IST-4-027756 WINNER II

D6.13.14 version 1.1

WINNER II System Concept Description

Contractual Date of Delivery to the CEC:	*November 30, 2007*
Actual Date of Delivery to the CEC:	*January 18, 2008 (updated public version)*
Author(s):	**Gunther Auer, Martin Döttling, Klaus Doppler, Jörn von Häfen, Niklas Johansson, Kari Kalliojärvi, Thierry Lestable, Jijun Luo, Johan Nyström, Magnus Olsson, Afif Osserian, Mikael Sternad, Tommy Svensson, Carl Wijting**
Participant(s):	**EAB, NSN, NOK, CTH/UU, SEUK, DoCoMo**
Workpackage:	**WP6 System Concept**
Estimated person months:	*32*
Security:	**PU**
Nature:	**R**
Version:	1.1
Total number of pages:	172

Abstract:

This document contains a top-down view of the WINNER system with respect to logical node architecture, protocol architecture and cooperation architecture to give a basic understanding of the WINNER concept and as a complement to performance evaluations and design examples. The document also describes some of the important cross-layer optimizations that have been made. In addition to describing the concept, several example reference designs have been studied and are described.

Keyword list:

System concept, system architecture, protocol architecture, logical node architecture, cooperation architecture, cross-layer optimization, system design, deployment scenarios

Disclaimer:

Executive Summary

The WINNER system concept has been developed within the IST-WINNER II project of the FP6 research program.

This deliverable is started with a short at-a-glance- overview of system capabilities, and key features of the WINNER concept.

The deliverable shows the top-down framework for the final WINNER concept. Furthermore, for a number of important functionalities that are not part of the concept as such, but that will enhance the performance significantly, some best-choice reference design algorithms for various scenarios are described.

The concept consists of the logical node architecture and protocol and service architecture as well as cross-layer interactions and key RRM mechanisms.

These architectures form a common abstract description of functional relationships for all possible designs optimised for a certain scenarios.

The interfaces between the logical nodes are described and some basic functional interactions are provided. Especially the MAC and PHY protocol layer descriptions are given in some detail so that a basic understanding of the lower layer functionality is provided.

The solutions for the logical node architecture are partly inspired by the 3GPP LTE/SAE architecture, but also contain many novel features such as a spectrum server for optimised spectrum management, and optional relaying functionality.

These architectures are designed to accommodate the WINNER requirements over a large span of scenarios including Wide Area, Local Area and Metropolitan Area deployments with or without relaying.

Reference designs are given for these scenarios that will serve as examples of best use, and to show the flexibility of the WINNER system to handle many types of scenarios.

Authors

Partner	Name	Phone / Fax / e-mail
Ericsson AB		
	Johan Nyström	Phone: +46 8 757 05 86 Fax: +46 8 585 314 80 E-mail: johan.nystrom@ericsson.com
	Niklas Johansson	Phone: + +46 8 50877860 Fax: +46 8 757 57 20 E-mail: niklas.j.johansson@ericsson.com
	Magnus Olsson	Phone: +46 8 585 307 74 Fax: +46 8 585 314 80 E-mail: magnus.a.olsson@ericsson.com
	Afif Osseiran	Phone: +46 8 585 32670 Fax: +46 8 757 57 20 E-mail: afif.osseiran@ericsson.com
Nokia Siemens Networks GmbH & Co. KG		
	Jörn von Häfen	Phone: +49 89 636 46228 Fax: +49 89 636 1346228 E-mail: joern.von_haefen@nsn.com
	Jijun Luo	Phone: +49 89 636 45268 Fax: +49 89 636 45591 E-mail: jijun.luo@nsn.com
	Martin Döttling	Phone: +49 89 636 73331 Fax: +49 89 636 13 73331 E-mail: martin.doettling@nsn.com
Nokia		
	Kari Kalliojärvi	Phone : +358 40 483 6232 Fax: +358 7180 35935 E-mail: kari.kalliojarvi@nokia.com
	Carl Wijting	Phone; +358 50 486 0564 Fax: +358 71 803 6857 E-mail: carl.wijting@nokia.com
	Klaus Doppler	Phone: +358 50 48 76673 Fax: +358 71 803 6857 E-mail: klaus.doppler@nokia.com
CTH/UU		
	Mikael Sternad	Phone: +46 704 250 354 Fax: +46 18 555096 e-mail: mikael.sternad@signal.uu.se
	Tommy Svensson	Phone: +46 31 772 1823 Fax: +46 31 772 1782 E-mail: tommy.svensson@chalmers.se

SEUK		
	Thierry Lestable	Phone: +44.1784.428600 (Ext.720) Fax: +44.1784.428624 E-mail: thierry.lestable@samsung.com

DoCoMo (DoC)		
	Gunther Auer	Phone:+49 89 5682 4219 Fax: +49 89 5682 4301 E-mail: auer@docomolab-euro.com

Table of Contents

Terminology

General

Radio Access Technology	**RAT**	The radio access technology (RAT) is the air interface that is used to allow the link between User Terminal and Base Station or Relay Node of the RAN. This includes also multi-hop/relaying elements.
Deployment Concept		The term “Deployment Concept” describes network element types and their functions (i.e. logical network elements), (a) how these network element types are linked in a network topology, (b) how logical network elements are mapped onto physical network elements and (c) where physical network elements are deployed according to the radio propagation scenarios for which the deployment concept is applicable.
WINNER Radio Access Network	**WRAN**	The radio network composed of WINNER Base stations, Relay stations, Gateway nodes, RRM servers, and Spectrum servers.

Physical Network Elements

Physical Network Element		A physical network element denotes a physically existing device in the RAN that incorporates certain functionality, thereby representing one or possibly even more logical network nodes.
(Physical) Base station	**BS**	A stationary physical network element serving relay nodes or user terminals via its radio access capabilities. Base stations are interconnected with network elements belonging to the RAN. A physical base station contains one or more base station logical nodes.
(Physical) User terminal	**UT**	A physical network element used by the end user to access a service or set of services.
(Physical) Relay node	**RN**	A physical network element serving other RN or UT in a given geographical area via its radio access capabilities. It is wirelessly connected to a base station, another relay node and/or a user terminal and forwards data packets between these network elements.
(Physical) Radio access point	**RAP**	A physical network element in the radio access network responsible for radio transmission and reception to or from the user terminal via its radio access capabilities. A RAP can be either a relay node or a base station.
Access System		The access system is used to connect the WINNER user terminals to the base station either directly or via relay nodes. The elements of the access system are the WINNER base stations and the WINNER relay nodes.
Feeder System		The feeder system is the transport system used to feed the base stations. The distinctive characteristic compared to the access system is that WINNER users shouldn't connect to this network directly. The transmission technology used by the feeder system could be wireless or wired and is irrelevant and transparent for the final user.
Site	**-**	A site is defined as the physical co-location of base station hardware serving a set of antennas. Users may be connected to a site either directly or through relay nodes.

Logical Nodes

Logical Node	**LN**	A Logical Node is defined by the service (or group of services) it provides towards other nodes and the service (or group of services) it requires from other

		nodes. Identical Logical Nodes terminate an identical set of protocols and provide/require the same group of services. One physical element can comprise one or several LNs.
Base Station Logical Node	**BS_{LN}**	A logical node terminating the transport network layer protocols on the network side as well as the radio protocols on the UT and RN side. It contains a single MAC entity corresponding to a single cell, and it manages the logical relay nodes connected to it.
Relay Node Logical Node	**RN_{LN}**	A logical network node with relaying capabilities that is connected to a BS_{LN}, UT_{LN} or another RN_{LN}. Like the BS_{LN} it terminates the radio protocols (MAC and PHY) on the UT side as well as on the BS side and, in case of more than two hops, also on the RN side. The RN_{LN} does not terminate the transport network layer protocols. It contains a single MAC entity corresponding to a single cell.
User Terminal Logical Node	**UT_{LN}**	A logical node comprising all functionality necessary for it to communicate directly with another UT_{LN} or the RAP.
Gateway nodes	**GW_IPA_{LN}** **GW_C_{LN}**	The gateway functionalities are performed by GW_IPA_{LN} and GW_C_{LN}. Generally, the former logical node is an IP anchor (IPA) for the user, where also the data is routed, whereas the latter node is a control node that controls various aspects of the data flows in the former node. Thus, certain functionalities are represented in both logical nodes, however different aspects are captured in the different types of nodes. Each **UT_{LN}** is normally associated with one **GW_IPA_{LN} - GW_C_{LN}** pair regardless of serving BS.
Spectrum Server Logical Node	**$SpectrumServer_{LN}$**	The $SpectrumServer_{LN}$ is a logical node that contains the centralised Spectrum Sharing and Spectrum Assignment related functions.
RRM server Logical Node	**$RRMserver_{LN}$**	Optionally, a $RRMserver_{LN}$ can be used. It terminates only network internal protocols (i.e. does not have a direct relation to the terminals). It may carry load sharing or resource partitioning functions and is for further study.
Cooperative RRM	**CoopRRM**	The CoopRRM will be responsible for the decision making process of the cooperation mechanisms (handover, admission control and QoS management).

Links, Flows, Cells and Handovers

Flow		A flow is a packet stream from one source to one or several destinations, classified by QoS requirements, source and destination(s).
Flow Class		A flow class is identified by its attributes (guaranteed bit rates versus non-guaranteed bit rates, maximum bit rates, delay budget, loss tolerance etc). These attributes are configurable and reconfigurable as new flows are set up or multiplexed onto an existing flow class and are governed by filters in the BS and UT that map flows onto flow classes and its attributes.
Cell		A cell is defined by the geographical coverage area of its broadcast channel.

Relay enhanced cell	**REC**	The geographical area covered by the broadcast channels of cells whose resources are managed by a single base-station and its connected relay nodes.
Multi-Homing (Multi – RAN Transmission)	-	Multi-homing means that a UT is associated to more than one RAN simultaneously.
Multi-Mode-Transmission	-	Multi-mode-Transmission means that a UT is connected by more than one link to different cells of one WINNER RAN. These cells use different spectrum bands.
Handover	**HO**	A Handover is a change in the set of links between a RAP and a UT. This includes a hard "switch" from one cell to another, moving into and out of a multi-mode-transmission and a changing of the links used for multi-mode-transmission.
Intra-system HO	-	Intra-system HO is a handover between two different radio cells within the same WINNER system.
Inter-cell HO	-	Inter-cell HO is an intra-mode-handover between WINNER cells operating in the same system mode at the same frequency.
Inter-frequency HO	-	Inter-frequency HO is an intra-mode handover between WINNER cells operating in the same system mode but at different frequencies.
Inter-system HO	-	Inter-system handover: An inter-system handover is a handover between two different radio systems e.g. WINNER <->WLAN, UMTS <->WINNER. Two subcategories are distinguished: inter-system handover of radio networks belonging to the same operator and inter-system handover of radio networks belonging to different operators.

Logical Channels

Logical channels		Logical channels are defined as the interface between the RLC and the MAC.
Broadcast Control Channel	**LBCCH**	A downlink channel for transmitting broadcast control information.
Paging Control Channel	**LPCCH**	A downlink channel for transferring paging information whenever the network does not know the location cell of the terminal.
Common Control Channel	**LCCCH**	A bi-directional channel for control information between the network and such terminals that are not having radio resource control connection with the network.
Dedicated Control Channel	**LDCCH**	A point-to-point bi-directional control channel that is established after RRC connection setup and it transmits dedicated control information between a terminal and the network.
Multicast Common Traffic Channel	**LMTCH**	**A** unidirectional point-to-multipoint traffic channel for transfer of dedicated user information for all or a group of specified terminals.
Multicast Common Control Channel	**LMCCH**	**A** unidirectional point-to-multipoint control channel for transfer of control information for all or a group of specified terminals.
Dedicated Traffic Channel	**LDTCH**	A point-to-point (both uplink and downlink) traffic channel that is dedicated to one user for the transfer of user information.
Protocol Data Unit	**PDU**	Output from a protocol layer
Service Data Unit	**SDU**	Input to a (protocol) layer. A packet in a logical channel is a MAC SDU and a RLC PDU

Transport Channels

Transport channels		Transport channels are defined as the interface between the MAC and the PHY.

Transport Broadcast Channel	**TBCH**	A downlink channel for broadcasting system information, etc., to all terminals inside the cells coverage area.
Transport Paging Channel	**TPCH**	A downlink channel used for broadcast of paging information into an entire cell.
Transport Random Access Channel	**TRAC**	A contention-based uplink channel for initial access to the network.
Transport Shared Channel	**TSCH**	A point-to-point data channel for both the uplink and the downlink for both user and control data.
Transport Multicast Channel	**TMCH**	A separate transport channel for multicast (MBMS).
Protocol Data Unit	**PDU**	Output from a protocol layer.
Service Data Unit	**SDU**	Input to a (protocol) layer. A packet in a transport channel is a PHY SDU and a MAC PDU.

Physical Channels

Physical Broadcast Channel	**PBCH**	Downlink physical channel for broadcasting system information etc. to all terminals inside the cell coverage area.
Physical Downlink Frame Control Channel	**PDFCC**	A physical downlink channel that is used to signal information in a downlink slot related to downlink data transmission within that slot and uplink transmission in the subsequent slot.
Physical Downlink Control Format Indicator Channel	**PDCFC**	Indicates layout of the PDFCC. Is broadcasted.
Physical Frequency-Adaptive Data Channel	**PADC**	Physical channel for frequency-adaptive transmission for point-to-point user-data transfer in downlinks and uplinks.
Physical Non-Frequency Adaptive Data Channel	**PNDC**	Physical channel for non-frequency-adaptive transmission for point-to-point user-data transfer in downlinks and uplinks.
Physical Multicast Broadcast Channel	**PMBC**	Is used to carry multicast and broadcast service in downlinks. Uses non-frequency-adaptive transmission with safe selection of modulation and code rate, optionally combined with cooperative multi-RAP transmission. Physical resources of same type as for PNDC.
Physical Uplink Control Channel	**PUCH**	Controls urgent uplink control signalling. Well protected by FEC, does not use HARQ process.
Physical Random Access Channel	**PRAC**	A contention-based uplink channel used for acquiring time alignment.

MAC and PHY specific terms

Medium Access Control	**MAC**	
Physical Layer	**PHY**	
Cyclic Redundancy Check	**CRC**	Parity bit sequence used for error detection. Comprises an inner error detection code.
Forward Error Correction encoding	**FEC**	
Low Density Parity Check	**LDPC**	The preferred FEC scheme for large FEC blocks in the WINNER II reference design is quasi-cyclic block LDPC encoding.
Multiplexing	**MUX**	Concatenation of several RLC PDUs that belong to different flow classes. Is optionally performed at the MAC layer before transmission.
MAC Re-transmission unit, Transport Block	**RTU**	Name of the MAC PDU. Represents the segment to be protected by incremental redundancy retransmission by the HARQ process. Each RTU is protected by one CRC sequence.

FEC block		Part or whole of a RTU that is encoded by a forward error correction (FEC) code. Normally, the whole RTU + the CRC code is one single FEC block.
Hybrid ARQ	**HARQ**	Link retransmission scheme, which uses incremental redundancy.
HARQ segment or code block		A transmitted (punctured) FEC block. Is alternatively denoted a code block.
IR block		Incremental redundancy block, HARQ segment that adds reduncancy bits.
Frequency-adaptive transmission	**FA**	PHY transmission scheme that utilizes the frequency-selective small-scale fading.
Non-frequency-adaptive transmission	**NFA**	PHY transmission scheme that averages over the frequency-selective fading.
Resource division		Pre-allocation of transmission resources to be used for FA and NFA transmission, respectively.
Resource partitioning		Pre-allocation of transmission resources (power and frequency) for use by BS and RN.
Generalized multicarrier	**GMC**	Basic transmission technology of WINNER. The special case of ordinary Cyclic-prefix OFDM is used in downlinks. The DFT-precoded special case of block-IFDMA is used in non-frequency-adaptive uplinks.
Channel symbol		One subcarrier by one OFDM symbol duration. Smallest time-frequency resource that can be individually modulated.
Chunk		Rectangular block of time-frequency transmission resources. Consists of n_s subcarriers by n_t OFDM symbols, or $n_s n_t$ channel symbols. Represents the basic allocation unit for resource partitioning and resource division.
Layer		Spatial degree of freedom in generic transmitter
Chunk layer		Chunk within one spatial layer of the spatial transmit scheme. Represents the basic allocation unit used for frequency-adaptive transmission.
Block		Rectangular block of time-frequency transmission resources, with size ≤ one chunk. Represents the basic allocation unit used by the B-EFDMA and B-IFDMA schemes for non-frequency-adaptive transmission.
Frequency division duplex	**FDD**	Uplink and downlink use different, paired, frequency bands.
Time division duplex	**TDD**	Uplink and downlink use the same frequency band.
Half-duplex FDD		Slotted FDD transmission where UT does not transmit and receive simultaneously.
Slot		Time interval for uplink or downlink transmission in half-duplex FDD and in TDD.
Frame		Temporal resource unit.A frame contains two slots in time. The frame duration equals one downlink slot and one uplink slot for half duplex FDD and for TDD transmissions. It also includes duplex guardtimes in TDD.
Super-frame	**SF**	Time-frequency-spatial unit on the physical channel. Consists of a preamble followed by *n* frames.
Scheduling		Over-all control of the transmission within a frame. Scheduling is executed per slot and is composed of *payload selection, resource allocation, resource mapping and link adaptation.*
Payload selection		Decision on what flow classes to transmit in the scheduled slot.
Resource allocation		Allocation of time-frequency spatial transmission resources within a slot to MAC PDUs.
Resource mapping		Decision that determines the mapping of HARQ segments onto a set of transmission resources, using a set of link adaptation parameters.
Link adaptation		Decision that determines the set of modulation parameters, the spatial layer/beam and the code rate used for transmitting

		a code block.
Spatial Division Multiple Access	**SDMA**	Allocation of transmissions from/to several UTs to different spatial layers within the same time-frequency chunk.
Signal to interference and noise ratio	**SINR**	
Channel Quality Indicator	**CQI**	Measurement of SINR.
Channel State Information	**CSI**	Measurement of complete complex channel gain matrix.
Paging Indicator	**PI**	Indicates range of global UT addresses for which paging requests may follow in the following Paging Message.
Paging Message	**PM**	
Mutual Information based Adaptive Coding and Modulation	**MI-ACM**	Reference design technique for frequency-adaptive transmission.
Spatial Adaptation		Selection of appropriate spatial transmit scheme, individually per flow class.
Configuration table	**CT**	Part of downlink out-band control signalling for frame allocation. Broadcasts structure of subsequent tables.
Length indicator	**LI**	Part of downlink out-band control signalling for frame allocation. Contains pointer to start of AT. Is broadcasted.
Allocation table	**AT**	Part of downlink out-band control signalling.
Transport format table	**TFT**	Part of downlink out-band control signalling.

1 Introduction

The WINNER project Phase I (2004-2005) and Phase II (2006-2007) are ambitious research projects aiming at identification and assessment of key technologies for Beyond 3G mobile systems, and to define a system concept as well as suitable reference designs for a wide range of scenarios. The goal of the WINNER mobile access network is a system that is highly flexible and efficient and can provide a wide range of services to a multitude of users in many different environments.

The main capabilities of the WINNER system are highlighted in Chapter 2 to serve as a 'teaser' and provide reading instructions and refernces to relevant sections and other WINNER deliverables.

This document describes the WINNER system concept in terms of logical node architecture and protocol architecture and provides such reference designs. Furthermore, some of the most important design criteria are motivated in Chapter 2.

An important requirement for communication systems and networks is interoperability between different vendors' equipment. The main advantage of interoperability is that resources are exploited in an efficient way and that economies of scale can prevail. This contributes to the benefits of subscribers, service providers, manufacturers, and to the entire economy and society as a whole.

For this purpose, this document uses a top-down approach to describe the WINNER Radio Access Network (WRAN) based on logical nodes and interfaces. Furthermore, this document describes the framework for protocol and service architecture that is necessary for interoperability. Together, they represent the WINNER II ***system concept***.

Numerous aspects and results from WINNER are important for the performance and flexibility of the concept, although they would not necessarily have to be standardized from an interoperability point of view. The collection of algorithm proposals and best design guidelines represents the WINNER II ***reference design***.

The document is arranged as follows:

Chapter 2 contains a brief overview of the main capabilities of the WINNER system as well as motivations for some of the main design decisions for the system concept and the elements of the reference design.

Chapter 3 describes the logical node architecture and interfaces between logical nodes

Chapter 4 describes the protocol and service architecture using a layered approach.

Chapter 5 describes some important key technology issues that cannot readily be described in a per-layer manner. This includes cross-layer topics and some functionality in the network that is not terminated in the UT.

Chapters 3, 4, 5 and 6 together define the system concept. It provides comprehensive, scalable and flexible access network architecture.

Chapter 7 then contains the reference design elements. That is, the reference design contain examples of scenario specific algorithms, parameter settings or other vendor or operator specific settings based on the tools given in the concept. The chapter also illustrates how the WINNER logical nodes can be mapped onto and combined into physical network elements (physical nodes) in different deployment scenarios.

Hence, the material in Chapter 7 is not to be standardized in a standardization body, it merely provides 'best use' example or preliminary guidelines.

The WINNER system has been assessed in three example scenarios: Wide Area (see [WIN2D61310]), Metropolitan Area (see [WIN2D61311]), and Local Area (see [WIN2D61312]). A substantial part of the WINNER research effort has gone into assessment of various scheduler principles, algorithms for adapting to the environments, and other areas that are part of the reference designs.

Finally, the document is concluded in Chapter 8.

2 Design principles and main characteristics

2.1 WINNER concept elements and reference design

In order to meet the requirements [WIN2D6114], the WINNER radio access network has been designed as a packet-oriented, user-centric, always-best concept.

An always-best solution, which provides competitive performance in a wide variety of situations, is a challenging design goal. Here, parameterisations can be used to provide added flexibility and maximum efficiency depending on the particular radio environment, usage scenario, economic model, etc. The always-best solution is further supported by the flexible protocol architecture of the WINNER radio interface and incorporates mechanisms for both long-term and short-term adaptation. Both relaying and advanced spatial processing is integrated elements of the system architecture, to be used in a way appropriate in each particular scenario.

The WINNER access network targets fulfilment of the ITU-R M.1645 recommendation according to the views and amendments specified in the WINNER deliverable [WIN1D71].

The always-best solution is enabled by several innovative key components. We here provide an overview of where the descriptions of these components can be found in the concept part (chapters 3, 4, 5, 6), the reference design (chapter 7), and the other main WINNER project reports.

Key concept components include:

- ***A Flexible logical node architecture*** that uses as few logical nodes as possible to keep the number of interfaces small. The function grouping is defined such that it enables and encourages flat, flexible, scalable and cost efficient physical node implementations. The logical nodes are defined such that physical RAN implementations can efficiently use them as pooled resources (to avoid single point of failure, for load balancing and trunking gain purposes).
- ***A flexible protocol architecture*** enabling efficient interworking between different system parameterizations is outlined in section 4.1 and described in chapter 4. This architecture focuses on the three lowest layers of the OSI stack. The two lowest layers, represented by the physical (PHY), Medium Access Control (MAC) and Radio Link Control (RLC) sub-layers are assumed present in all base station (BS), user terminal (UT) and relay (RN) logical nodes. This enables an efficient co-design of these layers and supports low latencies over the air interface.
- ***Relay-enhanced cells*** are an integrated part of the concept. Wherever cost efficient, a deployment can utilize advanced decode-and-forward layer 2 relay nodes to optimize the deployment, cost, extend the range of transmission, cover shadowed areas, and re-distribute the offered capacity between centers of cells and cell borders. Relay-enhanced cells have been developed and studied in [WIN2D351] and [WIN2D352].
- ***Design and support for operation in shared spectrum and using inter-system coordination,*** are likely to become important elements of future broadband wireless systems. System functionalities that support shared spectrum use and inter-system coordination have been developed within WINNER [WIN2D481], [WIN2D591], and [WIN2D592]. The required concept elements are here described in the chapter 6 and Annex B on RRM mechanisms. Important elements of the reference design are interference mitigation mechanisms (section 7.8), handover mechanisms and technologies for sharing spectrum between systems and operators (section 6.1 and section 7.11).
- ***Highly optimized user plane processing with respect to overhead***. Inband packet control signalling is cross-layer optimized to provide small overhead by means of segmentation and concatenation at RLC, and multiplexing at MAC providing one single transport block per frame per UT, see section 5.1. Cross layer design of HARQ and RLC-ARQ minimizes overhead and status reporting. For both case, it is ensured that multi-hop operation is supported efficiently, see section 5.1.
- ***MAC designed for short radio interface delays***. A low latency over the air interface is desired for several reasons. It enables adaptivity with respect to fast channel variations, it facilitates high-throughput TCP/IP traffic, low latency services, and it enables fast link retransmission,

which is of advantage for the performance perceived at higher layers. The system concept provides a small minimal delay over the air interface, 1 ms in downlinks and 2 ms in uplinks over single hops, by a combination of short frame durations and tight feedback control loops, see sections 4.7.1 and 5.2.

- ***The use of advanced interference control mechanisms***. Multi-antenna transmission and reception is integrated into the WINNER system concept. It offers versatile tools for suppressing interference at receivers [WIN2D473] and intelligent transmit processing [WIN2D472]. The WINNER reference multi-antenna transmit scheme for wide area deployment, a grid of fixed beams, results in a significant interference reduction in neighboring cells [WIN2D472]. The WINNER concept also enables the use of additional mechanisms such as resource allocation targeting interference avoidance by coordinated scheduling across base stations and relay nodes or the use of (spatial) precoding over distributed antennas. An overview of the interference mitigation tools of the reference design are provided in section 7.8.
- ***An optimized physical layer design***. The WINNER physical layer concept is based on generalized multicarrier transmission. This technique is based on cyclic-prefix OFDM in combination with extra DFT precoding steps. Both single-carrier and multi-carrier waveforms can be generated and received within one transmitter/receiver chain with low computational complexity [WIN1D210]. The basic resource elements and transmission techniques assumed in the WINNER concept are outlined in sections 4.7.1 and 4.8. The general transmission technique can be applied in different configurations to ensure low complexity, high spectral efficiency, and high granularity of resource elements. The configurations selected for the reference design are described in sections 7.2, 7.3, and 7.4. The physical layer processing is described in detail in [WIN2D223], [WIN2D233] and [WIN2D341].
- ***A spatial multi-user link adaptation concept,*** which can be applied to a wide range of deployments, operational scenarios, propagation channel, service requirements, and terminal capabilities. The core elements are:
    - A scheduler that uses tight inter-layer interaction between the RLC, MAC, and PHY layers to accomplish joint link adaptation channel-aware scheduling (section 4.7.2).
    - Two main resource allocation principles (section 4.8.1): Frequency-adaptive transmission uses individual link adaptation within rectangular time-frequency units (chunks), supported by efficient control signaling. Non-frequency-adaptive transmission uses smaller time-frequency blocks for diversity-based transmission. Both use advanced channel coding schemes outlined in section 4.8.3 and Annex A.2, with complete descriptions in [WIN2D223].
    - Control signaling that supports adaptive transmission with reasonable overhead (Annex A.1).
    - A generic multi-antenna transmit/receive scheme, which can be configured into various diversity, multiplexing and multi-user MIMO configurations (section 4.8.2). The spatial transmission can be adjusted individually to the needs of different packet flows to/from a user terminal.
    - Pilot schemes that support various types of adaptive transmission and multi-antenna transmission with acceptable pilot overhead (section 4.8.4 and Annex A.3).

