if adjacent Resource Blocks (RBs) are assigned to a user, in order to minimize the PAPR. We propose a novel Hybrid Automatic Repeat Request (HARQ) scheme combined with the usage of the Adaptive Modulation and Coding (AMC). This approach uses the channel information feedback, and in that way the transmitter can estimate the probability of correct decoding at the receiver with corresponding coding and modulation schemes. Another proposal foresees multiple access with network coding optimized for the

transmission of short messages. In contrast to legacy random access procedures, our scheme doesn’t waste considerable amount of energy as a result of collision resolution or long contention periods. Since the message is very short there is no need to establish a connection-oriented communication.

The optimization of the LTE-M Radio Resource Control (RRC)-Layer is closely related to the achievements in the MAC-sublayer. Also in the RRC protocol design, reduction of signalling overhead and energy consumption as well as a more efficient usage of the network hardware are envisaged. Several principles have been identified that can definitely help to achieve these aims, one concerning the behaviour during inactive periods, one concerning the schemes used for short, infrequent transactions, and one describing a way to deal with extreme scenarios where the Cell Radio Network Temporary Identity (C-RNTI) space is depleted. In this section we also present an energy efficient relaying scheme. First simulation results show that we can minimize the power allocated to each relay node for a given Quality-of-Service (QoS) constraint.

Also Broadcast- and Multicast Services are seen as fundamental aspect of LTE-M since it is expected that a sizeable portion of downlink traffic in LTE-M will be of broadcast or multicast type. This traffic will typically be of moderate or high volume and will exhibit very high coverage/availability requirements and very low error tolerance. One solution for this could be the application of Low Density Parity Check (LDPC) -like rateless codes. A first design approach is presented in this report. Future work will be focused on suitable channel codes and parameter optimisations. As a second approach, a collaborative broadcast architecture is presented.

The remainder of this report describes different methods that facilitate M2M communications but are not part of standardized LTE-M signal processing and protocols. They are based on a dynamic adaptation of the system parameters according to measured signals. Firstly, spectrum sensing is discussed. Important parameters like the sensing delay and their minimum requirements have been identified, and a first comparison of candidate algorithms is presented. The second proposal is about energy harvesting. This approach has recently been made possible by introducing rechargeable batteries that can harvest energy from the environment in order to extend the lifetime of the system. Accordingly, it is essential to design the system operation taking into account the nature of the energy harvesting process to increase energy efficiency. The third proposal deals with energy reduction in the network by adaptively switching base stations on or off. This could be particularly useful for the Intelligent Transportation System (ITS) use case in rural areas along a highway. Finally, the last proposal builds the bridge to the end-to-end architecture including capillary networks. A generic MAC scheduling solution is proposed that addresses the coexistence of the cellular and the capillary network and can be seen as one interface between both domains. It takes into account the traffic characteristic of M2M communication and foresees a prioritisation scheme.

In the next months, the proposed LTE-M algorithms and concepts will be further elaborated with respect to performance and their ability to fit into the given LTE system architecture. Here we expect close interactions with other EXALTED work packages with respect to transferring our proposals to one of the test beds, submitting contributions to the ongoing specification of LTE Release 11 and refining the overall EXALTED system architecture.

Table of Contents

1. Introduction and Background

Scalability and efficiency – these are the key issues in the transition of M2M communications from various legacy systems to 3GPP LTE and its future releases. The EXALTED project has already identified the bottlenecks in the LTE specification that prevent the adequate support of M2M today, and is now going to develop novel solutions enabling the vision of cheap and simple integration of sensors and other machine devices, all of them characterized by low data rate applications, into the LTE framework. The totality of these enhancements is referred to as LTE-M.

The report at hand gives a first overview of LTE-M algorithms that are being developed in EXALTED. The presented work reflects the efforts spent during the first eleven months of the project. The main focus of this report is to describe ideas on how M2M can be supported in an LTE network, to establish a relationship between these ideas and the technical requirements discussed in the report D2.1 [1] and to give an outlook to the next steps and possible implementation options. First simulation results are added in some sections, where they have been available. This report must be understood as a toolbox of isolated candidate solutions and not yet as the specification of a complete system. This is an intermediate step that will help EXALTED to make the right selection of algorithms. Future reports will pick up these pieces as well as similar input from other work packages and integrate them to an overall EXALTED system.

Section 2 is about the overall LTE-M system architecture. We start with an inventory of the already specified LTE releases and briefly summarize their most important features. We will give reasons why this current system is by far not sufficient for the support of the presumed multitude of short messages in M2M communications and show how it can be evolved towards LTE-M. Basically, we will adopt the existing LTE system architecture, but enhance it with additional functionality and configuration options that will consider the particular requirements and constraints of M2M communications. A definition of terminology and a systematic description of possible types of communication within the LTE-M system conclude this section.

After the presentation of this top-down view, the report proceeds with a more focussed discussion of the individual innovation proposals. As justification for this approach, each of our proposals will be introduced with a brief description stating which of the shortcomings of current LTE will be solved with it, and clarifying how the performance evaluation will be done. Section 3 covers physical layer (PHY) aspects. Firstly, possible alternative radio access methods are discussed. Novel concepts for the LTE-M PHY related to antenna techniques conclude this section.

The analysis of Layer 2 enhancements is up to now limited to the MAC sub-layer, which is described in section 4. The LTE-M system architecture foresees two main functional units in the MAC layer. The first is the actual controller with the LTE-M scheduler as the most important element. We discuss how signalling overhead in LTE-M could be drastically reduced. Moreover, we propose a cross-layer optimization. The second functional unit in the LTE-M MAC handles retransmissions. HARQ scheme optimized for M2M communications is presented as well as a proposal for a combination with network coding.

Section 5 is dedicated to the RRC layer within the LTE-M system architecture. We show possible improvements in the RRC protocol and discuss the relation to traffic aggregation, which is crucial aspect in low data rate communications. A novel relaying approach, primarily aiming at energy efficiency, complements this part of the report.

The remainder of the report is not anymore sorted according to the layer structure. Section 6 integrates proposed innovations with respect to Broadcast- and Multicast Services. Firstly the exploitation of Low Density Parity Check (LDCP) codes in LTE-M is evaluated. Afterwards, a collaborative broadcast architecture is presented.

All proposals presented in sections 2-6 require more or less adaptations and extensions to the standard. Beyond that, EXALTED partners have also worked on concepts that are expected to be applicable almost independently from the ongoing specification in 3GPP. These approaches are described in section 7. They deal with spectrum sensing and energy harvesting. The final topic relates scheduling options in capillary networks to LTE-M and can be seen as one initial approach for a jointly optimized end-to-end architecture.

Finally, section 8 summarizes the most important findings and achievements and gives an outlook on how the investigations on LTE-M will proceed.

2. LTE-M System Architecture

This first part of the report presents the outcomes of a top-down approach for the design of LTE-M. The starting point in section 2.1 is an overview of main functionality and characteristics of the LTE Releases 8, 9 and 10. We give several reasons why these specifications are by far not sufficient for an efficient and scalable support of M2M in LTE (section 2.2). In section 2.3 we briefly list different proposals for improvements and extensions that can help to overcome these problems. We also indicate to which parts in the LTE system architecture the different proposals belong. A more detailed discussion of these algorithms and concepts can be found in sections 3-7. Finally, in section 2.4 we describe the functional units of the new consolidated LTE-M system architecture and explain how the communication in this system works.

2.1 Overview of 3GPP LTE

In recent years, the data volume requirements in cellular mobile communication have multiplied. This comes along with a gradual shift from voice to broadband data services. Enablers for this trend are flat-rate pricing, web-friendly mobile terminals, increased penetration of smart phones with increased demand for multimedia and data traffic, USB data modems for PCs, and an evolved network infrastructure allowing for low cost per bit. In order to cope with this demand, 3GPP has standardized LTE.

3GPP Release 8 LTE is a scalable system supporting 1.4 MHz, 3 MHz, 5MHz, 10 MHz, 15 MHz and 20 MHz bandwidth options. LTE supports both half and full Frequency Division Duplex (FDD) operation in addition to Time Division Duplex (TDD). LTE supports peak data rates of 300 Mbps in downlink (DL) and 75 Mbps in uplink (UL). LTE is optimized for packet switched services and provides an evolution path for 3G systems. It adopts Orthogonal Frequency Division Multiple Access (OFDMA) in downlink and Discrete Fourier Transform (DFT)-spread Orthogonal Frequency Division Multiplexing (OFDM), also referred to as Single Carrier Frequency Division Multiple Access (SC-FDMA) in uplink. OFDM allows combating multipath interference and supports flexible and robust scheduling mechanisms such as frequency dependent scheduling and advanced Multiple Input Multiple Output (MIMO) operation, whilst the SC-FDMA allows for low PAPR. There are 8 MIMO transmission modes supported for various MIMO techniques (Single-User (SU)-MIMO, beamforming, transmit diversity etc). Low latency is achieved with support for small Transmission Time Intervals (TTI) of 1 ms. User plane (U-Plane) (one way) latency is the time delay for Internet Protocol (IP) packets transmitted from/to a Radio Access Network (RAN) edge node to be available at the RAN/User Equipment (UE). Control plane (C-Plane) latency is the latency for Idle to connected mode transition. LTE latency requirement are captured in [2]. Theoretical analysis of U-Plane and C-Plane latency have been performed in [3] which satisfy the requirements of less than 5 ms for U-Plane and less than 100 ms for C-Plane. Handover interruption time for intra Evolved Universal Terrestrial Radio Access (E-UTRA) is of the order of 12 ms [3]. 3GPP Release 8 LTE supports several features such as Circuit Switched (CS) fallback in Evolved Packet System (EPS), Single Radio Voice Call Continuity (SR-VCC), Self-Organizing Networks (SON) and Home eNodeB (H-eNodeB).

LTE in Release 9 of 3GPP supports enhancements for support for dual layer transmission, and further enhancements for SON and H-eNodeB. Support for Multimedia Broadcast/Multicast Service (MBMS), IP Multimedia Subsystem (IMS) emergency calls, and Location Control Services (LCS) also was specified.

3GPP Release 10 LTE-Advanced (LTE-A) was specified by 3GPP as further evolution of LTE and requirements captured in [4]. LTE-A is an International Telecommunication Union (ITU) approved International Mobile Telecommunications (IMT)-Advanced technology with support for emerging technologies such as relay (inband and outband), interference management for heterogeneous cells, spectrum aggregation (up to 100 MHz) and UL (up to 4 layers) and DL (up to 8 layers) MIMO enhancements. As an enhancement to uplink access scheme, clustered SC-FDMA is specified in addition to SC-FDMA with dynamic switch between the two. Other enhancements supported are Minimization of Drive Test (MDT), H-eNodeB Local IP access, WiFi offloading, and support for Core network overloading for Machine type communication (MTC). LTE-A advertises a peak of 1 Gbps in DL and 500 Mbps in UL.

Cooperative Multipoint Tx/Rx (CoMP) system improvements for MTC and further enhancements to 3GPP Release 10 features are being considered for future releases (Release 11 and 12) of 3GPP.

3GPP Release 8 does not support Multimedia Broadcast Multicast in a Single Frequency Network (MBSFN) operation, but allows for signalling of the resources as being identified by network for the multimedia broadcast service. These set of radio resources could be termed fake MBSFN when mobiles do not expect any downlink transmission in the data region. The data region of these radio resources is then free for features not beneficial for legacy mobiles. This has proved to be an important tool for co-existence of several features that can only be supported by mobiles of later release (e.g. Relays (Figure 2-1), improvements to interference co-ordination, energy saving in eNodeB etc).

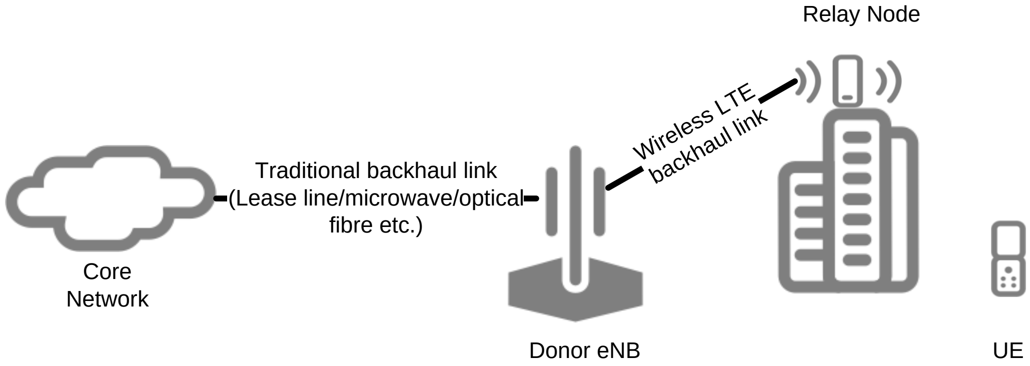

Figure 2-1: LTE relay scenario

Relays connect wirelessly to the donor cell served by donor eNodeB and can be classified as transparent or non-transparent. Relays can provide throughput enhancement and coverage extension. If a relay controls cells of its own (similar to eNodeB) then it is called a type 1 relay, alternately if the relay node is part of the donor cell then it is a type 2 relay. When a relay node controls cells of its own, it has unique physical-layer cell identity, the same Radio Resource Management (RRM) mechanisms as normal eNodeB is used and relay shall appear as a Rel-8 eNodeB to Rel-8 UEs. A type 2 relay node on the other hand is part of the donor cell and does not have a separate physical cell identity and is transparent to Rel-8 UEs.

Half duplex relays are by definition unable to transmit to the UE and receive from the donor eNodeB simultaneous and requires resource partitioning between the wireless backhaul link to the eNodeB and the access link to the UE.

Full duplex relays on the other hand can either operate as outband relays or as inband relays with enough spatial separation or with enhanced interference cancellation, thus requiring no specific resource partitioning.

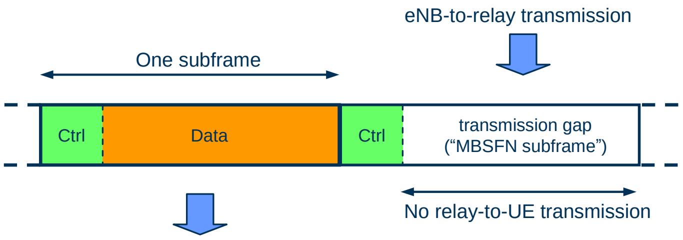

Figure 2-2: LTE relay-to-UE communication using normal subframes (left) and eNodeB-to-relay communication using MBSFN subframes (right)

Type 1, half duplex inband relay is not transmitting any signal to UEs when it is supposed to receive data from the donor eNodeB (see Figure 2-2). The relay then configures these subframes as MBSFN subframes (Fake MBSFN) when UEs (including Rel-8 UEs) are not supposed to expect any DL transmission avoiding any legacy UE measurement issues. The relay should still transmit control region (including the reference signals).

Table 2-1: Overview of relay types

3GPP Rel-10 relays		L3 Relay	L2 Relay
Inband	Half Duplex	Type 1	Type 2
	Full Duplex	Type 1b	Not defined
Outband		Type 1a	Not defined

Type 1a: ***Outband***
Type 1b: ***Inband full-duplex***

Support of Time domain Inter-cell Interference Coordination (ICIC) is performed with co-ordination of radio resource utilization across different cells in time through backhaul signalling. Similar to type 1 relay operation, eNodeB ensures backwards compatibility towards UEs by transmitting necessary control channels and physical signals as well as system information in all the appropriate transmission opportunities.

LTE was primarily designed as a packet system to support mobile broadband. Evolution of LTE optimized/enhanced the system for coverage, capacity and peak data rate improvements. Whilst the improved spectral efficiency of LTE compared to other cellular systems such as Global System for Mobile Communications (GSM) and High Speed Packet Access (HSPA) is attractive for mobile broadband and M2M traffic alike, LTE as specified by 3GPP, if used for providing M2M service, would result in over dimensioning of network and may be inefficient for certain M2M applications depending on range and environment. Several of the LTE design characteristics would be an overhead for an M2M system. Unlike devices targeted for mobile broadband, the majority of M2M applications are expected to be low end devices with limited capabilities requiring long shelf life (e.g. 5 years battery life), low cost, low power devices supporting low data rates, e.g. sensor and tracking devices. M2M devices should support broadcast calls for handling large number of devices efficiently e.g. firmware updates. M2M system should also be optimized for energy saving at eNodeB and signalling in the network. Reachability, addressing and management of address space for a large number of devices are some of the challenges that are required to be addressed for LTE M2M system. In the following sections, we will describe shortcomings of the current LTE specification for M2M communication and give an outlook to possible system enhancements towards LTE-M.

2.2 Shortcomings of LTE with respect to M2M communication

This section aims to give a detailed discussion on the question why the existing specification of LTE cannot meet the stringent technical requirements [1] with respect to the efficient support of a multitude of short messages.

- **Small data transmission in M2M**: The majority of M2M applications transmit and receive only very small amounts of data, in the order of a few kbps. One important example is smart metering with a required data rate below 20 kbps. For E-health, the required data rate is even below (<5 kbps) [1]. The aim is to transmit these short messages with very efficient resource usage, in particular with respect to the control channels. The current LTE specification was designed for broadband applications with reasonable control information overhead to achieve the required high peak data rates and high mobility support. But as can be seen in Figure 2-3, the ratio between the payload and the control information becomes inacceptable, if the same specification is applied to short messages. This is further emphasized by the fact that we envision that the number of M2M devices located in one cell will be by a factor of 10-100 higher compared to the legacy LTE UEs. Hence, in LTE-M we envisage a reduction of signalling overhead to achieve again a reasonable ratio between payload of a short message and the required control information. For this we need the option to group devices together to simplify control signalling, management and charging (e.g. by aggregating the traffic in one common channel for all M2M devices), currently not foreseen in LTE Rel-10.

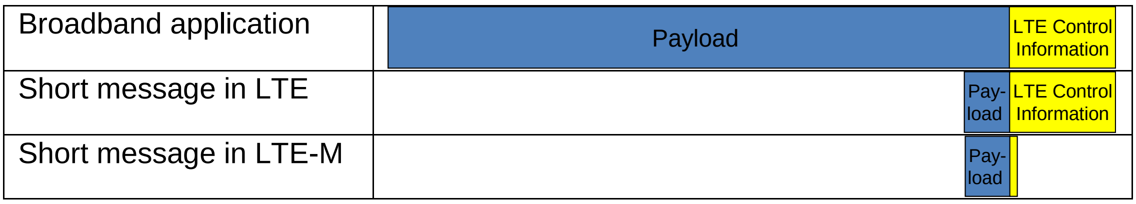

Figure 2-3: Ratio between payload and required control information in LTE and LTE-M

- **Radio access**: When in active mode, LTE devices use scheduled radio resources to transmit data. These resources are used in time and frequency and are utilized exclusively by a device to which they were allocated, hence, collisions cannot happen. The drawback is that scheduling information has to be distributed among all active devices in a cell, and that these devices have to listen to the control channel and to decode the information in order to know which resources they may utilize. This basic LTE system design paradigm is well-thought-out as long as the amount of data to be transmitted is big and the number of devices small. But in our EXALTED reference use cases [1], we presume the opposite situation: A big number of devices sending only small messages. One basic means to disburden the control channel in LTE-M (see Figure 2-3) would be the application of random access. But this alternative has a disadvantage as well: Without assignment of dedicated radio resources, collisions can occur and messages can be lost. The challenge is to find a trade-off. This requires more flexibility for the selection of the instantaneously best suited access method, a prerequisite that the current LTE specification doesn't provide.

- **Device cost issues**: Wide area M2M networks currently utilize GSM and/or (E)GPRS because their coverage is almost ubiquitous and the cost is sufficiently low. However, operators wish to reduce the number of supported radio access technologies in order to simplify maintenance of deployed hardware. Moreover, it is desirable to reuse GSM and GPRS spectrum for LTE because this technology is verifiably much more spectrum efficient. One crucial challenge for this migration is the manufacturing cost of the device. The cost of a LTE device is much higher than for a GSM or GPRS device. One focus in this work package must therefore be the provision of means to reduce the cost of a LTE-M device to the level of GSM/GPRS or even below. The main reasons for the comparably high manufacturing cost of a LTE Rel-10 device are the following:
  - **Bandwidth**: The device has to support scalable bandwidth with maximum bandwidth of the frequency band that it may utilize. This is typically 20 MHz. Naturally, this comes along with significantly higher complexity in both baseband processing and RF compared to a smaller bandwidth of e.g. 1.4 MHz.
  - **Data rate**: LTE devices must satisfy the requirements of one of the so-called UE categories [5]. This category then defines the peak data rates in both uplink and downlink the device has to support. The peak data rates are related to the required buffering capabilities of the device, the complexity of the decoder and other radio protocol processing functions, and in return, all of them impact the device cost. On the other hand, the peak data rates of the lowest UE category still seem overdesigned for M2M communications. Hence, it is obvious that existing LTE UE categories are not suitable for M2M and need to be extended.
  - **Transmit power**: According to the specification, an LTE device must be able to transmit with +23 dBm power [6]. This raises certain design requirements for the power amplifier, which makes it complex and expensive. A reduction of the maximum transmit power could simplify the power amplifier and reduce the cost. Any impact to coverage will then need to be addressed.
  - **Half-Duplex**: It is envisaged M2M applications such as fleet management may require support of roaming and hence support for multiple bands. Duplexer is band specific and with each additional band additional duplexer is required. A device operating only in half-duplex mode does not need the expensive duplexer, but can utilize cheaper filters and switches. Half-duplex M2M UE's do not transmit and receive simultaneously and can be operated in all the identified FDD bands. Whilst 3GPP has specified protocol requirements and UEs are able to indicate support for half-duplex operation to the network, 3GPP has not yet specified RF requirements for half-duplex operation.
  - **RF chains**: The number of receive antennas at the LTE UE is not specified. However, the receiver performance requirements presume that at least two receive antennas are available [6]. The second RF chain causes additional costs, not only due to hardware cost, but also due to increased buffer size and processing complexity, e.g. for channel estimation. Existing receiver performance requirements are therefore not suitable for M2M. A relaxation of these performance requirements would allow to design a device with only one single RF chain and to simplify signal processing.
  - **Signal processing accuracy**: The previous items have shown means how to reduce the device cost by simplifying the hardware and by relaxing

performance requirements. On the other hand, if cheap components are utilized, the computational accuracy of the signal processing with respect to e.g. channel estimation, decoding and synchronization may be affected. This in return entails the necessity to simplify the PHY layer functionality itself. As an example, a simple power amplifier requires a RF signal with extremely low PAPR in order to avoid a significant distortion on the radiated signal. Also the selection of modulation and coding schemes or even the multiple access method itself has to be adapted to the capabilities of cheap devices. All these aspects are currently not considered in the specification of LTE. One significant part of the research in EXALTED is therefore the investigation of M2M-friendly alternatives to LTE signal processing algorithms.

- **Network overload issues**: One of the key working assumptions in EXALTED is the support of a big number of M2M devices in one cell, i.e. by the factor of 10-100 more devices than legacy LTE UEs. Even if these devices are in idle mode most of the time, an external event could wake them up and let them connect or attach. The current LTE system is not prepared for such an overload situation, in particular the Mobility Management Entity (MME) and the Packet Data Network Gateway (PGN) may be vulnerable. Protection algorithms have to be specified that firstly prevent the network from a collapse, and secondly guarantee the QoS of the M2M application that has caused the massive device triggering. Ongoing discussions in 3GPP tend to focus primarily on the first aspect, e.g. by detaching or blocking M2M devices running certain applications, whereas EXALTED partners see the necessity to provide a suitable solution for both requirements.

- **Low mobility support**: LTE air interface is not optimized for low resource utilization for short transactions, and for low device power consumption, since M2M devices in LTE are treated in the same manner as other mobile subscribers (human), although their way of working and communication patterns are quite different. In LTE various radio measurements are performed to ensure best connection quality as the terminal is moving around. These procedures are relatively complex, use huge amount of radio resources and have significant impact on energy consumption. Actually, these procedures are necessary in the case of standard mobile terminals (phones), while for the case for many M2M terminals this is just an unnecessary overhead, since large number of M2M devices will be at fixed locations and will throughout their lifetime use one or two radio cells and consequently will not require nor use fast handovers. However, in some cases M2M devices will be mobile (e.g. deployed on vehicles) and mobility support will be required. Also, due to the characteristics of the M2M traffic (short transfers, mainly initiated from the devices towards remote servers), it is not required and necessary to have very smooth handovers and uninterrupted connections. Promising approach that will be investigated in EXALTED to overcome these issues would be to allow the M2M terminals to control the timing and to perform the required measurements only when needed, thus significantly reducing unnecessary signalling and saving the batteries of M2M devices.

- **Paging of M2M devices**: In LTE, when M2M device is in the IDLE mode, monitoring paging channel takes up substantial amount of time and consumes significant amount of energy. For mobile phones this is necessary to ensure quick response in case of incoming calls, while for M2M devices it is an overhead as the majority of traffic is terminal initiated or at least it is possible to tolerate delay in answering the incoming calls. To cope with this issue, for M2M devices should be possible to avoid monitoring paging and/or to increase the paging cycle (Discontinuous Reception (DRX) cycle), which will be further explored in EXALTED.

- **Addressing schemes of M2M devices**: IP header, even if header compression techniques such as Robust Header Compression (ROHC) are used, could present significant overhead for M2M applications with small payload. A mechanism is required to reduce overhead of the application layer addressing for M2M devices in an LTE-M system. One solution is to map the application layer identity to radio identifiers, e.g. to the UE specific C-RNTI. A mapping from/to IP address to/from RNTI would be required to be done at the eNodeB, and the application on the M2M device will not be IP centric. It should also be noted that Transmission Control Protocol (TCP) may not be the optimal protocol for the transport of M2M traffic with very long periods without data transfer due to retransmission timeout.

2.3 System enhancements towards LTE-M

Here, we give an overview on possible improvements that can overcome the shortcomings of the existing specification discussed in the previous section. Each individual proposal will be justified and studied in detail in sections 3-7.

The enhancements in the LTE protocol stack towards the full support for the requested LTE-M features can be done in each of its layers and sublayers. Their logical implementation is presented in Figure 2-4.

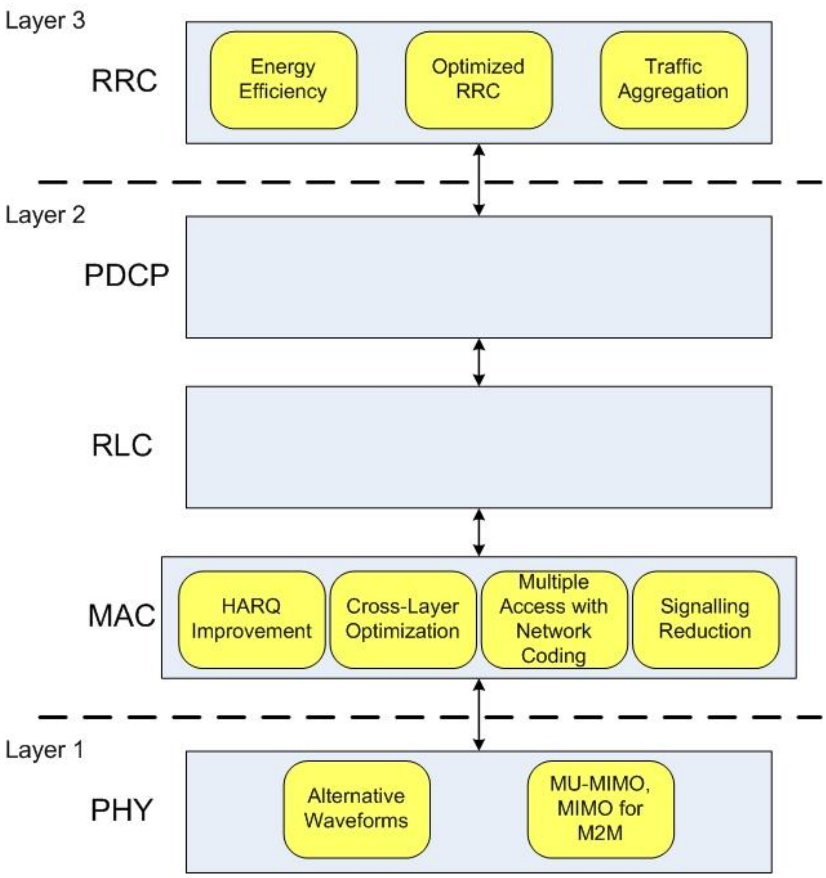

Figure 2-4: LTE-M Landscape

Physical Layer (Layer 1) improvements, within the scope of WP3, are the following:

1. **Introduction of alternative waveforms** – LTE-M technology requires simultaneous utilization of a large number of cheap, low power consuming end-devices. Standard orthogonal waveforms are not applicable in M2M systems, because of their lack of necessary spectral characteristics, such as low PAPR, high coverage range and spectral efficiency. Generalized Frequency Division Multiplexing (GFDM) and Wideband Code Division Multiple Access (W-CDMA)-like waveforms are two proposed modulation candidates for LTE-M, providing good spectral characteristics and low power consumption.

2. **Additional antenna techniques** – Improvements must be made in the area of signal coverage, greater spectral (energy) efficiency, and support for large number of simultaneously served M2M devices in LTE-M systems. The following two techniques are offered:
   a) MU-MIMO (Multi-User MIMO) technology: an efficient channel feedback scheme that meets QoS criteria, providing reduced signalling and avoidance of repeating of highly correlated information in the feedback link.
   b) MIMO for M2M technique: additional savings in signalling and computational resources are achieved by decreasing the number of channel estimations, i.e. only channels selected for data decoding are estimated. In this way, it is possible to gain energy savings and greater energy efficiency of the M2M system.

As specified by 3GPP, Layer 2 is divided into 3 sublayers: Media Access Control sublayer (MAC), Radio Link Control sublayer (RLC) and Packet Data Convergence Protocol (PDCP) sublayer. Among the several possible techniques for the improvements in each of the sublayers, the following are the most relevant:

1. **Hybrid Automatic Repeat Request (HARQ)** – In order to reduce number of required retransmissions, two types of HARQ retransmission scheme are proposed:
   a) Type I HARQ requests packet retransmissions when received data can't be decoded. "Bad" packets are discarded, and retransmissions are repeated until packet is correctly decoded or pre-set number of retransmissions is reached.
   b) Type II HARQ introduces incremental data redundancy. If the first packet cannot be correctly decoded, additional redundant information is sent in the following packets.

   Both types of the HARQ schemes are improving the spectral efficiency, but neither is a good solution for the LTE-M systems – type I is a high power consuming and low spectral efficient method, while type II requires a buffer at the receiver side. The proposed solution includes the usage of the Adaptive Modulation and Coding (AMC) schemes jointly with the HARQ, because the AMC uses the channel information feedback, and in that way the transmitter can estimate the probability of correct decoding at the receiver with corresponding coding and modulation schemes.

2. **Cross-Layer Optimisation** – This method is related to the uplink schedulers. For the purpose of EXALTED scenarios, where a large number of devices need to transmit data to the eNodeB, the uplink scheduler is assigning adjacent RBs to a user, in order to minimize the PAPR. The main difference from the existing LTE uplink schedulers is that LTE-M schedulers do not require the formation of different classes of devices, which leads to suboptimum solutions, since the actual QoS constraints (e.g. delay) may vary for each user/device within that class. An efficient scheduling may increase the percentage of satisfied users/devices, in terms of different QoS criteria, such as throughput or maximum tolerable packet-delivery delay.

3. **Multiple access with network coding** – This is one of the improvements for LTE-M systems, referring to a random access mechanism for short messages transmission. When a device needs to transmit an unscheduled short message to the network, there is a need for introducing a new random access scheme for the uplink of LTE-M, based on a collision resolution scheme (devices transmit short messages without previously listening to the channel). Legacy random access may waste considerable amounts of energy as a result of collisions or long contention periods. Since the message is very short there is no need to establish a connection-oriented communication.

4. **Signalling Reduction** – This is a key issue if a multitude of devices in one cell request radio resources. We present a method called Slotted Access that can effectively prevent RAN overload while achieving a high random access success rate, even if a lot of devices compete for RACH resources. The idea is to allow the random access procedure of a device only in the frame, where it is paged. Moreover, the exchange of scheduling information between eNodeB and UE can be significantly reduced by the so called persistent scheduling configuration. We investigate improvements with respect to M2M.

LTE-M Radio Resource Control Layer (Layer 3) has several improvements in comparison to the LTE:

1. **Energy efficient relaying** - Relay Nodes have been introduced to LTE-Advanced to enable traffic forwarding between eNodeB and UE. The proposed algorithm aims to provide a two-hop energy efficient communication scheme between UE and eNodeB by employing a set of Layer 1 (L1) relay nodes. In the proposed scenario, a source node communicates with the corresponding destination node with the help of one or more relay nodes. The objective is minimizing the total consumed power at the relays under specific performance constraints, while also considering the individual power limitation of each relay, and also determining the optimal transmission energy at each relaying node, while providing QoS assurance to the receiver.

2. **Optimized RRC protocol** - The main approach in designing such an optimization relies on utilizing the knowledge of an M2M device characteristics and the way the devices plan to use the mobile network, in order to avoid unnecessary signalling procedures and when possible to schedule such transfers to enable optimal utilization of available radio resources. With the simplification of the network procedures and protocols, the power consumption of M2M devices will be also reduced, thus gaining benefits on both sides.

3. **Traffic aggregation** - One of the main concerns for serving M2M traffic in LTE is low utilization of radio resources due to large overhead. There are several approaches to deal with the issue of radio resource utilization. First could be setting up of semi-permanent communication channels. During the initial request, a M2M device would need to provide additional parameters specifying the amount of data to be transferred during each session, periodicity of data transfers, tolerable delay, transfer speed, etc. eNodeB would then use these parameters in the scheduling process to set up a dedicated channel for this device. This channel would be put to a sleep state after the required amount of data has been transferred. While in the sleep state, radio resources will be released and available for other users. At a specified time, the channel is put back in the active mode using previous parameters and thus avoiding new signalling procedure. Another approach is to use a common channel for all M2M devices in a cell combined with downloadable components responsible for processing

data from M2M devices. LTE provides a default bearer for each terminal which is also assigned an IP address and the scheduler allocates appropriate radio resources to terminals. Possible approach is also to setup a default M2M bearer that all M2M devices in a cell can use, reducing in that way the amount of signalling, increasing amount of transferred user payload, but at the same time complicating the scheduling process.

2.4 Communication in an LTE-M network

The key elements of the EXALTED high-level architecture are presented in Figure 2-5. Furthermore, the elements are described, together with the types of communication with the Application Server, and between them.

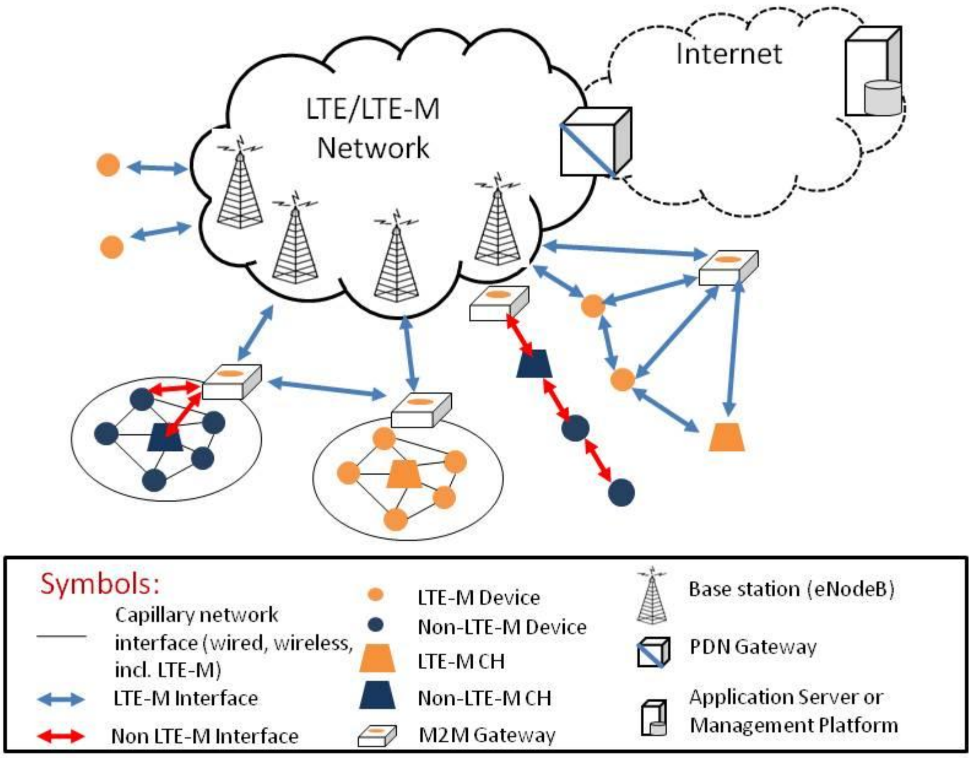

Figure 2-5: High-level EXALTED system architecture

The main M2M system components are the following:

- LTE-M Devices and Non-LTE-M Devices (or end-devices),
- M2M Gateway,
- LTE-M Cluster Heads (CHs) and Non-LTE-M Cluster Heads
- LTE-M enabled Base Station,
- Core Network,
- M2M Application Servers.

Both **LTE-M Device** and **Non-LTE-M Device** are M2M devices running one or more M2M applications regardless of the air interface used. Each can act as a standalone device, communicating with the Application Server over the LTE-M network, or it can form a capillary network with other devices of the same type, i.e. devices that use the same communication protocol. The capillary network may use LTE-M, but also other protocols (e.g. ZigBee, Bluetooth) for exchanging the information between devices. An **LTE-M Device** is a special M2M device that is equipped with an LTE-M radio interface. Non-LTE-M devices and Non-LTE-M cluster heads are used for explanation purposes only, and they are beyond the scope of EXALTED project.

Each capillary network must have a node which provides connectivity to LTE-M, and, if necessary, performs protocol translation, data aggregation and other functions for the M2M devices. This node is called **M2M Gateway**. The M2M Gateway can run the same M2M application as other devices, therefore an M2M Gateway is considered to be an advanced M2M device. An M2M Gateway can use arbitrary network interface for the communication with the capillary devices, but the communication with the LTE-M network is done via LTE-M interface. According to Figure 2-5, the M2M Gateway exhibits a key role in the EXALTED system architecture.

CHs are considered as more powerful M2M devices with some additional capabilities. Like regular M2M devices, they are also part of capillary networks and the communication from a regular M2M device will be directed through and managed by a CH. The functionalities of a CH may include data aggregation, device management, routing, etc. Depending on the network access technology they implement, they can be named as LTE-M CHs, and non-LTE-M CHs.

The network access to the LTE-M Devices and Gateways is provided by the **LTE-M Base Station** (LTE-M eNodeB). An LTE-M Base Station is defined as an LTE(-Advanced) base station that fully supports LTE-M functionality. Hence, it is part of the LTE-M network, but also of the LTE network, which co-exists with LTE-M according to the proposed architecture. The **Core Network** is considered to be the existing Evolved Packet Core (EPC). Hence the operator can utilize already deployed infrastructure for both LTE and LTE-M.

On the top of the underlying protocols and technologies, the particular **M2M Application Servers** communicate with M2M devices and Gateways that run the same application. Apart from the Application Servers, the architecture assumes other relevant servers, such as Device Management Server (DMS), which uses specifically designed protocol over the same network for communication with Devices and Gateways and servers needed to ensure security requirements.

Based on the above definitions, the communication between the elements can be one of the following two general types:

- Communication of Devices/Gateways with the Application Server (Type 1)
- Communication between Devices/Gateways (Type 2)

2.4.1 Communication of Devices/Gateways with the Application Server (Type 1)

In this communication type, the following communication modes are available (shown in Figure 2-6):

A. **LTE-M Device ↔ Application Server**: This is the simplest mode of communication, where the application data is directly encapsulated into the LTE-M protocol stack, and forwarded over the LTE-M network from a Device to a Server, and vice versa.

B. **Non-LTE-M Device ↔ Non LTE-M CH ↔ M2M Gateway ↔ Application Server**: In this case the Non-LTE-M Device is part of a capillary network that is not LTE-M, and the access to the LTE-M network is realized through a M2M Gateway. The application data is sent by the Non-LTE-M Device to the Non-LTE-M CH. The CH forwards the data to the M2M Gateway, which performs protocol translation from the capillary network protocol to the LTE-M protocol stack and vice versa. The application data is extracted and re-encapsulated by the M2M Gateway.

C. **LTE-M Device ↔ LTE-M CH ↔ M2M Gateway ↔ Application Server**: This type of communication has one major difference from the previous one – there is no protocol translation done by the M2M Gateway. However, the M2M Gateway has other functions that are not common with other devices, such as data aggregation, routing, etc. Actually, this scenario can be considered as a LTE deployment with relays.

D. **Multi-hop Device ↔ Application Server**: In case of a complex routed network of LTE-M capable devices, the multi-hop scenario is possible. This case differs from the previous one, since devices do not form a capillary network, and the communication between them is necessary only for routing purposes.

E. **LTE-M Device Group ↔ Application Server**: This type of communication corresponds to broadcast/multicast traffic (not shown in figure for the sake of clarity). Application data is encoded using a rateless coding scheme and encapsulated into the LTE-M protocol stack. A relay extension is possible using network coding.

F. **Non-LTE-M Device Group ↔ Application Server**: It corresponds to broadcast/multicast traffic (not shown in figure for the sake of clarity) to a group of Non-LTE-M devices. Application data is encoded using network coding.

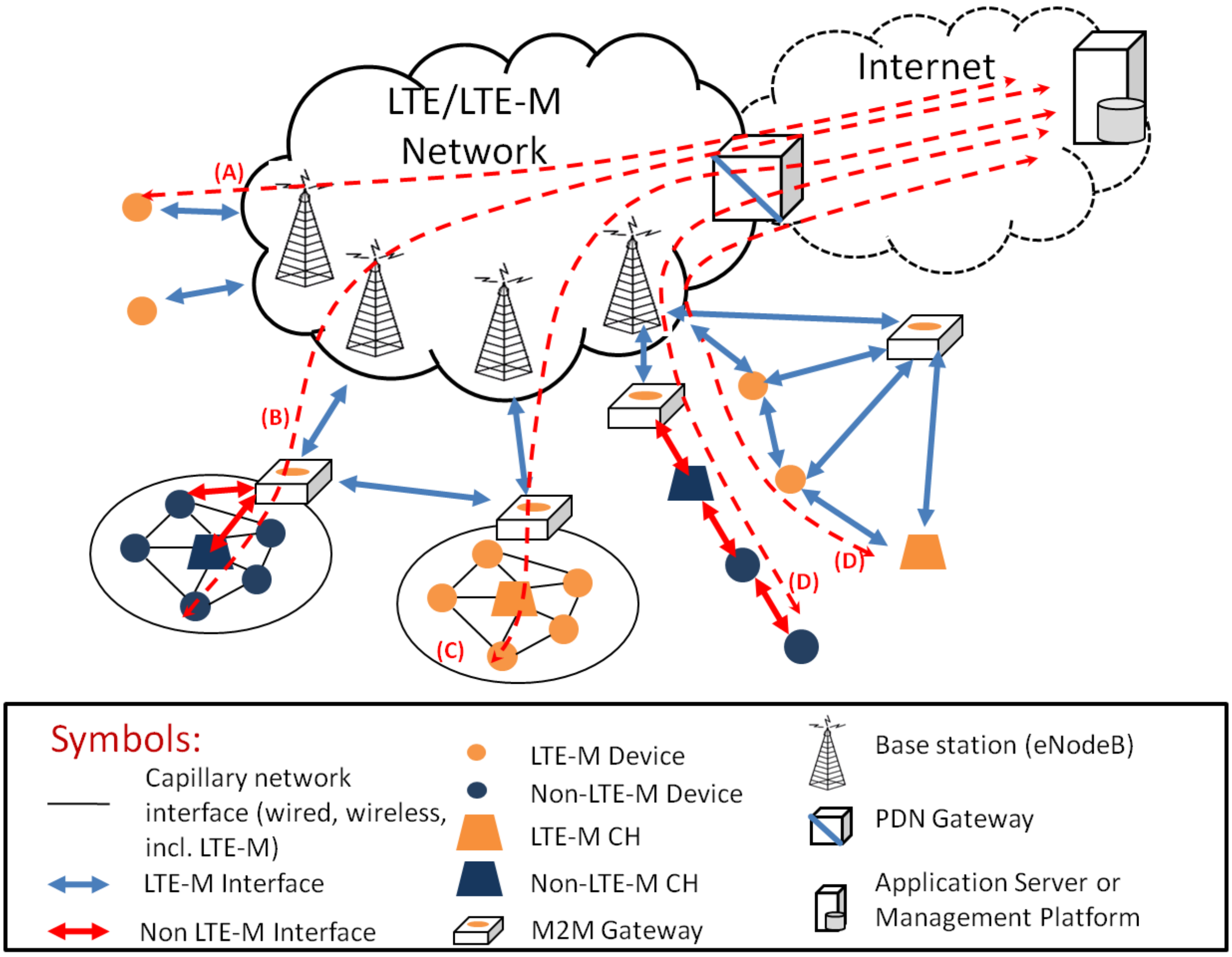

Figure 2-6: Device/Gateway – Application Server types of communication

2.4.2 Communication between Devices/Gateways (Type 2)

In this communication type, the following communication modes are available (shown in Figure 2-7):

A. **Device-to-Device within a capillary network**: In this simplest case devices exchange information directly, i.e. the core network is not aware of it.

B. **Device-to-Device outside a capillary network**: Only LTE-M Devices (and M2M Gateways) are allowed to communicate directly outside the capillary network. The purposes of this communication are numerous: optimal routing of information in large areas, peer-to-peer exchange of information between different devices or capillary networks of the same type, signalling reduction and traffic offloading in the core network.

C. **Device-to-Device over LTE-M network**: In case when LTE-M/Non-LTE-M Devices or M2M Gateways cannot establish a direct link (e.g. they are far away from each other), they communicate over the Base Station(s) and the LTE-M transport network.

D. **Non-LTE-M Device-to-Non-LTE-M Device in different capillary networks**: This is the most complex case of Device-to-Device communication. Devices exchange application data, but since they are not in the same capillary network they must forward the information to their M2M Gateways, which perform the protocol translation, route the traffic (over the LTE-M network or multi-hop) to the receiving M2M Gateway, who then again translates the information to the destination device. End-devices must be of the same type (i.e. using the same protocol).

E. **LTE-M Device-to-LTE-M Device in different capillary networks**: This case is similar to the previous one, with the only difference – there is no protocol translation needed by the M2M Gateways.

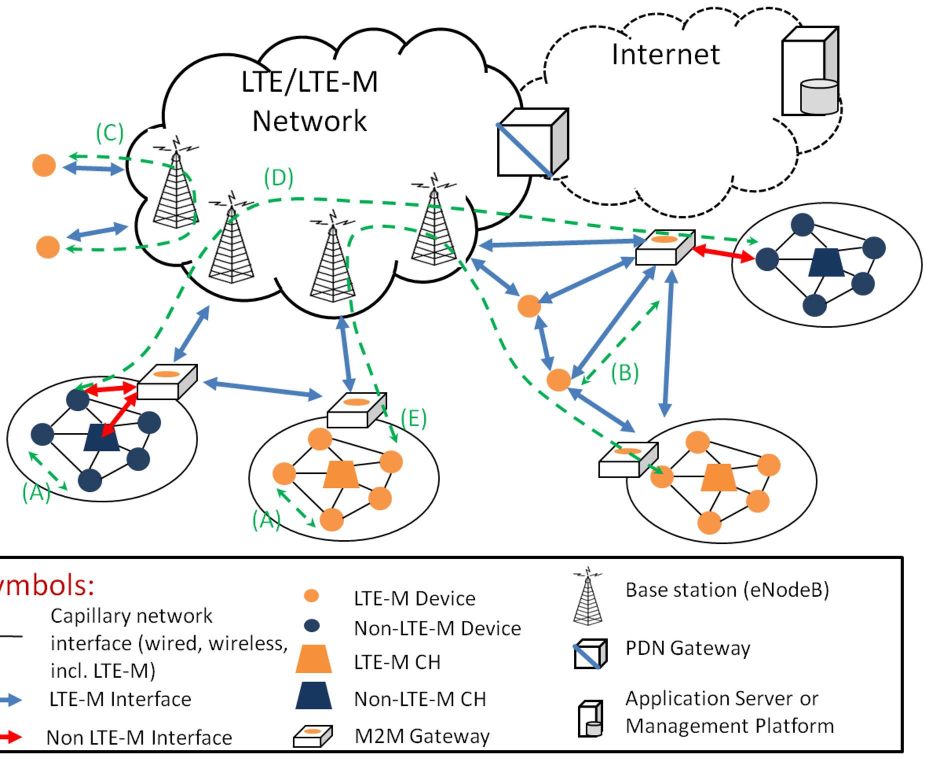

Figure 2-7: Device/Gateway-to-Device/Gateway types of communication

3. LTE-M Physical (PHY)-Layer

The LTE system and especially the physical layer of the LTE system are designed upon several criteria, i.e. high data rate and low latency. The envisaged rise of M2M type communication sets other design criteria for the development of a physical layer such as energy efficiency, low complexity and reliability. These are the scope of the work within the development of a suitable physical layer of the LTE-M system.

The following section introduces suggestions for algorithms that aim to accomplish the design criteria and requirements of a M2M system as described in [1]. The focus will be on new radio access methods described in Section 3.1 and on different antenna techniques enhanced for the requirements of the EXALTED system in Section 3.2.

3.1 Radio access

There are several reasons for not utilizing OFDMA/SC-FDMA for M2M, but applying alternative waveforms.

- We must assume that the transmit power of an LTE-M device is lower than that of a legacy LTE UE. Hence, LTE-M may have to cope with coverage holes. An alternative waveform should be applied if the range of the radio link can be extended with it.

- We must further assume that an LTE-M device is equipped with cheap and low-complex power amplifiers and other hardware components. An alternative waveform would be beneficial if the transmit signal could by synthesised without significant distortion. For this purpose, the PAPR of the waveform should be as low as possible.

- A third crucial requirement is the number of active LTE-M devices that can be supported simultaneously. An alternative waveform would be superior to OFDMA and SC-FDMA if this number could be increased, e.g. by assigning different (W-)CDMA codes to the devices.

- Another argument for the application of an alternative waveform for LTE-M would be the simplification of signal processing at the device that comes along with a reduction of energy consumption and cost. This could e.g. be achieved if only a small fraction of the bandwidth has to be decoded.

The performance of the alternative radio access methods presented in this section has to be evaluated with respect to these criteria.

Figure 3-1 illustrates two completely different LTE-M radio access approaches within the LTE frame structure. The bright blue rectangles symbolize LTE RBs, i.e. 12 physically adjacent subcarriers in frequency direction and 7 OFDMA or SC-FDMA symbols in time direction. LTE adopts OFDMA in downlink and SC-FDMA in uplink.

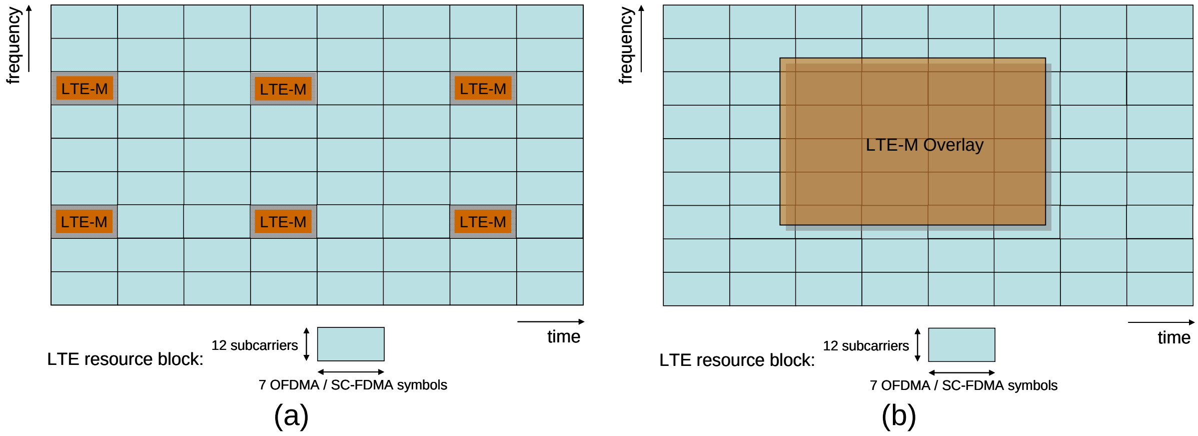

Figure 3-1: LTE-M radio access alternatives in the LTE spectrum: (a) orthogonal resources and (b) non-orthogonal resources

The basic idea in the left figure is that some RBs are not assigned to legacy LTE users. These empty spots in the frame can be utilized to implant an LTE-M waveform that does not necessarily have to be based on OFDMA or SC-FDMA. This is indicated by the red rectangles. Eventually, additional guard intervals in time and frequency between LTE and LTE-M resources have to be included to avoid mutual interference and to relax synchronization requirements of the LTE-M devices. The pattern in Figure 3-1a is just an example. As alternatives, a certain frequency resources, e.g. one RB, can be assigned permanently to LTE-M users, a pattern that could avoid the obligation for synchronized transmissions completely, or the complete frequency band can be dedicated to LTE-M for single RBs in time direction, thus relaxing frequency accuracy requirements. Common to all these patterns is the attribute that RBs are either utilized exclusively for LTE, or exclusively for LTE-M.

The right figure follows a different approach. Firstly, all available RBs are utilized for LTE. There are no empty spots like in the left figure. Simultaneously, LTE-M devices transmit LTE-M signals in the same frequency band. These transmissions can cover the complete frequency band or only a part of it. The principle can be seen as an overlay system. The advantage of this approach is that LTE-M devices don't need to synchronize to the LTE frame to fill exactly their dedicated RBs, which may relax the signal processing in the device and therefore reduce energy consumption, complexity and cost. Moreover, the transmission is not limited to the size of a RB, but may stretch across a flexible time-frequency region. However, LTE and LTE-M signals will superimpose on air, i.e. mutual interference will occur. The cross-talk from LTE-M to LTE must be kept as small as possible to maintain the quality of experience of the LTE user. Hence, the spectral power density of the LTE-M signal should be as small as possible. In fact, from the LTE perspective, LTE-M signals must not affect the LTE signal quality much more than weak additional noise. On the other hand, highly specialised low Signal-to-Noise Ratio (SNR) receivers at the base station, eventually based on signal correlation, can counteract cross-talk from LTE to LTE-M.