The key WINNER reference design elements (to serve as examples of “best-use”) include:

- Frequency-adaptive transmission [WIN1D24] can be supported at vehicular velocities in FDD downlinks, see section 7.3.2.4. A new scheme for efficient link adaptation for frequency-adaptive transmission, MI-ACM, has been developed within the WINNER project [WIN2D223]. It combines use of strong codes with fine-grained adaptive modulation to provide superior performance and it is included as part of the reference design.
- The allocation principles for frequency-adaptive and non-frequency-adaptive transmission are outlined in section 7.3.2 and Annex C on multiple access. For non-frequency-adaptive transmission, a specially developed parameterization of GMC is proposed for uplinks, block-IFDMA [WIN2D461]. It targets low terminal power consumption by reducing the signal envelope variations and facilitating power-saving terminal micro-sleep intervals.
- The principles for selecting optimized spatial transmit schemes, and the preferred schemes in different deployment scenarios, are described in section 7.3.3. A more detailed description of the spatial transmission can be found in [WIN2D341], with performance results in [WIN2D61310], [WIN2D61311] and [WIN2D61312].

- The preferred pilot schemes for different deployment scenarios are summarized in section 7.4. The pilot design supports the all the multi-antenna transmit schemes for each deployment scenario and all the multiple-access schemes with low power overhead.
- ***Self-organised synchronization of terminals and base stations is supported***. This mechanism also can be used for network-wide synchronization of base stations and relay nodes, to facilitate multi-cell coordination [WIN2D233]. The reference design schemes for link level synchronization and network synchronization are described in section 7.5.

2.2 WINNER system capabilities at-a-glance

In Table 2-1, some of the basic capabilities of the WINNER system are provided for quick reference.

Capability	Value	Comment
Spectrum related		
Carrier frequencies [GHz]	Any between 450MHz to 5000MHz, especially the bands: 450-470MHz, 698-862MHz, 2300-2400MHz, 3400-3600MHz	These bands were identified at WRC07 in various Regions.
System band width [MHz]	1.25-100 with a theoretical resolution on a sub-carrier basis	
Duplexing	FDD and TDD	
Type of spectrum	Optimized for controlled interference scenarios (licensed spectrum) but the system also works in unlicensed spectrum	
Flexible spectrum sharing with other WRANs	Yes	
Flexible spectrum sharing with other primary or secondary systems	Yes	
Link adaptation		
Modulation	BPSK, QPSK, 16 QAM, 64 QAM, 256QAM	
Channel coding	Convolutional coding, and LDPC codes, optimized for short and long block length, respectively	
Spatio-temporal processing	MU-MIMO w 2 terminal antennas and up to 32 base station antennas.	Single antenna UTs are also supported
Hybrid ARQ	LDPC based incremental redundancy, mother code rate =1/3	
Multiple Access Methods (See also [WIN2D6137] for more details)		
Multiple Access	TDMA/OFDMA, combined with SDMA when appropriate	
Subcarrier spacing	FDD: app.39 kHz TDD: app.49 kHz	The frequencies can be derived from a common clock
Superframe/frame duration	5.69 ms / 0.6912 ms	
Scenarios		
Cyclic prefix	Optimized for ca 1 km cell radius	
Mobility	Stationary, low, high, very high	
Deployment		
Relaying	Decode-and-forward relaying with cooperative relaying as an optional add-on. Optimised for 2-hop relaying, but extendable to more hops	
Handover	Inter system HO + highly optimized HO to at least one IMT-2000 system	
Service related		
Peak rates	Fulfils the requirements of 100Mbps with high mobility and 1 Gbps with low mobility.	

Table 2-2: WINNER capabilities at a glance.

3 WINNER Logical node architecture

The WINNER system is designed to work well over a wide range of operating scenarios, and proof of concept results are being produced using certain scenario specific designs and parameter settings. It is the purpose of the logical node architecture to show a high level functional abstraction of all of these exemplary designs and show the basic flow of user data, control data, and functional interactions without having to consider specific physical design implementations or scenario dependent parameter settings.

The logical node architecture is a framework for the description of the WINNER system concept that identifies functions, groups them in logical nodes and interconnects these by well-defined, open interfaces.
Logical nodes are therefore defined as the smallest entities in the radio access network including the user terminal for which inter-operable interfaces are defined that are independent of the implemenation of a specfic vendor. This enables operators to buy interoperable equipment from multiple vendors. In contrast to logical nodes, physical nodes are physical devices that may contain multiple logical nodes depending on the specific deployment scenario.

Another purpose of the logical node architecture is to offer a compact top-down overview of the WINNER system starting from an abstract level that is refined down to detailed protocol definitions.

3.1 Overview of WINNER Radio Access Network

Figure 3-1 shows the general overview of the WINNER architecture.

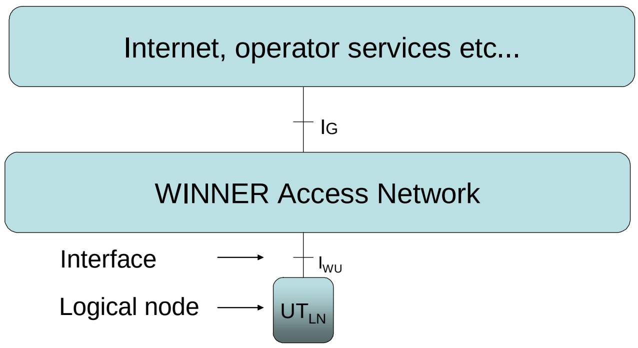

Figure 3-1: General overview of the WINNER architecture.

The WINNER Radio Access Network (WRAN) is connected to an external packet data network (e.g. Internet) via the I_G interface (see Section 3.5). The WRAN provides the I_{WU} interface that is a radio interface to connect WINNER terminals (“UT”).

Strictly speaking, the WRAN also contains functionalities that are more related to the core network rather than the radio network, such as gateway functionalities. Nevertheless, for simplicity we will use the term WRAN to denote the novel access network.

In order to accommodate handover and other cooperation functions (spectrum sharing) between WRANs and non-WINNER Access Networks, WINNER has to comply with a cooperation architecture (see Section 6.3) provided outside the WINNER RAN. Thus, each of the WRANs have to support respective functionalities, e.g. provide support functions such as measurements.

3.2 Logical node architecture

This section aims at providing a brief overview of the logical nodes (LN) and their interactions. The logical nodes will be described in greater detail in section 3.4, and the interfaces in section 3.5.

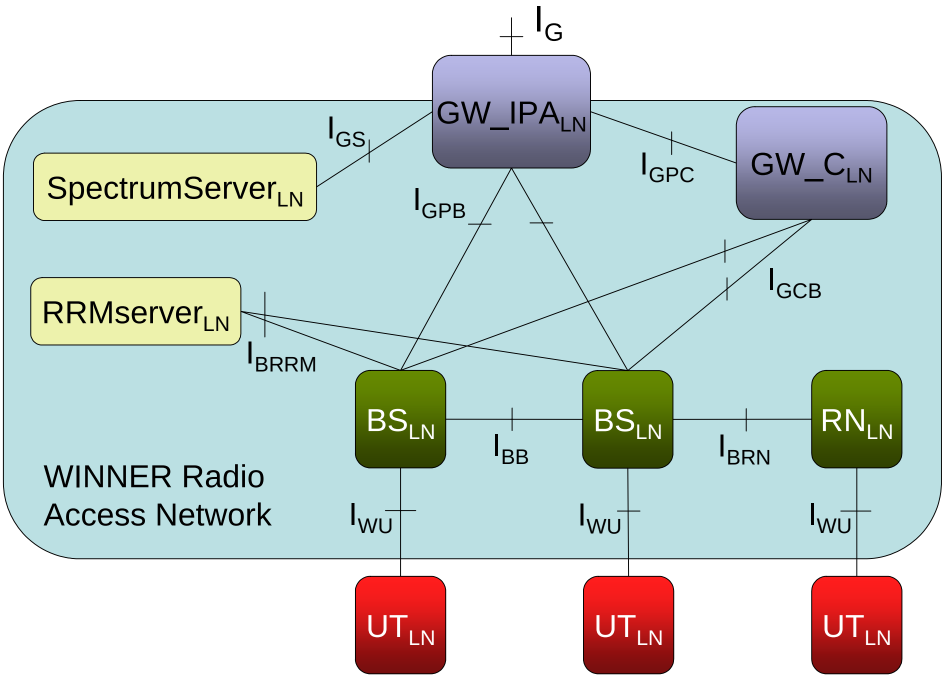

Figure 3-2: WINNER logical nodes and interfaces.

As is illustrated in Figure 3-2 the WINNER logical node architecture is made up of the base station Logical Node BS_{LN}, the Relay Node Logical Node RN_{LN} and User Terminal logical node UT_{LN} node, the logical nodes $SpectrumServer_{LN}$ and the $RRMServer_{LN}$ as well as two types of gateway logical nodes, the IP Anchor GW_IPA_{LN}, and the Control GW_C_{LN}.

The logical nodes are interconnected by logical interfaces denoted as "I". As the logical nodes are defined as the smallest entities in the radio access network including the user terminal for which inter-operable interfaces are defined, the logical interfaces constitute the respective interfaces. Protocols are defined by the termination points and the interfaces between them.

The GW_IPA_{LN} provides access to external data networks (e.g., Internet, Corporate Networks, operator controlled core networks) and operator services (e.g., MBMS=Multimedia Broadcast Multicast Service). It also terminates flows on the network side and serves as the anchor point for external routing. Thus, all functions that operate on user data traffic are located here. It is accompanied by the GW_C_{LN} that provides control functions for UT_{LN}s that are not active (i.e., terminals not sending data) and functions that control and configures the GW_IPA_{LN}. Note that in practice there will most likely be several GW $IPA\__{LN}$ present in the network as well as an independent number of GW C_{LN}.

The WRAN contains two points of radio access points, the BS_{LN} and RN_{LN}. The RN_{LN} is a logical network node with relaying capabilities that is wirelessly connected to a BS_{LN} or another RN_{LN}. The RN_{LN} communicates with the BS_{LN} using the I_{BRN} interface. A WINNER terminal UT_{LN} is connected to the WRAN via the I_{WU} interface and communicates either directly with the BS_{LN} or indirectly to the BS_{LN} via the RN_{LN}.

The BS_{LN} performs all radio related functions for active terminals (i.e., terminals sending data) and is responsible for governing radio transmission to and reception from UT_{LN} and RN_{LN} in one cell. The BS_{LN} also controls the resources the RN_{LN} uses. The BS_{LN} are connected to each other by the I_{BB} interface.

The BS_{LN} is connected to the GW_IPA_{LN} and GW_C_{LN} pair via the I_{GPB} and I_{GPC} interfaces, respectively. These interfaces are multinode-to-multinode capable interfaces meaning that one BS_{LN} can be connected to multiple GW_IPA_{LN} and GW_C_{LN} pairs and conversely, that one GW_IPA_{LN} and GW_C_{LN} pair can be connected to multiple BS_{LN}. The GW_IPA_{LN} and GW_C_{LN} form a pool of equipment that may cover large areas, e.g. cities.

A UT_{LN} is said to be associated with a GW_IPA_{LN} and GW_C_{LN} pair when the GW_IPA_{LN} and GW_C_{LN} pair provides an anchor point for the traffic to/from the $UT_{LN}$and control functionality for the UT_{LN}, respectively. Two UT_{LN} connected to the same BS_{LN} can be associated to different GW_IPA_{LN} and and

GW_C_{LN}. Conversely, two UT_{LN} associated with the same GW_IPA_{LN} and GW_C_{LN} pair can be connected to different BS_{LN}s

The $SpectrumServer_{LN}$ is a logical node that contains centralised spectrum sharing and spectrum assignment related functions, and enables sharing and co-existence with other radio access technologies and spectrum assignment between WRANs. The $SpectrumServer_{LN}$ interfaces with the GW_C_{LN} to obtain information about associated UT_{LN} and their characteristics. The $SpectrumServer_{LN}$ communicates spectrum sharing and spectrum assignment decisions to the BS_{LN}. The $SpectrumServer_{LN}$ is accessible via the GW_IPA_{LN} for spectrum negotiations.

The $RRMserver_{LN}$ is a logical node that contains centralised radio resource management related functions.

3.3 Main concepts and capabilites of the WINNER logical node architecture

3.3.1 Pool concepts and micro mobility

As is suggested in Figure 3-5 there is at any moment an association between a UT_{LN} and an IP Anchor GW_IPA_{LN} and between the UT_{LN} and a Control GW_C_{LN}.

The GW_IPA_{LN}, contains the user plane functionalities and serves as a mobility anchor. Each UT_{LN} is associated with one each of GW_IPA_{LN} and GW_C_{LN}. The association with a GW_IPA_{LN} and a GW_C_{LN} for one UT_{LN} may be different from that of another UT_{LN}connected to the same BS_{LN} That is, for example, two UT_{LN} may be associated with the same GW_C_{LN}, but with different GW_IPA_{LN} (Figure 3-3). This implies that one advantage with this approach is improved dimensioning and scalability of the gateway functionalities according to the need in the deployed network. Furthermore, it gives the vendors more freedom to choose how to build e.g. a physical base station product with the user plane gateway functionalities close to the BS_{LN}, while having a more centrally located control server. A drawback with the split of the gateway functionalities in this way is that there is a need for a new standardized interface I_{GPC}.

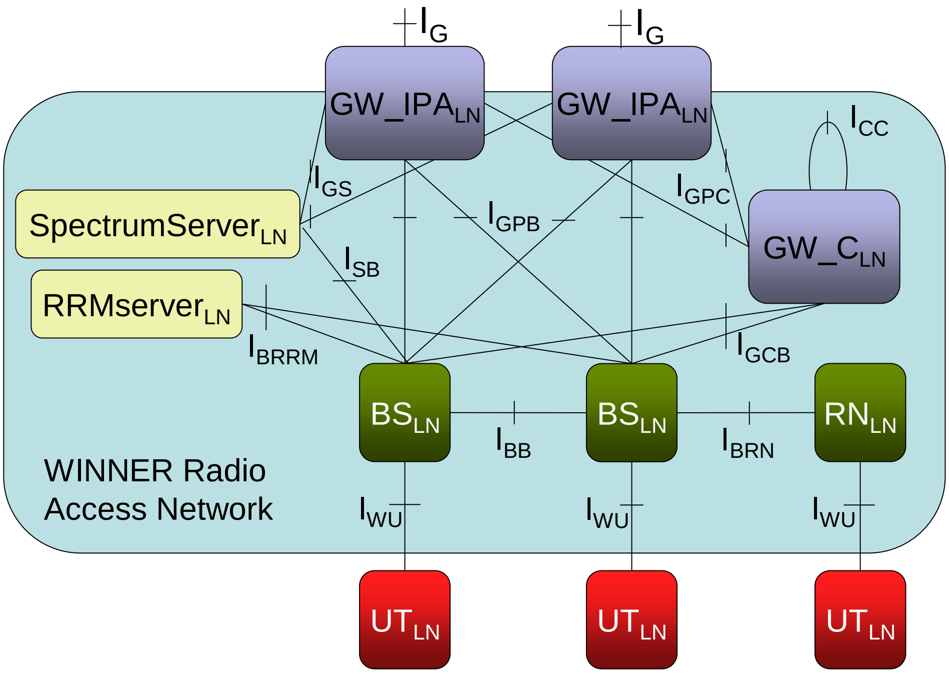

Figure 3-3: One set of GW_C_{LN} may serve a (larger) set of GW_IPA_{LN}
Each UT_{LN} is associated with one GW_C_{LN} / GW_IPA_{LN} pair.

There is also an interface I_{CC} between multiple instances of GW_C_{LN} for relocation and information transfer between two GW_C_{LN}. In case of relocation of GW_IPA_{LN}, and in case of simultaneous relocation of two BS_{LN} (a rare event since we have overlapping pool areas), lacking an I_{BB} interface (also a rare event) packet forwarding between these two BS_{LN} is performed via the two involved GW_IPA_{LN}.

Hence there is no need for an interface between the two GW_IPA_{LN} since no control signaling is performed between the two GW_IPA_{LN}.

For simplicity and visual clarity in the figures, when this split is not of essence for the discussion, the GW_IPA_{LN}/GW_C_{LN} pair is represented only with one logical node, the GW_{LN}. In addition, the interfaces I_{GPB} and I_{GPC}, can be represented with the interface $I_{GB.}$ In this overview and in most part of this document this simplified view is adopted (Figure 3-4).

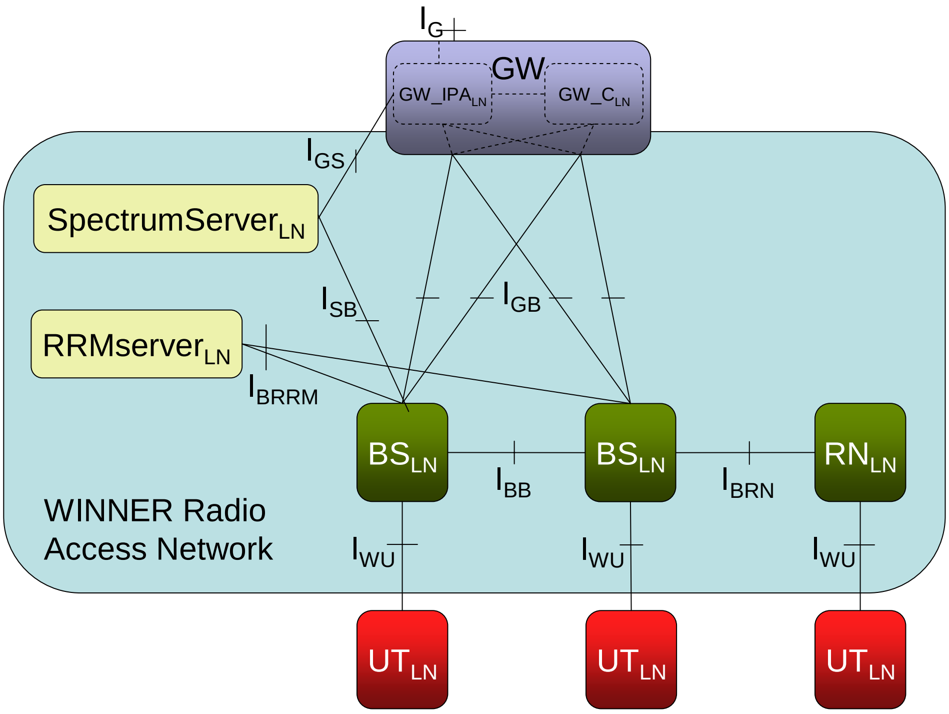

Figure 3-4: Simplified representation of the GW_IPA_{LN} and GW_C_{LN} as $GW\__{LN}$

The logical association between a UT_{LN} and the GW_{LN} is independent of the BS_{LN}s, and can be kept during a handover as depicted in Figure 3-5, and hence the set of GW_{LN} can be seen as a pool of resources. Within a pool, each BS_{LN} can forward user traffic to/from any GW_{LN}, and conversely each GW_{LN} can communicate with all BS_{LN}. Of course the UT_{LN} - GW_{LN} association can change if necessary for e.g. load balancing purposes. This also facilitates network scalability.

The GW_{LN} – pool concept decouples the physical relation between a GW_{LN} and a number of BS_{LN} in the pool area. Instead, each GW_{LN} is connected to each BS_{LN} in the pool area. Following that, the pool area is defined as an area in which the UT_{LN} may roam without the need to change the GW_{LN}. The GW_{LN} capacity of a pool area can be scaled simply by adding more GW_{LN}s (or more accurately, GW_IPA_{LN} and a GW_C_{LN} can be added independently). In contrast, in a hierarchical structure, each BS_{LN} is connected to one GW_{LN} only that serves its own location area.

There are several benefits of having the set of GW_{LN}s as a pool of resources:

- It reduces the requirements for macro mobility since the GW_{LN} is an anchor point for external routing.
- The pool capacity is easily optimized by adding or removing GW_{LN}s.
- Single point of failures is avoided, that is, should a GW_{LN} break down, the users can be handed over to any other GW_{LN}.
- Load balancing can easily be achieved, e.g., to support different services such as MBMS.

The pool concept includes logical associations from each GW_{LN} to each BS_{LN}, which means that fully meshed interconnections are needed (i.e. a switching or routing function is needed) (Figure 3-6). For the routing L3 technologies as IP may be feasible provided they implement virtual logical connections from each GW_{LN} to each BS_{LN} that allow micro mobility without reconfiguration of routing tables or change of IP addresses. An IP – based tunnelling protocol is foreseen that provides tunnels between GW_{LN}s and BSs. The GW_{LN} routes any data for a specific BS_{LN} by addressing the corresponding tunnel.

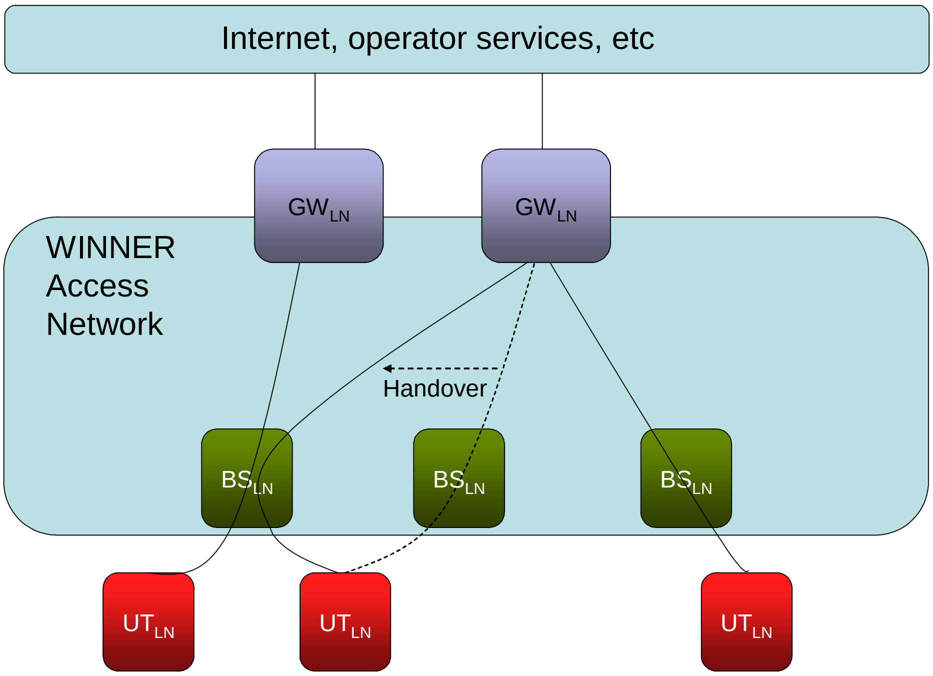

Figure 3-5: The UT$_{LN}$ - GW$_{LN}$ association is normally kept during handover.

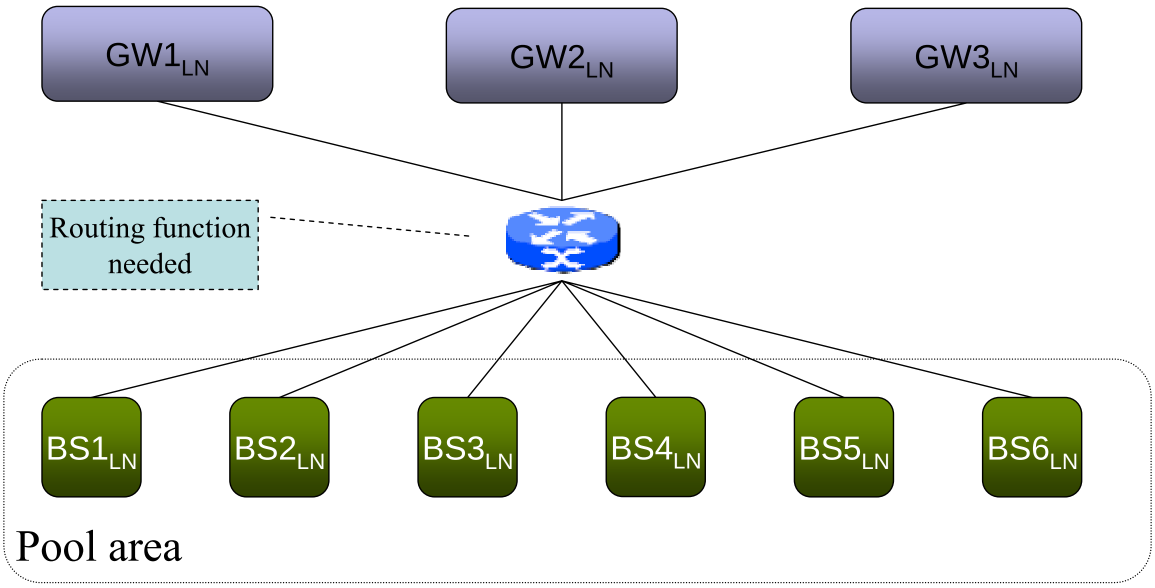

Figure 3-6: Transport network for meshed RAN structure (pool concept).

The targeted large size of the pool area may lead to a situation where a UT$_{LN}$ is associated to a GW$_{LN}$ that is located very far away from the BS$_{LN}$ that it is connected to. This may increase the signalling load compared to a hierarchical structure where the signalling is kept locally in the location area of a GW$_{LN}$.However, each vendor may provide intelligent functions for optimizing these associations w.r.t. delay and signalling load.

3.3.2 Equipment sharing

It is also an important requirement that multiple operators can share part of the same physical RAN, for example transport network, and the physical BSs and RNs. In order to support this in the logical node architecture the following multiple sets of GW$_{LN}$ each set connecting to different operator back bone networks. Each UT$_{LN}$ can then associate with the GW$_{LN}$ relevant for the operator in question (Figure 3-7). The operators will thus be able to share the physical BSs and physical RNs, as well as the transport network (not depicted) connecting the BS.

The operators thus must have different sets of GW_{LN}, but nothing prevents a vendor from manufacturing physical GWs that can host logical GW_{LN}s belonging to different operators, thus enabling also operators to in a sense share physical GWs. This is a design choice, however, and does not influence the logical architecture and is not treated further.

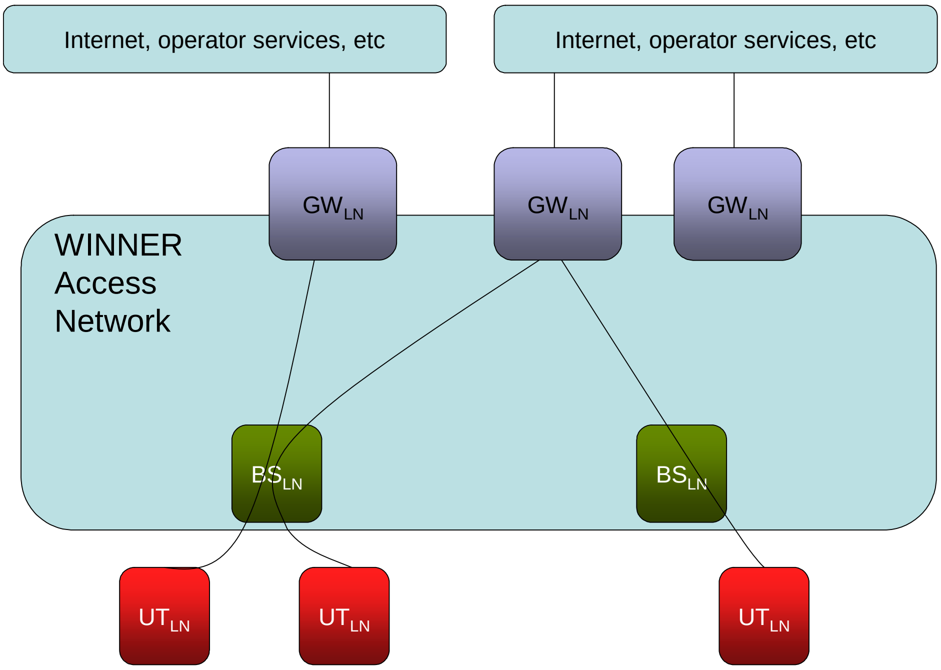

Figure 3-7: Equipment sharing support in WINNER LNA.