The following two subsections will describe candidates for LTE-M waveforms in more detail.

3.1.1 Generalized Frequency Division Multiplexing (GFDM)

The proposal in a nutshell:

We overcome the following shortcomings of the existing LTE specification (see section 2.2):

- Small data transmission: especially regarding the support of a large number of M2M terminals
- Device cost issue: especially the signal processing complexity in conjunction with the specific simplicity of M2M terminals and the corresponding accuracy conditions

The evaluation and the comparison with the benchmark LTE will be done presumably with the following performance metrics (definitions see Appendix A3):

- Bit error rate (BER): The metric is a typical assessment metric for the PHY layer meaning it aims on all EXALTED use cases. The performance should at least be the same as LTE. The evaluation will be done either analytically or via link-level simulations.
- Peak-to-Average Power Ratio (PAPR):.This metric also applies to all use cases as it can only be calculated at the common PHY layer of LTE-M. The PAPR is important for low-cost devices with relaxed requirements on the RF frontend so the PAPR of GFDM should be less than that of LTE.
- Out-of-Band Transmission (OOB): If the transmission of messages from M2M nodes happens within the spectrum of the primary LTE system, the radio access technology of the M2M system (LTE-M) has to ensure that the primary system will not be harmed above a certain threshold. To keep this influence as low as possible the respective out-of-band radiation of the used transmission scheme (GFDM) must stay under a certain threshold and should be below that of LTE. Therefore the OOB will be assessed via link-level simulations.
- Spectral efficiency: The efficient use of spectrum is very important in a regime where spectrum is expensive. LTE offered certain techniques to increase the spectral efficiency of the system but for M2M the design goal is different as the aim is low complexity and short message transmission. Therefore spectral efficiency might not be as high as LTE. The evaluation is done on a simulative basis.
- Complexity: Complexity is a very crucial issue in the implementation of e.g. low end sensor nodes as processing power is a limited resource especially with the aim to have long battery lifetime. The expected complexity of GFDM should stay at least the same as for LTE and the evaluation is done analytically.

Detailed description:

The following approach deals with the approach shown in Figure 3-1a where dedicated resources of LTE are used for the radio access of LTE-M.

GFDM is a digital multi-carrier transceiver concept, derived from the general idea of filter banks. It aims to extend traditional OFDM by introducing additional degrees of freedom that primary address the spectral shaping of the transmitted signal as well as the spectral efficiency of a wireless communication. In this context, GFDM provides means to contain out of band radiation through pulse shaping with adjustable matched filters. However, sharp edges in frequency domain response have to be traded for a greater spread of the signal in time domain. To avoid increasing the length of the cyclic prefix with the pulse shaping filter length, GFDM relies on tail biting. Further, by combining several multi-carrier symbols as they are known from OFDM to one GFDM block, the amount of Cyclic Prefix (CP) per data can be reduced and thus spectral efficiency increases.

The large amount of expected M2M terminals (Technical requirement FU.1, see [1]) shows the need for a scalable and spectrum efficient (FU.2) radio access scheme like GFDM. The capability of supporting M2M terminals with low data rate and also such with higher data rate serves the diverse M2M services envisaged within the EXALTED vision (FU.3). As GFDM is a very spectrum efficient radio access method, it can use scalable dedicated resources from the LTE resource grid while not harming the primary LTE system (NT.2). For the implementation of this radio access method within already deployed infrastructure the only need is to have software updates for base band processing as the signal processing part is done in the digital domain (NT.3).

The current used radio access techniques in LTE are not sufficient for the use within the diverse functionalities of the M2M context. OFDM performs well if the resource grid is perfectly in sync. If not, heavy distortion will destroy the benefits OFDM brings into play. As GFDM already deals with being non-orthogonal and therefore asynchronicity it conquers the shortcomings of OFDM. This is of special interest when very simple M2M terminals with low complexity come into play.

The GFDM approach is a generalized vision on single-and multicarrier access schemes and aims on a contemporary understanding of the mechanisms for radio access for different requirements. As M2M scenarios have such diverse requirements GFDM is the starting point for deeper investigations. It is a generalization of OFDM and SC-FDMA and therefore covers these solutions too while bringing into play a wide range of flexibility.

GFDM Model and architecture

Let $d[k,m]$ be a complex valued information symbol. The $K \times M$ matrix

$$\mathbf{D} = \begin{pmatrix} d[0,0] & \dots & d[0,M-1] \\ \vdots & \ddots & \vdots \\ d[K-1,0] & \dots & d[K-1,M-1] \end{pmatrix} \qquad \textbf{(3-1)}$$

will be addressed as an information block. Therein, $k = 0,\dots,K-1$ shall denote a subcarrier while $m = 0,\dots,M-1$ refers to a time slot. With the intention to distribute the data symbols in time and frequency, the discrete impulse response of the pulse shaping transmit filter $g[n]$ needs to be movable in those dimensions. Mathematically, the expression $g[n-mN]e^{j2\pi\frac{kn}{N}}$ accounts for these shifts, where given a sampling time T_S, the length of one symbol in time is NT_S and $\frac{1}{NT_S}$ denotes the spacing of two neighbouring subcarriers in frequency domain.

The transmit signal

$$x[n] = \sum_{m=0}^{M-1}\sum_{k=0}^{K-1} d[k,m]g[n-mN]e^{j2\pi\frac{kn}{N}}, \quad 0 \le n \le NM \qquad \textbf{(3-2)}$$

results for one block from the superposition of all shifted impulse responses that are weighted with the respective information symbols $d[k,m]$.

In order to be able to perform equalization at the receiver in frequency domain, $x[n]$ is prefixed with a cyclic extension and yields $\tilde{x}[n]$, which is the signal that is going to be sent through the radio channel. The received signal is given by

$$\tilde{y}[n] = \tilde{x}[n] * h[n] + n[n] \tag{3-3}$$

where $*$ denotes convolution with respect to n. Removing the CP, provides $y[n]$ and assuming the channel response $h[n]$, is known perfectly at the receiver, one block of $K \times M$ information symbols is equalized by

$$\bar{y}[n] = \mathrm{IDFT}\left(\frac{\mathrm{DFT}\ \ y[n]}{\mathrm{DFT}\ \ h[n]}\right) \tag{3-4}$$

with $\mathrm{DFT}\ \bullet$ being the Discrete Fourier Transform and $\mathrm{IDFT}\ \bullet$ denoting its inverse. However, in order to ensure the cyclic structure of $y[n]$ that is a prerequisite to (3-4), the cyclic prefix of the system requires to account for the channel, as well as the transmit and receive filter.

Assuming T_h denotes the length of the channel impulse response in time domain and T_g the length of the matched filter, then the cyclic prefix needs to be of length $T_{CP} = T_g + T_h + T_g$ to prevent interference between subsequent blocks and to make Frequency Domain Equalization (FDE) possible. The resulting decrease of the data rate is of factor $\frac{T_b}{T_b + T_{cp}}$ and the increase of the power required to transmit one bit of information is its reciprocal for $T_b = MT_d$. Clearly, from this point of view it is desirable to keep T_{CP} as short as possible, while at the same time for spectral shaping large values for T_g are favourable. Tail biting [7] has been introduced as one way to reduce the length of the CP without cutting short on the pulse shaping filter length. It is based on the idea of preserving a circular structure within each transmitted block, which allows to keep the length of the CP independent from the length of the transmit filter.

While in [7], tail biting is only used on the transmitter side, here the concept is also applied to the receiver. Therefore each subcarrier is received and processed using the matched filter $g[n]$ according to

$$\bar{y}_k[n] = \bar{y}[n] e^{-j2\pi \frac{kn}{N}} \ \#\ g[n] \tag{3-5}$$

with a circular convolution $\#$ with respect to n. By keeping every Nth sample, the information symbols $\bar{d}[k,m] = \bar{y}_k[mN]$ are retrieved and passed to the detector.

The complete GFDM system model is depicted in Figure 3-2.

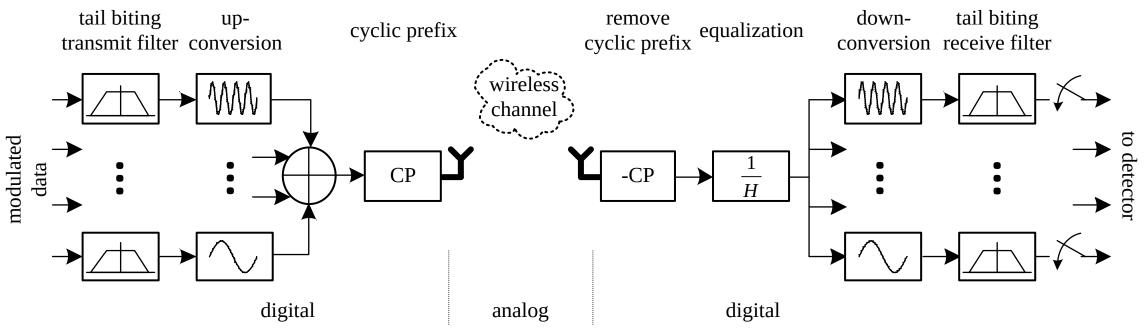

Figure 3-2: GFDM transceiver model

In order to investigate out-of-band leakage of OFDM and GFDM, an OFDM primary system is considered. First, a white space is artificially created within the bandwidth of the primary system by silencing a given number of subcarriers. Then, an asynchronous secondary system is inserted into the whitespace according to Figure 3-3. Two setups are investigated: In Setup 1, OFDM was used as a secondary system, while in Setup 2 a GFDM secondary system was employed.

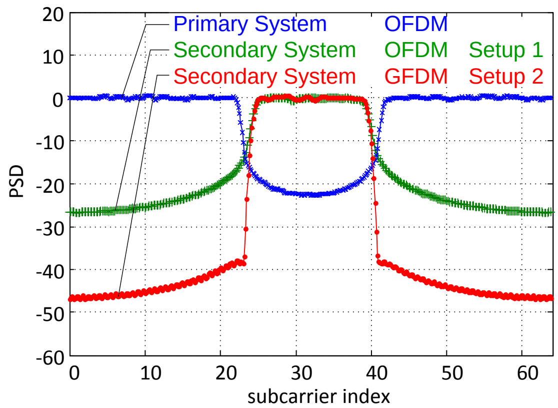

Figure 3-3: Power spectral density of primary and secondary system from Setup 1 and Setup 2.

The results shown in Figure 3-4 have been obtained through simulation of uncoded bit error rate performance under Additive White Gaussian Noise (AWGN) conditions and with Quadrature Phase Shift Keying (QPSK) modulation. A total bandwidth of 64 subcarriers was considered, where 2/3 were assigned to the primary system and 1/3 was assigned to the secondary system. One subcarrier was left silent as a guard band on each side. The GFDM system employed a Root Raised Cosine (RRC) pulse shaping filter with roll-off factor 0.25 and filter length of 11 symbols.

The conclusion that can be drawn is that while the secondary GFDM system performs worse in Setup 2 than the secondary OFDM system in Setup 1, GFDM offers a better protection to the primary system.

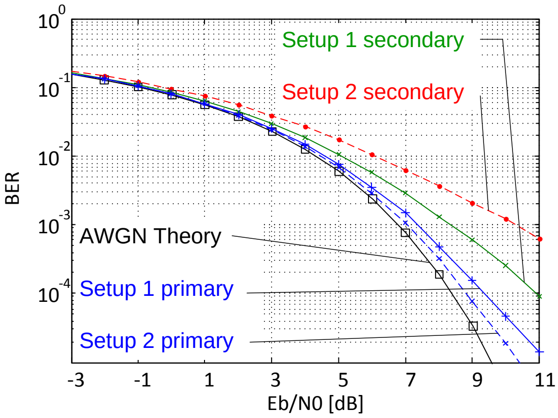

Figure 3-4: Bit Error Rate performance of primary and secondary system from Setup 1 and Setup 2

Further research might be carried out in various directions. One important problem is to deal with the non-orthogonality induces by filters and the channel. Equalization must be applied. To further understand the concept of GFDM various investigations on the system model will be carried out to search for better implementation strategies and coexistence behaviour with other systems like OFDM. As the system architecture for the EXALTED LTE-M system is not defined yet, the approach fits well the current status of development as its generality gives the opportunity to cope with different possible envisaged system solutions. Probable challenges on the system level are the synchronization of the GFDM system within the primary OFDM/SC-FDMA system. Several settings can be envisaged. Also if and how to use a power control in the uplink might be a challenge as a M2M system tries to minimize energy consumption and therefore signalling. Incorporating power control would need a certain amount of signalling due to reference/sounding signals embedded but can on the other hand reduce transmit power of the LTE-M devices when adapting to channel conditions.

3.1.2 (W-)CDMA-like waveforms

<u>The proposal in a nutshell:</u>

We overcome the following shortcomings of the existing LTE specification (see section 2.2):

- Device cost issues, in particular with respect to signal processing complexity and achievable computational accuracy
- Small data transmission, in particular with respect to number of supported devices in one cell and the respective signalling overhead

The evaluation and the comparison with the benchmark LTE will be done presumably with the following performance metrics (definitions see Appendix A3):

- Peak-to-Average Power Ratio (PAPR): This is of particular interest for all use cases, where we must assume very cheap devices with reduced computational accuracy, e.g. sensors. A small PAPR relaxes the required accuracy of the power amplifier. Therefore we aim at a PAPR that is smaller than in LTE. The evaluation will be done either analytically or by link level simulation.
- Out-of-Band Transmission (OOB): Again, this metric is of particular interest for all use cases, where we must assume very cheap devices with reduced computational accuracy. Low OOB relaxes the required accuracy of the synchronization of the transmission in time and frequency. With (W)-CDMA we aim at an OOB that is not higher than in LTE. The evaluation will be done either analytically or by link level simulation.

- Spectral efficiency: As spectrum is expensive it has to be utilized as efficient as possible. LTE is optimized for broadband applications and provides several means to boost the spectral efficiency, e.g. MIMO or Coordinated Multi Point (CoMP). In contrast, LTE-M in general, and (W)-CDMA in particular, focuses on transmission modes with low complexity and cannot achieve a similarly high spectral efficiency. But it should not be much lower than in LTE. We will provide system simulation results for the spectral efficiency.
- User per cell capacity: We expect that for some key use cases (e.g. monitoring applications) the number of LTE-M enabled devices in one cell is much higher (working assumption: factor 10) than the number of legacy LTE UEs. Alternative radio access methods like (W)-CDMA have to cope with this requirement. The user per cell capacity is therefore a key metric in the evaluation of this proposal. The evaluation will be done analytically.
- Range and coverage: EXALTED is based on the assumption that existing network infrastructure will be reused, i.e. operators will exploit their already deployed LTE base station sites for LTE-M as well. If we suggest an alternative PHY, in this case (W)-CDMA, we must ensure that at least the same range and coverage as in LTE will be achieved, however with the constraint that we presume cheap devices with reduced computational accuracy. This basic requirement is equally valid for all use cases. The evaluation will be done with system simulations.
- Complexity: This is a key metric for the evaluation of (W)-CDMA. We aim at a simpler signal processing, which can be realized with cheap hardware components. Therefore, this metric is relevant for all use cases, where we assume cheap devices. We will assess the complexity analytically by comparing the required number of multiplications and expect it significantly lower than in LTE.

Detailed description:

Firstly, we will address the more conventional case in Figure 3-1a, where radio resources are utilized exclusively either for LTE or for LTE-M. The idea is to apply synchronous or asynchronous Code Division Multiple Access (CDMA) as LTE-M waveform. The principle of CDMA is that several signals sent on the same resources in time and frequency, e.g. one LTE-M RB in Figure 3-1a, but modulated with different orthogonal codes, can be separated by a specialised correlation receiver. Beyond time and frequency we gain with the code dimension one additional degree of freedom for the resource allocation. This principle is shown in Figure 3-5.

In the left figure, the code dimension is not yet exploited. Four LTE-M users are distributed over four LTE-M RBs. They may apply e.g. SC-FDMA in uplink and OFDMA in downlink, similarly as the legacy LTE UEs. In the right figure, the same LTE-M RBs are utilized, but in the shown example the signals are modulated with two different orthogonal codes. Obviously, the resource allocation gains more flexibility that can be exploited e.g. for the adaptation of the number of resources per LTE-M with respect to the amount of data they have to transmit or the code rate they need for a reliable connection. The simultaneous transmission from more than just one single LTE-M device per available LTE-M RB can help to reduce the mean delay in the communication. A detailed simulation study and theoretical analysis is ongoing. In particular the following questions are investigated:

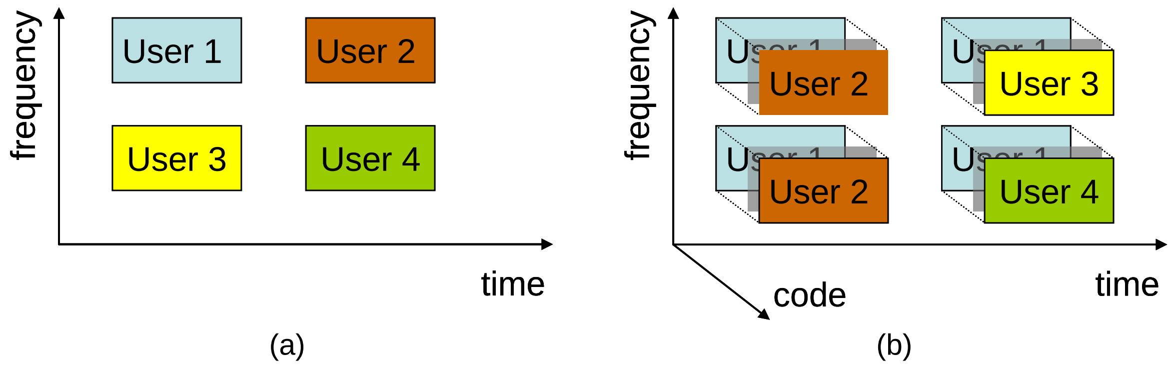

Figure 3-5: (a) Resource allocation in time and frequency in OFDMA / SC-FDMA and (b) a combined time-frequency-code allocation in LTE-M

- Is the number of active LTE-M devices that can be supported simultaneously with the CDMA approach higher than with OFDMA / SC-FDMA?
- Is it possible to minimize the mean PAPR, e.g. by selecting one particularly advantageous code sequence out of a reservoir of possible orthogonal codes for each device individually?
- Is it possible to increase the range with the CDMA approach that could help to save energy at the LTE-M device by reducing the transmit power?
- To which degree is the overall spectral efficiency affected?
- How does the system behave in non-ideal conditions, e.g. performance degradation due to cross talk between uplink LTE-M signals that are not well synchronized, and how sensitive is such a system?
- How does the LTE/LTE-M coexistence problem differ in uplink and downlink? Presumably the challenges are different.

A second study addresses the overlay system depicted in Figure 3-1b, where the radio resources for LTE and LTE-M are not orthogonal, i.e. they overlap in time and frequency. The basic idea is to spread the LTE-M waveform over the complete LTE frequency band by modulating the signal with a long code sequence, thus lowering the spectral power density to a level, where the LTE-M signals can be considered as weak additional noise from the perspective of an LTE receiver. A prerequisite for this approach is that the LTE-M device can be equipped with a cheap transceiver covering the complete band. The eNodeB operates with a low SNR decorrelation receiver to detect the LTE-M waveform in the received signal. As the waveform covers a relatively broad frequency range, it is referred to as Wideband CDMA (WCDMA). It was applied for UMTS. Figure 3-6 illustrates the principle of spreading. The transmitted power, represented by the area of the rectangles, is kept constant.

Several specialised code sequences exhibit particular capabilities with respect to flat spectrum, low PAPR, and robustness against missing synchronization. The study covers basically the same questions as mentioned above. Moreover the following specific topics are investigated:

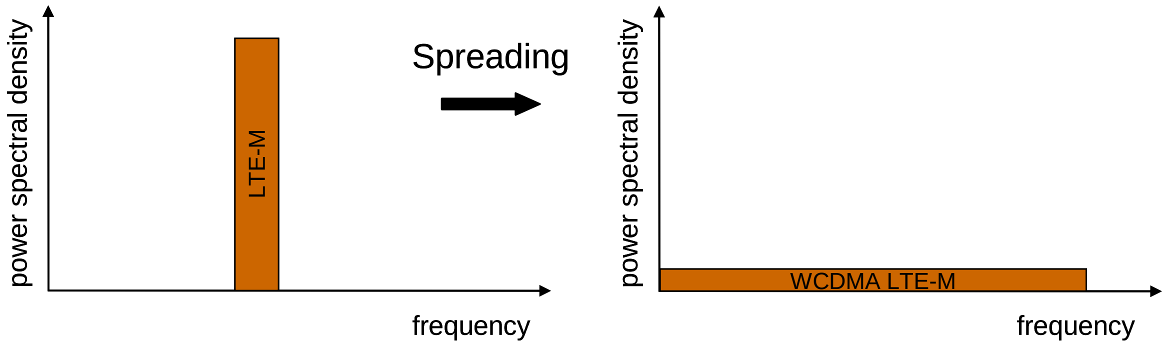

Figure 3-6: Power spectral density of the LTE-M waveform before and after the spreading

- Which code sequences are particularly adequate for M2M communications with respect to the EXALTED technical requirements [1]?

- To which degree is LTE performance affected due to interfering LTE-M signals?

Finally, both approaches shown in Figure 3-1 will be compared, and the respective advantages and disadvantages will be elaborated. Beyond a pure performance analysis, also implementation aspects have to be considered. Of particular interest is the question how easily an LTE-M transceiver can be included in existing base station hardware. Also the required modifications in the LTE specification are a crucial issue that can impact the recommendation for one or the other solution.

3.2 Antenna techniques

3.2.1 Directional antennas

<u>The proposal in a nutshell:</u>

We overcome the following shortcoming of the existing LTE specification (see section 2.2):

- Network overload issues: the algorithm aims at reducing the required bandwidth for communicating the feedback bits from the devices to the eNB.

The applicability of this algorithm is not restricted to a specific use case, since it satisfies the general requirement in M2M communications for relieving the network overload because of the large number of devices.

The evaluation and the comparison with the benchmark LTE will be done presumably with the following performance metrics (definitions see Appendix A3):

- Spectral efficiency (sum-rate): Number of successfully transmitted bits per time unit per frequency unit per cell in bit/s/Hz/cell. From the utilization of this algorithm, we expect a control over the trade-off between the sum-rate and the number of feedback bits that have to transmitted to the eNB. More specifically, the sum-rate is expected to be equal or less than that achieved with the LTE corresponding algorithm, and the number of feedback bits to be less than those required in LTE. The evaluation will be done via system-level simulations
- User per cell capacity**:** Maximal number of simultaneously active users per cell. Similarly, we expect a control over the trade-off between the user per cell capacity and the number of feedback bits that have to transmitted to the eNB. The evaluation will be done via system-level simulations.

Detailed description:

In LTE, downlink beamforming is supported resulting in a considerable improved the radio coverage. More specifically, with beamforming multiple users can be served at a time with reduced complexity, since each user stream is coded independently and multiplied by a beamforming weight vector for transmission through multiple antennas. The eNodeB generates a beam using an array of antenna elements and then applies the same precoding to both the data payload and the user equipment reference signal. In MU-MIMO transmit beamforming it is also necessary to include a feedback for the Channel State Information (CSI) that will be send back from the receiver to the transmitter. Such a CSI feedback can potentially incur excessive overhead due to the multiplicity of channel coefficients as well as the large number of devices that is considered in EXALTED. It is noted that in current LTE systems the feedback protocol that are employed are periodic, where a fixed number of bits in each reverse-link data block are allocated for CSI feedback. It becomes clear, that one important aspect of MU-MIMO precoding is the selection of UEs for transmission, since the throughput gain of naive linear precoding techniques very much depends on exploiting the multi-user diversity available in the system.

However a method for selecting UEs is not specified in LTE or other similar systems. Such a technique would be more than necessary in M2M applications, where the device density is expected to be considerably high. A quite general and very simple greedy strategy that has been found to be appropriate as a MU-MIMO precoding technique, consists of adding one UE at a time, as long as the additional UE increases the overall throughput, estimated according to a criterion such as the achievable sum-rate. This method gives the optimum performance with, however, increased complexity since that is proportional to the number of users. Next this method, which from now on will be denoted as algorithm 1, will be compare with the newly proposed one (Figure 3-7).

Due to the fact that the uplink and downlink take place on the same frequency in a TD-LTE system, the uplink sounding reference signals can be used directly to estimate the channel, which can then be used to derive the weighting for the downlink beamforming. A quite promising solution that has been also adopted in our analysis is to use multi-user diversity with zero-forcing beamforming (ZFBF) at the transmitter together with a selection algorithm that is based to an orthogonality criterion [8]. Furthermore, by using ZFBF the weight vectors are appropriate selected in order to avoid interference among user streams. Additionally, a user selection algorithm that is based on orthogonality, the semi-orthogonal user selection (SUS), and achieves near optimal sum rates, with considerable less complexity, is also included, and can be found in Figure 3-7. In this context, in the proposed system model the received signal at the kth device can be expressed as

$$\mathbf{y_k} = \mathbf{H_k x} + \mathbf{z_k}, \quad k = 1, \cdots, K$$

where $\mathbf{x} \in C^{Mx1}$ is the transmitted symbol, $\mathbf{h_k} \in C^{1xM}$ is the channel gain vector to the kth user, z_k is the AWGN at the kth user, while y_k is the received signal by the kth user. With respect to the downlink, we assume an error-free, zero delay feedback link from each node to the aggregation point. Here it should be noted that in ZFBF perfect channel state information is assumed to be available at the transmitter. However, this assumption is not always feasible in practical situations, where the feedback link has very tight throughput constraints. It is noted that with the proposed algorithm a noticeable reduction on the CSI that is send back to the BS is achieved, since only the users that satisfy a certain criterion, such that $||h|| \geq \beta$, where β is a threshold, report their CSI to the BS.

(a) Initialization : $i = 1,\ L_i = S_1, S_2 = \emptyset,$

(b) For $k \in L_i$, access point calculates

$$\mathbf{g_k} = \mathbf{h_k} - \sum_{j=1}^{i-1} \frac{\mathbf{h_k}\mathbf{g_{(j)}}}{||\mathbf{g_{(j)}}||^2}\mathbf{g_{(j)}}$$

For $i = 1 \rightarrow \mathbf{g_k} = \mathbf{h_k}$

(c) User is selected according to:

$$\pi(i) = \arg\max_{k \in L_i} ||\mathbf{g_k}||, \quad S_2 \leftarrow S_2 \cup \{\pi(i)\}$$

$$\mathbf{h_{(i)}} = \mathbf{h_{\pi(i)}},\ \mathbf{g_{(i)}} = \mathbf{g_{\pi(i)}}$$

(d) L_{i+1} is recalculated according to

$$L_{i+1} = \left\{ k \in L_i,\ k \neq \pi(i) | \frac{|\mathbf{h_k}\mathbf{g_{(i)}}^*|}{||\mathbf{h_k}||||\mathbf{g_{(i)}}||} < \alpha \right\}$$

where α is a small positive constant.

Figure 3-7: The selection algorithm.

Table 3-1: Requirements addressed with directional antennas techniques.