3.3.3 Support for multicast and broadcast service

Support for MBMS has been identified as important requirement for WINNER. While unicast traffic can be efficiently handled with a one to one GW_IPA_{LN} to UT_{LN} association the situation for MBMS is different.

An efficient implementation of a MBMS service includes a central entity that multicasts the traffic to all base stations with connected UT_{LN}s that are using this service. In a hierarchical concept, that is obviously simple because the MBMS traffic has only to be distributed to all GW_IPA_{LN}s where these UT_{LN}s are connected to. Forwarding within the location area is handled by the respective GW_C_{LN} and the respective signalling traffic is kept on the transport network that is logically behind this GW_IPA_{LN}. UT_{LN}s moving between location areas may require only an update of the routing to the GW_IPA_{LN}s

In the pool concept, one GW_IPA_{LN} has to take over the role of a central anchor point that distributes the MBMS traffic to all UT_{LN} within the pool area that subscribe this service.These may however be associated with different GW_IPA_{LN} for their unicast traffic. In addition, multiple MBMS sessions may be ongoing in parallel to many (partly overlapping) UT_{LN} groups. In order to support MBMS it is therefore required that a UT_{LN} is able to associate with multiple GW_IPA_{LN}. In this case it is assumed that unicast traffic of a UT_{LN} is terminated by a single GW_IPA_{LN} so that the one to one relation between GW_{LN} and the UT_{LN} is kept for unicast traffic, see Figure 3-8. The multicast flows for each UT_{LN} may however come from a different GW_IPA_{LN}. In the pool concept, this means that the GW_IPA_{LN} anchoring MBMS traffic has to be updated of all cell changes of UT_{LN}s that are using this MBMS service. As the forwarding of all these UT_{LN}s is handled by different GW_IPA_{LN}s in the general case (their unicast GW_{LN}s) and no inter-GW_IPA_{LN} interface is foreseen, the unicast forwarding procedures cannot be used to update the MBMS GW_{LN} about any cell changes of the UT_{LN}s. Thus, an additional forwarding update procedure for the MBMS GW_{LN} is needed that runs in parallel to the unicast forwarding procedure.

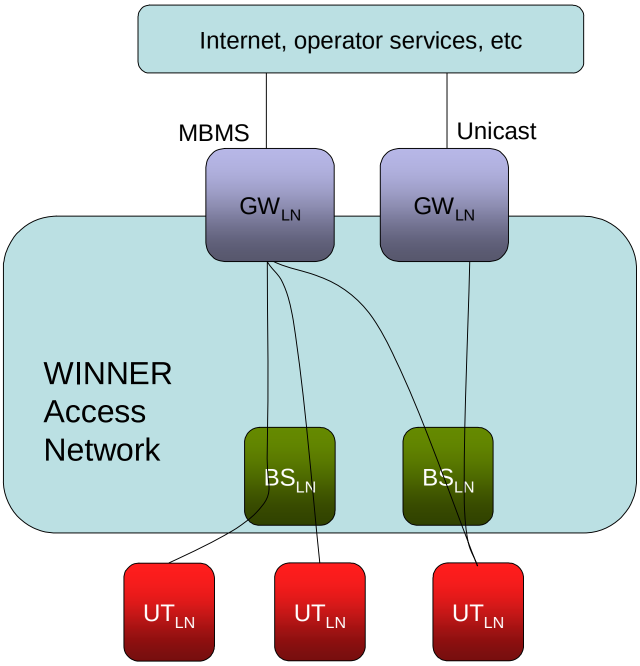

Figure 3-8: Multiple association between UT$_{LN}$ and GW$_{LN}$s.

Broad- and unicast traffic may be transmitted in a single frequency network by multiple BS$_{LN}$ in the same frequency band so that the UT$_{LN}$(s) are able to utilise the multiple signals, see Figure 3-9 . In this case the broad- and/or unicast traffic is transmitted from the GW$_{LN}$ to the serving BS$_{LN}$ (which is the one that handles also unicast traffic).

For WRAN it is assumed that the GW_C$_{LN}$ controls the respective BS$_{LN}$ in the single frequency network area.

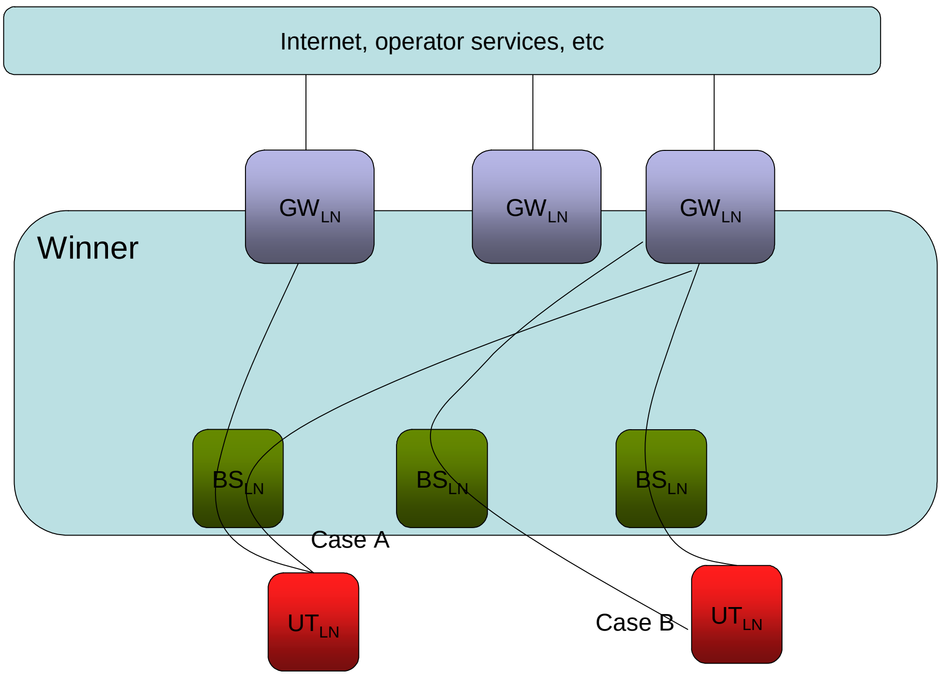

Figure 3-9: Case A: Combined unicast and multicast; Case B: Multiple BSs form a single frequency network.

Case A: the UT is associated with one GW for unicast traffic, and another for dedicated MBMS

Case B: The UT is associated with one MBMS GW, and receives MBMS data from multiple BS forming an SFN.

3.3.4 Multiple band transmission from different BSs

The LNA also supports simultaneous transmissions from multiple BS_{LN} s to one UT_{LN} as is shown in Figure 3-10 to support for overlay networks. As an example, it enables that low data rate but delay and jitter sensitive flows (e.g. VoIP) are transmitted by an overlay wide area cell to avoid frequent HO situations while the delay and jitter insensitive high data rate flows are transmitted by a local area cell. The traffic is distributed by the GW_{LN} to the BS_{LNs}. It also provides gains in load and admission control.

In the figure, the physical implementation of the user terminal and base station nodes are denoted by UT_{PN} and BS_{PN}.

It should be noted that this option addresses the case if both cells operate in different bands and the physical UT supports multiple transceivers that are modelled by different UT_{LN}. Furthermore, the network will view a physical terminal with two logical nodes as two separate entities. Transmission from two BS_{LN} in same band to the same physical UT_{LN} is at the moment foreseen for MBMS traffic.

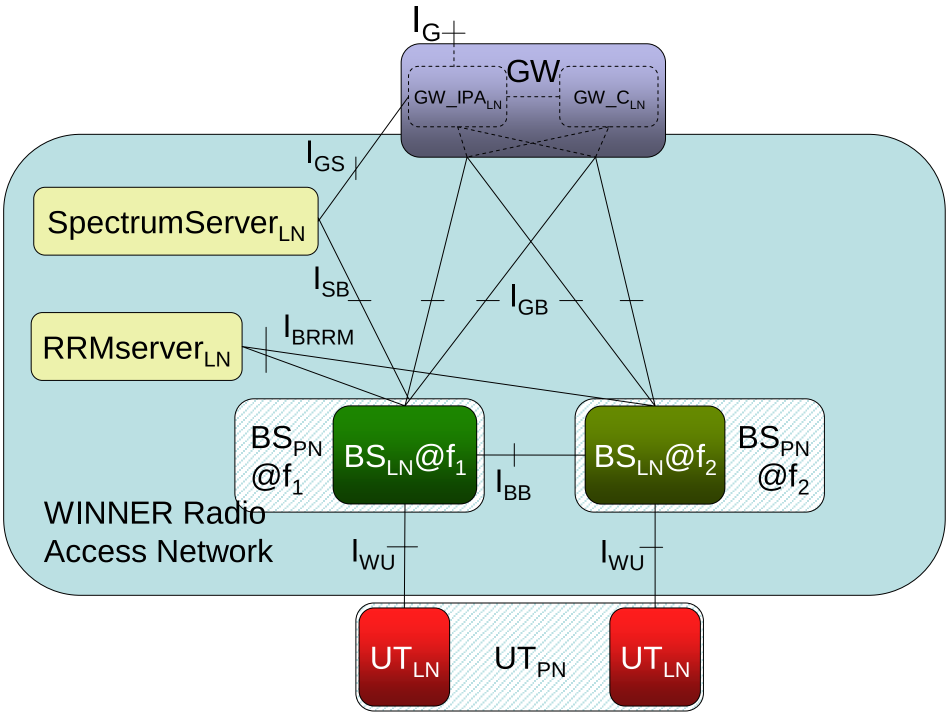

Figure 3-10: Multi-band-transmission from two physically separated BS_{LN} to one physical UT containing two UT_{LN}.

3.4 Description of logical nodes

This section provides a detailed description of the WINNER logical nodes and the functions that they perform. These functions are defined in more detail in [WIN2D6138].

3.4.1 The gateway nodes: GW_IPA_{LN} and GW_C_{LN}

The gateway functionalities are performed by GW_IPA_{LN} and GW_C_{LN}. Generally, the former logical node is an IP anchor (IPA) for the user, where also the data is routed, whereas the latter node is a control node that controls various aspects of the data flows in the former node. Thus, certain functionalities are represented in both logical nodes; however different aspects are captured in the different types of nodes. For example, policy enforcement is performed on the data by the GW_IPA_{LN}, but the policy enforcement is controlled by the GW_C_{LN} (i.e. GW_C_{LN} configures the GW_IPA_{LN} w.r.t. policy enforcement).

It is possible to pool a number of GW_C_{LN} or GW_IPA_{LN} nodes together in order to eliminate the risk that one node failure will cause parts of the network to be out of service as well as to enable load sharing in between different GW_C_{LN} or GW_IPA_{LN} nodes. This is possible since there is a many-to-many relation interface between the BS_{LN}, GW_C_{LN} and GW_IPA_{LN}, where each BS_{LN} is associated with a set of GW_C_{LN} or GW_IPA_{LN} called a GW_C_{LN} or GW_IPA_{LN} pool. When a UT_{LN} attaches to the network, it is assigned to one GW_C_{LN} and GW_IPA_{LN} in this pool. No change of GW_C_{LN} or GW_IPA_{LN} is normally required while the UT_{LN} moves around among BS_{LN}.

The GW_IPA_{LN} and GW_C_{LN} provide the interface to the outside world. As such, the GW_IPA_{LN} acts as the IP anchor point for external routing functionality. The same GW_IPA_{LN} to UT_{LN} association is generally maintained for unicast traffic as long as the UT_{LN} is located within the pool area. The GW_IPA_{LN} contains therefore functions that perform data forwarding in case of user mobility.

In order to ensure data integrity and confidentiality within the WRAN header compression and then ciphering are also performed in the GW_IPA_{LN} (but configured by the GW_C_{LN}).

The flow class setup and release procedures are performed by the GW_C_{LN}. QoS is also addressed in this node to configure respective access network elements (e.g. routers within the WRAN – the details are however beyond the scope of this document). In addition, the GW_C_{LN} communicates with an external AAA server to handle authentication and authorisation requests by the UT_{LN}. Once the UT_{LN} has been authorised the GW_C_{LN} is responsible to forward charging related information to the external AAA server.

The GW_C_{LN} supports UT_{LN} power saving by providing the support for UT_{LN} idle mode. The GW_C_{LN} is informed by the UT_{LN} about the current paging area in case of UT_{LN} movement and updates an internal database with this information. It initiates the paging procedure to initiate the UT_{LN} state change. It is for further study to determine the paging areas and their potential relation to the pool concept.

When the UT_{LN} is in active state GW_IPA_{LN} forwards the data towards the UT_{LN} (using the BS_{LN} where the UT_{LN} is attached). In case of active state user mobility the path switch function in the GW_C_{LN} is informed by the serving BS_{LN} about handover decisions. The GW_C_{LN} then reconfigures the GW_IPA_{LN} so that henceforth the flows are routed via the target BS_{LN}.

MBMS traffic is supported by the GW_C_{LN} and GW_IPA_{LN} in a similar way except that a MBMS flow instead of unicast flow is used. Both flow types differ with respect to the maximum number of destination UT_{LN} . For each MBMS session one MBMS flow is established to all BS_{LN} within one single frequency network area (i.e. several BS use coordinated transmission in order to obtain downlink macro diversity, for more details see [3GPP08]) in which at least one UT_{LN} subscribes to that session. It is up to the BS_{LN} to select an appropriate transmission channel to transmit the data to the UT_{LN} given in the MBMS flow.

The following functions are performed by the GW_IPA_{LN}

- Forwarding for Unicast and Multicast/Broadcast services
- Header compression
- Ciphering

- Policy enforcement
- Traffic measurements (for charging purposes)

In addition to the control and configuration of the functions performed in the GW_IPA_{LN}, the following functions reside in the GW_C_{LN}

- Idle mode mobility control
- Macro mobility support
- Authorisation and authentication support
- Flow class establishment/release and QoS signalling
- Flow class admission control
- Integrity protection
- Paging

3.4.2 BS_{LN}

The BS_{LN} performs all radio related functions for both active and idle terminals and is responsible for governing radio transmission to and reception from UT_{LN} and RN_{LN} in one or more cells. The BS_{LN} is in control of relays (if used) and determines routes, forwards packets to the respective relay and takes care of flow control for the relays to ensure that they can forward the data to their associated UT_{LN}.

User mobility in UT_{LN} active mode is handled by the BS_{LN}. It determines necessary handovers based on measurements obtained by the UT_{LN} and respective information governed by neighbouring BS_{LN}. The handover decision between relays is drawn in similar manner based on measurements obtained by the UT_{LN} to collect the link information. Furthermore the load situation at all involved nodes and BS/RN to RN link quality is taken into account in this process.

Each UT_{LN} that is served by the BS_{LN} is represented by a user context that is kept in the BS_{LN}. Each flow class that is transmitted via the BS_{LN} is represented by a flow class context in the BS_{LN}. In the regular case a UT_{LN} handover involves the transfer of the user context and the transfer of all UT_{LN} flow class contexts to the target BS_{LN} by using the I_{BB} interface. Hence if a UT_{LN} is only connected to one BS_{LN} (the regular case) all flow class contexts and the user context are linked and kept in the serving BS_{LN}.

In the special case of a dual transceiver (physical) UT that is served by two BS_{LN} operating in different bands a handover (flow class context transfer) of individual flow class is possible. In this case a separate user and flow class context is kept in each BS_{LN} for each UT_{LN} to which the physical UT is connected to.

In order to detect overload situations load supervision is located in the BS_{LN}, and to prevent this situation countermeasures like load balancing and flow class admission may be performed.

The following functions reside in the BS_{LN}. Detailed descriptions of these functions are given in [WIN2D6138]. The function list below only contains functions that are important for the understanding of the characteristics of the interfaces within the logical node architecture. The lower layer functions such as scheduling, channel coding, etc are more or less restricted to reside in the BS_{LN} (on the network side) for delay and processing reasons and are not in the list.

- Flow class admission control
- Flow control between RAPs
- Packet scheduling over the radio interface
- Outer ARQ
- Buffer management
- Lower layer QoS configuration
- Route establishment within an REC
- User context transfer
- Flow class context transfer

- Load supervision
- Load sharing/control
- Forwarding
- Horizontal spectrum sharing with coordination
- Horizontal spectrum sharing without coordination
- Short term spectrum assignment
- Micro mobility (for BS and RN)
- Resource partitioning

3.4.3 RN_{LN}

The RN_{LN} is a logical network node with relaying capabilities that is wirelessly connected to a BS_{LN}, UT_{LN} and/or another RN_{LN}. As such it contains a forwarding function and schedules packets on the RI.

Furthermore system information broadcast, provided by the BS_{LN} is relayed by the RN_{LN}.

In WINNER, RN_{LN} is assumed to operate as a decode-and-forward layer 2(L2)-relay. Decode-and-forward L2-relays allow advanced forwarding and can take advantage of adaptive transmission with different modulation and coding schemes on the different hops.

The relaying solution is not optimised to reside in a physical mobile device, but rather in a fixed installation. However, there is nothing in the logical node architecture that prevents mobile relays.

The RN_{LN} serves UT_{LN}s the same way as the BS_{LN}, which means the UT_{LN} does not need an extra mode or reconfiguration.

The relaying concept shall be primarily designed and optimised for two hops (BS_{LN}-RN_{LN}-UT_{LN}) in order to achieve a high performance relay deployment. Nevertheless, the concept aims at supporting any number of hops in order to allow a high degree of deployment flexibility, e.g. in areas where coverage has priority over performance.

The tree topology is used, as it is less complex than, e.g. a mesh topology. In the case of node failure the RN_{LN} should autonomously connect itself to another RAP in its range. Although the re-association to the network is not seamless and can lead to some lost connections, the tree topology can still be assumed as self-healing.

The function list below only contains functions that are important for the understanding of the characteristics of the interfaces within the logical node architecture. The lower layer functions such as scheduling, channel coding, etc are more or less restricted to reside in the RN_{LN} for delay and processing reasons and are not in the list. Many of those functions can be found in the appropriate layer in Chapter 4.

- Packet scheduling over the radio interface
- Flow control between RAPs
- Forwarding in the REC
- Outer ARQ
- Horizontal spectrum sharing mechanisms related to low layer access in a shared medium

3.4.4 UT_{LN}

UT_{LN} comprises all functionality necessary to communicate directly with the network, i.e. a BS_{LN} or a RN_{LN}. It contains functions to handle UT_{LN} mobility in active and idle states, as well as functionality to perform an initial access to the network. It furthermore contains functionality to initiate a flow establishment.

The function list below only contains functions that are important for the understanding of the characteristics of the interfaces within the logical node architecture. The following functions reside in the UT_{LN}.

- Initial Access
- Paging area update
- Flow class establishment request

3.4.5 RRMServer$_{LN}$

The RRMserver$_{LN}$ terminates only network internal protocols (i.e. does not have a direct relation to the terminals). The following functions are included in a RRM server$_{LN}$.

- Load sharing/control
- Micro mobility
- Admission control
- Resource partitioning

However, it is for further study to decide if these functions also can reside in a distributed way in the mandatory logical nodes. Thus, in case if the RRMserver$_{LN}$ is present the central instances of these functions can take over from the distributed instances and make better decisions using information from a larger part of the network.

3.4.6 SpectrumServer$_{LN}$

The SpectrumServer$_{LN}$ is a logical node that contains centralised spectrum sharing and spectrum assignment related functions. The SpectrumServer$_{LN}$ enables the sharing and co-existence with other radio access technologies and spectrum assignment between WRANs. The SpectrumServer$_{LN}$ interfaces with the gateway.

The SpectrumServer$_{LN}$ monitors the load, contains constraints on the available spectrum (e.g. information about exclusion zones), and keeps track of the available spectrum from spectrum sharing and spectrum assignment. The spectrum functions in the SpectrumServer$_{LN}$ interact with the spectrum functions residing in the base station, so that spectrum availability can be communicated and local optimizations of the spectrum allocation can be made by mechanisms residing in the BS$_{LN}$.

The following functions reside in the SpectrumServer$_{LN}$, they are detailed further in Section 6.1.

- Vertical sharing 1
- Vertical sharing 2
- Centralised component of Horizontal sharing with coordination
- Spectrum Register
- Long term spectrum assignment

For Horizontal sharing with coordination most functions reside in the BS$_{LN}$, the SpectrumServer$_{LN}$ only contains functions related to centralised coordination related to Horizontal Sharing with coordination.

3.5 Description of interfaces

3.5.1 I_G interface

The I_G interface incorporates, beside the user data traffic, the protocols needed for communication with external functions, and cooperation and spectrum functions outside of the WRAN. This is the interface to external packet data networks (e.g., Internet) and contains the end-user's IP point of presence. All user- and control-plane functions that use the I_G interface are handled above the end-user's IP layer. All terminal mobility will be handled below the I_G interface.

Since I_G interface connects the entities outside the WRAN controlled by the GW_{LN} and the GW_{LN} itself, therefore, it carries the following signalling:

- GW_{LN}-IP based public data network signalling
- GW_{LN}-Home Subscription Server signalling
- GW_{LN}-GW_{LN} signalling
- GW_{LN}-non WINNER system signalling
- GW_{LN}-non operator enterprise network signalling
- GW_{LN}-Policy and charging server

Through those collections of signalling, at least the following functions are supported by the I_G interface:

- Enhanced service provisioning, e.g., the IMS, AAA control
- Mobility management for idle UT_{LN}
- Support of non WINNER system (e.g. legacy systems) anchor with providing inter-system HO, which is compliant with Ambient Network ARI interface (see [AN2006]) and connect to CoopRRM/MRRM server.
- Unified QoS control between heterogeneous networks,
- External spectrum managers
- WINNER $SpectrumServer_{LN}$
- Roaming service
- Policy and charging enforcement

In addition to signalling described above, all data traffic to external networks goes through this interface.

3.5.2 The I_{GPB} and I_{GCB} interfaces

The I_{GPB} is the interface between GW_IPA_{LN} and BS_{LN} while the I_{GCB} is the interface between GW_C_{LN} and $BS_{LN.}$ Both of them are multinode-to-multinode interface. All user data is transported via the former interface. The I_{GCB} interface provides means to establish a flow class context in the BS_{LN} by the GW_C_{LN}. Furthermore it contains means to handle UT_{LN} mobility. The BS_{LN} informs the GW_C_{LN} about user mobility like handover and paging updates. Vice versa the BS_{LN} provides means to request paging for a certain UT_{LN}.

On the user plane, all user data traverses the I_{GPB} interface between the GW_IPA_{LN} and BS_{LN}.

It has been assumed that the transport technology can be selected by the network operator, and therefore the transport technology is not specified in the WINNER concept. However, the meshed nature of the GW-pair/B_{LN}s, inter-connections is a challenging transport solution, but at least IP tunnelling is seen as a potential solution.

3.5.3 I_{WU} interface

This is the interface between UT_{LN} and the WRAN. It defines the necessary functionalities and communication formats that enable the radio connection between UT $_{LN}$ and network. It is assumed that the same interface is used between UT_{LN} and network, regardless whether the RAP in the network side is a relay or a base station.

This interface is used in transferring all user data to and from UT_{LN}. Moreover, all control information between network and UT_{LN} are carried through this interface. Such control information includes, for example, any measurements made by UT_{LN} and reported to network, connection set up, scheduling allocations, mobility related signalling, link adaptation and quality signalling, etc.

3.5.4 I_{RN} interface

This is the interface between BS_{LN} and $RN_{LN.}$ This interface is a subset of I_{WU} described in the previous section, but it also contains control traffic between the RN_{LN} and BS_{LN}. This includes all control signalling between RN_{LN} and BS_{LN} (or other RN_{LN} in a multi-hop deployment). For example intra-REC routing and relay specific initial access signalling, advance forwarding of resource partitioning information to relays, and flow control signalling are supported by this interface.

3.5.5 I_{BB} interface

This is the interface between BS_{LN} nodes and is a multi-to-multi interface. The interface supports distributed RRM functions, like active mode mobility, interference management schemes, and other distributed inter-BS_{LN} RRM functions (for example, load balancing). Moreover, several spectrum functions are located in BS_{LN} and will use this interface.

Active mode mobility requires at least user and flow class context transfer during handover.

3.5.6 I_{CC} interface

This is the interface between GW_C_{LN} nodes that contains functions for GW_C_{LN} relocation and GW_C_{LN} to GW_C_{LN} information transfer.

3.5.7 I_{GPC} interface

This is the interface between the GW_C_{LN} and $GW\_IPA_{LN.}$ with some functions for paging coordination, security and mobility. This is a multi-to-multi interface interface.

3.5.8 I_{GS} interface

This is the interface between the $SpectrumServer_{LN}$ and the GW_IPA_{LN}. This interface supports communication between the $SpectrumServer_{LN}$ and external nodes such as other network's spectrum servers, and external spectrum managers.

3.5.9 I_{SB} interface

This is the interface between the $SpectrumServer_{LN}$ and the BS_{LN} . This interface supports the exchange of spectrum sharing, co-existence and spectrum assignment information to the BS_{LN}.

3.5.10 I_{BRRM} interface

This is the interface between the $RRMServer_{LN}$ and the BS_{LN} . This interface defines the control plane control messages. This interface fulfils the following tasks:

- Requesting measurement reports from the BS, where the measurement reports including the load status of the current BS, the traffic volume information, buffer status, handover request rate, etc.
- Commands for release of resources in source BS and reservation of resources in target BS
- Warning of the potential added load and the source BS
- Handover command
- Admission control messages including requests, resource allocation requests to the BS and the rejection/admission/alternative BS command
- Commands for spectrum reallocation and recommendation of the candicate spectrum sharing peers
- Resource partitioning commands

4 Protocol and service architecture

In this chapter, the radio protocol architecture is described for the user plane and the control plane.

4.1 Overall user and control plane protocol architecture

The overall layer structure consists of the three lowest layers of the OSI reference model that are network layer (or L3), data link layer (or L2), and physical layer (or L1). Following common conventions, the protocol architecture is further divided into user (U) plane, composed of those protocols that implement data transfer services of the actual user-data, and control (C) plane, composed of those protocols that are aimed for controlling transfer services and connections between the access network and UT_{LN}. The overall structure is illustrated in Figure 4-1 and Figure 4-2 for the single-hop and 2-hop case, respectively. In Figure 4-1 the user plane (grey boxes) and control plane (red) are shown for the single hop case. In Figure 4-2 the user plane (grey boxes) and control plane (red) are shown for the 2-hop case. Note that the RRC in the BS terminates protocols initiated at both the RN and the UT. The non-access stratum (NAS) is a functional layer in the protocol stack that supports signalling and traffic between the core network and user terminal.

The network layer (or L3) C plane signalling is handled by Radio Resource Control (RRC) protocol. The control and configuration of the RN_{LN}, is performed by the sub layer RRC2. The data link layer (or L2) provides transfer services towards both U plane and C plane. The data link layer is partitioned into IP convergence layer (IPCL), Radio Link Control (RLC) and Medium Access Control (MAC) protocol sub-layers. The physical layer (PHY or L1) provides transfer of data over the radio link.

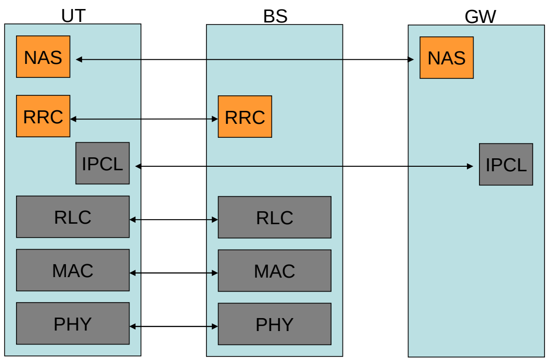

Figure 4-1: WRAN radio interface user and control plane architecture and termination (single hop).