Requirement	Solution description	LTE Solution
FU.1	The efficient support of a large number of devices in a MU-MIMO scheme cell requires a limited feedback link that takes also into consideration the fact that the served users are much less than the active.	The existing LTE uplink feedback signalling uses two different mechanisms for reporting the required information, one for limited amount of info and one when large amount is required. However, a two stage feedback algorithm is not taken into consideration
SV.1	QoS criteria as well as priority issues have been taken into consideration for the feedback signalling selection.	See FU.1
NT.2 and NT.3	The proposed algorithm complies with the existing LTE signalling specifications and no modifications are required.	N/A
NT.17	This two stage feedback algorithm will significantly help to reduce the signalling	There are mechanisms for reducing the signalling in the feedback link but not so adaptive.
NF.2	Reduced feedback will save the transmitting power of the nodes	N/A

3.2.2 MIMO for M2M

The proposal in a nutshell:

We overcome the following shortcoming of the existing LTE specification (see section 2.2):

- Device cost issues: More specifically, the algorithm aims at reducing the number of CSI estimations at the receiver (eNB), by replacing the CSI estimator with a simple energy detector at each antenna element. Moreover, the number of required RF chains at the receiver is expected to be reduced, by selecting those antennas, which will utilize them.

The applicability of this algorithm is not restricted to a specific use case, since it satisfies the general requirement in M2M communications for reducing the complexity of devices.

The evaluation and the comparison with the benchmark LTE will be done presumably with the following performance metrics (definitions see Appendix A3):

- BER: Bit error rate at the output of the decoder. The BER is expected to be comparable with that achieved with the conventional method, i.e. when the antenna selection is based on the CSI. The evaluation will be done via link-level simulations.
- Number of CSI estimation: the utilization of this algorithm aims at reducing the number of CSI estimations required for supporting MIMO in LTE-M. The evaluation will be done via link-level simulations.
- Number of active antennas: the expected number of active antennas is expected to reduce. The evaluation will be done via link-level simulations.

Detailed description:

In LTE and LTE-A systems, MIMO configurations are supported for both downlink and uplink connections, with two different modes, i.e. spatial multiplexing and transmit diversity. MIMO configuration provides a more efficient usage of available spectrum and offers higher data rates.

In this subsection, a MIMO configuration for transmit diversity is presented, with 2 transmit and M>=1 receive antennas. This technique can be applied to both the uplink and the downlink, depending on the complexity constraints of network elements involved. The basic feature of this transceiver is that in contrast to the conventional one, it does not require to estimate the instantaneous channel gain for each of the receive antennas, but only for those that are eventually selected for the data decoding. In this way some bits necessary for performing channel estimation, as well as computational resources can be saved. Moreover, in addition to previous receiver, in order to mitigate the effect of excessive switching at the receiver, which may lead to synchronization problems and larger delays, an adaptive receive filter can be applied, according to the measured Doppler spread.

The proposed uplink scheduler addresses the following requirements as described in [1].

Table 3-2: Requirements addressed with MIMO techniques

Requirement	Solution description	LTE Solution
NT.2 and NT.3	The proposed algorithm complies with the existing LTE signalling specifications and requires no modifications.	N/A
NF.2	Due to less channel	MIMO in LTE requires full channel

	estimations at the receiver side, energy savings are possible.	estimation.

In more details, in the signal-plus-noise-based MIMO scheme, the antennas that participate at the detection stage are determined according to the energy of the received signal. In other words, only those L antennas are selected, with the highest sum-of-amplitudes of the received signal on the jth antenna, i.e.

$$s_{Cj} = |r_{j,n}| + |r_{j,n+1}| \tag{3-6}$$

where $r_{j,n}, r_{j,n+1}$ are the received signal enveloped at the receiver. In this way unnecessary signal-to-noise ratio estimations can be avoided. In Figure 3-8, the transceiver's structure is depicted. At the transmitter a conventional Alamouti space-time block code is utilized. At the receiver, the signal-plus-noise (S+N) at each antenna is measured, in order to determine those antennas that will be selected for the decoding stage. After the antenna selection, the CSI at each selected antenna is estimated, which will be used for the decoding and combining stage.

The receiver however, can switch between the two selection methods, (signal-to-noise ratio or signal-plus-noise) according to the measured Doppler spread and some predetermined performance results for particular scenarios. The normalized switching rate (SWR) of the two antenna selection schemes versus the SNR, for various values of the normalized bandwidth of the band-pass filter at the receiver is depicted in Figure 3-9. The normalized SWR of the S+N-based selection scheme decrease as the ratio of the bandwidth of the receiver filter to the maximum Doppler frequency decreases. The SWR of the SNR-based scheme is independent of the filter's bandwidth, since we assume that the estimation of the fading amplitudes for branch selection does not depend on the noise process. Again, it is observed that larger number of available antennas results in higher switching rates, for any value of the filter's bandwidth. However, the most important result is that the SWR of the S+N-based scheme can be either greater or less compared to that of the SNR-based scheme, depending on the ratio bandwidth/Doppler spread. This result indicates that an adaptive receiver filter could be utilized to avoid excessive switching.

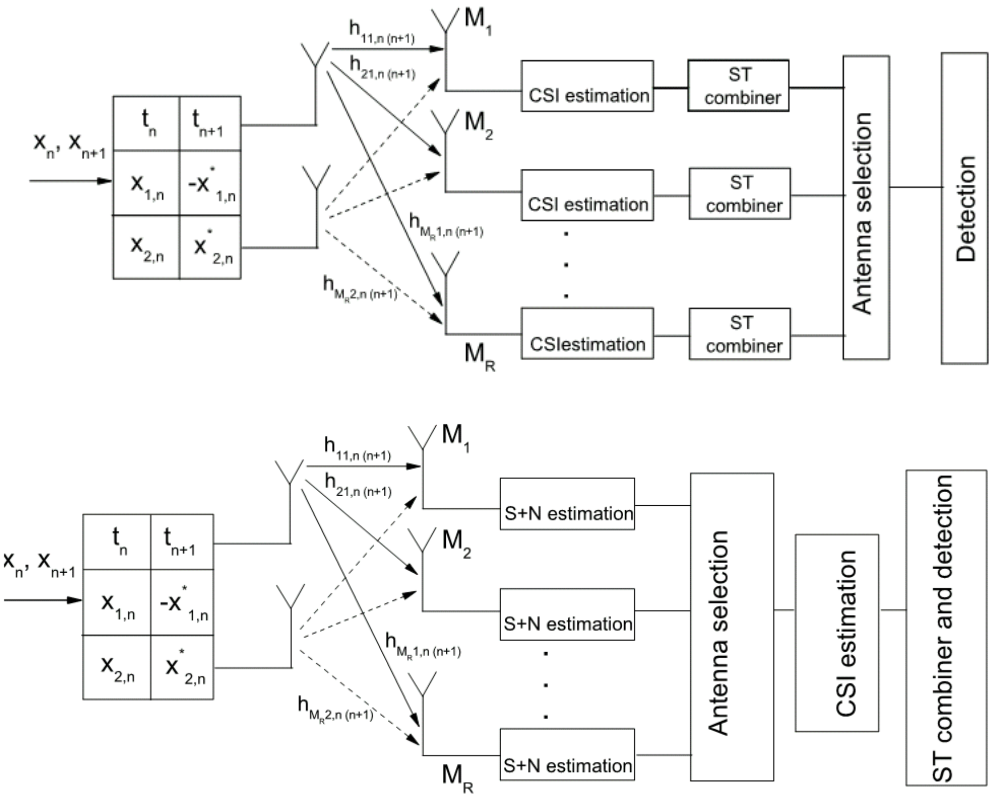

Figure 3-8: CSI and S+N-based MIMO transceivers

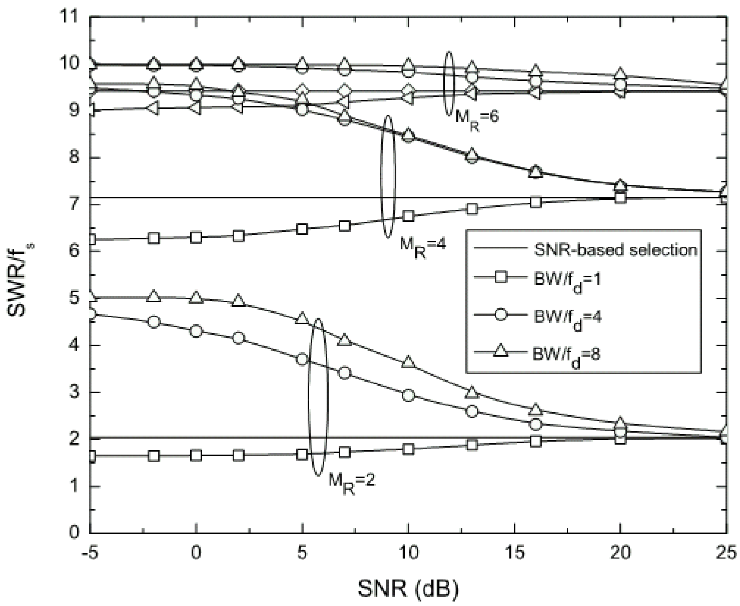

Figure 3-9: Performance indicators for determining the MIMO mode

4. LTE-M Medium Access Control (MAC)-Layer

Efficient MAC-Layer protocols are expected to play a key role in LTE-M, which is necessary to support a significantly larger number of devices, compared to the existing cellular networks. The reduction of signalling and stricter scheduling algorithms, compared to LTE, are indispensable in order to support the large number of M2M devices, while not jeopardizing the service of the LTE users. Moreover, improved retransmission schemes are needed, which shall take into account the requirements of M2M networks. The algorithms and techniques described in this section aim at enhancing LTE MAC-Layer towards supporting M2M communications.

4.1 Scheduling

The scheduler is a functional unit in the base station that decides about the usage of radio resources in both uplink and downlink. The available resources are distributed among UEs and LTE-M devices located in one cell, and among the different bearers of each UE or LTE-M device. The details of the scheduling algorithms are left to the implementation of the base station, but the required signalling is standardised. In the following we describe how this signalling could be reduced in LTE-M and present a cross layer optimization.

4.1.1 Reduction of signalling

<u>The proposal in a nutshell:</u>

We plan to overcome the following shortcomings of the existing LTE specification (see section 2.2):

- Small data transmission, in particular with respect to signalling overhead reduction
- Network overload issues.

The evaluation and the comparison with the benchmark LTE will be done with the following performance metrics (definitions see Appendix A3):

- Access delay: With this proposal, we primarily address use cases with relaxed access delay constraints such as smart metering and environment monitoring. We will compare and evaluate the access delay for various traffic load situations (number of active users per cell, average message lengths and idle times) either analytically or with a simplified system simulation chain. We expect that the access delay will become only slightly higher than in LTE.
- User per cell capacity: This is one of the key metrics of this proposal. It is primarily important for all use cases with a much higher number of LTE-M enabled devices per cell compared to legacy LTE UEs (working assumption: factor 10), e.g. environment monitoring. Expected outcome is that our approach can avoid network overload situations, even if a lot of LTE-M devices start transmitting simultaneously in an event-driven application. The evaluation will be done by system level simulations.
- PHY control channel and pilot overhead: This metric is of particular interest for all use cases with very short messages and sparse transmissions, but with the assumption of a very high number of devices per cell showing this behaviour, e.g. smart metering. The control channel overhead of our proposal will be compared with LTE analytically, and we expect a significant improvement.

<u>Detailed description:</u>

In contrast to high broadband services like file transfer, where the additional exchange of control information is negligible, it captures a relatively big fraction of the overall amount of data in low rate applications in general and many types of M2M communication in particular. Hence, the reduction of control information is an efficient mean to decrease the system load, particularly if a large number of devices requests access. As a consequence, the number of supported devices increases, or in return, the network capacity is exploited more efficiently. Moreover, the energy consumption per device shrinks significantly. These are main technical requirements addressed in this section, where we focus on backward compatible approaches aiming at a reduction of the signalling overhead in LTE-M.

The number of LTE-M devices in one cell can be very large (in the order of thousands). Although these devices may belong to several independent groups, a worst case can happen, where a lot of devices perform access to the network at about the same time. The sudden surge of M2M traffic and respective exchange of signalling information may cause traffic congestion in the radio access network and even in the core network. Moreover, the limited capacity of the RACH in a fully-loaded cell must be taken into account. The RACH capacity is further reduced by inter-cell interference and non-ideal intra-cell detection. The access failure rate would become very high in this situation.

The LTE-M MAC layer must ensure to avoid this situation and to ensure unrestricted operation of legacy UEs. The fact that most M2M applications simply require automated short data reporting either triggered by timer or event in a medium- or long-term period can be exploited to define an appropriate solution for both LTE-M devices and LTE UEs. Slotted access is a very simple solution for this. The basic principle is shown in Figure 4-1. It is following the existing framework of UE terminated access after the UE is paged. The M2M application is allowed to trigger the LTE-M device performing access during its paging slot. The method forces the arrival distribution being flat over the paging cycle. This is a pushing approach, i.e. Physical Downlink Control Channel (PDCCH) signalling information that would be required for pulling a large number of devices, but may be already fully occupied by prioritized legacy LTE connections, can be avoided. However, in the case of traffic congestion or for high priority M2M applications, the pulling method could be applied as an additional option and complement slotted access. The expected performance impact on legacy traffic is small.

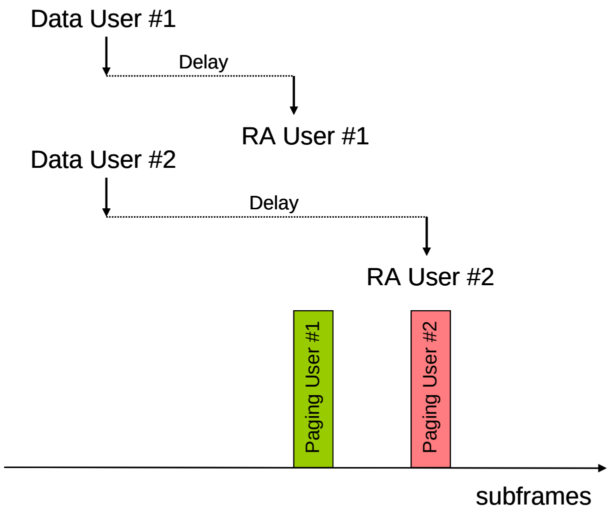

Figure 4-1: Slotted access

Another option to reduce signalling overhead is utilizing the persistent scheduling configuration instead of the usual dynamic scheduling. In this operation mode requested radio resources are semi-statically assigned to LTE-M devices. In contrast to dynamic scheduling, respective information exchange on the PDCCH on a per sub-frame basis is not needed. This configuration is helpful if the amount of data to be transmitted and the respective latency constraints are known a priori, or if the data is more or less equally distributed over time, thus assuming a constant data rate. At least one of these prerequisites should be valid for many M2M services. Hence, this simple method could be applied quite widely. The persistent scheduling configuration is already available in LTE, however designed for broadband services. We investigate means aiming at further improvements with respect to low data rate M2M traffic, in particular energy saving at the LTE-M device by letting it sleep as long as possible without the obligation to decode control information.

4.1.2 Cross-layer optimization

<u>The proposal in a nutshell:</u>

We overcome the following shortcoming of the existing LTE specification (see section 2.2):

- Radio access: More specifically, the algorithm aims at increasing the number of served devices per cell, by taking into account the different delay QoS of M2M devices in contrast to LTE, where the number of delay QoS is restricted.
- Network overload issues: the utilization of the algorithm reduces the probability that the network is overloaded, in the case that a large number of devices requests access.

The applicability of this algorithm is not restricted to a specific use case, since it satisfies the general requirement in M2M communications for increasing the number of served devices per cell.

The evaluation and the comparison with the benchmark LTE will be done presumably with the following performance metrics (definitions see Appendix A3):

- BER: Bit error rate at the output of the decoder. Here the average BER for all devices is calculated. The goal is to achieve comparable BER with that of the LTE scheduler. The evaluation will be done via system-level simulations.
- Percentage of satisfied users: the utilization of this algorithm aims at increasing the number of served devices per cell. The evaluation will be done via system-level simulations.

<u>Detailed description:</u>

The scheduler in the LTE system is considered as a key element, which is responsible for the efficient resource sharing among the active UEs/Devices, on the shared data channel. For scenarios where a large number of devices needs to transmit data to the eNodeB (e.g. EXALTED scenarios), the uplink scheduler plays an essential role, since an inefficient scheduling can easily result in a high percentage of unsatisfied users/devices, in terms of different QoS criteria, such as throughput or maximum tolerable packet-delivery delay. A particular feature of every LTE uplink scheduler (as well as of the one proposed in this subsection) is that the assigned RBs to a user must be adjacent in frequency, in order to minimize the PAPR.

In the following, we describe the basic principles of uplink scheduler, designed to serve M2M devices with a significant diversity in QoS criteria in terms of throughput and tolerable packet-delivery delay. The main feature of the proposed uplink scheduler is that it takes advantage of the channel quality information and the delay tolerance for each user/device, while not requiring the information of different user/devices classes. Note that the channel quality information is fed to the scheduler, with the same way as in any conventional LTE scheduler, without needing any extra signalling. The algorithm determines not only the specific RBs assigned to each user/device but also the number of RBs per TTI per user, taking into account the requested (or expected) RBs within a specific time interval.

The proposed uplink scheduler addresses the following requirements as described in [1].

Table 4-1: Requirements addressed with cross-layer optimization

Requirement	Solution description	LTE Solution
FU.1	Supporting a large number of M2M devices in a cell requires that the Uplink scheduling takes advantage of the QoS requirements for each device. For example, skipping serving devices with high delay tolerance (e.g. minutes or hours), decreases the possibility for traffic congestion and outages.	Because of the currently limited number of users/devices and diversity of services (e.g. Voice over IP (VoIP), internet browsing), the existing LTE uplink schedulers (note that the scheduling algorithms are not part of the LTE specifications ant it is up to the manufacturer to implement any signalling-compatible algorithm), are not designed for serving devices with tolerable delays varying from some ms to several hours.
FU.3	The proposed algorithm takes into account the diverse M2M classes.	See FU.1.
SV.1	Enables Classification and prioritization of services is necessary in M2M networks.	See FU.1
NT.2 and NT.3	The proposed algorithm complies with the existing LTE signalling specifications and requires no modifications.	N/A
NF.4	The proposed algorithm supports real-time constraints.	LTE also supports real-time constraints but not M2M-type ones (e.g. simultaneous low-rate transmissions with diverse delay constraints).

In more details, each device, which can be uniquely identified in the network, sends a request to the eNodeB (or the Gateway) for packet transmission, along with its currently QoS constraints, e.g. the total number of requested RBs and the corresponding maximum tolerable delay it can experience (in TTI time) for transmitting the corresponding data packets. These parameters from all the users/devices in the cell can be stored in the eNodeB/Gateway memory. This information is stored for each device as long as it has not sent all its packets. If some devices ask for additional RBs, this information is updated accordingly. Then the scheduler starts assigning RBs to those users/devices having the best Channel Gain (CG) (increasing the total throughput), but only if the corresponding tolerable delay is lower than a predefined threshold. In this way, users/devices that can be served in subsequent TTIs can be skipped, giving priority to other users/devices with more stringent delay constraints.

The main differentiation from existing LTE uplink schedulers [9], [10] is that it is not necessary to form different classes of devices, which leads to suboptimum solutions, since the actual QoS constraints (e.g. delay) may vary for each user/device within that class (e.g. a class including devices with delay varying from 1 ms to 10 ms). However, sending more detailed values for the delay constraint requires different feedback size.

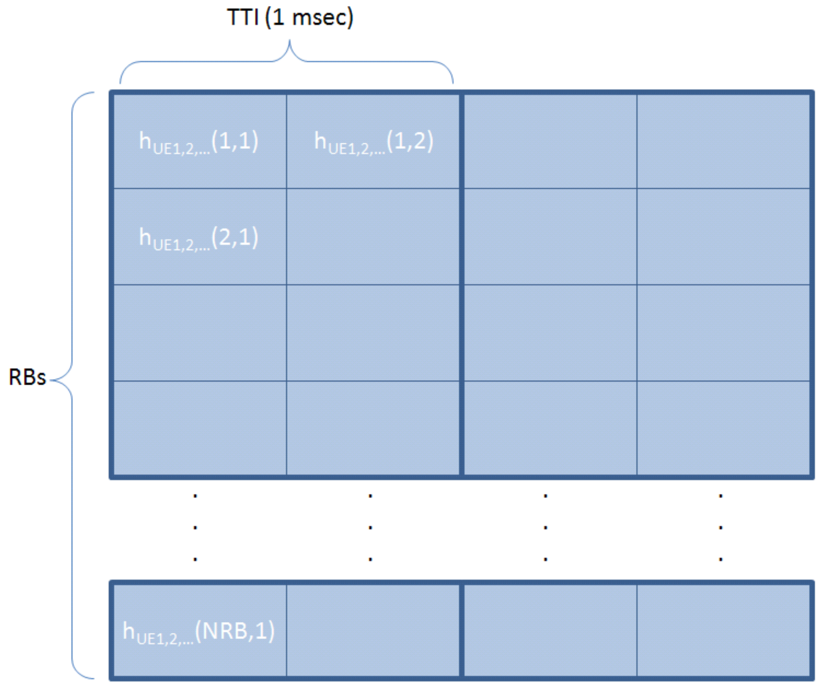

Figure 4-2: Main assumptions of the proposed Uplink Scheduler

The proposed algorithm takes advantage the exact real-time constraints for each user/device. The element that performs the scheduling requires the following information (Figure 4-2):

1. The channel gains $h_{UE1,2,\ldots(NRB,t)}$ at each RB for each user. This information is needed for calculating the expected BER for each user/device.
2. The matrix that includes the initially requested RBs by each user/device. This matrix is updated as soon as new users/devices request transmissions, or previous users are served. For 10 k devices with maximum delay of 24 h approximately 160 kB memory is required.
3. The matrix including the remaining number of RBs for each user/device. The matrix is updated as soon a RB is assigned to a user.
4. The matrix including the maximum tolerable delay per user in terms of TTI time.

The utilization of the proposed algorithm increases the percentage of satisfied users/devices, i.e. those meeting their QoS constraints (Figure 4-3. No modifications are needed to the current LTE-System, since the same signalling is required as in the usual LTE uplink schedulers. The next steps include the development of a new algorithm that puts more weight on the delay constraint rather than on the channel quality of each user, increasing in this way significantly the percentage of satisfied users. Moreover, efforts are put on decreasing the feedback required for informing the eNodeB about the delay constraints.

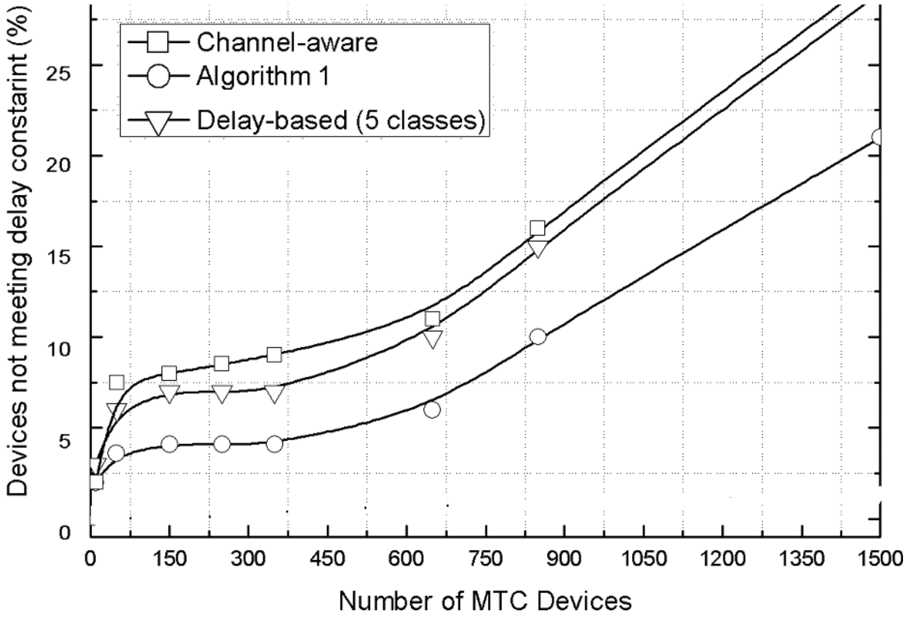

Figure 4-3: The performance of the proposed algorithm (Algorithm 1)

4.2 Retransmission schemes

Forward error correction does not guarantee error-free transmissions in the event of deep channel fades or high interference levels when several devices are transmitting simultaneously. Therefore, an ARQ mechanism is necessary in order to ensure the correct delivery of a message. In this section we discuss two mechanisms. First, a HARQ scheme that deals with decoding errors by requesting additional redundancy. Second, a mechanism to recover packets involved in a collision in a random access scheme is proposed. In both cases, high energy efficiency is sought by taking advantage of the original transmission in the decoding process.

4.2.1 HARQ

<u>The proposal in a nutshell:</u>

We overcome the following shortcoming of the existing LTE specification (see section 2.2):

- Small data transmission, in particular with respect to transmit short messages with very efficient resource usage.

The evaluation and the comparison with the benchmark LTE will be done presumably with the metric of spectrum efficiency in terms of the payload to signalling overhead ratio. It is valid equally for all use cases. LTE is not optimised for transmitting short messages with acceptable payload to signalling overhead ratio, and it is expected that the proposed HARQ scheme can largely improve the ratio. Computer simulations will be carried out to demonstrate the spectral efficiency.

<u>Detailed description:</u>

HARQ is a combination of Forward Error-Correcting (FEC) coding and ARQ error detection method and it can be implemented for UL in the LTE system. The basic principle for HARQ is to reduce the number of retransmission by introducing redundancy to the data to be transmitted. Two types of HARQ protocols have been proposed. Type I HARQ is the simplest HARQ. In this scheme, if the received data cannot be correctly decoded, the receiver discards the packet and requests a retransmission of the packet until the packet can be correctly decoded or a pre-set number of retransmission is reached; also the retransmissions take place at either the same or lower code rate. In Type II HARQ, incremental redundancy is introduced. This scheme is to transmit additional redundant information (i.e. FEC parity bits) in each retransmission if the first received packet cannot be correctly decoded, as shown in Figure 4-4, only the shadowed parts are transmitted on retransmissions. Since retransmissions are activated when necessary, HARQ is quite effective in improving spectral efficiency comparing to using only the FEC at the physical layer.

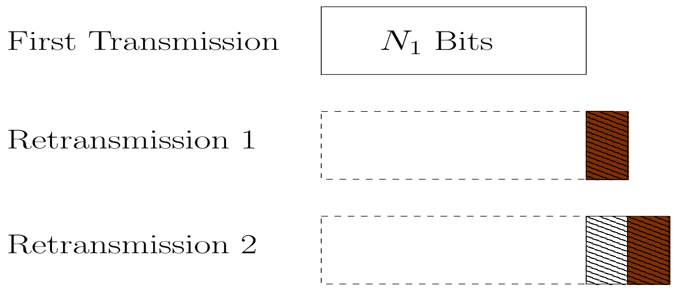

Figure 4-4: Type II HARQ Transmissions

Due to a delay constraint imposed by the service, the design of HARQ is important to provide reliability and spectral efficiency. Although Type I HARQ is easy to implement with lower complexity, this scheme is not efficient since every retransmission carries the whole code word. Because the devices in LTE-M are sensors with limited power, Type I HARQ scheme won't be suitable for this scenario. With Type II HARQ, only the redundant information is transmitted on the retransmission. However this scheme requires a buffer at the receiver for soft combining the original packet and the incremental parity packet. In EXALTED system, the receiver could be an eNodeB, and then the hardware limitation is not a big problem for implantation.

Since the improvement in spectral efficiency decreases as the maximum number of retransmissions increases, a limited number of retransmission is essential to provide a good delay-throughput trade-off which will be investigated in the future work. Also to further improve the spectral efficiency, AMC schemes can be employed jointly with HARQ to reduce the number of retransmissions, the objective is to overcome the power limitation of the device in LTE M2M network and achieve a good trade-off between the power consumed and the quality of service. AMC and successive decoding are promising techniques for this objective and will be investigated in the future work. With the aid of the channel information feedback, the transmitter can estimate the probability of correct decoding at the receiver with corresponding coding and modulation schemes. LTE utilizes Incremental Redundancy HARQ with a 1/3 turbo encoder used for forward error correction. The Transport Block (TB) Cyclic Redundancy Check (CRC) is used to detect errors. The receiver only receives different punctured versions of the same turbo-encoded data; each of these retransmissions is self-decidable. Considering the power limitation of the device in M2M communication, we should reduce the power for retransmission with more efficient coding schemes i.e. fountain code. Compared with the HARQ scheme already in LTE, the HARQ for EXALTED is more power efficient at the mean time maintains the quality of service. The computer simulations and performance analysis will be carried out in the future.

4.2.2 Multiple access with collision recovery

The proposal in a nutshell:

Certain types of M2M applications generate a large number of short, unscheduled messages. As an example, monitoring applications responding to sudden events will generate unscheduled messages. Random access is an interesting solution for this type of traffic since it has minimal overhead. However, throughput and energy efficiency of random access suffer when there are collisions on the radio channel. A multiple access scheme with collision recovery can have the advantages of random access without suffering an excessive penalty due to collisions. This scheme is particularly relevant if one takes into account that M2M traffic is prone to generate large numbers of collisions. For example, if an event triggers a widespread alarm (e.g. a power outage triggering an alarm from all smart meters in a large area), it is likely that all devices will attempt a transmission almost simultaneously to report it. If an efficient collision recovery mechanism is in place, such transmissions may be successful, increasing the resiliency of the system.

In our proposal, devices transmit unscheduled packets following a slotted random access scheme. A carefully designed precoding scheme, together with collision recovery decoding, allows partial decoding of multiple packet collisions. Then, with partial retransmissions, all packets may be decoded. We overcome the following shortcomings of the existing LTE specification (see section 2.2):

- Small data transmission in M2M: The proposed scheme is expected to improve the performance of random access schemes for short message transmissions, both in terms of throughput as well as energy consumption.
- Radio access: Efficient random access schemes require much less signalling than a connection-oriented communication. Since a large fraction of M2M traffic is expected to consist of short messages, the amount of required signalling can be reduced.

The evaluation and the comparison with the benchmark LTE will be done presumably with the following performance metrics (definitions see Appendix A3):

- Packet Loss Rate: This proposal seeks to reduce the number of packets lost to collisions and in turn improve energy efficiency of the devices by reducing the number of required retransmissions. Such a mechanism represents an improvement with respect to random access as implemented in LTE, which lacks the collision recovery feature. The degree of improvement in packet loss rate will be evaluated in upcoming deliverables.
- Throughput: Reduced number of retransmissions will improve the throughput with respect to random access as implemented in LTE, which will be evaluated in upcoming deliverables.
- Access delay: It is expected that average delay performance of this scheme is comparable or better to delay performance of random access as implemented in LTE.

Detailed description:

The main advantage of the proposed scheme is the support of a large number of devices, since it provides a solution to collisions, and energy- and spectrum-efficient operation. Requirements FU.1, FU.2, NT.10, NF.1 [1], are therefore supported in terms of large number of devices. Requirement DV.3 is addressed in terms of energy efficiency. The proposed scheme also results in reduced signalling of the multiple access phase (Requirement NT.17).

The most widespread random access mechanisms, namely Aloha and Carrier Sense Multiple Access (CSMA), are not energy-efficient when a large number of devices are attempting to access the medium. In Aloha systems, whenever a collision occurs, packets are lost and all devices need to retransmit. If the offered traffic is high, the average number of transmission attempts per successful transmission can be quite large, and the throughput diminishes to zero. This represents a considerable penalty in energy per packet with respect to scheduled access. CSMA represents an improvement since each terminal senses the medium before transmitting and if the medium is busy it does not transmit. However, CSMA is not very adequate for LTE-M system because it becomes more effective when transmitted packets are long, which is not typically the case for device communications, and because it suffers from the hidden node problem, which causes collisions when devices are not able to sense another device transmitting. Since the LTE-M uplink is highly asymmetrical, with devices placed at very low elevations and communicating with a BS tower, it is expected that the hidden node problem will occur very frequently. Moreover, in case of high traffic volumes, CSMA requires long listening periods, which is very wasteful in terms of energy.

Our approach is based on a collision recovery scheme for Aloha-like access. When needed, devices access the channel to transmit a packet. In case a collision occurs, the eNodeB attempts to decode the XOR sum of packets. In this scheme we are not trying to recover each individual packet, which would require successive or parallel interference cancelation schemes and a considerably higher SNR, but the XOR operation on the transmitted encoded bits. The eNodeB needs to know which devices are transmitting, therefore we assume that packet headers are CDMA-encoded, and that the eNodeB is able to decode the headers even in case of a collision. Note that encoding the message body with CDMA, which might also resolve collisions, would require much longer messages due to the CDMA processing gain. Therefore, we assume that CDMA encoding is restricted to small packet headers only, and that the processing gain obtained is sufficient to recover all headers even in the event of moderate and large collisions. In order to prevent two devices from picking the same spreading code, pseudo-random sequences with random offset can be used. The following example illustrates the idea for the case of Binary Phase Shift Keying (BPSK) transmission without channel coding[1]. Consider the following mapping:

b_i	x_i
0	1
1	-1

The receiver tries to decode b_1 XOR b_2 from the received symbol, which is given by

$$y=h_1x_1+h_2x_2+n \quad \textbf{(4-1)}$$

where h_i are the channel coefficient and n represents Gaussian noise. The corresponding mapping is given in Table 4-2:

Table 4-2: Symbol mapping at the receiver

b_1	b_2	b_1 XOR b_2	Y
0	0	0	$h_1 + h_2$
0	1	1	$h_1 - h_2$
1	0	1	$-h_1 + h_2$
1	1	0	$-h_1 - h_2$

[1] The same principle would work for higher order constellations but, for the applications envisioned in EXALTED, BPSK is more appropriate due to lower error rate and better performance in case of amplifier non-linearity.

Assume, for simplicity, that h_1=h_2=1. The maximum likelihood receiver will decode 1 for the shaded decision region and 0 for the whitened decision regions shown in Figure 4-5.

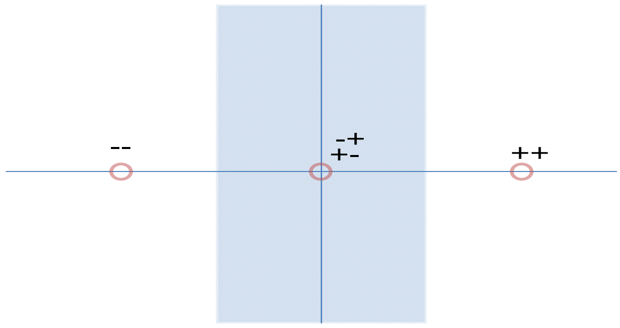

Figure 4-5: Example of decoding regions for collision resolution scheme.

Figure 4-6 shows the decoding performance of the XOR of collided packets for different collision sizes. In the simulation, an LDPC code was used, and the equivalent log-likelihood ratios for sets of bits resulting in the same XOR value were derived. As it can be seen, even large collisions can be recovered with less than 3 dB loss with respect to a single packet transmission. An alternative approach to resolve collisions consists in decoding both packets by means of successive interference cancellation. However, the resulting error rate is significantly higher than in the proposed approach, since more information needs to be extracted from the incoming signal.

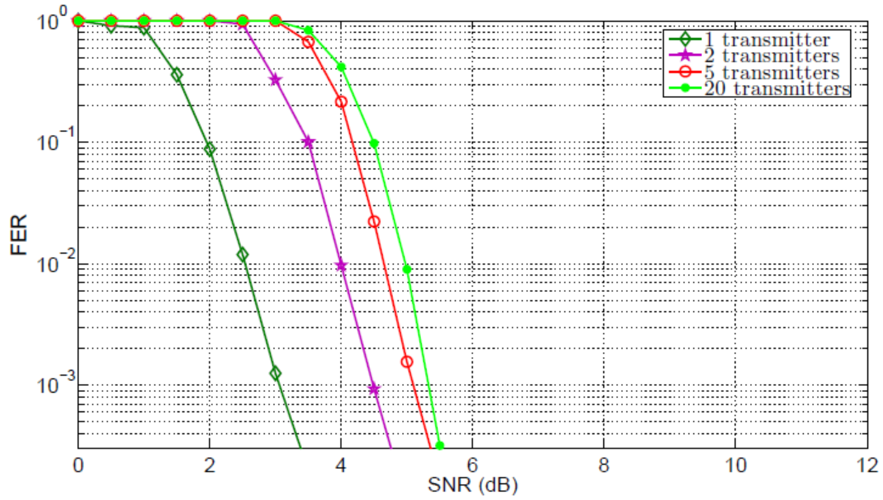

Figure 4-6: Frame (packet) error rate for the XOR of received packets as a function of the SNR. Several collision sizes are compared to a single packet transmission.

Recovered packets are stored and represent linear combinations of transmitted packets over Galois Field (GF2). The eNodeB then sends a negative acknowledgment to collided devices, which will attempt to retransmit the packet with probability q<1 on successive time slots, until the collision is resolved. As a result of successive retransmissions, a sufficient number of linearly independent combinations of packets will accumulate at the eNodeB receiver, which can then decode the packets solving a set of simultaneous linear equations in GF2, in a process that is equivalent to decoding of a network or fountain code. Encoding vectors are given by the CDMA-encoded packet headers.

The retransmission probability can be optimized with regards to throughput and energy consumption. In order to maximize throughput the retransmission probability must maximize the probability that received collisions are linearly independent, i.e., they involve different sets of packets. Setting retransmission probabilities close to one or zero results in higher probability of linearly dependent combinations of packets, since vectors with high number of zeros or ones become more likely. On the other hand, if we aim at minimizing energy consumption then the retransmission probability should be set close to zero to ensure that a small number of retransmissions occur, while achieving good decoding probability.

A preliminary performance evaluation is shown in Figure 4-7. The figure shows the performance of the proposed scheme for M=20 and M=50 users. In the options analyzed so far, a collision triggers a collision recovery phase (CRP) which ends when the collision is resolved. During the CRP, only nodes involved in the collision retransmit packets. The CRP can be of fixed size if the eNB estimates the number of slots necessary to resolve it and transmits this number to the devices. On the other hand, the CRP can be of variable size if the eNB signals the end of the CRP once the collision is resolved. The two methods have implications in terms of energy efficiency and delay that must be analyzed and compared.

In the figure, Φ stands for normalized throughput (number of successfully decoded packets per slot), λ stands for normalized offered load (number of new packet transmission attempts per slot), and M is the number of devices attempting access to the channel. As it can be seen, the variable CRP size scheme achieves higher throughput at the expense of more control signalling. Details on the proposed scheme, evaluation methodology, scalability and extensive evaluation results will be provided in upcoming deliverables.

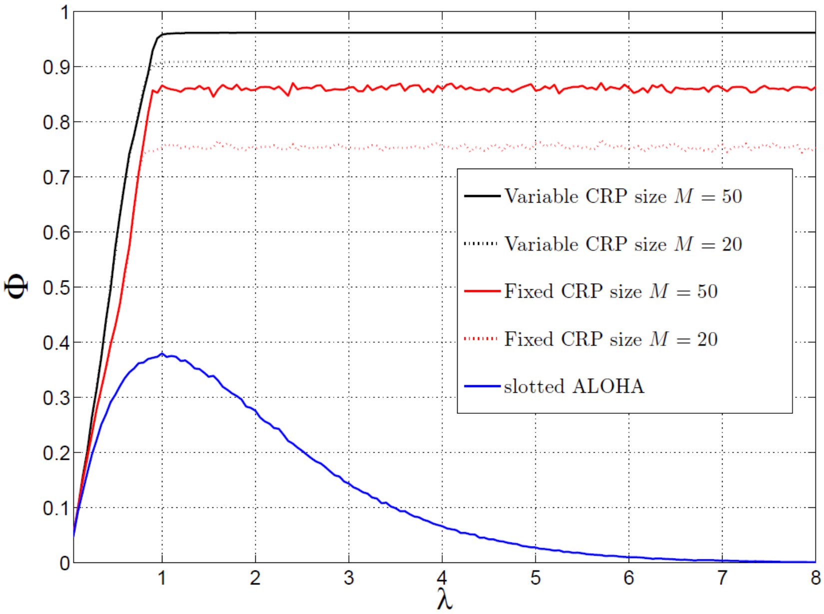

Figure 4-7. Normalized throughput of the proposed collision recovery multiple access scheme for Poisson traffic arrivals. Slotted Aloha is shown for comparison.

As a final remark, the proposed random access scheme assumes symbol-synchronous channel access by the LTE-M devices. This can be achieved if the devices listen periodically to synchronization beacons. However, it may be desirable that devices transmit directly upon wakeup in order to save energy. To that end, asynchronous mechanisms for collision resolution will be addressed in the future. Moreover, system-level simulations to test the robustness, efficiency and scalability of the proposed scheme will be carried out.

5. LTE-M Radio Resource Control (RRC)-Layer

Mobile networks are being increasingly used for M2M applications. These applications primarily collect information from a number of sensors attached to a mobile network and transfer them through to a M2M operator which then stores the received data and makes it available to the M2M application users. Today, these M2M devices that collect information are treated in the same manner as other mobile subscribers (human), although the way of working and communication patterns these two groups rely on are different (human mobile subscribers can at any time make a connection to any remote party, while M2M devices communicate with predetermined remote parties and mainly at predefined times). Cellular air interface such as LTE and 3G are not optimized for low device power consumption, and low resource utilization for short transactions.

In this chapter focus is on the structure and performance of LTE access part. After identifying the most challenges part in the access network, proposals for optimizations are presented. Different approaches for improvements of the RRC Layer are discussed, mainly focussing on a reduction of signalling overhead and power consumption at the LTE-M device. This is complemented by a proposal for an energy efficient relaying scheme.

From the radio access point of view the following issues can be considered as the most important:

- Device power consumption
- Radio resources
- Network energy consumption
- Network hardware resources

Generally, there are a few areas where improvements can be achieved: one concerning the behaviour during inactive periods, one concerning the schemes used for short, infrequent transactions, e.g. for sensor type devices and one describing a way to deal with extreme scenarios where the C-RNTI space is depleted.

5.1 Optimized RRC protocol

<u>The proposal in a nutshell:</u>

We plan to overcome the following shortcomings of the existing LTE specification (see section 2.2):

- Support for low mobility UEs, in particular with respect to LTE air interface which is not optimized for low resource utilization for short transactions, and for low device power consumption. The M2M devices in LTE are treated in the same manner as other mobile subscribers (humans, high-end devices, etc.), although their way of working and communication patterns are quite different. Full mobility support is not necessary when devices are on fixed locations and this can be used to simplify certain network procedures.
- Paging of M2M devices when these are in the IDLE mode and monitoring paging channel, takes up a substantial amount of time and consumes a significant amount of energy which is not necessary. For mobile phones this is necessary to ensure quick response in case of incoming calls, while for M2M devices it is an overhead as the majority of traffic is terminal initiated or at least it is possible to tolerate delay in answering the incoming calls.

The evaluation and the comparison with the benchmark LTE will be done presumably with the following performance metrics (definitions see Appendix A3):

- Paging efficiency: Percentage of specific control channel information for paging procedures measured by bit/user. This metric is of particular interest for use cases when traffic is terminal initiated, or at least it is possible to tolerate delay in answering the incoming calls, in which case paging cycle could be increased. The goal is to reduce control channel signaling in comparison with the usual LTE signaling. The evaluation will be done analytically.
- Mobility management efficiency: Percentage of specific control channel information for mobility procedures measured by bit/user. This metric is of particular interest for use cases when the M2M devices are located at fixed positions, thus not requiring mobility support. The goal is to reduce control channel signaling in comparison with the usual LTE signaling. The evaluation will be done analytically.

Detailed description:

The majority of M2M devices are expected to be battery operated, hence the low power consumption is one of the crucial requirements in this domain. Low power consumption of individual M2M devices will enable their long lifetime which is required to minimize maintenance and costs for replacing batteries as extensive manual labour is needed having in mind the potential number of these devices. Additionally, many of the M2M devices will be located at places difficult to access (for example parking sensors buried in the ground in each parking slot or attached to lamp posts) which further complicates battery replacement procedure.

However, not all these devices will actually be connected directly to the LTE-M network. Instead, they will usually use some sort of short range wireless access technology (e.g. IEEE 802.15.4) to communicate with a gateway which will in turn provide wide area connectivity via mobile networks (or any other network). The gateways can be battery operated as well (for example mobile phones can act as gateways), but quite often the gateways will have more or less uninterrupted power supply (for example, in the SmartSantander installation [11], the gateways are powerful mini computers attached to the mains power supply). M2M devices connected directly to the mobile network will in many cases monitor certain electrical devices (e.g. vending machines) or vehicles (thus making it possible to use the vehicle's battery as the power supply) etc.

In all these cases, the power supply is not the crucial problem from the performance of the device or the system point of view. The power consumption can be seen as the overall energy efficiency problem as a large number of such consumers might increase the overall power consumption, but this aspect is out of scope of this document and the project.

However, with a huge number of M2M devices attached to mobile networks, LTE in particular, the optimization of the network procedures and protocols is required to ensure serving of as many as possible M2M devices by the network without affecting the existing users (e.g. mobile phones) and without requiring significant mobile network infrastructure hardware extensions. The rationale for this stems from the fact that today, mobile networks treat M2M devices equally as any other mobile device. As the majority of M2M devices will generate small amount of data, much less than a mobile phone, the average revenue per device will be relatively low (assuming the current business models which rely on charging the amount of data transferred or included in a quota). Therefore, it will not be possible to justify the increase in the network capacity in the same manner as if the extension was required due to additional mobile phones; hence the requirement for optimization of LTE networks capability to serve M2M devices.

The main approach in designing such an optimization relies on utilizing knowledge of the M2M device characteristics and the way the devices plan to use the mobile network in order to avoid unnecessary signalling procedures and when possible to schedule such transfers to enable optimal utilization of available radio resources. With the simplification of the network procedures and protocols, the power consumption of M2M devices will be also reduced, thus gaining benefits on both fronts.

In many M2M solutions, the M2M devices are fixed at a location so mobility support is not required. This can lead to several optimizations as a number of measurements and reporting the terminals usually have to perform to ensure uninterrupted connections can be completely removed or at least executed less frequently.

Further to this, advance knowledge of the traffic pattern used in a M2M solution can be used to plan and schedule transmissions. This applies to the uplink traffic, but maybe even more to the downlink traffic as in many M2M scenarios the server side initiated connections with the M2M devices are rare and often not required to take place in real time. This allows M2M devices to avoid a number of procedures that regular mobile terminals have to follow to enable quick connection setup in case of incoming calls.

5.1.1 Registering information about the terminals

Today, mobile networks treat all terminals in the same manner, assuming that all of them are used in the same or similar way and have the same or similar requirements. To enable the networks to differentiate between different terminals and their requirements, a procedure that will enable terminals to describe themselves and their capabilities, traffic generation patterns and service requirements are required. One approach to implement this is to extend the information that is already stored in the network about the terminals with new information. These additional descriptions can be included in the existing signalling messages by extending some of the fields and stored in networks' Home Subscriber Servers (HSS), which of course have to be extended as well.

Using this information, the network will know the type of terminals attached to a radio cell and will be in position to use that information to configure the terminals and the cell in an appropriate manner. Extension of messages used by the terminals when registering to a network is required.

5.1.2 Monitoring paging channel and mobility support

Mobile terminals receive information from the network specifying parameters and actions the terminals are supposed to use when interacting with the network (for example cell selection and reselection, measurement management, location and routing registration, handover, power controls etc.) Today, these parameters are the same for all terminals in a cell. This, one-size-fits-all approach is fine when the network is serving one type of mobile terminals, e.g. mobile phones. However, for M2M devices some of these parameters are not suitable, and a solution is required to enable differentiation of terminals and optimization of parameters distributed in the system information blocks based on the type of terminals.

By extending the system information messages, and using the knowledge about the type of terminals attached to a cell (see Section 5.1.1), the network will be in position to specify different parameters for different groups of terminals thus enabling different treatment of M2M devices and making it possible to use modified protocols and procedures like the frequency of radio measurements. For example, monitoring paging channel takes up substantial amount of time and consumes significant amount of energy in the IDLE mode.

While for mobile phones this is necessary to ensure quick response in case of incoming calls, for M2M devices it is an overhead as the majority of traffic is terminal initiated or at least it is possible to tolerate delay in answering the incoming calls. Therefore, for M2M devices it is possible to avoid monitoring paging and/or to increase the paging cycle (DRX cycle).

Similar reasoning applies to the mobility support. As already mentioned, a large number of M2M devices will be at fixed locations and will throughout its lifetime use one or two radio cells and consequently will not require nor use fast handovers. Today, various radio measurements are performed to ensure the best radio conditions and thus the best connection quality as the terminal is moving around. Such procedures are relatively complex, are using up radio resources and have significant impact on energy consumption. While necessary in the case of standard mobile terminals (mobile phones), for many M2M terminals this is just an unnecessary overhead.

However, in some cases M2M devices will be mobile (e.g. deployed on vehicles) and mobility support will be required. Due to the characteristics of the M2M traffic (short transfers, mainly initiated from the devices towards remote servers), it is not required to have very smooth handovers and uninterrupted connections. Therefore, a possible approach could be to allow the M2M terminals to control the timing of and to perform the required measurements only when needed.

This way the network is relieved from continuously monitoring radio conditions for a large number of terminals that actually do not move, or if they move, they can afford short interruptions in connectivity.

5.1.3 Radio resource usage

Many machine-type applications generate low traffic volumes per device, compared to interactive services. But even with low traffic volume, the radio resource utilization can be quite high, because of transmit of small amounts of data frequently. This is due to control channels and signalling messages, which can be a significant part of the air interface load for such applications. Another potentially limiting resource is the identifier space. On the LTE radio interface, the 16-bit C-RNTI is the identifier for connected devices. With an extremely large number of machine devices per cell, it may not be possible to keep devices in *Connected* state between their transmissions. On the other hand, if the C-RNTI is released during periods of inactivity, signalling is required to establish a new C-RNTI for each transaction. Both approaches, keeping M2M devices in the RRC_CONNECTED state or in the IDLE, have advantages and disadvantages:

- If the M2M devices are kept in the *Connected* state, the eNodeB has to maintain a UE context, occupying memory and identifier space, but requiring relatively little processing for each packet.
- If the devices are instead released to the *Idle* state between the transactions, no resources are occupied in the eNodeB, but more processing is required for each transaction.

Further analysis is required to find optimal solution that might be dependant on the type of M2M devices, the traffic they generate or some other parameters.

One possible solution that can help in reducing signalling overhead is embedding M2M user payload into LTE signalling messages and thus avoiding the full signalling exchange. For this approach to efficiently work, the eNodeBs would need to know the size of the M2M payload and some other transmission parameters. Some of these parameters could be transferred a priory, during the first registration of the device to a network (and stored in HSS for example) as in many M2M scenarios, the payload will be constant for each transmission.

5.2 Traffic aggregation

The proposal in a nutshell:

We plan to overcome the following shortcomings of the existing LTE specification (see section 2.2):

- Small data transmission sessions in M2M: The majority of M2M applications transmit and receive only very small amounts of data in the order of a few kbps. Current LTE specification was designed for broadband applications with reasonable control information overhead to achieve the required high peak data rates and high mobility support In LTE-M we envisage a reduction of signalling overhead to achieve again a reasonable ratio between payload of a short message and the required control information. For this we need the option to group traffic into the new proposed common or default channel for all M2M devices to simplify control signalling, and management currently not foreseen in LTE Rel-10.

The evaluation and the comparison with the benchmark LTE will be done presumably with the following performance metrics (definitions see Appendix A3):

- Transmission Payload Size: New approach will lead to reduced amount of signalling, while relatively increasing amount of user payload transferred. This metric is relevant for all use cases, where we assume that a large number of short data transmissions, each transferring a small amount of data, will dominate the traffic patters. We will compare the total amount of signalling decrease and payload increase which are obtained by the proposed schemes in comparison to standard LTE approach. The evaluation will be done analytically.

- Complexity: proposed approach could introduce additional complexity in the scheduling process, so it should be analysed. This metric is relevant for all use cases, where we assume that a large number of short data transmissions, each transferring a small amount of data, will dominate the traffic patterns. We expect that the introduced complexity will not be significantly higher than in the LTE. The evaluation will be done analytically.

Detailed description:

As already stated, one of the main concerns in relation to serving M2M traffic in LTE is low utilization of radio resources due to large overhead, i.e. signalling required to setup a communication link is usually much higher than the actual user payload.

One approach to combat with this issue is to introduce a sort of semi-permanent communication channels. These channels would be setup on the first request of a M2M device. During this initial request, the M2M device would need to provide additional parameters specifying the amount of data to be transferred during each session, periodicity of data transfers as well as some other parameters of importance (tolerable delay, transfer speed, etc.). eNodeB would then use this information in the scheduling process to set up a dedicated channel for this device. This channel would be put to a sleep state after the required amount of data has been transferred. While in the sleep state, radio resources will be freed up and available for other users. At a specified time (based on the information provided in the initial request), the channel is put back in the active mode using previous parameters and thus avoiding new signalling procedure. It is for further investigation how such a channel can be implemented and what are the implications for other LTE procedures and components.

Another possible approach to deal with the issue of radio resource utilization is to use a common channel for all M2M devices in a cell combined with downloadable components responsible for processing data from M2M devices. LTE provides a default bearer for each terminal which is also assigned an IP address. Depending on the needs of each terminal (QoS) and the link status (CQI) at a given moment, the scheduler allocates appropriate radio resources to terminals.

Alternative approach is to setup a default M2M bearer that all M2M devices in a cell can use. This would make it possible to reduce the amount of signalling, while relatively increasing amount of user payload transferred, but would complicate the scheduling process. It is for further investigation if the a priori knowledge of the link status of M2M devices can be used to simplify gathering of information about the link status (if the devices are static, the link status should not change significantly over the time) and what other implications such approach would have.
Further to this, eNodeBs could be enhanced with new functionality responsible for processing of data received from different M2M devices. This functionality would be implemented in software packages that M2M applications would upload to eNodeBs based on the presence of M2M devices they use in a given cell. Such packages would enable processing of raw data at the network edges and thus make it possible to intelligently decide what data has to be forwarded through the network or how to compress the data (combine inputs from several M2M devices, postpone transfers until certain amount of data is available while making sure that only delays tolerable by an application are introduces). It is for further investigation how such functionality can be introduced and what impact would it have on the eNodeB hardware and the overall capacity.

5.3 Energy efficient relaying

The proposal in a nutshell:

We plan to overcome the following shortcomings of the existing LTE specification (see section 2.2):

- Energy efficiency: the algorithm aims at minimizing the power allocated to each relay node for the given LTE-M device QoS constraint. The applicability of this algorithm is not restricted to a specific use case, since it satisfies the general requirement in M2M communications for energy consumption reduction.