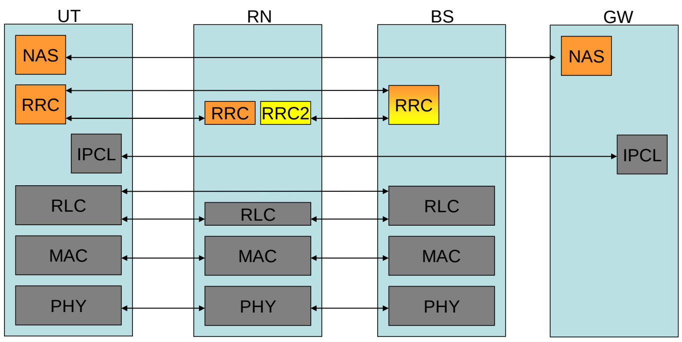

Figure 4-2: WRAN radio interface user and control plane architecture and termination (2-hop hop).

4.2 Mapping of user plane functions to protocol layer

In order to increase the understanding of the the text in subsequent sub sections Figure 4-3 introduces the protocol architecture frame work on a high level.

Figure 4-3 also gives an overview of functions mapped to the protocol layers. The scheduler (located at the MAC in this figure) is a (vendor specific) cross-layer functions, controlling actions at the RLC, MAC and PHY layers. Details of the scheduler are not part of the concept, but example high performance scheduling functions are described in Section 7.3.

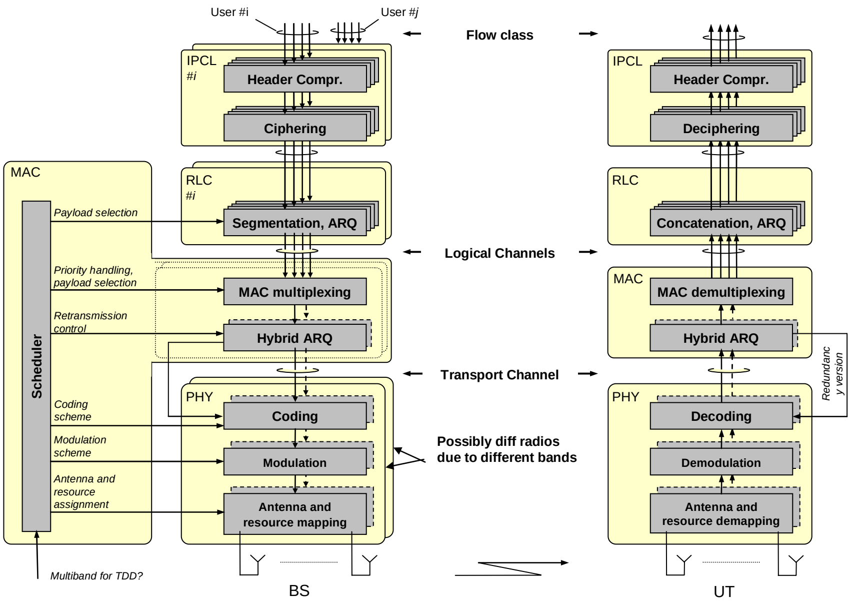

Figure 4-3: Mapping of user plane functions onto protocol layers.

Service access points (SAP) for peer to peer communication (that is, between two peer layers in the communicating protocol stacks) are marked with ellipses at the interface between layers. The SAP between the MAC and PHY layers provide the transport channels, and the SAPs between the MAC and RLC layers provide the logical channels.

4.3 RRC

The Radio Resource Control Layer controls the radio resources and configures the user terminal accordingly. The RRC includes measurement, exchange and control of radio resource-related indicators and commands between the WRAN and the UT_{LN}s.

The **measurements** include the determining values of standardized radio resource indicators that measure or assist in estimation of the available and potentially available radio resources.

The **exchange** of radio resource related indicators includes the procedures and primitives between logical entities used for requesting and reporting such measurements or estimations. The resulting information from exchange may be made available within the measuring stations using proprietary procedures and primitives that are not subject to standards (implicit), or, to a remote functional entity using standardised procedure and primitives (explicit).

The **control** mechanism refers to the decision made by the measuring station or remote entity to adjust radio resources based on the reported measurements, other information or using functions defined by the radio resource management (RRM) functions, and communicating such adjustments to logical entities using standardized primitives. More detailed RRM functions are explained in Chapter 6 and Chapter 7.

The control and configuration of RN_{LN} is carried out by introducing a new sub layer (RRC2) within the RRC layer. Similarly to RRC, the signalling from RRC2 is routed through an own RLC instance for the sake of reliable data transfer. Observe that RRC2 instances are available only in BS_{LN} and RN_{LN}, i.e. UT_{LN} functionality is not involved. This is also the case for the placement of those RLC instances that are used by the RRC2 protocol.

4.3.1 Basic RRC Functions

The control plane functions to be performed by the Radio Resource Control (RRC) layer are proposed and explained as follows.

- Broadcast of System Information
    - Basic cell-identification and cell-specific information that changes frequently are candidate for transmission by the broadcast channel
    - Other Possible candidates:
        - Spectrum sharing restriction parameters (slow)
        - Shared band availability
        - Cell ID, operator ID, TX power mask, FDD/TDD duplex mode info.
        - Pointer to next important control channel (super-frame allocation table)
        - Inter-system handover: Coupling point (gateway level versus IMS), slow.
        - Basic properties of the RACH (changes slowly)
- Admission control
    - The admission control functions performs admission of new flow classes, renegotiations of flow classes, or flow classes handed over from another BS. For more details of a possible implementation, see section 7.10.3.
- Establishment, maintenance and release of an RRC connection between the UT and WRAN including:
    - Allocation of temporary identifiers between UT and WRAN
    - Configuration of radio resources for RRC connection
- Security functions including:
    - Integrity protection for RRC message;
    - Ciphering of RRC messages;
- Establishment, maintenance, reconfiguration and release of flow classes for both unicast and multicast services
- Mobility functions including (for details see section 6.4):
    - Paging.
    - UT measurement reporting and control of the reporting for inter-cell and inter-RAT mobility;
    - Inter-cell handover;
    - UT cell selection and reselection and control of cell selection and reselection;
    - UT Context transfer between BSs, including
        - Control plane

    - User plane with WINNER context transfer at RLC SDU level. Another alternative solution is described in Chapter 7.
- UT measurement reporting and control of the reporting;
- Flow control in between BS and RN
- Cross layer configuration of layer 2 protocol entities within the same node

In case of relaying, at least the following RRC functions are included in the RN_{LN}:
- Broadcast of system information
- QoS management functions
- Paging

The RRC 2 protocol includes the following control plane functions:

- Establishment, maintenance and release of an RRC2 connection between the RN and WRAN
- Conveying broadcast information to the RNs
- Conveying information about establishment, maintenance, reconfiguration and release of flow classes for both unicast and multicast services for UTs served by the RN
- Conveying information related to resource partitioning
- Conveying paging information
- Cross layer configuration of layer 2 protocol entities within the same node, based on the above information

4.3.2 WINNER RRC states

In WINNER, active and idle states are two basic RRC states for the UT_{LN}s. Inside the active main state, there will be a separation of the states according to the UT activity, namely 'Active' and 'Dormant' sub-states. Besides that, detached is another state. The inter-transition among them is shown in Figure 4-4.

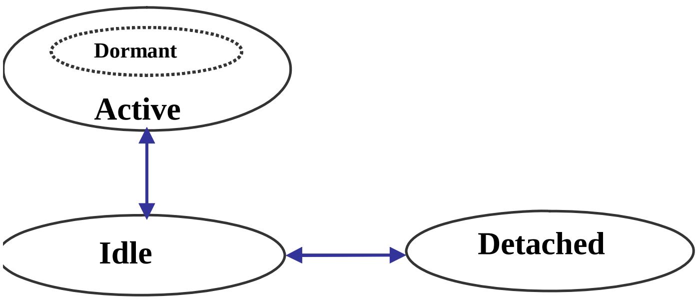

Figure 4-4: RRC States of the UT_{LN}

The RRC states are explained as follows:
- UT_{LN} Detached:
    - The network is not aware of this UT_{LN} (e.g. UT_{LN} is switched off or operating on another radio system for single band UT_{LN}s)
- UT_{LN} Idle:
    - After switch on the UT_{LN} camp in the most appropriate cell (for example: a cell with the highest signal strength). The UT_{LN} is able to send and receive system information and cell information
    - The UT_{LN} is maintained in this state until it is going to send or receives user information
    - Inherently power saving state: UT_{LN} should maintain this state several days
    - UT_{LN} performs cell selection and reselection autonomously
    - UT_{LN} monitors broadcast and paging channels
    - UT_{LN} is handled by the GW_{LN} node
    - Tracking/paging area: UT_{LN} belongs to the group of cells controlled by a GW_{LN} (after the initial connection the UT_{LN} can change position without notify it to the network)
- UT_{LN} Active:
    - UT_{LN} is moved to this state from idle state when transmitting/receiving data on traffic channels
    - UT_{LN} monitors the directed/common control channels continuously

  - UT is handled by BS
- There is a power saving sub-state within the Active state. This is the dormant sub-state where the UT_{LN} is ready to transmit or receive data on traffic channels and monitors the control channels discontinuously.

The change from the Active to the Dormant sub-state will be activated when no user plane transmission or reception is expected by the UT_{LN}. The aim is to reduce power consumption in the active state, with reduced activity cycle. The user plane connection between BS_{LN} and UT_{LN} is maintained.

4.3.3 WINNER RRC Interworking with Other Layers

The RRC layer manages and controls the use of radio resources and therefore has links over Control Service Access Points (SAP) to all other layers. The interworking enables RRC to control and configure other layers. By definition, the SAPs are solely used for interworking between protocol entities between layers of the same protocol stack and the RRC, instead of inter nodes. The WINNER system allows connections between RRC and lower layers that enable the receipt of measurement data from the lower layers as well as the control functions in the different layers.

The RRC layer provides the WRAN portion of signalling connections to the upper layers to support the exchange of upper layer's information flow. The signalling connection is used between the user equipment and the core network to transfer upper layer information. For each core network domain, at most one signalling connection may exist at the same time. The RRC layer maps the signalling connections for one UE on a single RRC connection.

4.4 Layer 2 overview

Layer 2 is split into three (sub) layers: IP Convergence Layer (IPCL), the Radio Link Control (RLC), and the Medium Access Control (MAC). Figure 4-5 and Figure 4-6 show the main functions performed at these layers for the downlink and uplink, respectively.

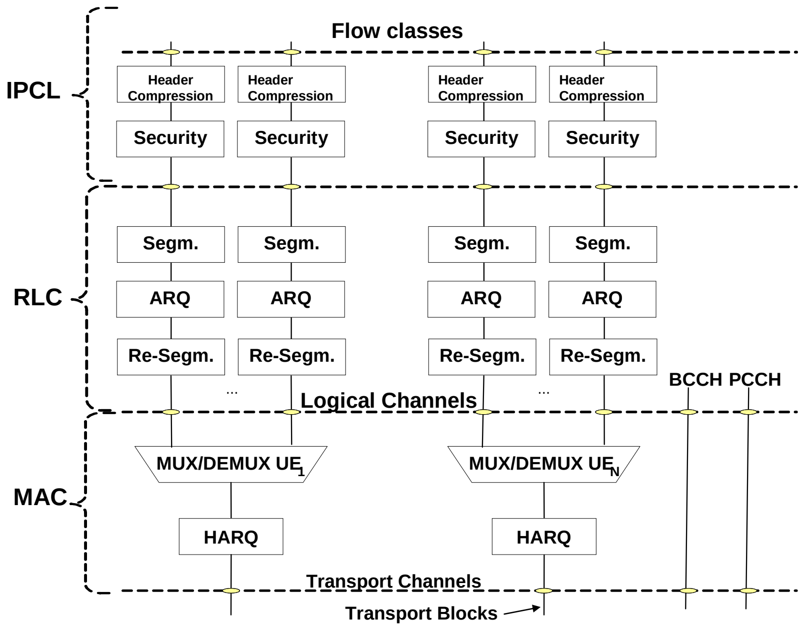

Figure 4-5: Layer 2 overview in the downlink.

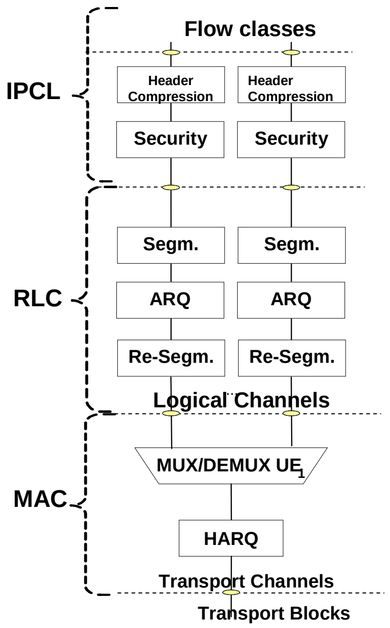

Figure 4-6: Layer 2 overview in the uplink.

4.5 L2: IP Convergence Layer (IPCL)

The IPCL layer is a pure user plane protocol, and hence does not exist for the control plane.

The IPCL layer supports **transfer of user data**; transmission of user data means that IPCL receives an IPCL SDU and forwards it to the RLC layer and vice versa.

The IPCL protocol is assumed to support **header compression** and **decompression**. Furthermore, IPCL supports **ciphering**. The IPCL shall also support **in-sequence delivery** of upper layer PDUs. Such reordering is performed by the GW_{IPA} in order to handle potential out-of-order packets that may happen during handover due to different transmission delays of the packets from the involved BSs.

Additionally, the IPCL performs **duplicate detection** of lower layer SDUs. Duplication of packets may occur both in the uplink and downlink direction if a correctly received packets ACK is not received by the transmitter, which then performs a retransmission of the same packet.

This layer also supports **reordering** of the downlink RLC SDUs at least during handover. Such reordering is performed by the UT in order to handle potential out-of-order packets that may happen during handover due to early forwarding of packets from the GW_{IPA}.

4.6 L2: RLC

The RLC layer is a pure user plane protocol, and hence does not exist for the control plane. The RLC protocol supports an unacknowledged (UM) mode and an acknowledged mode (AM). Whether UM or AM is used needs to be configured per flow class.

Furthermore, the RLC layer supports **segmentation** and **concatenation** of RLC SDUs. Depending on the scheduler decision a certain amount of data is selected from the RLC SDU buffer and segmented and/or concatenated depending on the size of the SDUs. This selected data block becomes the RLC PDU to which a sequence number is assigned. This means that one transport block contains only a single RLC PDU per flow class.
However, there are two exceptions:

i) if a retransmission PDU does not fit into the new transport block, or
ii) if a received PDU at a RN does not fit entirely into the new transport block.

Then, the RLC PDU is re-segmented. In these two cases, a new sequence number is appended to the re-segmented packets by means of an extension header.

The number of re-segmentations is not limited. In WINNER phase 1 it was shown that it is beneficial to keep the same RLC sequence number space on the entire path between the base station and the user terminal (see 5.1.1 for further details) to enable fast retransmissions after a user terminal has moved from one node to another node within the relay enhanced cell (REC). (In other words, intra-REC mobility is support by RLC-ARQ.) Hence, RLC SDUs are not re-assembled in the relay node, and hence only re-segmentations are allowed, see Figure 4-7.

In order to allow the RLC SDU reassembly at the receiver, the RLC header carries the required segmentation, re-segmentation and concatenation information. The RLC sequence number will also be used at the receiver **for in-sequence delivery** to the RLC SDU reassembly entity. Details of the segmentation/reassembly process and its interaction with multiplexing at the MAC layer is described in Section 5.1.1 and details on associated signalling is provided in Section 5.1.2.

In AM, RLC is responsible for correcting residual HARQ errors by operating another ARQ protocol, the RLC-ARQ. The ARQ retransmission units are RLC PDUs or RLC PDU segments. If an RLC retransmission is required and the radio quality has changed significantly compared to the original RLC transmission, then the RLC protocol is able to perform a re-segmentation. In that case, RLC segments a PDU into smaller PDU segments. In case of relaying, the ARQ operation needs to be enhanced (e.g. with relay ARQ) in order to be able to determine what packets have been correctly received by the UT and RN, respectively, see Figure 4-7. Details of the ARQ operation including its interactions with HARQ is described in Section 5.1.3.

Moreover, in case of relaying, the RLC also performs forwarding (i.e. mapping of an incoming SAP to an outgoing SAP), see Figure 4-7. Finally RLC provides means for protocol **error detection** and **recovery** (e.g. reset), **duplicate detection**, and **SDU discard**.

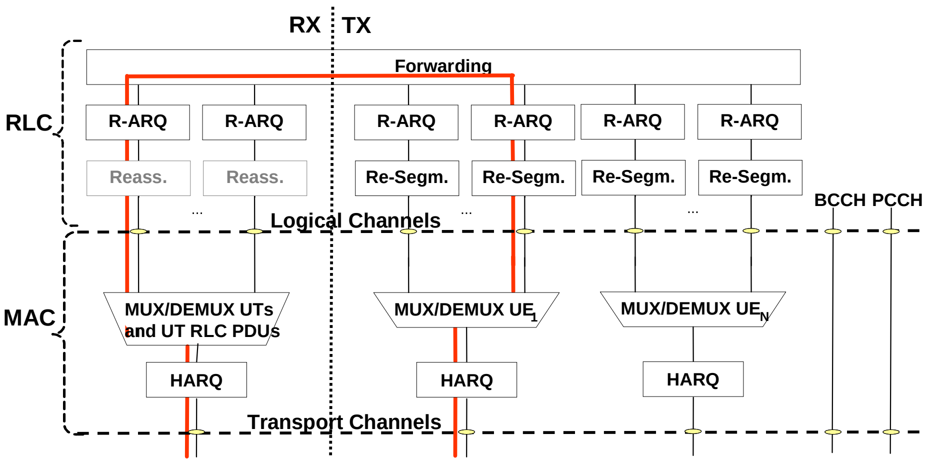

Figure 4-7: RLC layer at the relay node.

4.7 MAC

The WINNER Medium Access Control (MAC) layer performs three main tasks:

- S**cheduling,** which controls the transmission on the time-scales of the frames,
- **Multiplexing/demultiplexing,** and
- **HARQ**, i.e. a retransmission procedure over one hop that uses incremental redundancy.

The allocation of transmission resources is controlled by the network, not by the terminals, since we thereby attain the highest spectral efficiency. At the shortest time-scales, this control is performed by schedulers at the MAC layer of the BS_{LN} and RN_{LN}. The schedulers receive inputs and constraints from the RRM functions, which perform resource allocation and flow control at slower time scales.

The scheduler controls the complete transmission chain on a packet-by packet basis: It controls the segmentation at the RLC layer, the multiplexing at the MAC layer and the coding, modulation, multi-antenna processing and mapping onto transmission resources that are performed at the physical layer. The physical layer itself is completely controlled by the scheduler in the MAC layer. This fast and tight interaction is made possible since RLC, MAC and PHY layers of a node are always assumed to be co-located and can therefore interact with negligible delays.[1]

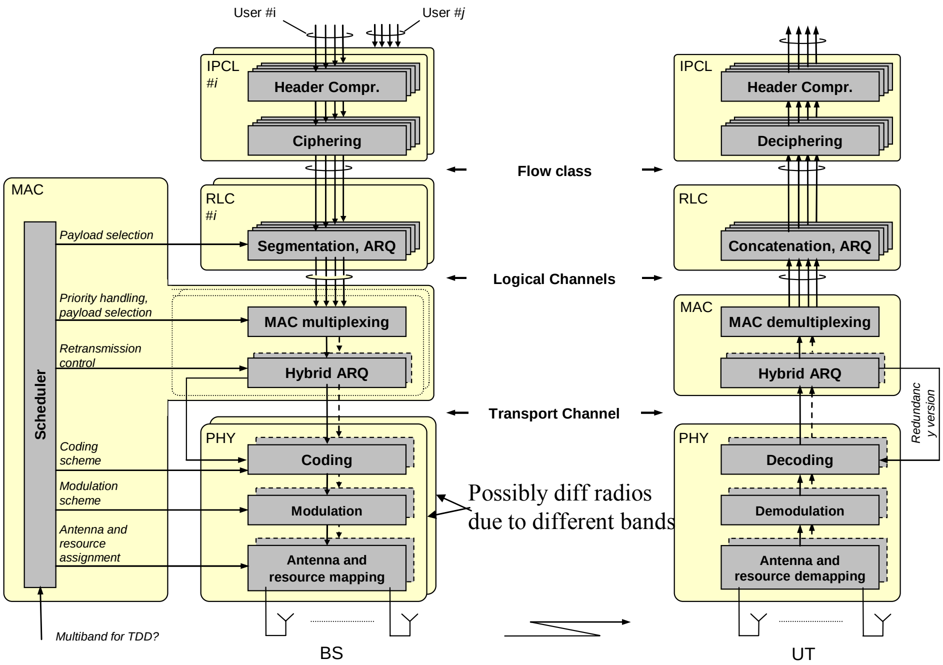

Figure 4-8: Overview of the MAC layer and the relation to other layers

If, in addition, several radios that use multiple bands are co-located, the scheduler can control the transmission resources within all the radio bands by controlling the multiple co-located physical layers.[2] This particular functionality is called a multi-band scheduler, see Section 4.6.3. All these aspects are outlined in Figure 4-8. The function of the scheduler and a proposed partitioning of it into functional entities will be discussed further in the reference design section 7.3.

[1] The scheduler of a BS or RN will also control the resource allocation used for uplink transmission from UTs.

[2] To be more precise, MAC layer scheduler can control all PHY layer resources that can be controlled on a frame time scale without additional signalling delays. This could be used for co-located base stations that use different spectrum resources (FDD and/or TDD). This is a departure from the WINNER phase 1 system concept, described in [WIN1D210] and [WIN1D76]. There, WINNER modes were defined as having different MAC layers. Co-located TDD and FDD access points would, for example, had different MAC layers and separate schedulers. This is no longer the case in the WINNER II concept.

4.7.1 Super-frames, frames, slots and chunks

The super-frame is a time-frequency-spatial resource allocation unit of the WINNER system concept. It fulfils two purposes:

- Each super-frame contains uplink and downlink synchronization pilots that are used for self-organizing network synchronization [WIN2D233]. The super-frame duration therefore scales the reaction time of this mechanism.
- The ***resource partitioning*** (the division of transmission resources between cells and between RNs and BSs) is assumed to be specified one super-frame in advance. This provides a stable background for the resource allocation that is performed by the scheduler.

The super-frame is assumed to have equal duration in the WINNER FDD and TDD modes. This facilitates inter-mode cooperation and multi-band transmission.

A super-frame consists of a short preamble with OFDM symbols used for pilots, followed by n ***frames.*** The n parameter is for further optimization, and is for now set to 8. In the frequency dimension, the super-frame comprises all frequencies (not necessarily adjacent) that are used within a cell.The spatial dimensions are denoted ***layers***, see Section 4.8.2.

A ***frame*** is a temporal resource unit. The frame duration has been set equal in the WINNER FDD and TDD modes. The frame duration contains two ***slots:***

- In a TDD mode, a frame consists of a downlink transmission slot followed by an uplink slot, separated by duplex guard-times.
- In an FDD mode, we likewise define a frame of duration equal to the TDD frame, that consists of two slots. In the cell, one set of half-duplex terminals would receive downlinks in the first slot and transmit in the uplink in the second slot. A second set of terminals could do the opposite, since an FDD base station can be assumed to use full duplex. Full-duplex FDD terminals could transmit and receive in both slots, which doubles the maximal data rate.

The super-frame used for performance evaluations of WINNER II is described in [WIN2D6137]. Its properties were proposed in WINNER phase 1. In the final WINNER II System Concept, the super-frame has been modified as follows:

- No contention-based uplink (DAC) channel is used, so separate transmission resources are not set aside for this purpose.
- The downlink physical broadcast channel (PBCH) and the uplink physical random access channel (PRACH) are no longer included in the preamble of each superframe. This increases the flexibility, since the time-scale of broadcast transmission and random access opportunities are decoupled from the superframe timescale. The broadcast channel is transmitted at a pre-specified position every m:th super-frame. The random access channel is available in every super-frame ($r = 1$), to reduce the delays for UT initial access to the network.[3]
- To eliminate unused frequency resources, the downlink and uplink network synchronization pilots have been moved into the first frame of the superframe, in the beginning of the downlink and the uplink slot, respectively of that frame. As in the case of the PBCH and PRACH, they utilizes only spectral bands that are available over a wide geographical area, to facilitate multi-cell coordination. They each comprise three consecutive OFDM symbols. This corresponds to the minimum B-EFDMA block size of the reference design (cf. Section 7.3.2.2), so it fits well into the WINNER II frame structure.
- Both the FDD and the TDD super-frame preambles include one OFDM symbol that contains uplink pilots.

The so modified super-frames thus consist of an uplink pilot preamble of one OFDM symbol followed by n frames. It is illustrated in Figure 4-9 below for the case of TDD transmission with uplink:downlink asymmetry ratio 1:1. In the performance evaluations of WINNER II, $n = 8$ frames per super-frame have been assumed.

[3] The exact position within the super-frame, the utilized spectral band and the repetition frequencies m have not been further specified.

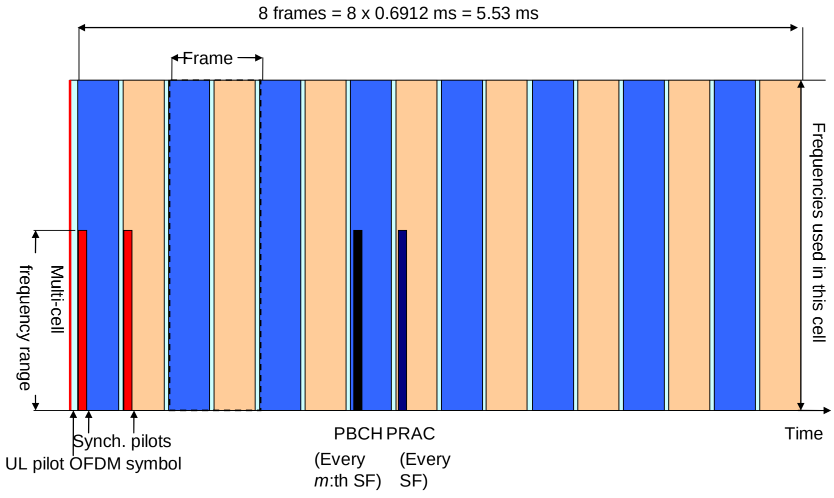

Figure 4-9: Super-frame for the WINNER TDD mode, with uplink:downlink assymmetry ratio 1:1 Light blue slots are for downlinks, beige are uplink slots. White bands represent duplex guard times between downlink and uplink slots.

With cyclic-prefix OFDM, the smallest time-frequency resource unit consists of one subcarrier by one OFDM symbol duration, here denoted a ***channel symbol***. Rectangular sets of n_s subcarriers by n_t OFDM symbols are assumed to be grouped into (time-frequency) ***chunks***. A chunk within one spatial layer is denoted a ***chunk layer***. The dimensions of the chunks that have been used for evaluations in WINNER II are illustrated in Figure 4-10 below. The frame duration is assumed equal in the FDD and the TDD modes and has been set to 0.6912 ms throughout the WINNER projects. The slot duration is half a frame in FDD and in TDD with asymmetry ratio 1:1 and is thus 0.3456 ms. These durations are the basic scales of delays and reaction speeds of the WINNER MAC layer.

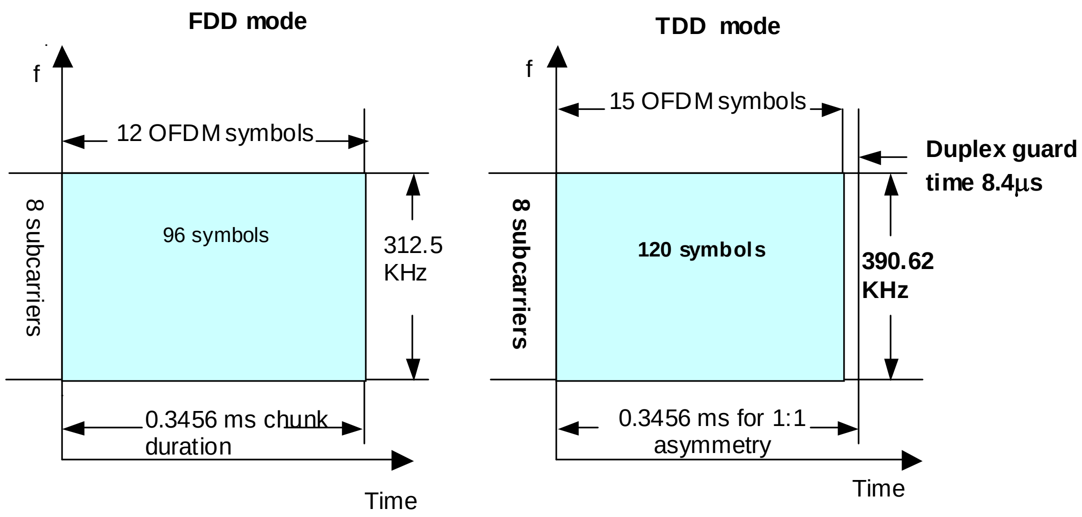

Figure 4-10: Example chunk sizes, used for evaluation purposes in WINNER II [WIN2D6137]. The figures show a slot (half of a frame) in each case, assuming 1:1 TDD asymmetry.

4.7.2 The main MAC layer functionalities

The MAC layer may multiplex RLC PDUs that belong to different flow classes. The resulting packets form retransmission units handled by the hybrid automatic repeat request (HARQ) function that works over each hop in a multihop transmission. The WINNER system supports small retransmission delays allowing HARQ to be invoked for most flow classes, also delay-sensitive flows, see Section 5.2.3. The whole procedure, as well as the physical layer processing, is controlled by the multi-layer scheduler. These three main MAC layer functions are described below.

Multiplexing. Packet processing proceeds by the following steps:

- Multiplexing of segments that belong to different flow classes can be performed by the MAC layer. The aim is to reduce the overhead, in particular for smaller packets and in transmissions over relay links. Multiplexing is allowed for segments (RLC PDUs) that are transmitted to the same logical node, for example a relay node. Segments to be transmitted to different logical nodes are not multiplexed.
- A MAC MUX header, (cf. Section 5.1.2) is appended to the optionally multiplexed RLC PDUs, resulting in the ***MAC PDU***.
- If link retransmission is to be used, a Cyclic Redundancy Check (CRC) code sequence is added to the MAC PDU. This represents the MAC ***Retransmission unit (RTU)***, also denoted a ***transport block***.

HARQ. The retransmission scheme used over one link is described in more detail in Section 4.2 of [WIN2D223]. It combines forward error correction and automatic repeat request (ARQ) in a scheme that uses incremental redundancy.

- The RTU is encoded as one or several Forward Error Correction code ***(FEC) blocks***, The preferred WINNER coding schemes are used:
  - Rate ¼ convolutional coding is used for blocks of < 200 payload bits,
  - quasi-cyclic block LDPC with mother code rate 1/3 is used for larger block sizes.

  Both are assumed to be systematic codes, so the FEC blocks consist of uncoded bits followed by redundancy bits. Assume that the first N_3 bits of the FEC block are systematic bits (the uncoded segment), the remaining bits are redundancy bits.
- A ***HARQ segment*** is then transmitted. The initial HARQ segment then uses the first $N_2 \geq N_3$ bits of the FEC block, where N_3/N_2 is the appropriate code rate that is specified by the link adaptation.
- After reception of the initial HARQ segment for all FEC blocks that comprise one RTU, the CRC code of the RTU is used to detect transmission errors.
- If retransmission is required, additional parity bits are transmitted by additional HARQ segments, denoted Incremental Redundancy ***(IR) blocks***. Each IR block uses a link adaptation that is appropriate for that transmission. Soft bit combining is used at the receiver. Since the CRC code spans the whole RTU, incremental redundancy transmissions have to be performed for all FEC blocks that belong to the RTU.
- When reaching the end of the FEC block without a correct reception, the HARQ process uses the FEC block cyclically from the beginning, producing additional IR blocks until a maximum allowed number of retransmissions is reached.

This scheme provides a seamless transition from the use of incremental redundancy at few transmissions to Chase combining for many retransmissions. The size of the IR blocks can be adjusted to a fraction of the initial HARQ segment size, see [WIN2D223], Section 4.2. This choice is signalled as out-band information at the initial transmission.

N-channel stop and wait with 1-bit feedback is used as the retransmission protocol for a flow class. This means that we allow each flow class to use up to N parallel retransmission channels. A new RTU can be transmitted if there are less than N outstanding unacknowledged RTUs, i.e. if at least one channel is not in use.

Scheduling:

A scheduler coordinating the RLC, MAC and PHY layers (but here described in the MAC context) is located in each BS and RN, see Figure 4-8, to provide efficient resource allocation.

The resource allocation allocates all available transmission resources within the slot (i.e. distributes time-frequency-spatial resources to different UTs). Some resources may be pre-allocated over multiple slots for transmission of flows whose packets have regular properties.

In short, the BS always performs the channel aware resource allocations, whereas the flow class prioritization always is performed by the transitting part. This means that:

- For the downlink the scheduler performs resource allocation and prioritizes between different UT flow classes within those resources.
- For the uplink, the scheduler in the BS (or RN) performs resource allocation, however, a simpler scheduler in the UTs prioritize between the different uplink flow classes for that UT.
- In case of cooperative relaying (downlink transmission from multiple RAPs), the scheduler at one coordinating RAP performs the scheduling (payload selection resource allocation and mapping, link adaptation). It signals these decisions to the schedulers in all participating RAPs.

The scheduler furthermore selects the transport format and the spatial transmit scheme individually for each HARQ segment/code block. The WINNER concept supports individual link adaptation within the different chunk layers that are used for transmitting one HARQ segment.

The scheduler has the following main types of inputs:

Static and slowly time-varying constraints and inputs.

- The antenna resources at the network side, the multi-antenna and other capabilities of the BS, RNs and UTs involved.
- Restrictions on the resource allocation that originate from the RRM functions (spectrum functions, resource partitioning between RN/BS, interference avoidance between cells, use of cooperative transmission).
- Properties and QoS constraints as defined by the flow class of the packets.
- The total load situation and distribution of users in the cell and the restrictions and possibilities for using SDMA (Spatial Division Multiple Access) that this implies.

Inputs and constraints that vary on a frame time-scale:

- Transmission requests for uplink transmissions.
- Queue lengths/transmission demand for each flow class.
- CQI and CSI information for time-frequency-spatial transmission resources within the frame to be scheduled.
- The need to transmit packets with high priorities, such as IR blocks (retransmissions).
- Scheduling constraints arising from cooperative relaying: One node then determines the transmission parameters; the schedulers at the involved RAPs receive these decisions as constraints on the scheduling allocation.
- Time-to-live information of individual packets in delay critical transmissions. (Depends on number of hops, delay budget, queue occupancy and relay link flow control input).

The scheduling has the over-all aim of satisfying the quality-of-service (QoS) constraints of all flow classes. By channel-aware scheduling, it can also allocate transmission resources that are advantageous for each transmission, to optimize the network capacity or the terminal power consumption.

The scheduling structure supports ***joint*** decisions that solve the following tasks (the exact implementation is manufacturer dependent):

Payload selection: A decision on what flow classes to transmit in the scheduled slot.

Resource allocation: Allocation of time-frequency-spatial transmission resources within a slot to specific MAC PDUs.

Resource mapping: Decision that determines the mapping of HARQ segments onto a set of transmission resources, using a set of link adaptation parameters. This decision is optimized jointly with a decision on segmentation/concatenation and the choice of link adaptation parameters.

Link adaptation: Decision on the modulation parameters, the spatial layer/beam and the code rate used for transmitting a code block.

As output, the scheduler provides

- A detailed allocation of packet segments to physical time-frequency-spatial transmission resources, including spatial link adaptation. This includes segmentation/concatenation commands to the RLC layer, multiplexing commands to the MAC layer and instructions on the modulation, encoding/puncturing, spatial scheme and time-frequency resource mapping performed at the physical layer. The structure of packet headers is outlined in Section 5.1.
- Out-band (i.e. control information not carried within any protocol header) frame control information that enables the receiver to decode the relevant part of a frame. This information is transmitted on the physical channels PDFCC and PUCH for the downlink and uplink directions, respectively, see section 4.8.6. The out-band frame control information and its corresponding overhead is also discussed in more detail in Annex A.1.

Figure 4-11 below illustrates the main input and output parameters of the scheduler in downlinks. It illustrates how the scheduler can be partitioned into a constraint pre-processor that handles various constraints on the transmission resources, a main scheduler and a sub-function that focuses on optimising the link adaptation. The allocation problems, principles and solutions will be discussed further in Section 7.3 below.

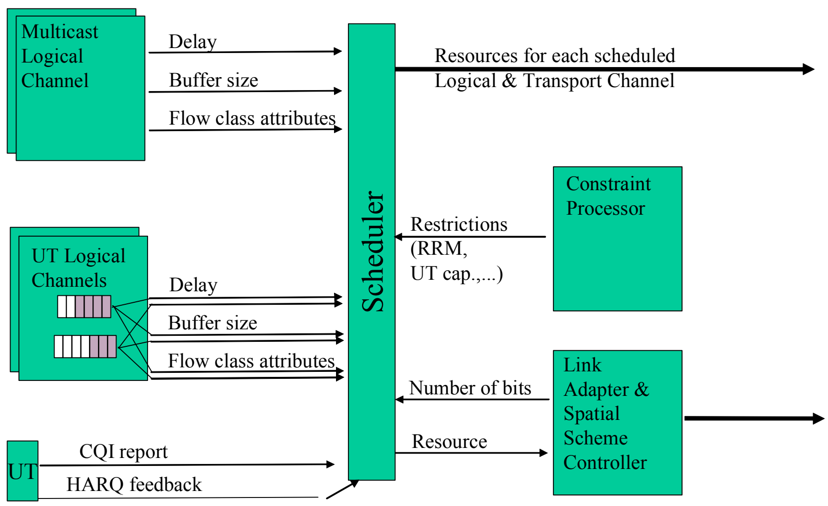

Figure 4-11: Scheduler: main inputs and outputs.

Transmission sequences and transmission timing are discussed further in Section 5.2.

4.7.3 Multi-band scheduling

When the MAC layer controls several physical layers that are co-located so that control transmission delays are negligible, then these multiple bands will just represent a widened resource pool as seen from the scheduler. One constraint imposed by the WINNER multi-band scheduling concept is that the UT should never have to transmit/receive simultaneously in several bands that are not covered by the same radio (same FFT), to simplify the terminal design. This means that MAC PDUs and their retransmissions should be transmitted in either band, not in all bands simultaneously. The scheduler may however decide to switch the transmission to/from one UT to another band, to obtain better load balancing.

A HARQ channel is regarded as associated with one particular radio band. When the transmission is switched to another band, the HARQ channel/state is associated to this new band. Then, HARQ transmissions for previously transmitted packets proceed over the new band.[4]

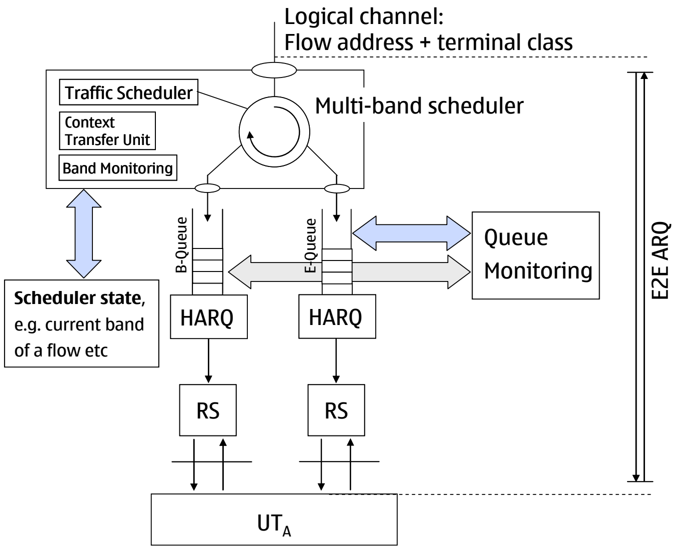

Figure 4-12: Functional Description of the Multi Band Scheduler, enabling operation on multiple bands and fast context transfers between these bands (RS= Resource Scheduler).

4.7.4 Description of logical channels, transport channels and their mapping

The *logical channels* define types of communication between RLC peers. They are characterized by the type of information transmitted. Some are for downlink information, some for uplinks only, while some are bi-directional.

The *transport channels* define types of communications between MAC peers. They are characterized by how and with what characteristic the information is transmitted. The logical channels used in the WINNER concept, their use for uplink/downlink transmission and their mapping on transport channels are outlined in Figure 4-13 below.

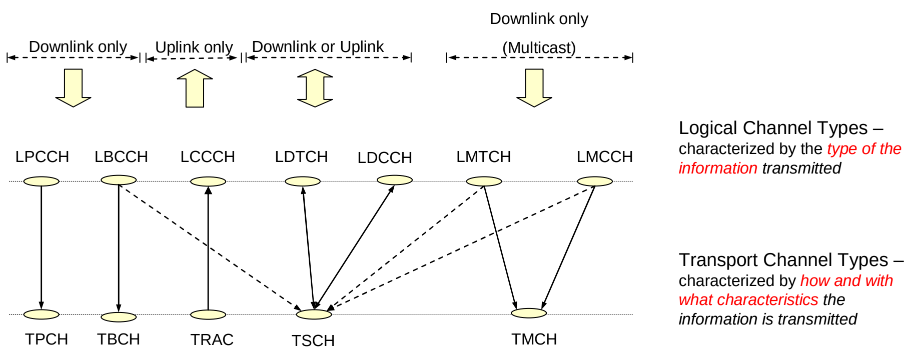

Figure 4-13: Mapping from Logical channels to/from Transport channels.

[4] This is possible with insignificant delays due to the assumed physical co-location of the radios/physical layers.Multi-band transmission that uses non-colocated radios has to use the RRM handover functionality.

The *physical channels* describe different distinct sets of physical resources that are required by various transmissions.

The logical and transport channels are described briefly below, while the physical channels are described in Section 4.8.6.

Logical Broadcast Control Channel (LBCCH)

The LBCCH is a downlink channel for broadcasting system control information. The LBCCH information is split in two parts: the static part and a dynamic part. The static part is always transmitted using the same pre-allocated physical resources every *m*: th superframe. The location of the dynamic part is signaled in the static part. The dynamic part is transmitted every *y*:th superframe (where *y* is a multiple of *m*), using the shared transport channel TSCH.

The static part contains information needed for an unknown UT to identify basic cell properties and to be able to initiate communication, e.g cell ID, operator ID, FDD/TDD duplex mode info. Other candidates for inclusion in the static part are spectrum sharing restriction parameters and RACH properties. In the dynamic part, parameters like transmission power mask are signalled.

Logical Paging Control Channel (LPCCH)
A downlink channel that transfers paging information (Section 6.4.1). This channel is used when the network does not have information about the location cell of the UT.

Logical Common Control Channel (LCCCH)
This channel is used by the UTs having no RRC connection with the network. It is used at initial access to the network.

Logical Dedicated Control Channel (LDCCH)
A point-to-point bi-directional channel that transmits dedicated control information between a UT and the network. Is used by UTs having an RRC connection.

Messages for setup of links from/to a RAP to/from a UT use the LDCCH. Various RRM mechanisms (Section 6) also use the LDCCH.

Logical Multicast Control Channel (LMCCH)
A point-to-multipoint downlink channel used for transmitting multicast and broadcast message service (MBMS) control information from the network to the UT, for one or several transport multicast channels (TMCHs). This channel is only used by UTs that receive MBMS.

Messages for the setup and modification of multicast groups use the LMCCH.

Logical Dedicated Traffic Channel (LDTCH)
A Dedicated Traffic Channel (LDTCH) is a point-to-point channel, dedicated to one UT, for the transfer of user information. A LDTCH can exist in both uplink and downlink.

Logical Multicast Traffic Channel (LMTCH)
The LMTCH is a point-to-multipoint downlink channel for transmitting traffic data from the network to the UT. This channel is used only by UTs that receive MBMS services.

Transport Broadcast Channel (TBCH)
The TBCH is a downlink channel for broadcasting system information, etc., to all terminals inside the cells coverage area.

Transport Paging Channel (TPCH)
The TPCH is a downlink channel used for broadcast of paging information into an entire cell.
The Paging Indication (PI) is a short message that indicates the range of UT addresses being paged within this cell.

The paging message (PM) contains the paging reason (>3 bits), the paging domain (1 bit) and the paging identity (32 bits).

Transport Random Access Channel (TRAC)

The TRAC is a contention-based uplink channel for initial access to the network. As such this channel is used to obtain timing synchronization (asynchronous random access).

Transport Shared Channel (TSCH)

This is a point-to-point data channel for both the uplink and the downlink for both user and control data.
It is possible to broadcast this channel in the downlink over the entire cell.

Transport Multicast Channel (TMCH)

This is a separate transport channel for multicast transmission. It is to be utilized by Multicast Broadcast Messaging Services (MBMS). This channel is broadcast in the entire coverage area of the cell. Combining of multicast transmissions from multiple cells is supported

4.8 Physical Layer PHY

The PHY layer handles the physical transmission of flows and of measurements and control signalling that is directly related to the radio interface. The PHY layer is not separated into user plane and control plane, since it is assumed that all control functionality of the PHY layer resides within the MAC layer.

This section introduces the main characteristics of the physical layer of the WINNER concept. Additional aspects on PHY-MAC interactions, in particular the transmission timing and control loops are discussed in more detail in Section 5.2. The parameterisations and algorithms that represent the WINNER reference design will be summarized in Section 7.

4.8.1 Introduction and overview of the physical layer processing

The transmit technique:
In the WINNER system concept, the transmission within cells is performed by (close to) orthogonal use of time and frequency resources. This requires some form of multicarrier modulation or the use of orthogonal time-frequency basis functions. Out of several alternatives investigated in WINNER phase 1, the following were selected (cf. Section 2.2.2 of [WIN1D210]):

- Cyclic-prefix OFDM is used in downlinks and in uplinks that are not power limited.
- DFT-precoded CP-OFDM is used in uplink transmissions in power-limited scenarios.

At the end of WINNER phase 1, several possible types of DFT precoding were under consideration. In WINNER II, it was decided to focus on a technique that DFT-precodes blocks of regularly spaced subcarriers. This technique, which reduces the signal envelope variations, harness the frequency diversity of the channel and provides an appropriate resource granularity for small packets, is denoted Block-Interleaved Frequency Division Multiple Access, or B-IFDMA [WIN2D461],[SFF+07]. Aspects on its design are discussed further in the reference design, in Section 7.3.2.

Modulation and coding:
As modulation schemes, BPSK and square M-QAM are used, using M = 4, 16, 64 and 256. As coding schemes, convolutional coding, using tail-biting and with mother code rate ¼ is used for FEC blocks of size up to 200 bits. For larger FEC block sizes, quasi-cyclic block LDPC codes have been selected as a mandatory scheme (it is included in the reference design), while duo-binary turbo codes (DBTC) is an optional scheme that has been used in various investigations [WIN2D223]. These coding schemes have been analyzed, refined and improved in WINNER II and they are described briefly in Section 4.8.3 and Annex A.2 below and fully in [WIN2D223].

At low SINRs, the coding may be combined with repetition coding. The lower limit SINR that has been assumed when evaluating the WINNER concept is -8 dB, corresponding to QPSK rate ¼ with repetition code rate 6.

Adaptive transmission:
WINNER phase 1 introduced two basic types of adaptive transmission (cf. Section 3.1.6 of [WIN1D210]):

- ***Frequency-adaptive transmission***, where payload bits from flows are allocated to chunk layers (rectangular time-frequency resources within a spatial dimension/layer). The corresponding multiple access scheme is denoted as ***chunk-based TDMA/OFDMA***. Individual link adaptation may be performed within each chunk layer. This link adaptation is adjusted to the frequency selective small-scale fading.

- ***Non-frequency-adaptive transmission*** averages over the frequency-selective fading. A code block is here interleaved and mapped onto a dispersed set of time-frequency-spatial transmission resources. The same link adaptation parameters (modulation and code rate) are used for the whole code block.

Frequency-adaptive transmission is an important feature of the WINNER concept that provides high spectral efficiency for single flows and significant multi-user scheduling gains for multiple flows [WIN1D24], [WIN2D223], [WIN2D61310]. Non-frequency-adaptive transmission provides lower spectral efficiency than frequency-adaptive transmission, but is available as a backup scheme whenever the former cannot be used. It is the main transmit scheme for most forms of control signaling. It is also used for multicasting transmissions, in which case the link adaptation parameters are adjusted to the recipient with worst SINR within the multicast group.

These basic techniques have been extended and refined in the following way in WINNER II:

- A reference design for frequency-adaptive transmission has been proposed and evaluated; see [WIN2D6137] and [WIN2D223]. It is denoted Mutual-Information based Adaptive Coding and modulation, or ***MI-ACM,*** and maps code blocks onto multiple chunks in a simple and efficient way. Briefly, link adaptation is used that applies constant transmit power but adjusts the modulation per chunk layer. An average puncturing and code rate for the whole block is calculated for the so link-adapted resources. The punctured block is interleaved and mapped onto the chunk layers. In this way, strong codes that work best with large code blocks can be combined with fine-grained link adaptation of resources within code blocks.

- Frequency-adaptive transmission was in WINNER phase 1 restricted to users with downlink SINR > 5 dB, to restrict the associated downlink control overhead to acceptable levels. With an improved principle for control signal transmission, described in Annex A1, this restriction no longer applies. This significantly increases the pool of potential users who can utilize frequency-adaptive transmission, thus improving the spectral efficiency and capacity of the system.

- The resource mapping used for non-frequency-adaptive transmission has been refined, see [WIN2D461] and Section 7.3.2. The selected mapping schemes are denoted B-EFDMA in downlinks and B-IFDMA in uplinks. The mapping uses orthogonal mapping of different code blocks onto small time-frequency blocks that are regularly spaced in frequency. The B-IFDMA scheme for uplinks uses DFT-precoding over a frequency-dispersed set of resources, to limit the signal envelope variations.

Principles for multi-antenna transmission

WINNER phase 1 introduced a generic multi-antenna transmitter, that can be configured for diversity transmission, beamforming, spatial multiplexing and multi-user MIMO transmission, cf. Section 4.8.2 below, or [WIN1D210] for more details. It involves a large set of possible transmit schemes as special cases. The multi-antenna transmit scheme can be selected individually, per flow class. In WINNER II, criteria and techniques for selecting the appropriate multi-antenna transmit scheme per flow class have been refined [WIN2D341] and are outlined in Section 7.3.3. This type of adjustment is denoted ***spatial adaptation***. In addition, preferred multi-antenna transmit schemes for different scenarios have been identified, and are given in Section 7.3.3.2.

Synchronization, channel estimation and pilots

Self-organizing network synchronization is an integral part of the WINNER concept, to avoid the need for other means to synchronize RAPs and UTs. The synchronization pilots have been introduced in Section 4.7.1 and the proposed synchronization scheme is outlined in Section 7.5.

A distributed regular pilot grid supports channel estimation at receivers and prediction of appropriate channel quality information for adaptive transmission. The design principles are outlined in Section 4.8.4 and Annex A.3 below, with particular parameterizations for different scenarios specified in Section 7.4. With this parametrization, the preferred multi-antenna transmit schemes combined with frequency-adaptive or non-frequency-adaptive transmission can be supported without excessive pilot overhead.

4.8.2 The Generic Transmitter

The generic configurable transmitter, depicted in Figure 4-14 was first introduced in [WIN1D27] and later refined in [WIN1D210] and [WIN2D341] to include also generalized multicarrier (GMC) processing. The generic transmitter is a flexible MIMO transmission concept which can realise both multiplexing, diversity and directivity gains utilising three main components:

- Per stream rate control, or code block segmentation, enable multi-level coding techniques and layered techniques such as Bell Labs Layered Space-Time Architecture (BLAST), Per Antenna Rate Control (PARC) and Per Stream Rate Control (PSRC) [WIN1D27] which may use successive interference cancellation after channel decoding.
- Linear dispersion codes [HH02], also known as matrix modulation or vector modulation, are used to realise flexible trade-offs between spatial multiplexing and spatial diversity. Spatial diversity gain is desirable in cases when diversity gain from other dimensions is not enough, while spatial multiplexing gain is important to improve system throughput.
- Linear precoding, also referred to as beamforming, apply a complex transmit weight coefficient per antenna, thus adjusting phases and amplitudes of transmit antennas. Physically this results in

changing the array beam pattern on the transmitter side, thus introducing directivity gain to the transmitted signal.

By configuring the building blocks of the generic transmitter from the MAC layer it is possible to realise a vast number of different multi-antenna transmit schemes, such as beamforming techniques, transmit diversity techniques, SDMA, multi-user MIMO precoding, etc. [WIN2D341].

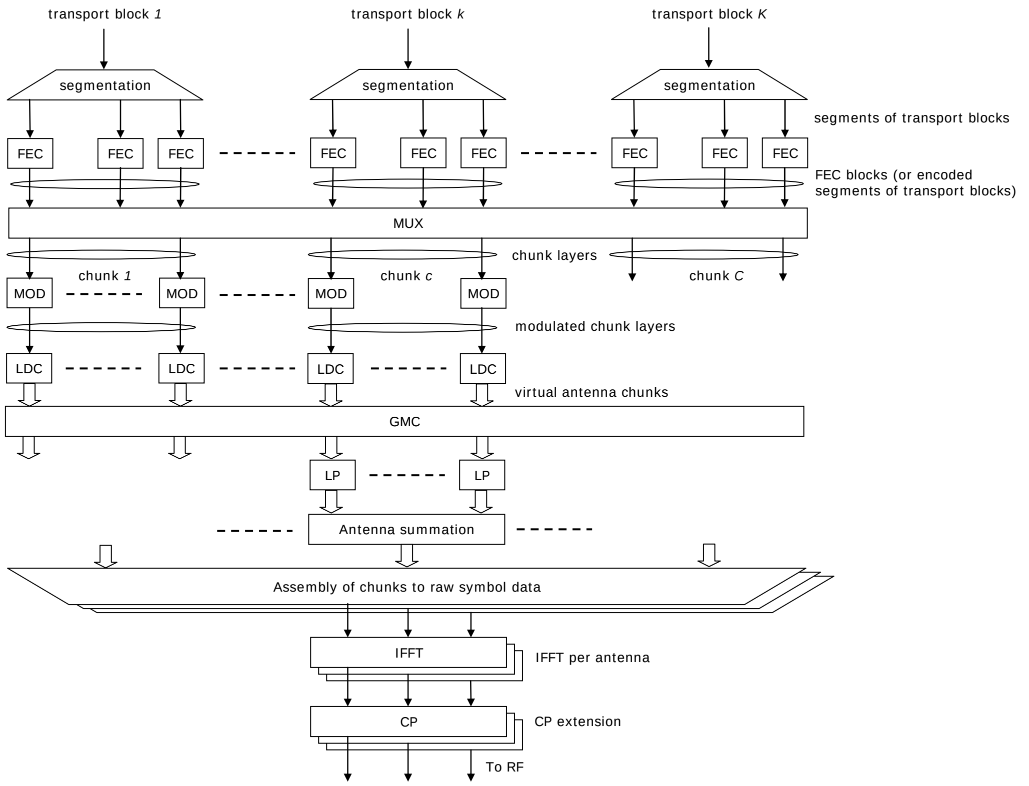

Figure 4-14: Generic Transmitter for GMC.