The evaluation and the comparison with the benchmark LTE will be done with the following performance metrics (definitions see Appendix A3):

- BER: Bit error rate at the output of the decoder. Here the average BER for all devices is calculated. The goal is to achieve comparable BER with that of the LTE scheduler. The evaluation will be done via system-level simulations.
- Consumed energy per message: the objective of the proposed algorithm is to minimize the total consumed energy at the relays under specific performance constraints, while also considering the individual energy limitation of each RN. We assume that the RNs amplify the signal transmitted by the UE, and transmit it to eNB. We seek to determine the optimal transmission power consumed at each RN while providing quality of service (QoS) assurance. The goal is to reduce the total consumed energy compared to the LTE case where no relay selection is foreseen. The evaluation will be done via system-level simulations.

- PHY Control channel and pilot overhead: Percentage of radio resources utilized for signalling, control channels and pilots on PHY layer. The energy savings should be followed by a sensible increase of the required signalling. This trade-off will be investigated via simulations.

Detailed description:

Relay nodes have been introduced to LTE-Advanced to enable traffic forwarding between eNodeB and UE [12]. Type 1 relays assist a remote UE unit to reach an eNodeB, which is far away from it. On the other hand, Type 2 relays can improve the service quality of a UE that has direct access to the eNodeB, by achieving multipath diversity and transmission gains. While LTE relaying aims to enhance coverage and capacity, issues such as energy savings, cooperative transmission and assurance of specific quality of service to the receiver need to be studied further. In M2M communications energy efficiency plays an essential role, especially for devices that are equipped with batteries. Moreover, power efficiency cannot be considered separately from the requested QoS. The proposed algorithm extends the relaying functionalities in LTE by:

- Enabling two or more Type 2 relays to forward data to the LTE-M device, increasing in this way the multipath diversity and hence the QoS.
- Ensuring the best achievable QoS for the LTE-M devices under specific transmission power constraint for them.
- Minimizing the power allocated to each relay node for the given LTE-M device QoS constraint.

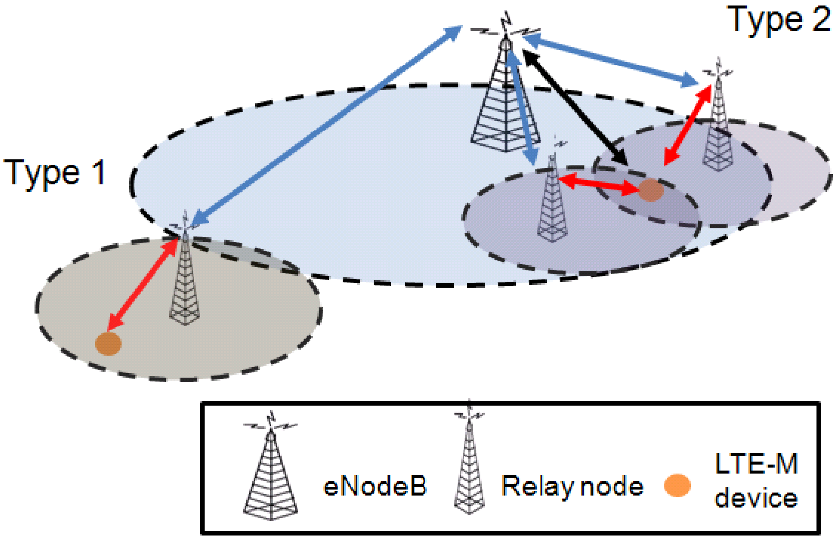

Figure 5-1: Power allocation in Type 2 relaying in LTE-M.

As shown in Figure 5-1, we consider a scenario where a source node (LTE-M device or eNodeB) communicates with the corresponding destination node (eNodeB or LTE-M device) with the help of one or more relay nodes in between. We assume that the relay nodes use the Amplify-and-Forward (AF) scheme, i.e. a relay node receives the signal from the source node at the first phase, amplifies it, and forwards it to the destination in the second phase. As opposed to the Decode-and-Forward (DF) process, the AF process is easy to implement through Radio Frequency (RF) amplification, has low latency and requires less energy.

Distributed transmit antenna selection, i.e. the selection of only a subset of the set of available relays, is a concept that has been extensively studied in literature recently [13].
Using the proposed technique, instead of finding the optimum relaying set, we seek to determine the optimal transmission energy at each relaying node, while providing QoS assurance to the receiver. Since the end-to-end SNR determines both the maximum achievable rate and probability of error, the optimization problem studied, is the minimization of the total energy consumed at the relays, subject to meeting a specific constraint on the target SNR. We assume that the transmitter has full knowledge of the source-relay and relays-destination channel conditions. In order to solve the problem, we transform the original problem into an equivalent convex optimization problem.

The proposed algorithm addresses the following requirements as described in [1]:

Table 5-1: Requirements with respect to relaying

Requirement	Solution description
NF.2	The proposed algorithm takes into account the energy consumption of each relaying node
SV.1	The proposed algorithm ensures the QoS of the LTE-M device by determining the set of transmitting relaying nodes and taking into account their energy consumption

Performance assessment of the proposed technique is carried out via simulations. We consider networks with 2, 4, and 8 relay nodes. We assume the transmitted symbols are BPSK modulated. The source-relay and relay-destination channel amplitudes are considered independent and Rayleigh distributed. Furthermore, we assume that the relay nodes can provide equal amount of maximum energy. In Figure 5.2 we present the average bit error rate of the proposed method and the best relay path assuming that the constraint on the target SNR is equal to 10^{-2} .The selected best path relay corresponds to that one with the highest end-to-end SNR. As illustrated, the Average Bit Error Probability (ABEP) of the best path selection scheme decreases as the SNR increases, while the ABEP of the proposed scheme always equals to the target one. This implies that no energy is wasted for ABEPs that is not required. Thus, the proposed method seems to achieve a well balanced tradeoff between energy consumption and error performance.

A next step is the determination of the optimal energy consumption at each relaying node, taking into account its remaining energy. Furthermore, future improvements may include the use of multiple antennas at the relaying nodes and the impact of interference.

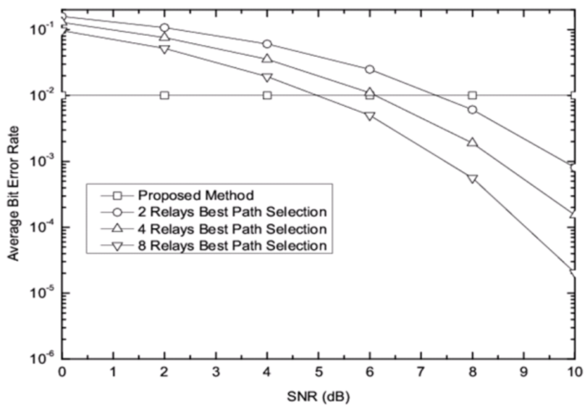

Figure 5-2: Energy saving with the proposed resource allocation algorithm

6. LTE-M Broadcast- and Multicast Services

Effective support of broadcast and multicast services is a fundamental aspect of LTE-M since it is expected that a sizeable portion of downlink traffic in LTE-M will be broadcast/multicast. In particular, many device management functions may imply the transmission of blocks of information to a set of devices, for example particular common configurations for a specific LTE-M operator or software update. Broadcast/multicast traffic of these characteristics will typically be of moderate or high volume, it will have very high coverage/availability requirements and very low error tolerance. The techniques described in this section specifically address these aspects in a much more effective way than in LTE systems.

6.1 LDPC-like rateless codes

<u>The proposal in a nutshell:</u>

We overcome the following shortcomings of the existing LTE specification (see section 2.2):

- Small data transmission: especially regarding the redundancy overhead per user together with signalling reduction
- Multicast capability: a low level multicast procedure is not defined in LTE release 8. The proposed solution fits into this gap by offering reliable efficient multicast transmission to a large number of devices

This is related to the following preliminary list of KPIs (definitions see Appendix A3):

- Redundancy overhead: To provide a reliable multicast mechanism, the transmission has to be secured with redundant data but for energy efficiency reasons, the redundancy overhead must be kept to a minimum. The redundancy overhead will be evaluated via link-level simulations showing the performance over a wide range of channel conditions while adapting the overhead per user. The overhead will be less than LTE.
- Number of served users: The goal of integrating a vast amount of users into the system will be assessed with the metric of served users. The evaluation is also done via extensive simulations showing that the number of served users is higher than with a traditional LTE system.
- Complexity of encoding and decoding: In the context of energy efficient M2M communication, complexity is a crucial issue. As the encoding and decoding of channel codes is a significant part in the signal processing of a transmitter/receiver the complexity of the algorithms should be at least equal to that of LTE. The comparison is done analytically.

<u>Detailed description:</u>

To ensure reliability in broadcast and multicast services, packet transmissions need to be protected against channel errors. Originally, the protection against errors is accomplished by channel codes – mostly convolutional ones in interaction with an additional block code. In this section, the idea of rateless coding for broadcast and multicast transmission based on LDPC-like coding structures is introduced. As traditional rateless codes are described via a tanner graph of the generator matrix, the new approach focuses on the design of rateless codes on the tanner graph of the parity check matrix. A typical tanner graph for rateless codes is shown in Figure 6-1.

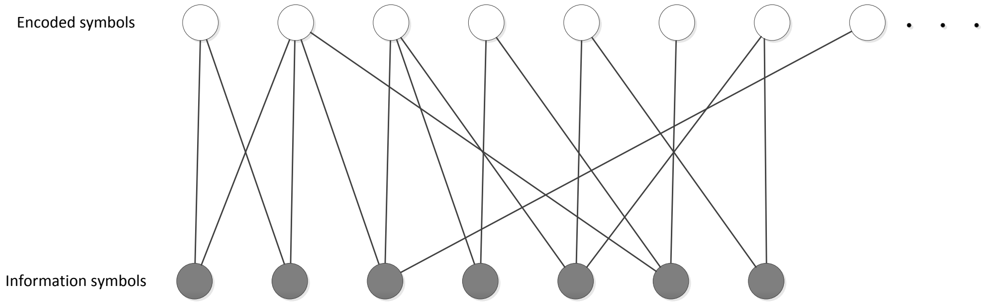

Figure 6-1: An example graph of the generator matrix of an LT code

The presented design approach shall lead to applicable codes that focus on several requirements of the EXALTED system vision. The amount of devices supported by the proposed approach can be very high, which aims at the requirements FU.1 and NF.1 [1]. In conjunction with that, requirements NT.10 and NT.17 are addressed as the acknowledgement procedures in a rateless coded broadcast and multicast transmission can be simplified extensively. By embedding a data stream into dedicated downlink broadcast and multicast resources such as the MBMS functionality of LTE Release 10, backward compatibility is ensured. In fact, the approach is more general and has minimal constraints on resource usage. As the approach focuses on protection of broadcast and multicast traffic against channel distortion, requirement NT.9 and NT.13 are addressed completely. Having an optimized duty cycle in decoding of a broadcast or multicast message, the approach aims also at the energy efficiency claimed in NF.2. At least from an application point of view within the M2M context, requirements DV.10 and DV.11 are addressed.

Traditional rateless coding schemes are applied on higher layers like network layer where they show a performance that achieves Shannon capacity for the Binary Erasure Channel (BEC) (on a packet basis, BEC is a good model for packet losses). They are suitable for high data rate applications in the LTE internetworking context where data has to be streamed to a client and the amount of information is extremely high. In the M2M context data (e.g. configuration updates) the amount of data to be transmitted might not be as high as in the LTE context. But having the rateless coding approach defined on a higher layer induces an immense portion of signalling and protocol overhead that would be wasteful in the energy limited M2M context. The new approach focuses on the design of a rateless coding scheme on the physical layer to avoid adding tremendous amounts of additional information per packet which cannot be exploited at an M2M terminal.

The proposed approach provides designs of new rateless codes with suitable properties for M2M broadcast and multicast communication by leaving the traditional approach of designing rateless codes. At first traditional rateless codes are defined on their generator matrix. While the encoding process takes place, the corresponding tanner graph can be built simultaneously since the design of the degree distribution used for encoding the graph is sparse. That is of crucial importance for the decoding which is also carried out on the graph assigned to the generator matrix as the complexity scales with the density of the graph. In contrast to this view, the new approach tries to bring together two different paradigms of coding: Original block (and convolutional) coding will be brought together with the rateless characteristic. This incorporates the description and design of a channel code in a rateless fashion on the parity matrix.

The approach is inspired by a generalization of rateless codes through Low Density Generator Matrix (LDGM) codes. In [14] a contemporary overview on LDGM codes is given. Those codes are also described by a sparse generator matrix and they are known as the dual parts of a LDPC code. As described earlier, the decoding complexity of a channel code defined on a tanner graph is directly proportional to the amount of edges which connect the tanner graph. Unfortunately, for an LDGM code where the generator matrix is sparse most probably the parity check matrix is dense. That leads to an unreasonable high decoding complexity which cannot be spent on an M2M device. When designing the channel code on the parity check matrix, one could make sure that the matrix is sparse and has low complexity. Other than that, a sparse matrix is extremely important for the performance of the decoder as with a traditional belief propagation algorithm the messages passed on the tanner graph need to be independent for successful error correction performance. In practical code design this is only achieved approximately. High independency of messages is ensured via the design of a graph with a high girth which means to have only cycles in the graph that have the maximum possible length.

Recent development has shown that there is a tremendous interest in coding structures like irregular LDPC codes [15] and LDPC convolutional codes [16] instead of the traditional use of a generator matrix for code design purposes. There is a general understatement that under certain circumstances, LDPC codes are capacity approaching. Additionally, with the design of suitable LDPC convolutional codes, the decoding algorithms can be simplified with techniques like window decoding.

The approach, rather in its beginning is now to use the design approaches of LDPC and LDPC convolutional codes for designing a rateless counterpart, suitable for the broadband and multicast transmission of M2M devices in the EXALTED system. Additionally, the approach benefits from the available tools used for the design of LDPC codes, e.g. Density Evolution and Extrinsic Information Transfer Chart (EXIT) charts.

The future work to be focused on is first to investigate the structure of a channel code for fitting the rateless coding paradigm. In particular, this task will force the principle design of a tanner graph and corresponding parity check matrix for the definition of the broadcast and multicast efficient channel code.

In addition, techniques like density evolution and EXIT charts are used to optimize the parameters and performance of the code for implementation in broadcast and multicast settings.

6.2 Collaborative broadcast architecture

The proposal in a nutshell:

As mentioned earlier, high coverage/service availability is expected to be a crucial requirement for broadcast/multicast traffic. In this section, a technique is proposed, which leverages the existence of an LTE-M network with relaying capabilities as well as a capillary network that improves robustness and availability of broadcast services. In this proposal, a network code is implemented at the base station and at a subset of destination devices, named Cooperative Gap Fillers (CGF). Network-coded schemes, also including multiple radio interfaces, have been proposed in the literature e.g. in the satellite context [17]. However, the deployment of such a scheme in an energy-constrained heterogeneous (cellular and short range) network represents a new challenge requiring detailed system-level simulations.

Traffic in M2M systems can be highly asymmetric. While uplink traffic typically consists of short messages, downlink traffic could involve large transmissions for device management purposes. For example, a software upgrade for an embedded Linux device may involve the transmission of even more than one Gbyte. It is critical that such large messages are transmitted efficiently in order not do deplete the device battery. Moreover, device management messages will typically be addressed to a group of devices. Therefore, an efficient and robust multicast mechanism is needed. For this traffic pattern, it has been acknowledged that rateless codes and, as a generalization, network codes, can be very effective. While a rateless code already provides these advantages, the use of network coding can, in addition, increase the availability of encoded packets to devices. A device with good reception capabilities may relay encoded packets to shadowed devices, improving coverage beyond what is available through the current LTE deployment. We emphasize that this proposal is addressed to large, multicast messages transmitted in the downlink of the LTE-M system. Use of network coding for short packets is not considered, as the overhead resulting of breaking such packets into smaller packets would not compensate for the advantages of network coding.

We plan to overcome the following shortcomings of the existing LTE specification (see section 2.2):

- Device cost issues: Network coding is a very flexible encoding scheme. Based on the principle of rateless codes, devices need only listen to a fraction of the encoded message to decode it entirely. Thus, devices may have longer sleep cycles and even temporarily interrupt reception to transmit in half duplex mode.
- Network overload issues: Rateless and network codes require less feedback than conventional coding schemes. For example, a single acknowledgment is sufficient, at the end of the transmission, for each receiver in the multicast group. This enables network coded applications to use unreliable link layer schemes.

The evaluation and the comparison with the benchmark LTE will be done presumably with the following performance metrics (definitions see Appendix A3):

- Redundancy overhead: the proposed method is based on using a secondary air interface to introduce redundancy. This may cause strain on capillary networks using that interface for non-multicast purposes. Nevertheless, it is expected that this proposal will require less redundancy overhead in the primary air interface than current methods for multicast used in LTE.
- Number of served multicast users: the simplicity of the encoding scheme, better coverage, the low amount of feedback required to terminate the transmission will allow the system to serve a very large number of multicast users.
- Coverage: the proposed system achieves higher coverage than conventional singlecast codes used in LTE, as it is shown in preliminary simulations.

<u>Detailed description:</u>

The proposed system is schematically shown in Figure 6-2. In the figure, the eNodeB (labelled as BS) transmits an LTE-M broadcast signal to a set of devices. Some devices function as CGF and are labelled accordingly. Typically, CGF are devices with extended functionalities, such as a capillary gateway, or devices with abundant energy resources available, as they need to selflessly relay packets to other devices. In-Band (IB) gap fillers use the same LTE-M air interface and frequency band, while Out Of Band (OOB) gap fillers use capillary air interfaces and frequency bands. The implementation of the network code is standard: a message is broken into several packets which are linearly combined at the source. Then, intermediate nodes, in this case the CGF, receive packets, store linearly independent ones, and retransmit new linear combinations of the received packets. A global encoding vector is inserted at each packet header so that end nodes can decode the packet [18]. If the device has multiple radios, it may recombine packets from both LTE-M and capillary interfaces. Once a device decodes the message, it sends a positive Acknowledgment message (ACK) via the uplink channel. The multicast stops when ACK from all devices have been received. An alternative stopping criterion may be established, such as stopping the multicast after the source has sent a fixed number of encoded packets.

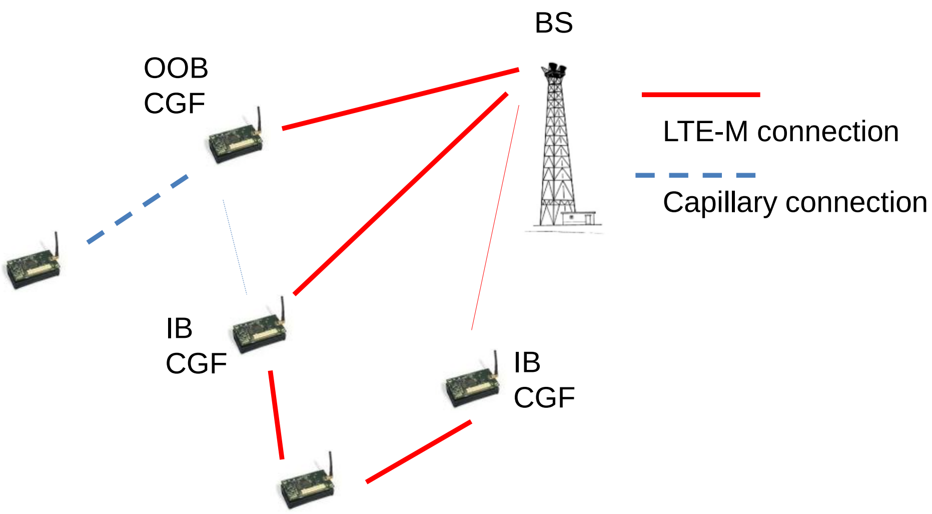

Figure 6-2: Collaborative Broadcast Architecture.

CGF retransmits packets in order to maximize the reliability of the multicast signal, either through the LTE-M or capillary interface. This scheme provides path diversity, since a node may receive packets either from the BS or from a CGF and recombine them later. The network code is implemented above the link layer, since packets from both LTE-M and capillary radio interfaces must be recombined.

The solution addresses several requirements of transmitting multicast information. It is a scalable, spectrum efficient multicast solution (addressing requirements FU.1, FU.2, NF.1, NT.10 [1]), and improves the efficiency of many remote device-management functions (addressing requirement FU.5), A few examples of such functions are provisioning of sets of M2M equipment (addressing requirement SV.3) and software updates over the air (addressing requirement DV.11). Finally, the proposed system improves the reliability of multicast transmissions (addressing requirements NT.9, NT13).

Support for multicast transmission is provided in LTE Release 10 as MBMS functionality [19], [20], which addresses efficient broadcasting of multimedia contents. In particular, the single frequency network option, MBSFN, is related to our approach. In MBSFN, several eNodeB nodes synchronize in order to transmit OFDM signals of a particular broadcast, using the same time-frequency resources. Due to a longer prefix, OFDM signals can be recombined at the mobile. The MBSFN approach is compatible with our solution in that several eNodeB nodes may transmit broadcast signals synchronously. However, the MBSFN approach may not solve coverage problems, and certainly does not reach devices connected to capillary networks. Moreover, the LTE solution does not allow efficient usage of capillary air interfaces for multicast. Therefore, our solution extends beyond LTE MBMS specifically for devices.

The implementation of the network code is severely constrained, in terms of complexity, by the processing power of the devices. Therefore, a low complexity implementation is appropriate. Encoding and decoding itself is rather simple, since a random linear encoder is employed, and the size of the finite field can be small (e.g. 8) in order to minimize complexity. Since retransmissions will take place over a capillary air interface, the risk of flooding the network with an excessive number of encoded packets is limited, since capillary interfaces have very short range. Moreover, if an out-of-band (i.e. non-LTE-M) capillary interface is used, retransmissions caused by the network code do not affect the LTE-M network.

Nevertheless, the number of retransmissions must be kept to the minimum necessary since excessive retransmissions will drain the device batteries and cause collisions, reducing the effectiveness of the code. A simple criterion that may be adopted by cooperative gap fillers is to retransmit only innovative (i.e. linearly independent) packets with a fixed probability. More redundancy can be achieved if each innovative packet is transmitted more than once on average. To analyze performance of the system as a function of the degree of redundancy we define parameter ξ as the average number of times an innovative packet is retransmitted. This scheme is extremely simple to implement and already fulfils the goal of improving coverage of the multicast transmission. Collision recovery mechanisms such as those described in Section 4.2.2 are also applicable to the broadcast environment. Moreover, the design of advanced codes together with the proposal in Section 6.1 also brings extra benefits.

A preliminary performance evaluation of the described collaborative network coding implementation is described in this section. The ITS scenario was considered, and the newly defined IEEE 802.11p vehicular networking standard was adopted as capillary interface. Basic simulation parameters are reported in Table 6-1. Further details and simulated scenarios will be available in upcoming deliverables. The objective of this preliminary simulation is to verify the potential usefulness of the proposed scheme.

Table 6-1. Simulation Parameters.

Simulation environment	Urban
Carrier frequency	2 GHz
Device (vehicle) speed	50 kph
Time interleaving	Optional
Modulation	OFDM
Channel coding	Turbo Code rates ½ and ¼
Network code	Random Linear, GF(8)

Two different iterleaver sizes (short and long) were used. In the ITS scenario, which was the one chosen for evaluation, a long interleaver reduces the probability of a device being in a deeply shadowed position. Figure 6-3 and Figure 6-4 show the performance of the proposed collaborative broadcast architecture based on network coding. In Figure 6-3, the normalized throughput is shown as a function of the average number of retransmitted packets. As it can be seen, a moderate number of retransmissions (around 0.8) increases the throughput by around 30% when a short interleaver and a rate ½ code are used. Similar improvements can be seen for long interleaver sizes and rate ¼ code. The performance of a simple relaying scheme, which consists in retransmitting received packets without the network coding layer, is shown for comparison. As it can be seen, simple relaying barely improves the performance of the one-hop scheme.

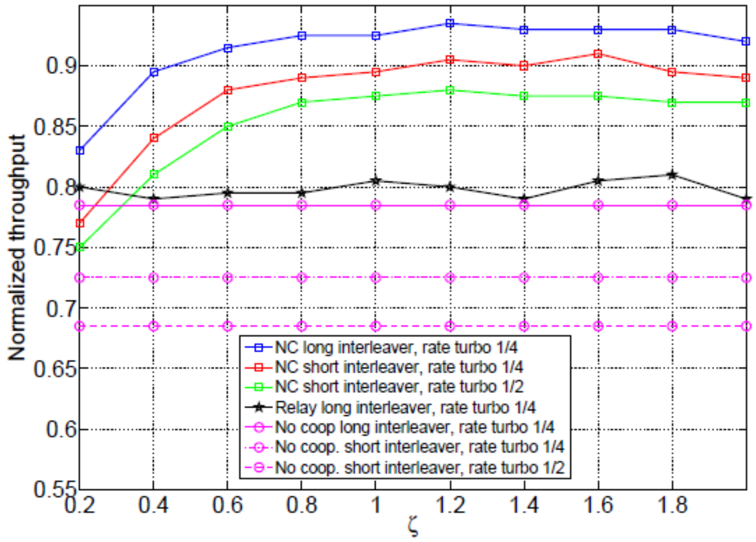

Figure 6-3. Normalized throughput (average number of correctly received source packets) for different modulation and coding schemes and cooperation level. Simple relaying scheme shown for comparison.

Figure 6-4 shows the level of coverage, or service availability, of the system. It represents the percentage of covered nodes as a function of the average number of retransmitted packets. As before, gains up to 30% in service availability can be obtained.

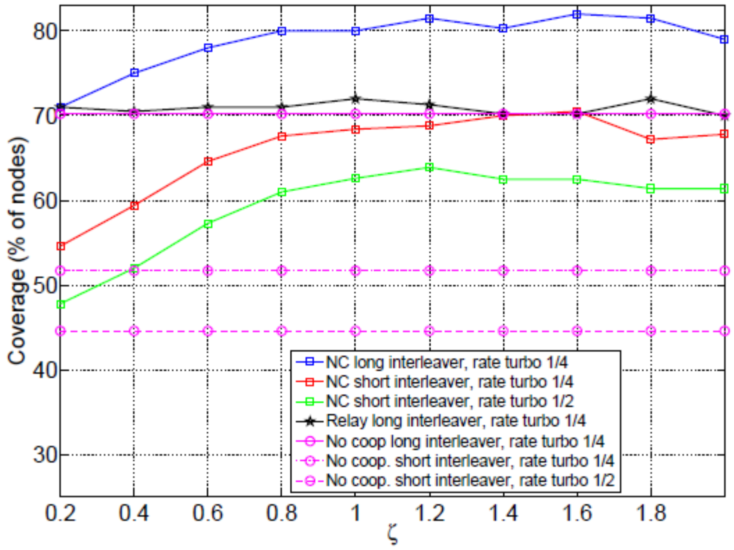

Figure 6-4. Percentage of service coverage for different modulation and coding schemes and cooperation level. Simple relaying scheme shown for comparison.

Additional aspects are left for future work within the project. One important aspect is how to handle acknowledgment messages and send them reliably to the eNodeB. Upon successful decoding, a device sends back a positive ACK. Transmission at the eNodeB stops when an ACK has been received for each device in the multicast. However, it is not clear when an intermediate node must stop its transmission. One possibility is to wait for a stop message from the eNodeB; a more efficient solution may consist in the CGF deciding on when to stop retransmitting upon reception of a number of ACKs.