The generic transmitter is essentially invoked once per slot (i.e. time-resource unit) when incoming (L2) data units, referred to as transport blocks, are to be transmitted. Each transport block is segmented and each segment is separately channel encoded in a forward error correction (FEC) entity. These encoded segments of transport blocks are referred to as FEC blocks. The FEC entity supports Hybrid Automatic Repeat Request (HARQ) functions so that for each (re)transmission slot the relevant subset of bits is selected, as determined by the HARQ protocol status. The retransmission entity is the transport block. More details on the channel coding and HARQ schemes in the WINNER system concept are given in Section 4.8.3 below.

An important property of the radio interface is that each chunk is partitioned into one or several layers denoted as chunk layers. The above-described FEC blocks, i.e. encoded segments of transport blocks, are multiplexed onto these layers. Note that a chunk layer may carry several consecutive FEC blocks belonging to the same transport block. The multiple layers of a chunk allow several users, e.g. target UTs in downlink, to share the resources of a chunk, for instance, by means of spatial re-use in the form of SDMA per chunk.

The maximum number of layers in a chunk can be different for different chunks and arbitrary in relation to the number of physical antennas. The time and frequency resources of a chunk may be re-used in the sense that several FEC blocks of one or several transport blocks may be mapped onto the same chunk. It is possible for the receivers to demodulate the data of the different layers by means of appropriate dispersion of the layers onto the time, frequency and antenna resources of the chunk. In its simplest form, a layer may be mapped directly to a physical antenna to support spatial multiplexing. It is also possible to disperse a single chunk layer over all physical antennas by using a linear dispersion code. It should be noted, that the concept of chunk layer herein is more general as compared to e.g. [WIN1D76], where a layer is said to represent only the spatial dimension.

The bits mapped to each chunk layer are separately modulated. The so formed modulated chunk layers are then dispersed or spread onto virtual antenna chunks with a linear dispersion code which is a three dimensional entity spanning the adjacent sub-carriers of the consecutive OFDM symbols in time and frequency corresponding to the chunk in addition to the spatial dimension which has been added. Each chunk layer is thus in the general case mapped onto a three-dimensional virtual antenna chunk.

All virtual antenna chunks are then subject to GMC processing. The GMC function operates on an OFDM symbol basis in the frequency domain over all chunks allocated to a transport block. More specifically, these modulated layers of chunks are jointly processed by an identity function (when OFDM is used) or a Discrete Fourier Transform (DFT) function (for the case of pre-coded OFDM) and then split and dispersed over the chunks again. The data link layer controls the segmentation and multiplexing (MUX) function and controls the modulation and GMC processing.

The virtual antenna chunk of each layer is further subject to linear pre-coding, which means that each virtual antenna chunk of each layer is linearly mapped onto a physical antenna chunk. Finally, the layers' physical antenna chunks are summed per antenna to form a three-dimensional antenna chunk, which is passed to assembly and OFDM modulation per antenna.

4.8.3 Channel coding

Only few of the increasing number of subsets of LDPC codes are seen as serious candidates for next generation wireless systems ([LZ04], [LR+06]). Indeed for realistic future systems, many different constraints have to be taken simultaneously into account, such as e.g. performance, encoding and decoding complexity, decoder throughput (parallelism), resulting into what is called lately "Adequacy Algorithm Architecture" approach ([Dor07]).

Among those candidates, *Quasi-Cyclic Block Low-Density Parity Check Codes* (QC-BLDPCC[5]) are among the most promising ones ([Fos04]). The WINNER system employs QC-BLDPCC.

The full parity-check matrices for base-model matrices from LDPC Codes can be found in Annex of [WIN2D223].

WINNER uses advanced Type-II HARQ, also called incremental redundancy (IR), that use a single low code rate mother code, that is punctured to obtain multiple higher coding rates (rate compatible punctured code or RCPC).

Although the overall evaluation process carried out during the project ended up with selecting the QC-BLDPC Codes as primary and mandatory coding scheme for data transmission, the use of duo-binary turbo-codes (DBTC) is a suitable candidate for medium block length and low coding rates. Hence, also this alternative is open for the further development of the WINNER concept.

The modulation and coding requirements for control channel signalling are different than the ones for user data transmission. The information sent through the control channel is very important for proper functioning of the advanced protocols of the WINNER concept. Although the proposed QC-BLDPCC and DBTC provide an excellent coding performance as shown in [WIN1D210], they can not efficiently be used for encoding the control information due to very short packet sizes being considered (in the order of 25 information bits). Therefore low rate tail-biting convolutional codes (CC) are used for these cases.

The above channel coding schemes are described in more detail in Annex A.2.

4.8.4 Pilot design

The WINNER pilot design is a modular concept consisting of basic building blocks defined on the chunk level. These building blocks are [WIN2D233]:

- The **pilot pattern** specifies the position of pilots on the chunk. The pilot positions is chosen such that a globally regular pilot pattern is obtained, i.e. a two dimensional (2D) grid with equidistantly spaced pilots in time and frequency, which is advantageous for channel estimation by interpolation.

[5] Alternatively abbreviated as BLDPCC or BLDPC codes in the following

- The **pilot type** specifies whether pilots include user specific transmit processing or not.
- The **orthogonal pilot set** specifies whether pilots associated to different spatial streams are orthogonally separated in time and/or frequency, or pilots are spatially reused, i.e. pilots of two spatial streams are placed on the same subcarriers.

This modular concept avoids that several pilot patterns corresponding to different pilot types are inserted within a frame. Instead only one pilot grid is inserted in the frame and the pilot type is determined by the chunk specific spatial transmit processing. Thus, a highly flexible and adaptive system concept can be supported with a modest pilot overhead.

4.8.4.1 Types of Pilots for Multi-Antenna Transmission

A scattered pilot grid with orthogonally spaced pilot symbols in time and frequency was proposed in [WIN1D21] and [WIN1D210]. This however, may result in prohibitive pilot overheads unless some form of reuse is deployed. For instance, in local area deployment a distributed antenna array with up to 32 antenna elements is foreseen. Fortunately, spatial precoding schemes forming beams that are spatially well separated allow to spatially reuse pilot symbols [WIN2D233].

Dependent on the transmit direction (uplink / downlink) and the kind of spatial processing being used, several types of pilots are being distinguished [WIN1D210], as detailed in Annex A.3.1: *common pilots* that can be shared between users, and user specific *dedicated pilots.* Furthermore, common or dedicated pilots may include spatial transmit processing, referred to as common pilots per beam (CPB) and dedicated pilots per beam (DPB). Otherwise, pilots that are transmitted unweighted are termed common pilots per antenna (CPA) and dedicated pilots per antenna (DPA).

4.8.4.2 Inband pilot patterns

A generic framework for inband pilot patterns is briefly summarized by [WIN2D233]:

- Pilot symbols in frequency and time, with respective spacings D_f and D_t should be placed sufficiently close to satisfy the sampling theorem [WIN1D21] allowing to reconstruct the channel response through interpolation. To allow for realizable filters oversampling factors should be at least 20% and 100% for common and dedicated pilots, respectively.
- The pilot pattern only determines the position of the pilots within the frame. The type of pilot being used (CPA, CPB, DPA or DPB as described in Annex A.3.1) is determined entirely by the spatial transmit processing scheme that is used in a particular chunk (e.g. grid of beams (GoB), MU MIMO precoding, or linear dispersion codes (LDC)). This accomplishes that only one pilot grid is necessary to support all flavours of spatial processing schemes, and is therefore the key to keep the resulting pilot overheads at an acceptable level. The receiver implicitly knows which type of pilots is transmitted, as it is uniquely determined by the spatial scheme selection and the transmission mode (FDD or TDD).
    - An important requirement to allow for a chunk specific selection of the pilot type, is that the pilot spacing in frequency D_f is chosen such that the chunk width is an integer multiple of D_f.
- Pilots from different spatial streams are reused if the associated beams are well spatially separated. In case of overlapping beams or unweighted transmit signals, pilots are to be orthogonally multiplexed in time and frequency.
- For TDD and half-duplex FDD systems pilots should be placed near the beginning and end of a frame in time direction. The rationale here is that interpolation between pilots exhibits a smaller estimation error than extrapolation near the beginning and end of the frame.
- Dedicated pilots should be placed near the corners of a chunk, as interpolation between pilots exhibits a smaller estimation error than extrapolation at the chunk edges.

The pilot spacings in frequency and time, D_f and D_t, as well as the corresponding overheads Ω_p of the WINNER pilot design are shown in Table 4-1 (see Annex A.3.2 for details). The pilot reference design for the FDD mode is described for the wide area deployment in Section 7.4.1. The pilot reference design for the TDD mode is described for the metropolitan area and local area deployment in Sections 7.4.2 and 7.4.3.

Table 4-1: Pilot spacings and overheads for the WINNER pilot design. The overhead is given as a function of the number of orthogonal pilot sets P_n.

	FDD[6]	TDD[7]	B-IFDMA
D_f	4	4	4
D_t	10	12	3
Ω_p	$4.16\% \cdot P_n$, $(5.2\% \cdot P_n)$, $P_n=\{1,2,3,4\}$	$3.33\% \cdot P_n$, $(1.67\% \cdot P_n)$, $P_n=\{1,2,3,4\}$	$8.33\% \cdot P_n$, $P_n=\{1,2\}$

4.8.4.3 Uplink superframe pilot preamble

To provide the BS with short-term CSI and CQI an uplink pilot preamble is inserted in the beginning of each superframe in both FDD and TDD modes, so to obtain estimates of the unweighted channel matrix. As uplink pilots over the full-band are very expensive in terms of overhead and UT power consumption, these full band pilots are inserted at a lower rate (once per superframe). This essentially limits the maximum velocity for frequency-adaptive transmission to 10 km/h in FDD uplinks. See also Section 7.3.2.4. Pilots are orthogonally multiplexed with a pilot spacing of D_f = 8, shown in Figure A-12 in Annex A.3.

4.8.5 Control Signalling

Two main principles have been used in the WINNER concept for selecting the physical format for transmitting control signalling and measurement reports:

- Is the control signalling urgent or non-urgent, in the sense that timing allows it to use existing HARQ schemes?
    - Any *non-urgent* control messages and measurement reports are in the MAC layer treated as packets to be transmitted using the TSCH transport channel. HARQ is used to reduce and control the error probability.
    - *Urgent* control messages need special FEC protection and may also have to be placed in special positions within the frame for feedback loop timing reasons. Special control channels (PDCFC, PDFCC and PUCH) have been defined for these urgent control messages. Here, we use non-frequency-adaptive transmission with small block sizes to maximize frequency diversity and to obtain precise message timing within the frames. Whenever possible, frequency diversity is to be combined with spatial/polarization diversity schemes to reduce the error probability.
- Does the control signalling need to be broadcasted or can it be multicast/unicast to UTs or groups of UTs? An optimisation is assumed to be performed to select the best type of transmission for each control message. Broadcasting results in transmission with the lowest spectral efficiency, since it has to be adjusted to the worst user, with possible not well known SINR. Two channels that need to be broadcasted have been defined, the physical broadcast channel PBCH and a channel PDCFC for frame control messages. To keep the control overhead acceptable, it is important to minimize the payload that needs to be transmitted over these two channels. Downlink control messages to groups of users (multicast groups) may be either transmitted individually to each user (unicast) or multicast. Multicasting is the most efficient scheme for downlink control information if the multicast groups contain sufficiently many members. Otherwise, control messages are preferably unicast.

The limit between urgent and non-urgent messages is determined by the possibility to perform at least one HARQ retransmission. With the delays indicated in Section 5.2.3, this limit is around 4 ms. The various

[6] Pilot overhead in brackets correspond to chunks of high velocity users with speed exceeding 150 km/h, where additional pilots are inserted.

[7] Pilot overhead in brackets correspond to chunks of low velocity users with speed below 10 km/h.

measurements required for RRM and for MAC/PHY control, and their delay requirements, have been summarized in section 5 of [WIN2D233] and in section 7.12 below.

WINNER supports two major transmission schemes that use either using frequency-adaptive or frequency-non-adaptive transmission. Although they partly have different control signaling requirements, it has proved possible to design one control channel with different parametrizations, that controls both schemes with acceptable downlink overhead.

Table 4-2 below gives a summary of the most important urgent control messages that needs to be transmitted, their information message sizes and their message frequency/urgency. Two physical control channels, PDCFC and PDFCC have been defined for supporting this frame control signalling.

Table 4-2: Control information and assumptions on required information bits and update rate

Information	Link Direction	Information bits per message	Message frequency	Comment
HARQ information	DL, UL	3 + 3	per HARQ channel use	3 bit HARQ-ID, 2 bit for redundancy version, 1 bit new data indicator (assuming asynchronous *N*-Channel Stop-And-Wait protocol supporting Incremental Redundancy)
Transport block size, Code Rate	DL, UL	5	every transport block	
Modulation Information	DL, UL	2	every transport block for non-frequency-adaptive tx, every chunk layer for frequency-adaptive tx.	
Spatial Processing Scheme	DL, UL	5	Per MAC PDU	
Cell specific user address	DL, UL	12	per user or per flow	(Implicitly signalled)
HARQ feedback	DL, UL	1	per HARQ channel use	
CQI feedback	UL		per user[8]	
CSI feedback	UL		per user	
Chunk allocation table	DL	flexible length	per frame	contains DL allocation and UL allocation of chunk layers to users, design principles discussed in detail in Annex A.1

It should be noted that a semi-static allocation of transmission resources can be used for some flow classes and users. This means that a fixed set of transmission resources (chunks/blocks and link adaptation parameters) is pre-allocated in multiple frames, which reduces the control overhead significantly as illustrated in Section 4.7 of [WIN2D61311]. This is in particular possible for voice-over-IP flows that are destined to slow-moving users.

4.8.6 Description of physical channels and the mapping between the transport channels and the physical channels

Physical Broadcast Channel (PBCH) is a downlink physical channel for broadcasting system information, etc., to all terminals inside the cell coverage area.

[8] Note, that update rates in frequency and time are not considered in this table, as they can be adjusted according to coherence properties of the channel as explained in [WIN1D24].

Transmission technology: B-EFDMA allocation with smallest block size (4 subcarriers by 3 OFDM symbols), to maximize frequency diversity. Cell-wide spatial transmission that uses (Alamouti) space-frequency coding to achieve spatial diversity. 4-QAM with rate ¼ convolutional coding and 6 x repetition coding is used, with target minimum downlink SINR 8 dB.

Position in Superframe: 3-OFDM symbol part of a downlink slot in frame *j* of every *m*:th super-frame. It uses only frequency resources that are available within a wide area (multiple cells). The PBCH transmissions from different RAPs can potentially use a frequency reuse pattern to minimize interference between PBCH of neighboring BS and RNs.

Physical Downlink Control Format Indicator Channel (PDCFC) is used for transmitting parameters indicates the location and layout of the PDFCC channel.

As outlined in Annex A.1, the downlink control information that signals the frame layout for downlinks as well as uplinks, for frequency-adaptive as well as non-frequency-adaptive transmission, is composed of several parts, to minimize the control overhead.
The PDCFC message contains the following parts:

- A broadcast configuration table, **CT**,
- An optional broadcast control message length indicator, **LI**,

The PDCFC message is broadcast to all involved terminals.

Transmission technology: The transmissions use B-EFDMA non-frequency-adaptive transmission with small (4x3) block sizes and spatial diversity where possible. 4-QAM with rate 1/4 convolutional coding with is used combined with 6 x repetition coding, to reach a target minumum downlink SINR of 8 dB.

Position in superframe: Uses the set of downlink resources within frame that are allocated to non-frequency-adaptive transmission. It is transmitted in the earliest 3 OFDM symbols of each downlink slot.

Physical Downlink Frame Control Channel (PDFCC) is a physical downlink channel that is used to signal information in a slot of a frame related to downlink data transmission within that slot and uplink data transmission in the subsequent slot. The PDFCC is used to define the resource allocation both for frequency-adaptive and non-frequency-adaptive transmission.

On the PDFCC allocation tables **AT** and transport format tables, **TFT** are transmitted. The following information is conveyed in this channel:

- DL scheduling control:
    - User terminal/relay node ID,
    - Resource allocation for PNDC: identifies what chunks/blocks have been allocated to the user terminal/relay node for DL,
    - Modulation and coding scheme,
    - HARQ related information, e.g. HARQ process ID, redundancy version, and ACK/NACK related to UL transmission.
- UL scheduling grant:
    - User terminal/relay node ID,
    - Resource allocation for PNDC: identifies what chunks have been allocated to the user terminal/relay node for UL (scheduling grant).
    - Modulation and coding scheme
    - (slow) uplink power control command.

All types of tables need not be used in a particular parametrization. Different table layouts are used for specifying frequency-adaptive and non-frequency-adaptive transmission. In the case of frequency-adaptive transmission, different table layouts are used in the case of few participating users and many users. For details, see Annex A.1.

Transmission technology: Uses B-EFDMA non-frequency-adaptive transmission with small (4x3) block sizes and spatial diversity where possible. Convolutional coding is used combined with repetition coding. The allocation tables and transport format tables are partitioned into sub-tables destined for different groups of users with different SINRs. Each such sub-table is encoded with an appropriate code rate, to limit the donwlink control overhead.

Position in Superframe: Uses the set of downlink resources within frame that are allocated to non-frequency-adaptive transmission. Tables for control of the downlink are transmitted in the earliest 3 OFDM symbols of each downlink slot. Tables for control of the subsequent uplink are transmitted in the following 3 OFDM symbols (no 4-6) of the slot.

Physical Frequency-adaptive Data Channel (PADC) is a physical channel for frequency-adaptive transmission for point-to-point user-data transfer in downlinks and uplinks. A PDFCC is associated to each PADC, to control the transmission.

Transmission technology: Modulation (BPSK, 4-QAM, 16-QAM, 64-QAM, 256-QAM) adjusted individually per chunk layer. An average code rate is computed for all chunk layers used, and then used for the whole HARQ block. Convolutional codes are used for FEC block sizes < 200 payload bits. Quasi-cyclic block LDPC codes are used for larger FEC block sizes.

Position within Superframe: PADC used a set of chunks within the slot that is pre-allocated for frequency-adaptive transmission by the resource division function.

Physical Non-frequency-adaptive Data Channel (PNDC) is a physical channel for point-to-point user-data transfer in downlinks and uplinks. A PDFCC is associated to each PNDC.

Transmission technology: One modulation (BPSK, 4-QAM, 16-QAM, 46-QAM, 256-QAM) and code rate is chosen for the whole HARQ block. Convolutional codes used for FEC block sizes < 200 payload bits. Quasi-cyclic block LDPC codes are used for larger FEC block sizes.

Position within Superframe: PNDC uses a set of chunks within the slot that is pre-allocated for non-frequency-adaptive transmission by the resource division function. The chunks are regularly spaced in frequency. Within that set, blocks of size ≤ one chunk are allocated to transmissions. One code block uses a set of blocks that are all of the same size and are regularly spaced in frequency within one single slot. In downlinks, the scheme is denoted B-EFDMA. In uplinks, DFT-precoding is used (B-IFDMA). Please see Section 7.3.3 for a further discussion.

Physical Multicast Broadcast Channel (PMBC) is used to carry MBMS services. Uses B-EFDMA non-frequency-adaptive transmission with a modulation and code rate adjusted to the user with worst SINR in the multicast group. The PMBC transmission can support cooperative relaying.

Transmission technology: As for PNDC downlinks, but the modulation and code rate is adjusted to the user with worst SINR within the multicast group.

Position within Superframe: As for PNDC downlinks. Uses the set of downlink resources within frame that are allocated to non-frequency-adaptive transmission.

Physical Uplink Control Channel (PUCH) is used for urgent uplink control messages. The PUCH resources are pre-allocated per UT and hence no UT ID needs to be conveyed in this message.

The PUCH channel contains

- HARQ ACK/NACK, triggered by DL data transmission
- CQI messages

- Scheduling requests. There are two cases for how this information can be transmitted physically:
  - If PADC or PDNC uplink resources are assigned, then scheduling requests may be multiplexed into the PADC or PDNC, respectively.
  - If no PADC or PDNC resources are assigned, then this information is transmitted on the PUCH.

Transmission technology: The transmission is protected by convolutional coding in combination with repetition coding, adjusted to the average uplink SINR of the UT. It does not use a HARQ process.

Position within the Superframe: Each active UT is given a small allocation in the uplink slot of each frame for the PUCH. The allocation is slowly time-varying with the SINR and the traffic load. B-IFDMA allocations are used. The allocation of PUCH resources is performed at initial access and is modified on a slow time-scale when the traffic load of the UT varies significantly. Since the PUCH is limited to only few bits, detailed information needs to be carried on other physical channels, e.g. PNDC.

Physical Random Access Channel (PRAC) is a contention-based uplink physical channel needed to acquire time alignment. The amount of resources used for PRAC is for further study, but it is important to schedule the PRAC resources in neighboring cells such that the PRAC channel does not cause severe interference to especially the PNDC and PADC channels.

Transmission technology: An uplink PRAC transmission uses 4-QAM in one of the two allowed OFDM symbols within the allocated uplink time-slot.

*Position within the S*uperframe Uses a time slot that comprises 3 OFDM symbol, that comprises 2 transmission OFDM symbol durations + synchronization error guardtimes in both directions. This slot is positioned in the uplink slot of the *j*:th frame of the superframe. Uses only frequency resources that are available within a wide area (same as PBCH).

Synchronization pilots. The downlink slot of the first frame in each superframe contains downlink network syncronization pilots. They utilize three OFDM symbols in the downlink slot.

The uplink slot of the first frame of each superframe contains uplink network syncronization pilots. They utilize the three OFDM symbols of the uplink slot.

These pilots utilize only spectral bands that are available over a wide geographical area, to facilitate multi-cell coordination. In the allocated OFDM symbols, they utilise the whole of these bands. Payload transmission may proceed simultaneously in other bands.[9]

Figure 4-15 shows the mapping between transport channels and some of the physical channels. The remaining physical channels that are described above are used for Layer 1/Layer 2 control signaling and are not shown in Figure 4-15.

The TBCH, TMCH and TRAC transport channels are directly mapped onto the corresponding physical channels.

The Transport Shared Channel TSCH may for its physical transmission use either frequency-adaptive transmission (PADC), non-frequency-adaptive transmission (PNDC) or, for urgent uplink control and CQI measurements, the Physical Uplink Control Channel (PUCH).

The paging transport channel (PTCH) uses non-frequency-adaptive downlink transmission at physical locations that are known to all UTs. It uses low rate coding and modulation that is appropriate for broadcast control messages. No HARQ process is involved.

[9] To reduce interference to such transmissions, the network synchronization pilots should use a transmit filter to suppress their interference within other sub-bands.

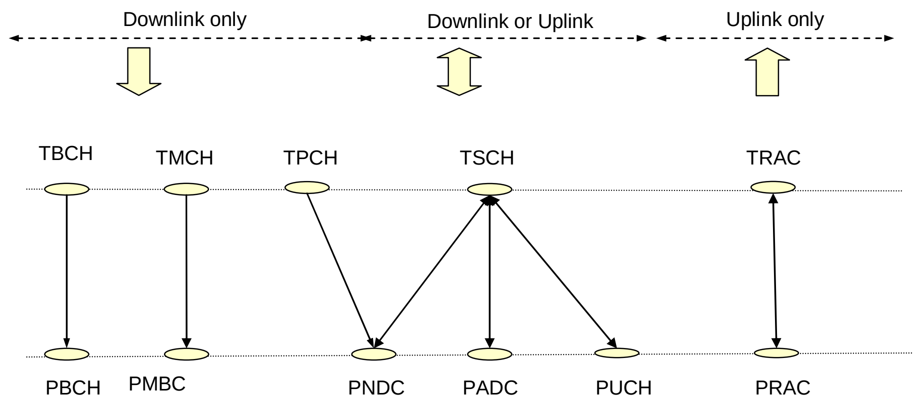

Figure 4-15: Mapping of Transport channels to/from physical channels. (Physical control channels not shown.)

5 Cross layer interactions

5.1 MAC & RLC Interactions

5.1.1 MAC & RLC Segmentation, and reassembly, concatenation and multiplexing

This section will focus on segmentation (and reassembly), concatenation and multiplexing of higher layer packets. Furthermore, we illustrate how these functions are mapped onto the protocol layers and the interplay between these functions/protocol layers to enable a failsafe and resource efficient (i.e. low overhead) cross layer design. An overview of this process is given in Figure 5-1.

In WINNER phase 1, it was proposed that the RLC should just add a header including a sequence number to the RLC PDUs for the purpose of supporting RLC-ARQ duplicate detection and in sequence delivery. As a result the MAC had to perform several actions. First an optional segmentation of MAC SDUs (i.e. RLC PDUs) was performed (requiring a new sequence number) and subsequently a CRC was added, resulting in a HARQ retransmission unit (RTU). Secondly, a further optional segmentation of RTUs into encoding blocks was performed (requiring another new sequence number) and the resulting encoding block was coded into a FEC block (forward-error-correction coded block). All these actions were performed prior to any scheduling decision and hence the resulting FEC blocks did not necessarily match the assigned chunks by the scheduler. This required puncturing of FEC blocks to a level where success in the first transmission attempt becomes highly unlikely.

In WINNER phase 2, it has been decided that MAC and RLC should be terminated in the same node. The common termination node allows further cross layer optimization by adapting the size of the RLC PDUs to the transport block size. Hence, based on the scheduler decision a certain amount of data is selected from the RLC SDU buffer and segmented and/or concatenated depending on the size of the SDUs and a sequence number is appended. Further, we propose that it should be possible to concatenate RLC SDUs from different flows but from the same flow class (see next section for a definition of flow class). Hence, one transport block contains one RLC PDU per flow class i.e. one RLC entity provides only one RLC PDU per frame. However, there are two exceptions: i) if a retransmission PDU and ii) if a received PDU at a RN does not fit entirely into the new transport block then the RLC PDU is re-segmented. In these two cases a new sequence number is appended to the re-segmented packets by means of an extension header.

The number of re-segmentations is not limited. In WINNER phase 1 it was shown that it is beneficial to keep the same RLC sequence number space on the entire path between the base station and the user terminal (see [WIN1D35] for further details) to enable fast retransmissions after a user terminal has moved from one node to another node within the relay enhanced cell (REC) (i.e. intra-REC mobility support by RLC-ARQ). Hence, RLC SDUs are not re-assembled in the relay node.

To minimize the amount of signalling required, MAC SDUs (RLC PDUs) from several flows destined for the same UT may be multiplexed at the MAC layer (multiplexing information has to be included in the MAC header). For the same reason, in case of relaying, MAC SDUs (RLC PDUs) from several users destined for the same RN may be multiplexed in MAC.

In the physical layer the MAC PDU corresponds to a transport block (TB) that is transmitted within one frame. A CRC is added to the transport block - note that this is the only CRC that will be added to the packet and hence it has to support both RLC-ARQ and HARQ operations. (further details will be given in section 5.1.3.). Further segmentation may be performed at the PHY layer prior to FEC encoding to support different spatial schemes (e.g. PARC) and limited FEC block sizes.

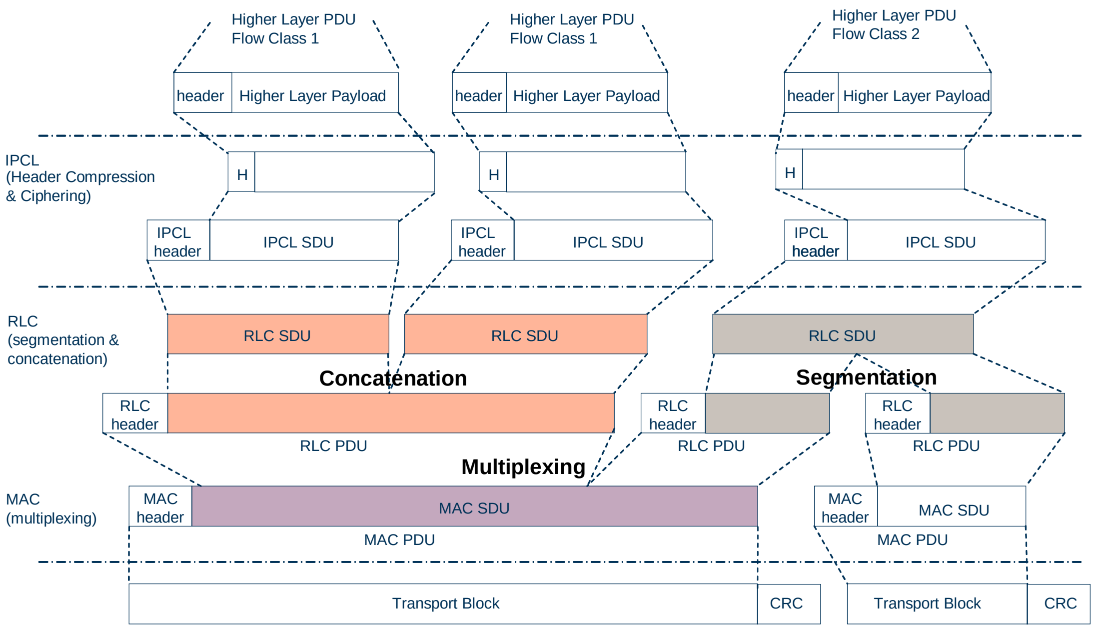

Figure 5-1: Segmentation, concatenation, and multiplexing process from higher layer PDUs to transport blocks.

It is furthermore proposed that the RLC retransmission unit is an RLC PDU or an RLC PDU segment while the HARQ retransmission unit is the MAC PDU.
Finally, the required information that needs to be included in the headers is summarized below:

- The RLC header includes: a sequence number (note that at most one new RLC PDU is transmitted per flow class and frame - therefore, the RLC sequence number space can be small and does not depend on the data rate) as well as required segmentation, re-segmentation (by means of extension header) and concatenation information to enable RLC SDU reassembly. N.B. the RLC header does not include any CRC.
- The MAC header includes: multiplexing information used to de-multiplex RLC PDUs to different flow classes (in case of relaying, multiplexing information to distinguish different users are also necessary (by means of extension header)) and flow class identifiers. N.B. the MAC header does not include sequence numbers since this is provided by RLC.

5.1.2 Proposed RLC/MAC header format

This section will focus on RLC/MAC header formats (i.e. in-band control signalling), however to understand the full context we will briefly discuss also out-band signalling (signalled at the beginning of every frame) relating to information that is required by the receiver to decode and combine the received packets. A high-level goal should be to minimize the amount of out-band signalling and should preferably be restricted to information that is beneficial for the receiver to know before decoding the received packet. The out-band signalling is discussed in Section 4.8.5.

The proposed RLC/MAC header formats will support dynamic flow classes. Thereby a flow class should not be mistaken for a service class. A flow class is identified by its attributes (guaranteed bit rates versus non-guaranteed bit rates, maximum bit rates, delay budget, loss tolerance etc). These attributes are configurable and reconfigurable as new flows are set up or multiplexed onto an existing flow class and are governed by filters in the BS and UT that map flows onto flow classes and its attributes. The motivation for multiplexing several flows onto the same flow class is to lower overhead, both in terms of headers as well as status reporting for both RLC and MAC. The only drawback would be that one flow may stall another flow mapped onto the same flow class. Some flow classes may be preallocated and initiated already at network attachment (e.g. for control signalling and best effort traffic).

- o N.B.1 the scheduler may still operate on individual flows.
- o N.B.2 the QoS concept has not been defined yet in WINNER and hence, what is described here may change.

Table 5-1depicts the RLC/MAC header with all the required fields. N.B. that in most cases only a subset of these fields is included in a PDU. Furthermore, the header is split into multiple sub-headers according to their functionality.

	MAC MUX Header				RLC PDU Identification Header				RLC Resegmentation Header		RLC SDU Reassembly Header					Payload
Short name	FC	LF	EX	…	T	SN	P	RF	SO	LS	SF	EF	EB	LI	…	
Length (bit)			1		1		1	1		1	1	1	1			
Long name	Flow Class ID	Length Field	Extension Flag	Further FC, LF, EX	Type Field	Sequence Number	Poll Bit	Resegmentation Flag	Resegmentation offset	Last Segment Flag	Start Flag	End Flag	Extension Bit	Length Indicator	Further E, LI	

Table 5-1: Generic RLC/MAC header with all required fields for communication with user terminals.

In the following the sub-headers are described in brief.

5.1.2.1 MAC MUX Header

The MAC Multiplexing header defines the number of contained RLC PDUs, the flow class to which a contained PDU belongs and the length of their payload in byte. The header fields are added per RLC-PDU (i.e. once per flow class and frame).

- **Flow class identifier (FC ID)**: The number of bits allocated to this field determines the number of parallel flow classes/pipes that may exist to a given user terminal and consequently the number of parallel RLC/MAC processes.
- N.B.1 some IDs need to be preallocated to RRC signalling.
- N.B.2 some IDs need to be preallocated to RLC/MAC control signalling.
- N.B.3 in order to minimize the amount of signalling both for the transport network as well as over the air one may predefine a (large) number of possible attribute combinations and associate that with a tag that then is the only thing that needs to be signalled.
- **Length field (LF)**: Length of the corresponding RLC PDU's payload in byte. With this concept no padding is introduced on RLC level which increases the protocol efficiency. Note that the separate length field for the entire RLC PDU replaces one length indicator in the RLC SDU Reassembly header. This Length Field must be large enough to cope also with very large RLC PDUs whereas the length indicator can be dimensioned according to the size of typically IP packets. The separate length field is required anyway in case of re-segmentation.
- **Extension bit (EX):** This 1 bit flag indicates if another length field follows, i.e., if MAC multiplexing is used.

5.1.2.2 RLC PDU Identification Header

The size of this RLC header part is constant and one such header exists for each RLC PDU or segment thereof.

- **Type flag (T):** This 1 bit flag defines if this RLC PDU is a status message or a data PDU.
- **RLC sequence number (SN)**: It is assigned to each RLC-PDU. Note that at most one new RLC PDU is transmitted per flow class and frame. Therefore, the RLC sequence number space can be small and does not depend on the data rate.
- **Poll bit**: If this 1 bit flag is set the RLC receiver is supposed to return a status message. It could be omitted for Unacknowledged Mode flow classes.
- **Re-segmentation flag (RF)**: This 1 bit flag indicates if this is a complete RLC-PDU or just a segment thereof.

5.1.2.3 RLC Re-segmentation Header

The RLC re-segmentation header appears only if the re-segmentation flag is set. It identifies the position of this RLC-PDU-segment's payload in the original RLC-PDU's payload.

- **Segment offset (SO):** This number identifies the first byte of the original RLC PDU contained in this segment. Note that it supports multiple re-segmentations of an RLC PDU or segment thereof while avoiding ambiguity between initial and consecutive re-segmentations.
- **Last segment bit (LS):** This 1 bit flag indicates that this segment contains the last byte of the re-segmented RLC-PDU.

5.1.2.4 RLC SDU Reassembly Header

The RLC SDU Reassembly Header contains information required to reassemble the contained RLC SDU(s). Its size depends on the number of contained SDUs (or parts thereof). In case of re-segmentation this header is only required in the first segment.

- **Start flag (SF)**: This 1 bit flag indicates if first contained RLC-SDU starts in this RLC-PDU, i.e., if it does not overlap to the previous RLC-PDU.
- **End of SDU flag (EF)**: This 1 bit flag indicates if the last contained RLC-SDU ends in this RLC-PDU.
- **Extension flag (E)**: This 1 bit flag indicates if another SDU is contained and if a length indicator follows.
- **Length indicator (LI)**: Specifies the length of the corresponding RLC SDU and thereby identifies the position of the first byte of the following RLC-SDU within the RLC-PDU. Note that no Length Indicator is required if the RLC PDU contains only one RLC SDU or a segment thereof. In other words, the number of Length Indicators is equal to the number of contained RLC SDUs (or segments thereof) minus one (N-1).

							RLC PDU Header 1									RLC PDU Header 2								
	MAC MUX Header						RLC PDU Identification Header				RLC SDU Reassembly Header					RLC PDU Identification Header				RLC SDU Reassembly Header			Payload, PDU1	Payload, PDU 2
Short name	FC	LF	EX	FC	LF	EX	T	SN	P	RF	SF	EF	EB	LI	EB	T	SN	P	RF	SF	EF	EB		
Length (bit)			1			1	1		1	1	1	1	1		1	1		1	1	1	1	1		
Long name	Flow Class ID	Length Field	Extension Flag	Flow Class	Length Field	Extension Flag	Type Field	Sequence Number	Poll Bit	Resegmentation Flag	Start Flag	End Flag	Extension Bit	Length Indicator	Extension Bit	Type Field	Sequence Number	Poll Bit	Resegmentation Flag	Start Flag	End Flag	Extension Bit		

Table 5-2: Example of an RLC/MAC PDU header with MAC multiplexing. The first RLC PDU contains parts of two SDUs. The second RLC PDU contains one complete RLC SDU.

5.1.2.5 MAC UT MUX Header

Table 5-3, Table 5-4, and Table 5-5 outline the extensions needed to support multiplexing of traffic associated with multiple user terminals on the base station to relay node link.

The MAC UT MUX header fields are added per UT and required to multiplex data of different UTs on the relay link between the BS and the RN.

- User terminals ID (UI):
- Extension indicator (EI): This 1 bit flag indicates if another UT address follows, i.e., if MAC UT multiplexing is used.

The rest of the sub-headers are the same as described above. Note, the relay node will remove the MAC UT MUX header and signal the UT address out-band for downlink traffic and will append the MAC UT MUX header for uplink traffic.

	MAC UT MUX Header			MAC RLC MUX Header				RLC PDU Identification Header				RLC Re-segmentation Header		RLC SDU Reassembly Header					Payload
Short name	UI	EI	…	FC	LF	EX	…	T	SN	P	RF	SO	LS	SF	EF	EB	LI	…	
Length (bit)		1				1		1		1	1		1	1	1	1			
Long name	UT ID	Extension Indicator	Further EI, LD	Flow Class ID	Length Field	Extension Flag	Further FC, LF, EX	Type Field	Sequence Number	Poll Bit	Resegmentation Flag	Resegmentation offset	Last Segment Flag	Start Flag	End Flag	Extension Bit	Length Indicator	Further E, LI	

Table 5-3: Generic RLC/MAC header with all required fields for communication between BS and RN.

					UT 1										UT 2											
	MAC UT MUX Header				MAC RLC MUX Header			RLC PDU Identification Header				RLC SDU Reassembly Header			MAC RLC MUX Header			RLC PDU Identification Header				RLC SDU Reassembly Header			Pay-load UT1	Pay-load UT2
Short name	UI	EI	UA	EI	FC	LF	EX	T	SN	P	RF	SF	EF	EB	FC	LF	EX	T	SN	P	RF	SF	EF	EB		
Length (bit)		1		1			1	1		1	1	1	1	1			1	1		1	1	1	1	1		
Long name	UT ID	Extension Indicator	UT Address	Extension Indicator	Flow Class ID	Length Field	Extension Flag	Type Field	Sequence Number	Poll Bit	Resegmentation Flag	Start Flag	End Flag	Extension Bit	Flow Class ID	Length Field	Extension Flag	Type Field	Sequence Number	Poll Bit	Resegmentation Flag	Start Flag	End Flag	Extension Bit		

Table 5-4: Example of an RLC/MAC PDU Header where no multiplexing/segmentation/concatenation is used.

					UT 1										UT 2											
	MAC UT MUX Header				MAC RLC MUX Header			RLC PDU Identification Header				RLC SDU Reassembly Header			MAC RLC MUX Header			RLC PDU Identification Header				RLC SDU Reassembly Header			Pay-load UT1	Pay-load UT2
Short name	UI	EI	UA	EI	FC	LF	EX	T	SN	P	RF	SF	EF	EB	FC	LF	EX	T	SN	P	RF	SF	EF	EB		
Length (bit)		1		1			1	1		1	1	1	1	1			1	1		1	1	1	1	1		
Long name	UT ID	Extension Indicator	UT Address	Extension Indicator	Flow Class ID	Length Field	Extension Flag	Type Field	Sequence Number	Poll Bit	Resegmentation Flag	Start Flag	End Flag	Extension Bit	Flow Class ID	Length Field	Extension Flag	Type Field	Sequence Number	Poll Bit	Resegmentation Flag	Start Flag	End Flag	Extension Bit		

Table 5-5: An example of an RLC/MAC PDU Header where two UTs are multiplexed but no further multiplexing/segmentation/concatenation is used.

5.1.3 RLC-ARQ & H-ARQ interactions

Information transfer over a radio interface is always exposed to uncertainties and thus error correction techniques are essential. Reliable information transfer is especially important when carrying data transmissions by the transmission control protocol (TCP), which may mistakenly interpret radio interface errors as congestion and therefore unnecessarily reduce the transmission rate.

It may be argued that a single hybrid ARQ (HARQ) retransmission protocol dimensioned for a single-hop is sufficient to achieve high reliability. However, since the WINNER system is designed for low end-to-end round trip time, it is desirable to have a very frequent ARQ feedback. Preferably the feedback messages should be sent every frame and with minimum size in order not to impose high peak power requirements on the UT. Moreover, in order to not incur too much overhead it is here proposed that the feedback messages are transmitted on pre-allocated resources (or piggybacked to data traffic in the opposite direction if possible).[10] However, even with this mechanism it is still costly to achieve a sufficient reliability [3GPP04], [3GPP07a], [3GPP07b], [3GPP07c], [3GPP08] , [3GPPR2-052057].

One solution for this problem is to include two (layered) retransmission protocols in WINNER (henceforth, these retransmission protocol will be termed **Outer-ARQ** (terminated in RLC) and **Inner-ARQ** (terminated in MAC, and assumed to be a HARQ process)). A second motivation for the layered approach is that it will ease handovers within RECs - the motivation for this may be found in the next section.

Although both the Inner-ARQ and the Outer-ARQ are ARQ protocols, their tasks are quite different. The Inner-ARQ shall provide means for correcting most errors quickly and support a radio-resource efficient transmission by incorporating soft-combining techniques (Hybrid-ARQ (HARQ)). In contrast to that, the Outer-ARQ has to be much more reliable, i.e. status reports with explicit sequence numbers are used, which are protected by a CRC[11]. It should handle all residual Inner-ARQ errors but does not need to be as fast as the Inner-ARQ, since outer-ARQ retransmissions are triggered more seldom and thus only have little impact on the overall delay and performance. There are three possible causes for a residual Inner-ARQ error:

- NACK →ACK error: causes data loss at the Inner-ARQ layer. (Inner-ARQ ACK → NACK error causes “only” an unnecessary retransmission.)
- DTX → ACK error
- Maximum number of allowed Inner-ARQ retransmissions is reached.

Furthermore, the common termination node allows a tight integration of Outer- and Inner ARQ by exchanging triggers between the layers in order to speed up the ARQ recovery. For instance, if the Inner-ARQ transmitter detects a failed delivery of a MAC PDU due to e.g. maximum retransmission limit it may notify the relevant transmitting ARQ by sending an RLC status report entities and potential retransmissions can be initiated. Moreover, if the HARQ receiver is able to detect a TB transmission failure (e.g. NACK to ACK error) it may notify the relevant transmitting ARQ entities and retransmissions can be initiated. As illustrated in section 5.1.1, the CRC is appended only to the Inner-ARQ retransmission units and hence, the Inner-ARQ has to provide error detection also for the outer-ARQ. In addition the terminating Outer-ARQ entities can poll for an RLC status report for example in cases where the receiving ARQ entity cannot detect NACK to ACK errors because it is the last packet in the sequence.

Since the uplink and downlink differ significantly in terms of control signalling we treat both transmission directions separately. Throughout the remainder of this section the following assumptions are made:

[10] The one bit feedback format adopted in HSDPA, E-DCH and LTE is a power and resource efficient method to achieve fast feedback.

[11] The reliability of the feedback information is ensured in two ways. First, the status messages are also transmitted by using the Inner-ARQ mechanisms at the MAC layer. But even if the transmission fails the reliable transmission of the status information is supported by the accumulative nature of the status messages, since each status message represents the full status of the receiver side at any given reporting time instance. All subsequent status messages following a not correctly received status report include the information of the lost status message.

- Synchronous Inner-ARQ feedback
- L1/L2 control signalling (assignment) sent in conjunction with data includes the Redundancy Version (RV), i.e. if an assignment is received it is known whether new data is sent.

Downlink transmissions

If the Inner-ARQ receiver has not successfully decoded a transport block after one or more Inner-ARQ transmissions it responds with a NACK. If a feedback error for this NACK occurs, i.e., the NACK is interpreted as ACK at the Inner-ARQ sender, there occurs a residual Inner-ARQ error that needs to be treated by the Outer-ARQ.

One example for an NACK➔ ACK error case is shown in Figure 5-2. If the NACK ➔ ACK error occurs, the sender has no means to detect that this error occurred. It stops transmitting that transport block and might assign the chunks to a new transmission for the same (as in this case) or another user. A dedicated outer NACK message is needed from the receiver to inform the sender that the error has occurred.

At the Inner-ARQ layer, the NACK ➔ ACK error can be detected either by the expiry of a timer at the Inner-ARQ receiver that expires when the maximum time until it should have received a retransmission has elapsed or by receiving a new assignment and transmission (like in this example). Then the receiver flushes the previous Inner-ARQ process and responds with an explicit outer NACK message informing the BS that the previous transmission for a particular process had failed. For this a resource request and corresponding grant might be needed.

The first possible Inner-ARQ-Outer-ARQ interaction for this error case consists of sending an Inner-ARQ control PDU that points to the failed transmission by means of a timing reference or the frame number. Upon reception of this reference the Inner-ARQ transmitter identifies the affected RLC PDU and triggers the corresponding Outer-ARQ transmitter. The Outer-ARQ transmitter may send a retransmission of the PDU or a segment thereof.

An alternative approach would be that the Inner-ARQ receiver triggers all its Outer-ARQ receivers, which then send status reports to their transmitting peers in the BS. Note that all active Outer-ARQ instances must send a status report, since the Inner-ARQ receiver does not know which Outer-ARQ instances were affected by the failed Inner-ARQ transmission. However, this approach involves more overhead compared to sending an Inner-ARQ control PDU and mapping that information to the RLC PDUs at the BS. Moreover, in this case unnecessary retransmissions might occur, since the Outer-ARQ status reports are likely to contain incomplete or imprecise information. In order to recover from those losses the BS must retransmit all not yet acknowledged RLC PDUs if it expected them to have arrived already (according to Inner-ARQ feedback) which may result in unnecessary retransmissions. This overhead can be reduced to some extend at the cost of extra recovery delay if the Outer NACK is delayed so that some pending Inner-ARQ processes can finish.

If no close Inner-ARQ-Outer-ARQ interaction is used, then Outer-ARQ would have to detect the Inner-ARQ failure on its own, i.e., by a missing Outer-ARQ sequence number. If RLC PDUs were multiplexed into the transport block that suffered from the NACK➔ ACK error, it is required that an RLC PDU for each affected Outer-ARQ instance is received to detect the error.

Finally, it is possible to speed up the recovery if Outer-ARQ polling is used. But since polling can potentially involve a huge amount of status messages, this option should be used with care and only in cases where this mechanism is required, e.g., for signalling traffic.

Uplink transmissions

If the NACK ➔ ACK error occurs, the UT has no means to detect that this error occurred. It stops transmitting that transport block. The UT might receive further NACKs from the Inner-ARQ receiver, but it does not react to them, since it can not be sure that it is the intended recipient. The resources might have been assigned by the BS to another user as it got an Inner-ARQ ACK. However, the BS is expecting retransmissions.

As was the case for the downlink, feedback may be provided by an Inner-ARQ control PDU, triggering Outer-ARQ or Outer-ARQ would have to detect the Inner-ARQ failure on its own, i.e. by a missing sequence number.

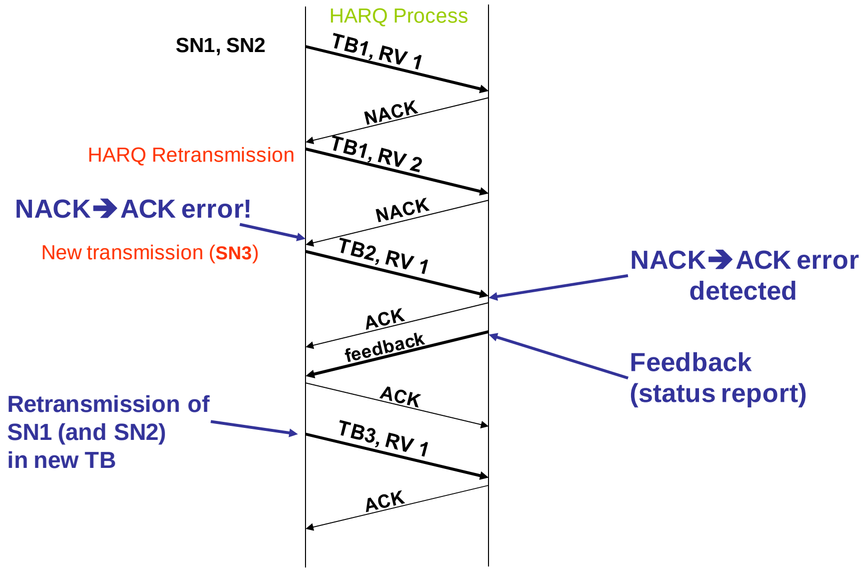

Figure 5-2: Response of the layered ARQ to "NACK to ACK" error.

5.1.4 Extensions for multi-hop support

Transmission in wireless multi-hop (MH) networks should result in the same reliability as in single-hop wireless networks to guarantee a required residual packet error rate. In order to achieve the required residual packet error rate over multiple hops with a single retransmission protocol on every hop, the residual packet error rate on each hop has to be even lower than in the single hop case. This would result in higher overhead than in the single hop case and it further emphasizes the need for two layers of retransmission protocols, Moreover, by incorporating an outer-ARQ scheme (that terminates in the BS and UT) resource efficient (and lossless) intra-REC handovers may be supported as retransmissions may be initiated from the BS after handovers.

Two proposals were presented in [WIN2D352] with and without RLC-ARQ at the RN. The current view within WINNER is to utilize RLC-ARQ at the RN. The benefit of this approach is that Relay-ARQ may be used for local error recovery over multiple hops which lead to faster recovery (errors on the second link will only lead to retransmissions on the second link). Moreover, the BS may poll the RN over 1 hop to release packets and during handover the BS will only need to poll the RN for UT status (if needed the RN may poll the UT for its status). In addition, these messages implicitely indicates the respective buffer status for each node along the path and may hence be used for flow control (e.g. if there is a large discrepancy between the number of acknowledged packets in the RN and the UT, it suggests that the transmission rate should be decreased).

Furthermore, there are two main proposals on how to include RLC-ARQ in the RN:

- Case 1: In this case 2 bits per entry are used to indicate the status for both the RN and the BS/UT status. That is, outer-ARQ is enhanced with Relay-ARQ in the RN (i.e. the RN may respond to the transmitter with either NACK (not received by peer), RACK (received by peer but not by the final receiver), or ACK (received by final receiver) - for details on Relay-ARQ the reader is referred to [WIN2D351]. This case is illustrated in Figure 5-3.
- Case 2: in this case the RN is polled for RN or UT status separately,using 1 bit per entry. In the status report one bit in the RLC-ARQ message indicatse whether the RLC-ARQ feedback is sent on behalf of the UT or the RN.

Depending on the frequency of polling, either case 1 or case 2 is preferred. In case of frequent polling the overhead is smaller for case 1, and vice versa. There is no decision in WINNER which frequency of polling that will be used and hence there is no decision on the preferred case.

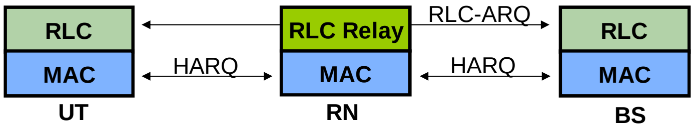

Figure 5-3: Outer-ARQ with relay part of the Outer-ARQ process.

5.1.5 Support of cooperative relaying

The proposed solution is based on the assumption that the link between the BS and the serving RN (first hop) is of very good quality and hence that retransmissions on the first hop happen quite rarely. If this is not the case, other solutions should be considered. An example of such a scheme is given at the end of this section.

In the WINNER system concept, cooperative relaying is an add-on to single-path relaying. One of the cooperating RAPs is the serving RAP, which is also responsible for the retransmissions. Further, in the two hop case, the BS schedules all the cooperative transmissions. If the BS waits for all the first hop acknowledgements and performs first hop retransmissions before scheduling the cooperative transmissions, the delays and the signalling overhead would be increased. Therefore we propose an ARQ protocol, where the BS schedules cooperative transmissions, assuming that the first hop transmissions were successful. The RN only needs to send NACKs for cooperative transmissions where it is the serving RAP. To further decrease the signalling load, the RN sends NACKs only for first hop transmissions for which they do not receive ACKs from the UT. Figure 5-4 illustrates the resulting flow chart for the Inner-ARQ procedure. Optionally, the BS can decide if it schedules a cooperative retransmission to the UT.

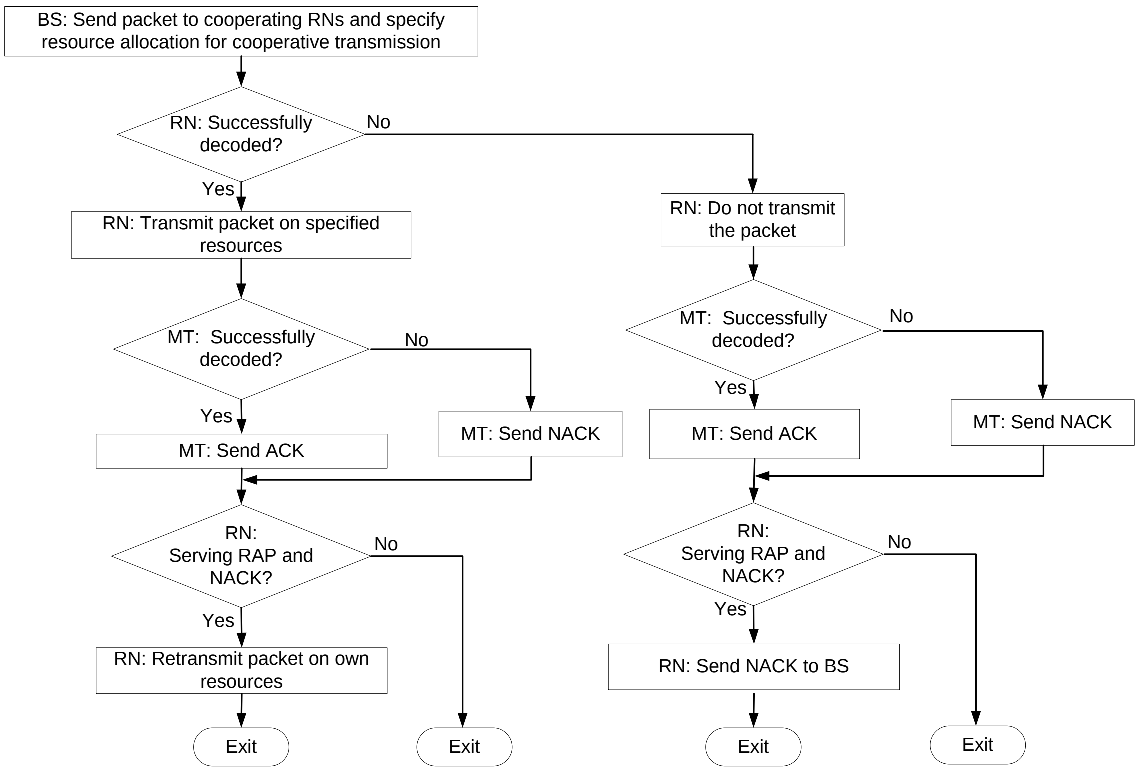

Figure 5-4: Inner-ARQ procedure for cooperative relaying.