7. Enablers

This section consists of methods that facilitate M2M communications but are not part of standardized LTE-M signal processing and protocols. They are based on a dynamic adaptation of the system parameters according to measured signals.

7.1 Spectrum sensing

The proposal in a nutshell:

We overcome the following shortcomings of the existing LTE specification (see section 2.2):

- Device cost issues, in particular with respect to signal processing complexity and achievable computational accuracy
- Radio access, in particular with respect to the random access and dynamic spectrum allocation.

The evaluation will be done presumably with the following performance metrics (definitions see Appendix A3):

- Spectrum sensing accuracy: key metrics for evaluating a spectrum-sensing algorithm are mainly probability of detection, which should be larger than 90%, and probability of false alarm, which should be smaller than 10%. It is suitable for all use cases of random access networks. Baselines for performance comparison are mainly state-of-the-art spectrum sensing algorithms, and the proposed algorithm is expected to offer more than 10 dB gain in low SNRs. Computer simulations would be carried out to demonstrate the advantages of the proposed algorithm.
- Sensing delay: it is another key metric for evaluating spectrum-sensing algorithms since a long sensing delay would make the sensing results less reliable. We target on very small sensing delay e.g. two symbols duration, which is never the case for state-of-the-arts.
- Complexity: In order to save device energy cost, we aim at a simpler signal processing, which can be implemented on cheap hardware components. This metric is suitable for all use cases of adopting low-cost devices. We will assess the complexity analytically by comparing the required number of multiplications and expect it lower than in state-of-the-arts.

Detailed description:

Spectrum sensing is often referred to as a signal-processing apparatus that can monitor user activities within a certain frequency band. It has become one of the key technologies taking us one step closer to fully self-organized networks, by allowing decentralized based spectrum allocation.

Since 1940's lots of effort has been paid towards developing reliable spectrum sensing schemes. Some of the well-known techniques including the match filtering [21], conventional energy detection [22], cyclostationarity based detection [23], wavelet based detection [24] and multi-antenna eigenvalue-decomposition [25] have been introduced in the literature for sensing purposes each having various requisite, advantages and disadvantages. Amongst the proposed detection techniques, matched filtering is the optimum solution given the pilot information and reasonable time/frequency synchronizations. Conventional energy detection is the most common and robust detector, as it does not require any a priori knowledge about the signal being detected. However the performance of the energy detection is limited by SNR wall, which is a phenomenon caused by noise uncertainty. The noise uncertainty issue becomes more severe as the SNR decreases making energy detection reluctant in low SNR environments. Cyclostationarity-based detection approaches can deliver an excellent performance with the trade-off of introducing a long latency into the system. Given the signals cyclic frequency cyclostationarity is able to significantly reduce the noise uncertainty factor. The wavelet detection, which has recently been introduced in the literature, is capable of performing coarse sensing with the aid of edge detection. Hence further processing is required in order to perform fine sensing. The eigenvalue-decomposition technique can only be implemented in systems that support multiple antennas. By exploiting the orthogonal subspaces of noise and signal, eigenvalue based detection can deliver reliable spectrum sensing. A matter of primary interest is that almost all existing detection techniques do not deliver an acceptable performance in low SNR range, i.e. (- 25,-10) dB, without introducing a large latency or complexity into the system.

A successful spectrum sensing unit must meet three requirements of latency, complexity and reliability. It is to our interest to minimize the latency of the sensing process to achieve higher throughput while low complexity unit reduces the overall energy consumption of the device. Reliability of a spectrum sensing technique is measured by two classical metrics: probability of detection and probability of false alarm. Meeting the mentioned requirements would be even a more challenging task when it comes to M2M communications due to the limitations of the devices. A typical sensing requirement in M2M scenario is summarized in Table 7.1. The main motivation behind performing spectrum sensing in M2M is to increase the spectral efficiency by reusing the available spectrum and consequently increasing the overall capacity of the system.

Table 7-1: requirement for the M2M sensing unit

Parameter	Minimum Requirement
Sensing delay	< 5 LTE-M symbols
Probability of detection	>90 %
Probability of false alarm	< 10 %
Complexity	$\leq N^2$
SNR Wall	-20 dB

In order to examine the performance of some of the most common spectrum sensing techniques, simulations based on the LTE-M standard were carried out and the results are shown in Figure 7-1. As it can be observed, the performance of two newly developed sensing techniques, namely, Differential Energy detection [26] and first order cyclostionarity based technique [27] are also shown in this figure. The Differential based detection performs sensing operation by exploiting the frequency diversity inherited in the communication channel. This method is developed for sensing OFDM sources and offers acceptable performance even in low SNR environments, while maintaining low complexity. First order cyclostatinarity based technique employs the pilot symbols energy in order to determine the vacancy of a spectrum band and it is proved to outperform all the existing pilot based detections. This technique can reliably detect a source even in low SNR environments given that the noise uncertainty factor is low. The performance shown in Figure 7-1 is for observation length of 7 OFDM symbols. For the purpose of assessing pilot based detection pilot information were embedded equal-spaced for every 32 sub-carriers.

It can be observed that Energy detection seems to be the optimum choice given the noise uncertainty factor, U=0. Since in practical scenarios it is not possible to obtain an accurate estimate of the noise power, there always exists a noise uncertainty factor and in fact it is shown in [28] that in most practical environments U=2. This would determinately affect the performance of the energy detection. Second order cyclostaionarity offers a reasonable performance but on the downside requires high computational complexity.

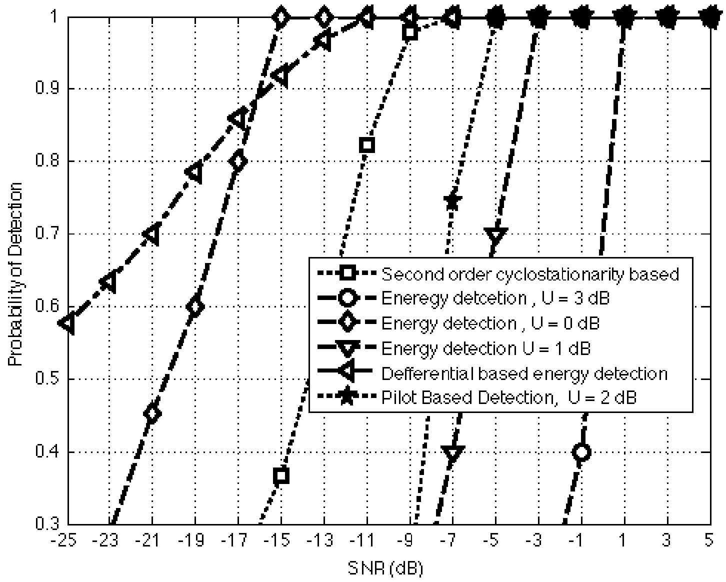

Figure 7-1: Performance comparison of same of the existing state-of-the-art techniques.

This would introduce high latency into the system and implies that a big portion of the energy should be consumed towards performing the sensing operation, which is not applicable in machine-to-machine communications. Hence [26] and [27] were introduced to not only operate in low SNR range but also minimize the sensing energy consumption by having low complexity.

7.2 Energy harvesting

The proposal in a nutshell:

The conventional metric used for assessing the performance of a communication standard like LTE is the spectral efficiency (in bit/s/Hz). In other words, the focus is set on maximizing the transmission rate that can be achieved per unit of bandwidth, *assuming a fixed value for the transmitted power*. Note that a dual (and thus identical) way of looking at the same performance measure is to analyze the minimum power needed to achieve a given throughput on a given bandwidth. LTE relies on several techniques to boost the achieved spectral efficiency, like e.g. MIMO, advanced error correction codes, adaptive modulation and coding (AMC), etc. The latter feature is especially interesting: ACM consists in adapting the instantaneous transmission rate to *the temporal variation in the channel state*. For instance, favourable instantaneous channel conditions are taken advantage of by transmitting a high order modulation with high coding rate. In this section, we instead study a feature which is disregarded in LTE. We consider the possibility of taking advantage *of the temporal variation in the battery state*. By battery state, we mean the amount on energy stored in the battery at a given time instant. In particular, in a node which is able to harvest energy from the environment, the battery state is directly dependent on the flow of harvested energy in the system, to which the evolution of transmission rate over time can be adapted. Note that the feature proposed here is compatible with LTE or LTE-M without the need for a specific standardization: the transmitting entity can decide independently on which transmission rate to adopt (i.e. which modulation and coding mode included in LTE/LTE-M to use) depending on both the battery and channel states.

We overcome the following shortcomings of the existing LTE specification:

- Energy efficiency, in particular we consider devices which are able to exploit the ambient energy for extending their lifetime.

The applicability of this algorithm is not restricted to a specific use case, since it satisfies the general requirement in M2M communications for energy efficiency.

The evaluation and the comparison with the benchmark LTE will be done presumably with the following performance metrics (definitions see Appendix A3):

- Throughput – number of successfully received bits per unit of time (bit/s).
- Energy efficiency - Ratio between transmitted power and achieved throughput (Joules/bit).

The objective is to demonstrate analytically the performance improvement (in terms of these performance metrics) of the proposed algorithms with respect to when the battery state (and thus energy harvesting process) is not taken into account to adapt the transmission rate.

<u>Detailed description:</u>

A significant proportion of M2M device are expected to rely on batteries. It would indeed be unrealistic to assume that all devices have direct access to a power supply. Moreover, in some cases, the replacement or manual recharging of the batteries is too costly. In such cases, the limited battery size (due to physical constraints or cost limitations) translates into a finite lifetime of the device, and possibly of the system. Consequently, energy efficiency is a key challenge in the sustainable deployment of battery-powered communication systems. A complementary approach has recently been made possible by introducing rechargeable batteries that can harvest energy from the environment in order to extend the lifetime of the system. Accordingly, it is essential to design the system operation taking into account the nature of the energy harvesting process to increase energy efficiency.

In this activity, the focus is set on a single point-to-point link, and the goal is to maximize, through power management, the amount of data that is transmitted to the receiver under various assumptions regarding the energy harvesting model as well as the battery limitations.

The basic idea that should motivate and dictate any energy efficient transmission scheme is the following: if one targets the transmission of a given amount of data, it is less energy consuming to transmit at lower power during a longer period of time (i.e. at lower rate). Mathematically speaking, the energy per bit $\varepsilon(q)$ (i.e. the energy needed for transmitting a bit) as a function of number of transmissions per bit q (i.e. the inverse of the rate) is a convex decreasing function, as illustrated in Figure 7-2. Note this property is satisfied in many common transmission models, such as the capacity of an additive white Gaussian noise channel, or suboptimal channel coding schemes [29]. However, in practice, one can obviously not let the transmission time tend to infinity (which would be the most energy efficient solution). Indeed specific deadlines are associated with the transmission of data packets, for instance in order to ensure the quality of service level needed by a particular application. In this context, the authors in [29] developed an algorithm for energy-efficient packet transmission taking into account the presence of such deadline constraint.

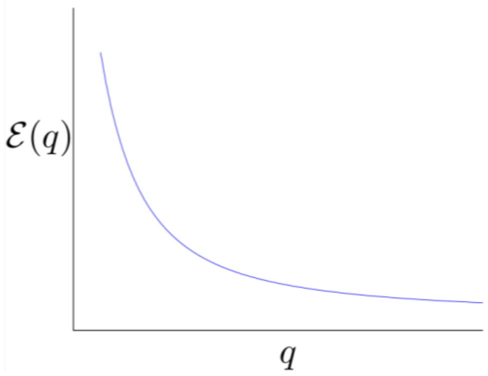

Figure 7-2: Energy per bit versus number of transmissions per bit

As mentioned above, harvesting energy from the environment is an important alternative for battery-limited systems to extend their lifetime. Accordingly, it is essential to design the system operation taking into account the nature of the energy harvesting process to increase energy efficiency. Optimization of the transmission scheme for energy harvesting systems has received a lot of recent attention, see e.g. [30], [31]. For instance, in [30], the considered problem is the following: assume that we want to maximize the number of bits transmitted by a node between time $t = 0$ and $t = T$ (deadline), given that the node harvests energy from the environment and that the energy arrives in packets. Also assume that the energy packets sizes and corresponding arrival times are known in advance. Given this information, one can come up with a power management strategy to maximize the throughput during the interval [0, *T*]. It is shown in [30] that the optimal power strategy satisfies the following properties:

- (*property 1*) the node never remains silent,
- (*property 2*) the transmission power remains constant between two energy packet arrivals,
- (*property 3*) the transmission power is non-decreasing. An example is provided in Figure 7-3 where: $H(t)$ denotes the cumulative harvested energy curve, and $E(t)$ refers to the cumulative transmitted energy curve (whose slope, at any time instant, gives the instantaneous transmission power). The above mentioned properties can be identified in this example.

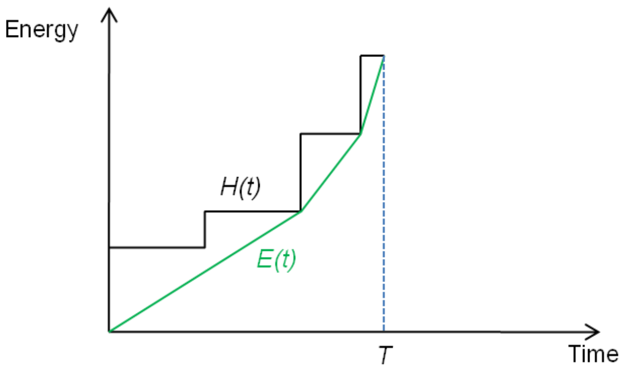

Figure 7-3: Transmission power management for packetized harvested energy arrivals

With respect to this existing work, this activity proposed to add practical battery constraints to the problem. In particular, at this point, the focus is set on a single point-to-point link, and the goal is to maximize the amount of data that is transmitted to the receiver under various assumptions regarding the energy harvesting model as well as the battery limitations. Two limitations are considered separately:

Limitation 1. A time-varying finite battery size constraint is considered, which models the degradation of storage capacity over time.

Limitation 2. A battery that suffers from energy leakage is considered. That is, energy is assumed to leak from the battery at constant rate, if the battery is non-empty. Note that such constraint can be used equivalently to model the minimum energy consumption needed to maintain a node awake (even when not transmitting).

The introduction of one of these two battery limitations modifies the above-mentioned properties of the optimal solution, as summarized here:

- With *Limitation 1*, *property 3* does not hold anymore. Indeed, it can be advantageous (in terms of total throughput) to transmit at higher power at the beginning of the time interval [0, *T*], for example, in order to empty sufficiently the battery such as to prevent the next energy packet arrival to exceed the battery capacity (in which case the excess of energy would be wasted).
- With *Limitation 2,* none of the three properties (*property1 to 3*) holds. Indeed, consider the case of a high leakage rate. In such case, it might be beneficial to empty the battery in between two energy packets arrivals. In fact, the longer you are transmitting, the more energy will leak. Consequently, the optimal power management strategy might be such that the node remains silent at times.

Next steps within this activity implies the detailed derivation of the optimal transmission strategies taking into account either of the above-mentioned limitations. For this, the powerful framework based on cumulative curves derived in [32] will be used. Later on, other scenarios can be considered, like the broadcast channel (in the parlance of information theory), or multiple hop transmissions.

7.3 Cross Layer MAC Scheduling for capillary LTE-M Networks

<u>The proposal in a nutshell:</u>

We overcome the following shortcomings of the existing LTE specifications (see Section 2.2):

- **Small data transmission** with heterogeneous types of traffic (MTC traffic vs. LTE compliant traffics)
- **Radio access** with coexistence of LTE devices of cellular networks and M2M devices within capillary networks
- **Device cost issues** and in particular **data rate**, since QoS requirements of M2M devices significantly differ from those of LTE devices
- **Network overload issues** with the necessity for LTE system to consider a huge number of additional M2M devices

The evaluation and the comparison with the benchmark LTE will be done presumably with the following performance metrics (definitions see Appendix A3):

- Packet Error Rate (PER): This metric is of particular interest for all use cases where some constraints are set on quality of service as well as on energy consumption. A small PER lets to reduce the number of retransmissions and hence avoid energy depletion. The evaluation will be done by system level simulations and it is expected that these values are similar or not much higher than those obtained with LTE.
- Packet Loss Rate (PLR): In comparison to PER, PLR is restricted to erroneous packets due to excessive latency. As a consequence, this is perhaps more relevant for eHealth use cases and for some ITS applications (such as collision avoidance) where high reliability and short latency may be required. Otherwise, as with PER, PLR will be assessed by system level simulations.

- Average number of retransmissions: As with PER and PLR, number of retransmissions could be an indicator interesting for all use cases that focus low energy consumption and especially those which are constrained in quality of services (in terms of latency, eHealth and some applications in ITS are concerned). A simulator of a whole system will let to assess this metric at the link level for all links and then compute a system average.
- Throughput and average packet call throughput: This is of particular interest for all use cases. Since EXALTED use cases mostly target low data rate communications, this metric for M2M links should be lower than throughput with LTE in most of M2M applications.
- Average packet delay: As with PLR, use cases more sensitive to strict quality of services constraints are more relevant for such a metric. A small packet delay lets the entity goes back quickly in an idle state with smaller energy consumption; hence energy waste is reduced. Link level assessments during system level simulations will be done. We expect higher delay for most of M2M applications than with LTE since MTC traffics are less stringent than those of LTE in terms of delay.
- Ratio between transmitted power and achieved throughput: This is a metric dealing with energy efficiency which studies the trade-off between the throughput and the requested power to achieve this throughput. It may be relevant for all use cases where cheap devices with low energy consumption are assumed. Link level simulations will assess if this metric for M2M communications is smaller than for LTE ones.

Detailed description:

The introduction of M2M devices as new actors of the LTE network raises several challenges that must be addressed. Indeed, two different worlds need to coexist and share the same resources: first, the LTE core network with LTE users directly connected to the base station (eNodeB) through a LTE interface; second, the capillary (and heterogeneous) network with LTE-M or non LTE-M devices connected to a M2M Gateway via any air interface (LTE-M or not), while the M2M Gateway is interconnected to the LTE base station via a LTE-M interface. In EXALTED scenarios, there is also the possibility for LTE-M standalone devices (i.e., LTE-M devices which do not belong to any capillary network) to directly access the base station via a LTE-M connection.

Techniques and algorithms exist for both networks but since issues differ completely from one world to the other, solutions designed for the core network do not apply to the capillary network, and vice versa. Furthermore, decisions taken by network operators and service providers that deal with vendor strategies and implementation are not necessarily in line with the coexistence of these two networks. Nevertheless, this challenge must be handled with great care.

In this section a generic solution is proposed to address at the scheduling level the coexistence of these two networks. Scheduling has already been handled in Section 4.1. Furthermore, Section 4.1.2 approaches this issue with a form close to the one which will be developed in this section. Nevertheless, Section 4.1.2 rather focuses on solutions compliant with LTE standardisation, while in this Section we propose a more generic proposition that tries to envision how both networks could operate at the scheduling level.

In the cellular network, the scheduler in charge of managing the access to the medium is designed for native LTE traffics. As it was said in Section 4.1.2, scheduling algorithms are not part of the LTE specifications. This scheduler must be adapted to integrate a new type of traffic generated by M2M communications. This new traffic is called Machine-Type Communication (MTC). From the core network point of view, this traffic is initiated by M2M

Gateways which are interfaces between the core network and the heterogeneous network, both in uplink and downlink modes (there is just the exception of LTE-M standalone devices which directly access the base station).

A M2M Gateway could aggregate all QoS requests coming from (LTE-M and non LTE-M) devices within its capillary network and then forwards them to the core network and its scheduler. In uplink mode (LTE-M or non LTE-M device → Core network), the M2M Gateway could negotiate with the central scheduler the access to the medium and then could schedule these resources within its capillary network to initiate the M2M communications. In downlink mode (Core network → LTE-M or non LTE-M device), the M2M Gateway would be in charge of addressing packets to the corresponding LTE-M device according to its scheduled resources and QoS constraints.

However, MTC traffic is completely different from native LTE traffics. Indeed, MTC traffic is characterized by a huge amount of LTE-M devices that access the medium for low data rate communications on a rather sporadic or bursty fashion. Except for specific scenarios (urgency, video monitoring), latency constraints are not so essential. On the other hands, LTE traffics target high data rate communications with stringent latency constraints. Furthermore, LTE traffics differ from MTC traffic in the sense that LTE traffics stand for human-to-human communications which are expectable. At network subscription, LTE customers indicates their type of traffic to the Call Admission Control (CAC) entity; each traffic is normalized (data rate, packet size, etc.) and hence the scheduler knows in advance how to schedule resources to LTE devices that request an access to medium. Such knowledge is not possible for LTE-M devices because of randomness and heterogeneity. As it was proposed in Section 4.1.2, scheduling algorithms should avoid excessive signalling while supporting the large number of M2M devices but without jeopardizing the services of LTE users.

New adaptive algorithms must be designed at the central scheduler to take into account dynamics of MTC traffic. Such aspects are not yet present in scheduling algorithms already designed for LTE networks. We propose to design an algorithm that adapts MAC of core network and capillary network to LTE traffics and MTC heterogeneous traffic. Different priority classes will also been considered with urgent and non-urgent traffics. A MAC algorithm specifically designed for M2M Gateways will also be proposed. Such algorithms should be integrated in our proposed “LTE-M” version of the standard.

8. Conclusion and Outlook

A detailed study of the functionality of LTE Releases 8, 9 and 10 has disclosed their inability to serve a multitude of small data rate devices in energy-, spectrum- and cost-efficient way. On the other hand, LTE is being deployed and mayor changes in its overall system architecture are unlikely to be accepted by 3GPP. Therefore, EXALTED will propose improvements that can be easily integrated in the existing system and maintain backward compatibility to previous Releases. In this sense, EXALTED partners intend to contribute to the Release 11 Work Item System Improvements to Machine Type Communications (SIMTC) and a Study Item proposed by Vodafone dealing with Low Cost MTC User Equipments based on LTE.

After this top-down view of LTE-M in section 2, the basic principles of the individual proposals were presented in the following sections together with a description, which of the technical requirements in [1] are addressed. We distinguished between proposals that need a modification of the LTE specification in sections 3-6, and those that are of a more generic nature and can thus be applied almost standard independently in section 7.

Section 3 was about the PHY-Layer of LTE-M. The current LTE system, and here especially its PHY, is primarily designed for high data rate and low latency. However, the envisaged rise of M2M sets other design criteria for the development of a PHY such as energy efficiency, low complexity and reliability. In this sense, the section introduced suggestions for algorithms that aim to accomplish these design criteria and requirements of a M2M system. The focus was on new radio access methods and different antenna techniques enhanced for the requirements of the EXALTED system. First simulation results underlined the initial expectations.

In section 4, we concentrated on the MAC-Layer. The MAC in LTE consists of the actual controller with the functions scheduling including the required information transfer between eNodeB and the UE, random access, timing alignment and DRX. Moreover, multiplexing / demultiplexing with logical channel prioritisation and HARQ are carried out. Efficient MAC protocols are expected to play a key role in LTE-M, which is necessary to support a significantly larger number of devices, compared to the existing cellular networks. It has been found out that the reduction of signalling is of particular importance to fulfil the technical requirements derived in EXALTED, primarily energy efficiency and network overload protection, and that stricter scheduling algorithms are indispensable in order to implement the envisaged scalability, Moreover, improved retransmission schemes are needed, which shall take into account the requirements of M2M networks, namely efficiency and reliability rather than low latency.

Section 5 was closely related to the previous section, but it described improvements of the protocols that belong to the RRC-Layer in the LTE system. Also in the RRC protocol design, reduction of signalling overhead and energy consumption as well as a more efficient usage of the network hardware are envisaged. Several principles were identified that can definitely help to achieve these aims, one concerning the behaviour during inactive periods, one concerning the schemes used for short, infrequent transactions, and one describing a way to deal with extreme scenarios where the C-RNTI space is depleted. In this section we also presented an energy efficient relaying scheme. First simulation results have shown that we can minimise the power allocated to each relay node for a given QoS constraint.

Section 6 was dedicated to broadcast and multicast services. This is seen as fundamental aspect of LTE-M since it is expected that a sizeable portion of downlink traffic in LTE-M will be of broadcast or multicast type. This traffic will typically be of moderate or high volume and will exhibit very high coverage/availability requirements and very low error tolerance. One solution for this could be the application of LDPC -like rateless codes. A first design approach was presented in this report. Future work will be focused on suitable channel codes and parameter optimisations. As a second approach, a collaborative broadcast architecture was disclosed.

Section 7 summarized different methods that facilitate M2M communications but are not part of standardized LTE-M signal processing and protocols. They are basically based on a dynamic adaptation of the system parameters according to measured signals. Firstly, spectrum sensing was discussed. Important parameters like the sensing delay and their minimum requirements have been identified, and a first comparison of candidate algorithms was presented. The second proposal was about energy harvesting. This approach has recently been made possible by introducing rechargeable batteries that can harvest energy from the environment in order to extend the lifetime of the system. Accordingly, it is essential to design the system operation taking into account the nature of the energy harvesting process to increase energy efficiency. Finally, the last proposal builds the bridge to the end-to-end architecture including capillary networks. A generic MAC scheduling solution was proposed that addresses the coexistence of the cellular and the capillary network and can be seen as one interface between both domains. It takes into account the traffic characteristic of M2M communication and foresees a prioritisation scheme.

In the next months, the proposed LTE-M algorithms and concepts will be further elaborated with respect to performance and their ability to fit into the given LTE system architecture. Here we expect close interactions with other EXALTED work packages with respect to transferring our proposals to one of the test beds, submitting contributions to the ongoing specification of LTE Release 11 and refining the overall EXALTED system architecture.

Appendix

A1. Technical Requirements

The basic working assumption in EXALTED has been the claim that the current LTE system does not support M2M adequately in manifold manner [1]. M2M raises technical requirements that were not considered during the standardisation of LTE, whose aim has been from the beginning the support of high rate data services. As the number of possible applications for M2M is enormous, and each application has its own particular technical requirements, EXALTED has identified the following key use cases [1]:

Intelligent Transportation System (ITS)

ITS represents communication methods and technologies that enable exchange of information between vehicles and transport infrastructure on one side and ITS applications on the other. ITS applications manage various parameters related to automotive industry, such as transportation time, traffic collision avoidance, on-board safety, fuel consumption, and many others. Use cases, based on ITS scenario, are divided into the 2 categories: vehicle-to-application server communication and vehicle-to-vehicle scenarios. The following automotive applications assuming vehicle-to-application server communication: Remote Monitoring of Vehicle Data (mileage, engine temperature, etc.), In-Vehicle M2M Diagnosis (in-vehicle wireless check, for example: rear light bulb), Railway remote monitoring and failure detection. EXALTED also examines the following vehicle-to-vehicle scenarios: Parking time check, Vehicle collision management, Gateway vehicle for Car-to-Car (C2C) communication.

ITS applications should deal with high mobility patterns. Most ITS applications (except for in-vehicle diagnosis) occur in outdoor environments especially are envisioned in urban environments. The mobility capability of vehicles envisions applications that require large geographical areas of coverage to monitor a fleet of vehicles. In-vehicle diagnosis and railway remote-monitoring scenarios may require high energy-efficient operation. Most ITS scenarios consider medium-to-high number of devices. In applications such as in-vehicle diagnosis, the number of devices might be fairly low. When safety is involved, reliability is a critical requirement in ITS. Applications such as parking time-check may not require strong reliability as retransmissions can be scheduled. However, real-time applications involving safety pose strong reliability requirements. Remote-monitoring application for the ITS scenario do not require real-time performance but require server initiated communications, as they may be based on a data polling scheme, from a central server to the devices monitoring the environment. Applications such as C2C communications may use multi-hop connectivity to create communication paths between vehicles, also multi-hop architecture applies to the railway application, where the devices installed along the rails may establish a multi-hop communication architecture.

Smart Metering and Monitoring (SMM)

Smart Metering and Monitoring (SMM) applications include several scenarios, mainly differentiated by the application scale (e.g. home, industry or environmental), which also determines each corresponding set of requirements. SMM describes several types of use cases in the EXALTED project, such as: Environmental Monitoring (EM), Energy Smart Metering (ESM), Industrial Monitoring (IM) and Security - Surveillance (SS) features. Energy Smart Metering (ESM) or Building Management means that the consumers might use a tool like an Energy gateway, to manage remotely a wide range of devices connected over short range RF, in order to read remotely their electricity consumption, shut down the heating in their house, etc. IM is focused on remote control and remote monitoring of industrial environments. Remote control is intended to interact with physical mechanisms, for instance lighting, access facilities or elements on an industrial assembly line. EM is referred to remote water quality monitoring, remote water level monitoring (e.g. for flood management control), remote optimum management of water usage for agricultural activities, remote air-quality monitoring and remote monitoring of sensitive areas (e.g. forests) for preventing environmental destructions. SS is relating to surveillance of buildings, prevention of theft, intruder detection, tracking of valuable objects or humans.

Remote management includes cases in which human-intervention is either too costly or too tedious or too time-consuming (e.g. isolated devices). There is a self diagnostics, i.e. devices may report failure or malfunctions. SMM applications use very low volumes of data due to bursty transmissions. The devices should communicate with LTE/LTE-M through a gateway, when possible. Due to the expected large volume of devices for such applications, expensive devices may not be suitable. Connection is not always triggered by the devices but by server. Connectivity, reliability and real time performance are application dependent.

E-health

E-health, as a type of M2M application, is growing independently and can be used for patient monitoring, remote diagnostics, activity monitoring, lifestyle suggestions, and personal security. E-health involves new forms of patient-physician interaction and poses new challenges and threats to ethical issues such as online professional practice, informed agreement, privacy and equity issues. E-health is increasing efficiency in healthcare, thereby decreasing costs but also improving quality. Directing patient streams to the best quality providers and on-line scheduling appointments are examples of improved quality of care. E-health creates knowledge bases of medicine and personal electronic records, accessible to consumers over the Internet, and enables evidence-based patient choice, education of physicians through online sources (continuing medical education) and education of consumers (health education, tailored preventive information). E-health enables consumers to easily obtain health services online from simple advice to more complex interventions or products such as pharmaceuticals. New technologies are easily integrated into existing scenarios when they address issues not yet resolved under the current implementations. Contemporary E-health reality consists of a combination of sensors for individual’s vital signs measuring or even tracking his position in case of an emergency. These sensors come in the form of lightweight portable (or even wearable) devices for enhanced user acceptability. Today’s implementations feature the combination of Bluetooth device for short range communications (hence constituting a Personal Area Network – PAN) and General Packet Radio Service (GPRS) access for fulfilling their primary target, that is patient monitoring.

Most E-health applications occur in urban operational environments. Remote management includes cases in which health practitioner intervention is required, for example - change in event recording frequency. E-health applications require small geographical areas of coverage and most scenarios consider small-to-medium number of devices. There is necessity for high reliability due to sensitive nature of data. E-health scenario considers event-triggered connections having benefits on power consumption, network signalling and scheduling. E-health applications may use mesh routing with multi-hop connectivity High mobility is expected because of tracking of moving objects/persons.

When the key use cases were selected, the respective technical requirements have been derived. The use cases served as the basis for the derivation of the technical requirements, but the requirements have general purpose, and they are not only related to these three use cases. These requirements are the background for design, specification and performance assessment of algorithms, protocols, and system concepts in WP3. The following Table A-1 lists all technical requirements. An identification number is allocated to each requirement. The meaning of the acronyms is as follows:

FU Functional requirements
SV Service requirements
NT Network requirements
NF Non-functional requirements
DV Device requirements

For more details, the interested reader is referred to [1].

Table A-1: Technical Requirements

ID	Title	Priority	Dependencies
FU.1	Support of large number of devices	Mandatory	NT.10
FU.2	Efficient spectrum management	Mandatory	
FU.3	Support for diverse M2M services	Mandatory	
FU.4	Network initiated packet-data communication	Mandatory	
FU.5	Local and remote device management	High	
FU.6	Unique identity for devices	High	NT.16
FU.7	Security and provisioning	Mandatory	
SV.1	Overall QoS	Mandatory	NT.14, NF.5
SV.2	Allow multiple service providers on M2M devices	Low	SV.3, NF.3
SV.3	Efficient provisioning of a set of M2M equipments	Mandatory	NF.1, NT.13, DV.1, DV.10
SV.4	Change of subscription	Mandatory	SV.3
SV.5	Delegation and distribution of functionality	Mandatory	NT.4
SV.6	Security	Mandatory	
NT.1	Heterogeneous networks	Mandatory	
NT.2	LTE-M backward compatibility	Mandatory	
NT.3	Minimum number of modifications in network infrastructure	Mandatory	NT.2
NT.4	Support of multi-hop communication	Medium	
NT.5	Half duplex operation of terminals	Mandatory	
NT.6	End to end device to device communication	Mandatory	NT.4
NT.7	Flexible addressing scheme	Mandatory	NT.3
NT.8	Mobility management	Mandatory	
NT.9	Reliable delivery of a message	High	
NT.10	High node density	Medium	NF.6
NT.11	Traffic aggregation	Medium	
NT.12	Self-diagnostic and self-healing operation	Medium	DV.1
NT.13	Multicast and broadcast communication	Mandatory	
NT.14	End-to-end QoS system	Mandatory	NT.6, NT.8
NT.15	End-to-end session continuity	Mandatory	NT.6, NT.8, NT.9
NT.16	Support for dual stack IPv4/IPv6	Mandatory	NT.7
NT.17	Reduced signalling	Mandatory	DV.3
NF.1	Scalability	Mandatory	NF.6
NF.2	Energy efficiency	Mandatory	DV.3, DV.9
NF.3	Extensibility and adaptability	Medium	
NF.4	Real time performance	Medium	
NF.5	Congestion control mechanism	Low	

NF.6	Address space scalability	High	
NF.7	Control signalling integrity protection and encryption	Mandatory	
NF.8	Service provisioning for MNO/SP customers	Mandatory	SV.3
NF.9	Roaming support	Mandatory	NT.8
DV.1	Self organized M2M equipments	Mandatory	
DV.2	Reliable M2M equipments	High	
DV.3	Energy efficient duty cycles	Mandatory	NF.2
DV.4	Location information	High	
DV.5	Location locked M2M equipments	High	
DV.6	Gateway detection and registration	Mandatory	NT.13
DV.7	Protocol translation at the gateway	Mandatory	NT.1
DV.8	Information routing at the gateway	Mandatory	NT.1
DV.9	M2M equipment wake-up	Mandatory	
DV.10	Remote configuration	Mandatory	NT.12
DV.11	Software update over the air	Mandatory	

A2. Traffic Models

In the document D3.2 Study of Commonalities and Synergies between LTE-M and the Heterogeneous Network (WP4) in Chapter 3, detailed description of the traffic modelling is given [33]. In this chapter we will focus on the traffic patterns exchanged between M2M node and server in the case of a payload transmission. In document [34] could be found more detailed description and traffic analysis of M2M applications.

Here, we will analyse one M2M game, i.e. the Virtual race (e.g. virtual bicycle race using real bicycles). In this game the opponents are on different locations, possibly many kilometres away. At the beginning of the race, the corresponding length of a race is agreed (i.e. 10 km or 20 min) between the peers. The measurements are taken by sensors (Global Positioning System (GPS), temperature, humidity, speed, terrain configuration etc.) and are exchanged between the opponents. They are used by the application to calculate the equivalent positions of the participants and to show them the corresponding state of the race (e.g. “you are leading by 10 m”). The number of competitors may be more than two, and all competitors must mutually exchange information, and the applications must present all participants the state of other competitors. For a large number of competitors (hundreds or more), a corresponding application server must be used. During the race they are informed about the place and the distances from each other (e.g. “you are the 3rd behind the 2nd by 10 m and leading before the 4th by 15 m”).

The traffic pattern in this scenario is comparable to a periodic constant low data rate transmission between M2M devices with shorter periods as the end of the race is getting closer. This traffic pattern can also be modelled as a classical ON-OFF traffic model with monotonic decreasing inter-departure time. We can assume that competitors are connected to the LTE network and that sensors are set for competition, so we will analyse only exchange of traffic between the competitors.

The packets containing GPS and sensors data are on the order of 1 kB, which reflects the P_{PL}. Taking into account the typical speeds (of bicycles) in this scenario (rarely higher than 50 km/h = 13.9 m/s), the packets should be exchanged approximately every 100 ms ($F_{Heartbeat}$), which corresponds to a resolution of 1.4 m. Also we can assume that competitors have the periods of low and medium speeds during the competition which corresponds to 10 and 30 km/h. This highly depends on the road topology, but we can assume that 20 % of the competition time riders will have a low speed, than 60 % of the time medium speed and finally 20% they will drive very fast.

If there are only two competitors (or a small number of them), there is no need for existence of an application server. In the case of a large number of competitors (or team competition) there will be a need for an application server. The application should be aware of the positions of all competitors with respect to the end of the race, and, when the competitors are close to the finish, packets should be sent every 70 ms ($F_{Heartbeat}$), which corresponds to a resolution of 1m (GPS accuracy). Data rates are normally not higher than 10 kB/s (about 15 kB/s at the final stage of the competition). A typical number of the competitors considered in this scenario is less than 100.

In Table A-2, a traffic model is presented for the case when there are two competitors only. One should have in mind that in this case there is no application server, so the competitors' devices are exchanging data directly, and the traffic is symmetric in the uplink and the downlink. We can expect that the competition will not take place in the city centres, but rather in the country sides, so we have to observe cell coverage for these areas.

Table A-2: Traffic parameters for Bicycle race - 2 competitors

	Low speeds	**Medium speeds**	**High speeds**	**Finish**
UPLINK – packet length	1 kB every 500 ms	1 kB every 150 ms	1 kB every 100 ms	1 kB every 70 ms
UPLINK- traffic per user	2 kB/s per user	7 kB/s per user	10 kB/s per user	15 kB/s per user
DOWNLINK – packet length	1 kB every 500 ms	1 kB every 150 ms	1 kB every 100 ms	1 kB every 70 ms
DOWNLINK – traffic per user	2 kB/s per user	7 kB/s per user	10 kB/s per user	15 kB/s per user

In the case that there are more competitors or team competitions, then there is a need for application servers who will process all competitors' data. Than all competitors are sending data towards the application server, which are processing data and informing competitors about their position in the competition every second. Please note that downlink traffic doesn't depend on the speed and on the time of the competition, i.e. it is the same amount of data sent every second towards competitors, since human perception and reaction time is about 1s, and there is no need to send data with higher frequency. In the Table A-3 is given traffic model for that case and in the Table A-4 are given values for traffic parameters.

Table A-3: Traffic parameters for Bicycle race

	Low speeds	**Medium speeds**	**High speeds**	**Finish**
UPLINK – packet length	1 kB every 500 ms	1 kB every 150 ms	1 kB every 100 ms	1 kB every 70 ms
UPLINK- traffic per user	2 kB/s per user	7 kB/s per user	10 kB/s per user	15 kB/s per user
DOWNLINK – packet length	1 kB every second	1 kB every second	1 kB every second	1 kB every second
DOWNLINK – traffic per user	1 kB/s per user	1 kB/s per user	1 kB/s per user	1 kB/s per user

Since application is continuously sending data from the beginning of the race without any trigger, we can treat it as "keep alive" traffic. With a small and medium number of competitors, the actual throughput is not critical as the amount of traffic generated by a user will be small. With a large amount of competitors, e.g. 100, the cell capacity limitations have to be considered.

Table A-4: Traffic parameters for Bicycle race - M2M model

Traffic	N	P_{OH}	P_{PL}	$F_{Keepalive}$ $F_{Heartbeat}$	$F_{Trigger}$	$F_{EventThreshold}$
Keep-alive(competitors)	2-100	50	1 kB	70-500 ms	NA	NA
Keep-alive (application server)	2-100	50	N*1kB	1s	NA	NA

Figure A-1 depicts the state diagram for the different states of the virtual race reference model. There are two states: Competitor Status and Virtual Game System. So, competitors are sending their data to the application server, i.e. Virtual Game System which is processing their data, makes the Ranking and after that sends the ranking data to the competitors.

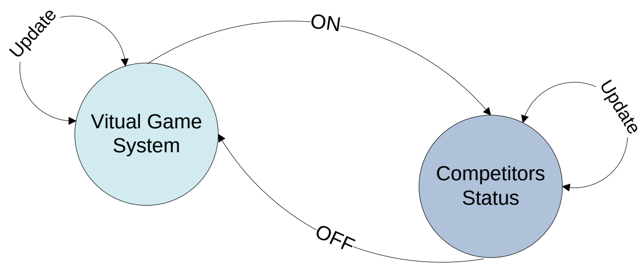

Figure A-1: Virtual Race Reference Model

A3. Summary of Performance Metrics

The performance metrics that will presumably be used for the evaluation of the proposals in Sections 3-7 are listed in Table A-5.

Table A-5: Summary of performance metrics

Metric	Definition
Bit Error Rate (BER)	Bit error rate at the output of the decoder
Packet Error Rate (PER)	A packet represents the information block protected by CRC at the MAC layer.
Packet Loss Rate (PLR)	As PER, but it only counts erroneous packets due to excessive latency.
Peak-to-Average Power Ratio (PAPR)	Ratio of peak power and average power of the transmitted signal in the time domain
Out-of-band radiation (OOB)	Radiated power outside the assigned spectrum
Average number of retransmissions	In case of erroneous transmissions, ARQ and HARQ mechanisms are used to retransmit packets until they are successfully received.
Redundancy overhead	Spent overhead per user for reliable multicast message reception
Throughput	Number of successfully received bits or messages per

	time unit in bit/s or messages/s
Average packet call throughput	It is defined as $$R_{pktcall}(i) = \frac{\sum_k good\ bits\ in\ packet\ call\ k\ of\ user\ i}{\sum_k t_{end\_k} - t_{arrival\_k}}$$ where k = denotes the k[th] packet call from a group of K packet calls where the K packet calls can be for a given user i , $t_{arrival\_k}$ = first packet of packet call k arrives in queue, and t_{end_k} = last packet of packet call k is received by the UE.
Spectral efficiency (sum-rate):	Number of successfully transmitted bits per time unit per frequency unit per cell in bit/s/Hz/cell
Average packet delay per sector	The averaged packet delay per sector is defined as the ratio of the accumulated delay for all packets for all devices received by the sector and the total number of packets. The delay for an individual packet is defined as the time between when the packet enters the queue at transmitter and the time when the packet is received successively by the device.
Access delay	Needed time in order to join the network
Percentage of satisfied users	It is the percentage of users whose packets arrive at the destination within their maximum delay tolerance time interval.
User per cell capacity	Maximal number of simultaneously active users per cell
Range	Maximal possible distance between a M2M device and base station to enable communication with a given QoS, either directly or via a gateway or relay
Coverage	Percentage of area, where M2M devices can connect to a base station, either directly or via a gateway or relay
PHY Control channel and pilot overhead	Percentage of radio resources utilized for signalling, control channels and pilots on PHY layer
Payload to signalling overhead ratio	Ratio between payload and signalling overhead
Paging efficiency	Percentage of specific control channel information for paging procedures in bit/user
Mobility management efficiency	Percentage of specific control channel information for mobility procedures in bit/user
Transmission payload size	Size of the message exchange between 2 peers (e.g. device, cluster head, gateway, device management server). The size depends on the data encoding scheme and compression.
Ratio between transmitted power and achieved throughput (energy efficiency):	Watt/(bit/s)=Joules/bit
Consumed energy per message	Sum of energy spent for signal processing and transmitted energy required for one message
Complexity	Number of operations (e.g. multiplications) to compute the output of functional unit (e.g. decoding)
Number of CSI estimation	Number of CSI estimation per decoded data bit
Number of active antennas	Number of activated antennas compared to the available ones
Number of served users	Number of users that are supported by the evaluated method/service, e.g. number of served multicast users

Spectrum sensing accuracy	Probability of detection of an existing signal and probability of false alarm (erroneous detection of not existing signal) in the sensing algorithm
Spectrum sensing delay	Sensing delay given in OFDM symbols

List of Acronyms

Acronym	Meaning
3GPP	3rd Generation Partnership Project
ABEP	Average Bit Error Probability
ACK	Acknowledgement message
AF	Amplify-and-Forward
AMC	Adaptive Modulation and Coding
AWGN	Additive White Gaussian Noise
BEC	Binary Erasure Channel
BER	Bit Error Rate
BPSK	Binary Phase Shift Keying
BS	Base Station
C-Plane	Control Plane
C-RNTI	Cell Radio Network Temporary Identifier
C2C	Car-to-Car
CAC	Call Admission Control
CDMA	Code Division Multiple Access
CG	Channel Gain
CH	Cluster Head
CRC	Cyclic Redundancy Check
CS	Circuit Switched
CSMA	Carrier Sensing Multiple Access
CGF	Cooperative Gap Filler
CoMP	Cooperative Multipoint
CP	Cyclic Prefix
CSI	Channel State Information
DF	Decode-and-Forward
DFT	Discrete Fourier Transform
DL	Downlink
DMS	Device Management Server
DRX	Discontinuous Reception
DV	Device requirement
E-UTRA	Evolved Universal Terrestrial Radio Access
EM	Environmental Monitoring
EGPRS	Enhanced General Packet Radio Service
eNodeB	Evolved NodeB
EPC	Evolved Packet Core
EPS	Evolved Packet System
ESM	Energy Smart Metering
EXALTED	Expanding LTE for Devices
EXIT	Extrinsic Information Transfer Chart
FDD	Frequency Division Duplex
FDE	Frequency Domain Equalization
FEC	Forward Error-Correcting
FER	Frame Error Rate
FU	Functional requirement
GF2	Galois Field 2
GFDM	Generalized Frequency Division Multiplexing
GPRS	General Packet Radio Service
GPS	Global Positioning System
GSM	Global System for Mobile Communications

H-eNodeB	Home eNodeB
HARQ	Hybrid Automatic Repeat Request
HSPA	High Speed Packet Access
HSS	Home Subscriber Server
IB	In-Band
ICIC	Inter-Cell Interference Coordination
IDFT	Inverse Discrete Fourier Transform
IEEE	Institute of Electrical and Electronics Engineers
IM	Industrial Monitoring
IMS	Internet Protocol Multimedia Subsystem
IMT	International Mobile Telecommunications
IP	Internet Protocol
ITS	Intelligent Transport System
ITU	International Telecommunication Union
LCS	Location Control Service
LDGM	Low Density Generator Matrix
LDPC	Low Density Parity Check
LTE	Long Term Evolution
LTE-A	LTE-Advanced
LTE-M	LTE for Machines
M2M	Machine-to-Machine
MAC	Medium Access Control
MBMS	Multimedia Broadcast Multicast Service
MBSFN	Multimedia Broadcast multicast service Single Frequency Network
MDT	Minimization of Drive Test
MIMO	Multiple Input Multiple Output
MME	Mobility Management Entity
MNO	Mobile Network Operator
MTC	Machine-Type Communication
MU-MIMO	Multi-User MIMO
NF	Non-Functional requirement
NT	Network requirement
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OOB	Out-Of-Band
PAN	Personal Area Network
PAPR	Peak-to-Average Power Ratio
PDCCH	Physical Downlink Control Channel
PDCP	Packet Data Convergence Protocol
PER	Packet Error Rate
PGN	Packet Data Network Gateway
PHY	Physical-(Layer)
PLR	Packet Loss Rate
QoS	Quality-of-Service
QPSK	Quadrature Phase Shift Keying
RACH	Random Access Channel
RAN	Radio Access Network
RB	Resource Block
RF	Radio Frequency
RLC	Radio Link Control
ROHC	Robust Header Compression
RRC	Root Raised Cosine
RRC	Radio Resource Control

RRM	Radio Resource Management
SC-FDMA	Single-Carrier Frequency Division Multiple Access
SIMTC	System Improvements for Machine Type Communications
SINR	Signal-to-Interference-and-Noise Ratio
SMM	Smart Metering and Monitoring
SNR	Signal-to-Noise Ratio
SON	Self-Organizing Network
SP	Service Provider
SR-VCC	Single Radio Voice Call Continuity
SS	Security - Surveillance
SU-MIMO	Single-User MIMO
SV	Service requirement
SWR	Switching Rate
TB	Transport Block
TCP	Transmission Control Protocol
TDD	Time Division Multiplex
TTI	Transmission Time Interval
U-Plane	User Plane
UE	User Equipment
UL	Uplink
VoIP	Voice over IP
W-CDMA	Wideband Code Division Multiple Access

References

[1] FP7 EXALTED consortium, “D2.1 – Description of baseline reference systems, scenarios, technical requirements & evaluation methodology,” project report, May 2011.
[2] 3GPP, “TS 25.913 – Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN)”.
[3] M. Lindström, “RP-070442 – LS on LTE latency analysis,” 3GPP TSG RAN Meeting #36, Busan, South Korea, June 2007, http://www.3gpp.org/ftp/tsg_ran/tsg_ran/TSGR_36/Docs/RP-070442.zip.
[4] 3GPP, “TR 36.913 – Requirements for further advancements for Evolved Universal Terrestrial Radio Access (E-UTRA) (LTE-Advanced)”.
[5] 3GPP, “TS 36.306 – User Equipment (UE) radio access capabilities (Release 10)”.
[6] 3GPP, “TS 36.101 - User Equipment (UE) radio transmission and reception (Release 10)”.
[7] G. Fettweis, M. Krondorf, and S. Bittner, “GFDM-Generalized Frequency Division Multiplexing,” invited paper, VTC2009-Spring, Barcelona, Spain, April 2009.
[8] Taesang Yoo; Goldsmith, A.; , "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," Selected Areas in Communications, IEEE Journal on , vol.24, no.3, pp. 528- 541, March 2006.
[9] L. A. M. R. de Temino, G. Berardinelli, S. Frattasi, and P. Mogensen, “Channel-aware scheduling algorithms for SC-FDMA in LTE uplink,” in Proceedings of the IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '08), September 2008.
[10] O. Delgado and B. Jaumard, “Scheduling and Resource Allocation for Multiclass Services in LTE Uplink Systems,” IEEE 6th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), pp. 355-360, October 2010.
[11] FP7 project SmartSantander, www.smartsantander.eu.
[12] 3GPP, “TR 36.814 – Further advancements for E-UTRA physical layer aspects (Release 9),” v9.0.0, March 2011.
[13] D. S Michalopoulos, G. K. Karagiannidis, T. A. Tsiftsis, R. K Mallik, “Distributed transmit antenna selection (DTAS) under performance or energy consumption constraints,” IEEE Trans. Wireless Commun., vol. 7, pp. 1168-1173, November 2008.
[14] T.J. Richardson, R.L. Urbanke, “Modern coding theory,” Cambridge University Press, New York, 2008.
[15] T.J Richardson, M.A. Shokrollahi, and R.L. Urbanke, “Design of capacity-approaching irregular low-density parity-check codes,” IEEE Transactions on Information Theory, volume 47, no. 2, pp 619-637, February 2001.
[16] M. Lentmaier, G.P. Fettweis, K.Sh. Zigangirov, and D.J. Costello, “Approaching capacity with asymptotically regular LDPC codes,” in Proc. Information Theory and Applications Workshop, pp. 173-177, San Diego, USA, 2009.
[17] G. Cocco, C. Ibars, O. del Rio Herrero, “Cooperative Satellite To Land Mobile Gap-Filler-Less Interactive System Architecture,” Proc. 5th Advanced Satellite Multimedia Systems Conference and the 11th Signal Processing for Space Communications Workshop (ASMS/SPSC), Cagliari, Italy, September 2010.
[18] P. A. Chou, Y. Wu, and K. Jain, “Practical network coding,” in Proc. 41st Allerton Conference on Communication, Control and Computing, Monticello, USA, 2003.
[19] 3GPP, “TS 36.300 – Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN) (Release 10),” December 2010.
[20] 3GPP, “TS 23.401 – General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (Release 10),” January 2011.

[21] H. S. Chen, W. Gao, and D. G. Daut, “Signature based spectrum sensing algorithms for IEEE 802.22 WRAN,” in Proc. ICC 2007, pp. 6487-6492, June 2007.
[22] H. Urkowitz, “Energy detection of unknown deterministic signals,” Proceedings of the IEEE, vol. 55, no. 4, pp. 523-531, April 1967.
[23] A. V. Dandawate and G. B Giannakis, “Statistical tests for presence of cyclostationarity,” IEEE Transactions on Signal Processing, vol. 42, no. 9, pp. 2355-2369, September 1994.
[24] Z. Tian and C. B. Giannakis, “A wavelet approach to wideband spectrum sensing for cognitive radios,” in Proc. International Conference on CrownCom, June 2006.
[25] Y. Zeng and Y. C. Liang, “Eigenvalue-based spectrum sensing algorithms for cognitive radio,” IEEE Transactions on Communications, vol. 57, no. 6, pp. 1784-1793, June 2009.
[26] P. Cheraghi, Y. Ma, and R. Tafazolli, “Frequency-domain differential energy detection based on extreme statistics for OFDM source sensing,” IEEE 73rd Vehicular Technology Conference (VTC), May 2011.
[27] Z. Lu, Y. Ma, and R. Tafazolli, “A first-order cyclostationarity based energy detection approach for non-cooperative spectrum sensing,” in Personal Indoor and Mobile Radio Communications (PIMRC), pp. 554-559, September 2010.
[28] S. Lal and A. Mishra, “A look ahead scheme for adaptive spectrum utilization,” in Proc. Radio and Wireless Conference (RAWCON ’03), pp. 83-86, 2003.
[29] E. Uysal-Biyikoglu, B. Prabhakar, and A. E. Gamal, “Energy-efficient packet transmission over a wireless link,” IEEE/ACM Trans. Networking, vol. 10, no. 4, pp. 487-499, August 2002.
[30] J. Yang and S. Ulukus, “Optimal packet scheduling in an energy harvesting communication system,” IEEE Trans. on Communications, submitted, June 2010.
[31] K. Tutuncuoglu and A. Yener, “Optimum transmission policies for battery limited energy harvesting nodes,” IEEE Trans. Wireless Communications, submitted, September 2010.
[32] M. A. Zafer and E. Modiano, “A calculus approach to energy-efficient data transmission with quality-of-service constraints,” IEEE/ACM Trans. on Networking, vol. 17, no. 3, pp. 898-911, June 2009.
[33] FP7 EXALTED consortium, “D3.2 - Study of Commonalities and Synergies between LTE-M and the Heterogeneous Network (WP4),” project report, August 2011.
[34] FP7 project, Achieving LOw-LAtency in Wireless Communications (LOLA), http://www.ict-lola.eu/deliverables/wp3-traffic-measurement-and-modelling.