The Outer-ARQ works in a similar way. If the RN is the serving RAP of the cooperatively served UT and receives an RLC status report, then it can initiate the required retransmissions if it has decoded the packets successfully. If not, it sends an RLC status report to the BS. Again, the BS can decide if it schedules a cooperative retransmission or if it retransmits the packet on the first hop and the RN performs the retransmission to the UT.

Another solution that could be employed when there is no guaranteed high quality link between the BS and any RNs, and hence giving a high rate of first hop retransmissions is described below.

In this case all retransmissions could be performed cooperatively by all cooperating nodes that receive a NACK, e.g. utilizing synchronous retransmissions. The benefits would be that the NACK to ACK errors would be somewhat alleviated and that cooperative transmissions could also be used for the retransmissions. However, the drawbacks would be that unnecessary retransmissions and collisions could happen in case of ACK to NACK errors.

Another solution would be for the BS to retransmit to the RNs in the first hop immediately if the BS receives a NAK from each cooperating RN in that hop, provided that the RNs send the NAKs immediately in response to the first hop transmission. Such a new cooperative retransmission can then take place already before or at the same time as when the RN was intended to transmit to the UT and would lower the total RTT between the BS and the UT. If the serving RN response is an ACK, this RN forwards the data to the MS, but if the serving RNs response is a NAK while another RNs response is an ACK, the BS sends a message to that RN instructing it to perform a restransmission on reception of a NAK from the UT.

5.2 MAC-PHY interactions

The multi-layer scheduler controls all aspects of the physical layer transmission and reception and the interaction between these layers is extremely tight. The outcome of this interaction is the scheduling decisions, resulting in the transmission and reception sequences for each segment. The transmission control sequences for downlink and uplink are outlined below. We then discuss the assumed timing of these sequences and the resulting minimal delays over the air interface.

The WINNER concept uses scheduled transmission in both uplink and downlink. Contention-based transmission modes are only supported for the RACH.

5.2.1 Transmission control sequences for downlinks

For downlinks the scheduler at the transmitter side performs the scheduling and controls the transmission (Section 4.7.2). Downlink transmission control for frequency-adaptive and non-frequency-adaptive transmission has been outlined in Sections 4.1 and 4.2 of [WIN2D461]. The transmission sequence is as follows:

1. The scheduler decides which flow classes are to share the downlink slot to be scheduled. Moreover, for frequency-adaptive transmission, channel quality predictions are obtained in the BS for relevant transmission resources and terminals. If necessary, the resulting predictions are reported to the transmitter over the reverse link.
2. The scheduling of all transmissions within the slot is finalized.
3. The downlink slot is transmitted, with associated out-band downlink frame control information (PDCFC and PDFCC) transmitted in the same slot.
4. All receivers receive and buffer the first three OFDM symbols of the slot where the PDCFC and PDFCC are located.
5. Each receiver then detects and decodes the PDCFC and relevant parts of the PDFCC.
6. The payload is then detected and decoded by the receiver.[12]
7. When all HARQ segments that comprise a RTU (a transport block) have been received, an ACK/NACK is generated based on the CRC code of the RTU. It is transmitted over the PUCH in the next available uplink slot.

5.2.2 Transmission control sequences for uplinks

Uplink transmission is controlled by the scheduler that is located at the RAP, so it has to be preceded by a transmission request. Initial transmission requests are in general transmitted over the PUCH. This uplink channel uses a dedicated transmission resource for each UT with active flows and it therefore generates a significant overhead.

To keep this overhead acceptable, only very small request messages can be transmitted within this resource. This request message only contains information that the UT has data to transmit, but not how much. This transmission request evokes a grant to send an initial transport block. The size of this block is a scheduling decision, and may be based on e.g. flow class, available resources, CQI/CSI, etc. For some flow classes, this is all that is needed to transmit all queued data. For others, additional transport blocks will need to be sent in subsequent frames.

The complete sequence is described as follows:

1. An initial request (one or two bits) is transmitted over the PUCH.
2. The scheduler may then grant the terminal to send an initial transport block within the next frame.

[12] HARQ segments do not span multiple frames, so decoding can be attempted directly after the slot has been received. Iterative channel estimation and decoding (turbo processing) can be used.

3. If granted, a transmission grant message is sent over the DL, using the PDCFC and PDFFC. It specifies the modulation, channel coding and utilised transmission resources for this initial transport block.
4. The UT sends the initial transport block over the UL.
5. If more detailed information is needed to send the rest of the queueud data, a detailed uplink transmission request is sent, piggybacked onto the initial transport block, specifying the queued data size per flow class.
6. The scheduler then sends a response to the detailed request over the downlink. The response specifies the transmission resources, modulation,coding scheme, etc for the next uplink slot.

The transmission then commences as outlined in Sections 3.1 and 3.2 of [WIN2D461].

5.2.3 Transmission and retransmission delays

The timing of the transmission control loops has as a target to attain a very short delay over the air interface. In the transmission control systems, a scheduling computation delay of max. 0.1 ms has been assumed. This is the delay from the arrival of the last information needed for scheduling a complete slot, until the scheduling is finalized. The computation delay for channel quality or state prediction is likewise assumed to be max. 0.1 ms, counted from the latest pilot symbol on which the prediction update is based. Both channel prediction and scheduling can then be finalized within one slot.

Regarding the delay of ACK/NACK for (link) retransmission, the delay of decoding is added. The results of Table B-2 of Appendix B in [WIN1D210] show that a delay of below one clock cycle per decoded bit is attainable, with appropriate parallel implementations of LDPC and DBTC decoders. For example, for an assumed FEC block sizes of max. 1200 bits, this corresponds to less than 6 μs when 200 MHz of the total clock cycles are allocated to decoding. We below allow the total *receiver processing delay to use up to one slot (345.6 μs).* Decoding of FEC blocks of size up to 1520 bytes should require less than 60 μs, so this provides ample time also for iterative turbo decoding /channel estimation.

In the TDD Physical layer mode, frames comprise a DL slot followed by an UL slot. In the FDD mode, half-duplex terminals are assigned to one of two groups. Group 1 transmits in the downlink in the first slot of the frame and in the uplink in the latter slot. Group 2 transmits/receives in the opposite way (Section 4.7.1).

Consider a transmission (either TDD or FDD) where the DL slot preceds the UL slot within the frame. Let "UL *i*" and "DL *i*" denote the uplink/downlink slots of frame number *i*, respectively. The resulting minimal delays over the air interface are then summarized by Table 5-6.
It is here assumed that

- uplink transmissions request resources on a slot by slot basis and do not make multi-slot reservations.
- The transmission requests refer to the initial requests, which determine the initial delay. If transmission of additional transport blocks are required, this adds to the delay.
- Furthermore, the whole RTU is assumed to be transmitted within one slot, so a decision for retransmission can be taken immediately after its decoding.
- A PUCH reserved resource, which contains reserved bits for ACK/NACK, is present in each frame.[13]

For example, in frequency-adaptive downlinks, the packet is assumed to arrive at the end of the downlink slot of frame *j-1*. Late during this slot and in the beginning of the uplink slot of frame *j-1*, the channel predictions are updated. In FDD systems, this is done by predictors in the UTs, and is based on downlink pilots up to the latter part of the downlink slot of frame *j-1*. CQI feedback is transmitted in the beginning of the uplink part of frame *j-1* and scheduling is then performend during the later part of this slot. The transmission then commences in the downlink slot of frame *j*. The downlink frame control tables (PDCFC and PDFCC) are transmitted in the beginning of the same slot. After the whole slot has been received, the decoding then takes place during the uplink slot of frame *j*. For a packet arriving at the BS at end of the downlink slof of frame *j-1*, the transmission delay over one hop, incuding decoding, is 3 slots or 1.5

[13] For a large number of UTs with active flows within the cell, this may create a large reverse link overhead. If PUCH slots are placed in every m:th frame to reduce the overhead, the retransmission delays are increased accordingly.

frames or 1 ms. This corresponds to the minimum delay over the air interface that has been targeted for the WINNER system. (Packets arriving up to one frame earlier cannot be transmitted earlier and they therefore experience correspondingly larger delays, up to 2.5 frames.)

Table 5-6: Transmission plus decoding delays.

	Initial Transmit Requests	CQI prediction Update (U) Scheduling (S)	DL frame control	Trans-mission	Decoding	1-hop delay incl. decoding (frames)	1-hop delay (ms)
Frequency-adaptive **uplink**	UL *j-2*	U,S:UL *j-1,*	DL *j*	UL *j*	DL *j+1*	3.0	2.1
Non-frequency-adaptive **uplink**	UL *j-2*	S: frame *j-1*	DL *j*	UL *j*	DL *j+1*	3.0	2.1
Frequency-adaptive **downlink**		U,S:UL *j-1*	DL *j*	DL *j*	UL *j*	1.5	1.0
Non-frequency-adaptive **downlink**		S: UL *j-1*	DL *j*	DL *j*	UL *j*	1.5	1.0

The attainable delays involved in a retransmission are summarized by Table 5-7 below. Note that uplink retransmissions are assumed not to require separate (re)transmission requests. They use sets of transmission resources and modulation and coding schemes that are assumed to be known beforehand by both transmitter and receiver.

Table 5-7: Retransmission delays

	RTU received	ACK/NACK transmission	Retransmission performed	Retransmission delay (frames)	Retransmission delay (ms)
Downlink	DL *j*	UL *j+1*	DL *j+2*	2.0	1.4
Uplink	UL *j*	DL *j+2*	UL *j+2*	2.0	1.4

6 RRM mechanisms

6.1 Resource partitioning process and time scales

There are several functions in different layers of WINNER system architecture which have significant impact on the overall spectral efficiency and resource utilization. The operation time scale and sequence of events in the spectrum assignment and resource partitioning process in WINNER is as follows (see Figure 6-1):

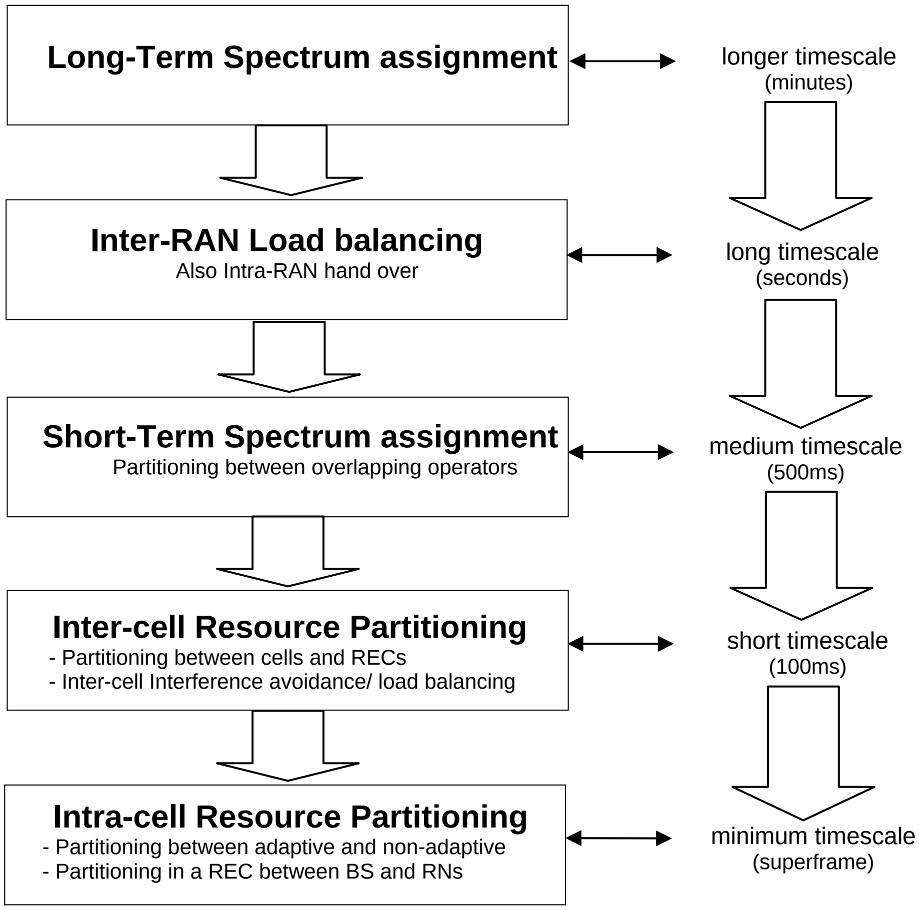

Figure 6-1: Spectrum assignment and Resource Partitioning time scales

On the highest level in the chain of events, comes Long-Term (LT) spectrum assignment. It is responsible for allocating the spectrum on long-term basis to overlapping network operators. The network's LT spectrum is not expected to be changed very often. It is foreseen that the operation time scale of LT spectrum assignment is in order of several minutes.

The next step is Inter-RAN load balancing which takes place to balance the load among overlapping WINNER networks with their existing available spectrum. This function operates in order of several seconds. Short-Term (ST) spectrum assignment re-adjusts the available spectral resources between overlapping WINNER operators from a shared resource pool on a temporary basis with faster time scale. As a result of resource negotiation between operators, a portion of shared WINNER spectrum will be dedicated to each one for duration of about five hundred milliseconds.

Inter-cell resource partitioning together with inter-cell load balancing tries to maximize the spectral efficiency without changing the available spectrum. The reason is to make sure we have made the most of existing spectrum before going for additional one. Inter-cell partitioning time scale is in order of 100 millisecond to allow the interference chain effect in neighbouring cells to settle down before triggering ST spectrum assignment.

Finally, resource partitioning in a Relay Enhanced Cell (REC) which is referred to as intra-cell partitioning, has to be carried out every super frame, more details are provided in AnnexB.1. This is the minimum time scale in resource partitioning and the lowest in the chain of events. Figure below illustrates the common view on time scale and sequence of events in the spectrum assignment and resource partitioning process in WINNER.

6.2 Spectrum management

6.2.1 Spectrum functions in a nutshell

A conceptual overview on the spectrum control functions is presented in Figure 6-2 [BKOL+07]. The most salient spectrum control functions are summarised and it illustrates the main interactions between the spectrum control functions and related RRM functions in WINNER.

Table 6-1 illustrates the main interactions between the spectrum control functions and related RRM functions in WINNER.

Table 6-1: Spectrum functionality classification.

Spectrum Sharing	**Flexible Spectrum Use**
Spectrum sharing functions	Spectrum assignment functions
This group includes: • Vertical Sharing 1 (VS1) • Vertical Sharing 2 (VS2) • Horizontal Sharing with Coordination • Horizontal Sharing without coordination	This group includes: • Long Term Assignment (LT) • Short Term Assignment (ST)

Under the Spectrum Sharing umbrella, four different schemes have been developed. They are based on the access rights of each system to the shared spectrum. If one system has a priority access to the spectrum over the other, then vertical sharing schemes are applied. In such a case, if the WINNER system is the primary system then this results in Vertical Sharing 1 scheme (VS1). If the WINNER system is a secondary system and has higher priority access rights compared to any other secondary systems, then Vertical Sharing 2 scheme (VS2) is defined. If both systems, i.e. WINNER and other, have same access rights to the spectrum then this results in horizontal sharing schemes. The horizontal sharing situation maps into two different sharing schemes. The first horizontal scheme assumes that the systems contending for spectrum can coordinate with each other to enable efficient spectrum allocation. This is called Horizontal Sharing with Coordination (HwC). The second scheme is considered when both systems do not coordinate with each other in the framework of spectrum sharing functionalities. This scheme is called Horizontal Sharing without Coordination (HwoC). The coordinated horizontal sharing introduces QoS agreement to the users compared to the uncoordinated horizontal sharing case.

Once the spectrum is allocated to the WINNER system, its allocation within the WINNER RANs is investigated in the spectrum assignment scheme. WINNER considers a Long Term (LT) and Short Term (ST) spectrum assignment strategy to take advantage of the changing nature of the spectrum availability and the traffic demand in different parts of a multi-operator environment. The LT scheme assigns the spectrum at a higher level of geographical granularity between multiple RANs. During the LT assignment functional procedure the spectrum is negotiated over a longer time scale, i.e., in the order of tens of minutes, over the WINNER deployment of operation. The ST assignment acquires the fine tuning of the spectrum assignment at the cell level. This is performed at shorter time scales than in LT assignment, i.e., the ST assignment negotiation of spectrum is performed over time periods as frames of several 100 ms in duration.

,

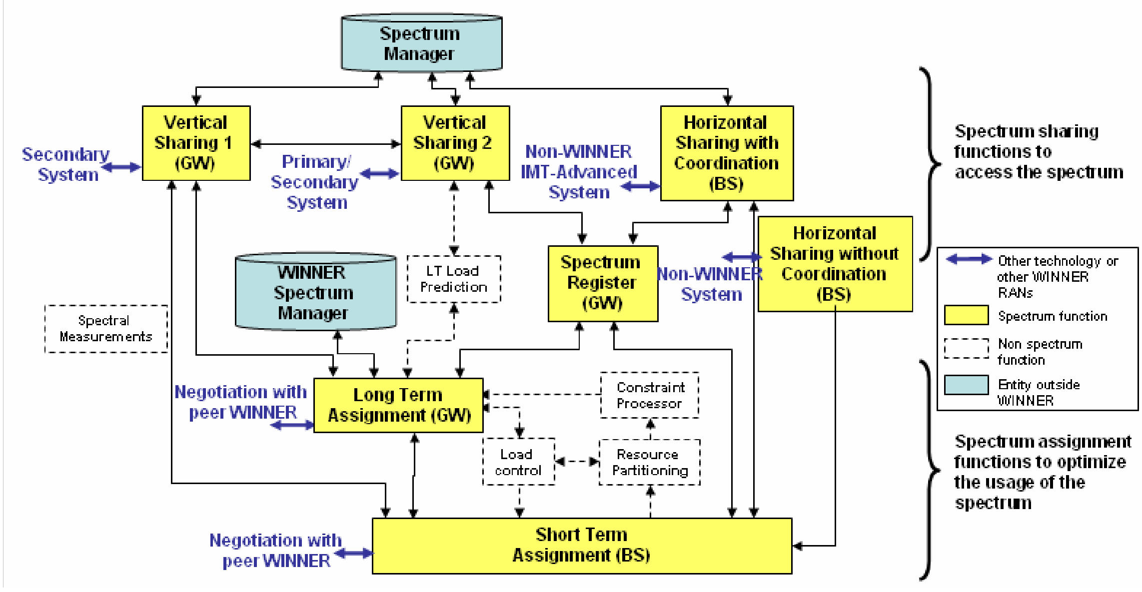

Figure 6-2: Illustration of the interactivity of the spectrum sharing and spectrum assignment function in the WINNER concept.

Due to the large difference between the wide area (WA), metropolitan area (MA) and local area (LA) deployment, the functional grouping of Figure 6-2 does not indicate any preferred mapping of functions to the logical entities in WINNER RAN. Functions belonging to the same group may be located to different logical nodes.

In general, spectrum sharing functions consist of concurrently triggered procedures. A normal spectrum control procedure (e.g. actualization of a given resource set, measurement collection, etc.) is the result of several individual action calls. Thus, in practice there is not a clear boundary among the spectrum control functions, and it is essential to consider the definition in details. Detailed signalling interactions between the various Sharing and Co-existence Functions and spectrum Assignment functions have been defined in [WIN2D591].

6.2.2 Sharing and Co-existence Functions

6.2.2.1 Vertical Sharing 1: WINNER is the primary system

This function is considered when WINNER is the primary system. WINNER may assist a secondary system by sharing its primary spectrum resources. This is feasible by signalling its unused spectrum resources via its broadcast channel (BCH), by means of a universal or customised broadcast radio beacon.

Depending on the expected incentives for the WINNER RAN, unused spectrum could be actively created. The responsibility for creating unused-spectrum belongs to the overall resource optimization which involves the LT and ST Assignment functions. Since the spectrum resources leased to the secondary systems are operator-wise, each operator can use part of its prioritised or its assigned common pool resources.

6.2.2.2 Vertical Sharing 2: WINNER is the secondary system

When the WINNER RAN is the secondary system, it has to control its emissions from both the BS and all the user terminals (UT) in order to avoid interfering with the primary system. More generally, the secondary system may adopt a dynamic, opportunistic use of the unused part of the spectrum. For that purpose, considerable knowledge about the deployed primary system may be required.

Besides the information acceded through the Spectrum Manager, spectrum related measurements done at physical layer (BSs and UTs) might be compiled. These measurements are facilitated when the primary system emits a standard beacon periodically. The compiled information, i.e. measurements, site information, location, protection requirements of the primary system downloaded from databases, etc is transformed into a set of transmission constraints defining and characterizing the shared spectrum, e.g. exclusion zones in the context of sharing with fixed satellite systems (FSS).

In non-loaded conditions, and depending on incentives for the WINNER RAN, the WINNER system may lease its identified unused spectrum to other secondary systems since WINNER may have better capabilities in accessing the spectrum compared to other secondary or sub-secondary system. In this case of spectrum "re-sharing", the regulatory perspective might require extra mechanisms.

6.2.2.3 Horizontal sharing with coordination

The involved systems (i.e. WINNER and other non-WINNER RAT) have equal access rights to the spectrum, and coordinate their spectrum access based on a set of predefined rules (spectrum sharing rules) that all the involved systems are submitted to. Each system adapts its transmission to mitigate interference to others by applying constraints issued from common policies shared by all the candidate systems or determined on the basis of the previous coordination phase. Location services and measurements of other systems radio activity might be useful for a better coverage estimation of the other system resulting in a better coordination in terms of actual mutual interference. These measurements are facilitated by using special purpose mutual beacon signals; therefore access to the BCH at MAC layer is expected for this function.

Similarly to VS2 function, the obtained shared spectrum indicators are conveyed to LT Assignment function via the Spectrum Register. In the case that the Spectrum Register entity is rather centralized and slow, for the more dynamic opportunistic LA approach, direct communication from the Horizontal Sharing function to the ST function might be implemented.

6.2.2.4 Horizontal sharing without coordination

The HwoC scheme would consider the RLAN licensed exempt band around 5GHz. It is known that the current unlicensed bands are limited from interference protection point of view. They are also likely to be congested with regard to the number of services allocated to them. Therefore, there is a need for the spectrum sharing functions to bring discovery mechanisms in order to reach for an opportunistic use of the spectrum. Nevertheless, efficient detection of white spaces, i.e., unused spectrum, is not easy task due to problems of the discovery of hidden nodes or energy-based detection mechanisms. Since QoS cannot be guaranteed for any system, it is important to understand whether WINNER can accept such QoS in LA only.

6.2.3 Spectrum Assignment

6.2.3.1 Long Term Assignment

The LT Assignment coordinates partly the usage of the spectrum between WINNER RANs. This function coordinates and negotiates the spectrum assignments between multiple WINNER RANs for large geographical areas with spatial granularity of cluster of cells. LT assignment entity is located at the GW. The spectrum assignments are updated periodically and at a slow rate, that is, in time frame of several tens of minutes. It can be used also to provide dynamic spectrum assignments between WINNER RANs. This entity is also responsible for the coordination between the WINNER TDD and FDD modes. Inter-RAN coordination is achieved through the direct negotiations between peer LT assignment functions in different WINNER RANs.

The LT Assignment could be extended to support the coordination of (static) spectrum assignments over the country borders. This requires signalling between the WINNER Spectrum Managers in the neighbouring countries, providing sufficient information to establish signalling between the related LT Assignments over the IP network, and over the country border. However, another natural option is that the over-the-border coordination is included fully to WINNER Spectrum Managers.

6.2.3.2 Short Term Assignment

This function controls the short-term and local, i.e. cell-specific, variations of the large-scale spectrum assignments. Hence, it enables faster adaptation to the local traffic load variations and geographically more accurate spectrum assignments than the LT Assignment. The assignments are performed in the time scale of several MAC super-frames, i.e. 200ms to several minutes. Due to the above cell wise functionalities the location of the ST assignment is at the BS.

The ST Assignment requests spectrum resources from other WINNER RANs after being triggered by the LT Assignment or by preventive load control. The fundamental reason for developing ST assignment is that in the case of two RANs providing the exact same services, traditional handover between RANs to support the load condition could be envisaged, but operators may be reluctant for such a handover of users , i.e., the operators would prefer to keep the UTs connected to their own RAN. Further in the case that increasing the spectrum of one RAN would result in enhancing the service capabilities, e.g. in terms of capacity provided to the UTs, spectrum assignment becomes a necessity for creating new services with better user QoS capabilities.

6.2.4 Description of the General spectrum functions needed for spectrum assignment

In this section, a description of the general spectrum functions needed for the spectrum assignment is presented. The spectrum functions belonging to the spectrum sharing group are presented in detailed in [WIN2D592] so they will not be repeated here.

6.2.4.1 WINNER Spectrum Manager

It manages the usage of the spectrum within the WINNER RANs only. It is a policy rules maker so that peer to peer negotiation between WINNER RANs follows the same rules. It contains relevant information on spectrum priorities as well as on fairness and/or cost metric and, thus, establishes a common control point on the spectrum assignment. It enables also simple introduction of new RANs by specifying the rules and limits. Also it may maintain logs of the parameters necessary for recovery, etc. It may be for instance a server accessible over the IP network maintained by a trusted party. It is located outside the WINNER RANs.

6.2.4.2 Spectrum Register

Each GW of a RAN has its own Spectrum Register. It conveys information on exclusion zones/available spectrum from spectrum sharing functions to the spectrum assignment functions within a WINNER RAN. The information in the register is dynamically updated. The introduction of the register is motivated by the sharing of the spectrum with FSS, i.e. related exclusion zones may be large. The purpose of the register is to provide an alternative for a centralized control entity while allowing for self-organizing deployment of LA cells. The Spectrum Register could be made accessible with IP (allowing access either through RANs own core network, or even through IP network in the case of small networks). It may also increase delays in the information flow between spectrum sharing and spectrum assignment functions. For this reason, the Spectrum Register logical entity is by-passed in the case of spectrum sharing schemes that require fast dynamical behaviour.

6.3 Cooperation mechanisms

The tasks that RRM must usually perform in cellular systems are admission control, channel assignment, power control and handover control. These tasks have to be performed on basis of an estimation of the current system state. Inputs to such estimation are system dependent and may include for example the Received Signal Strength (RSS), the Signal-to-Interference and Noise Ratio (SINR), Distance to base stations, Transmit Power or mobile velocity.

In heterogeneous environments radio resource management does not only target the resource utilization of a single radio access technology but introduces cooperation between technologies and operation modes. WINNER system supports the following cooperation scenarios:

- Inter-system cooperation
  - WINNER interworks with legacy system
- Inter-operator cooperation
  - Roaming and open coupling
  - Definition on SLA (Service Level Agreement)
- Inter-deployment cooperation between TDD and FDD
  - A pool of GWs inter-connect with a number of BSs with different deployments
  - Hierarchical cell structures scenarios and direct interworking between the wide area BS and Local area BS
- Inter-function cooperation
  - Inter-functions cooperation requires prioritization and triggering without invoking multiple functions being executed at the same time
- Inter-entity cooperation
  - Besides the interfaces introduced for important entities pairs as GW – BS, BS - relay node, BS – BS, WINNER supports, service server – WINNER RAN is of importance to provide the guaranteed service

To enable the cooperation scenarios, necessary functions are needed to be embedded into the WINNER system. To have a systematic view, the relevant functions selected from this document are depicted with their proper allocation into the logical nodes in the Figure 6-3.

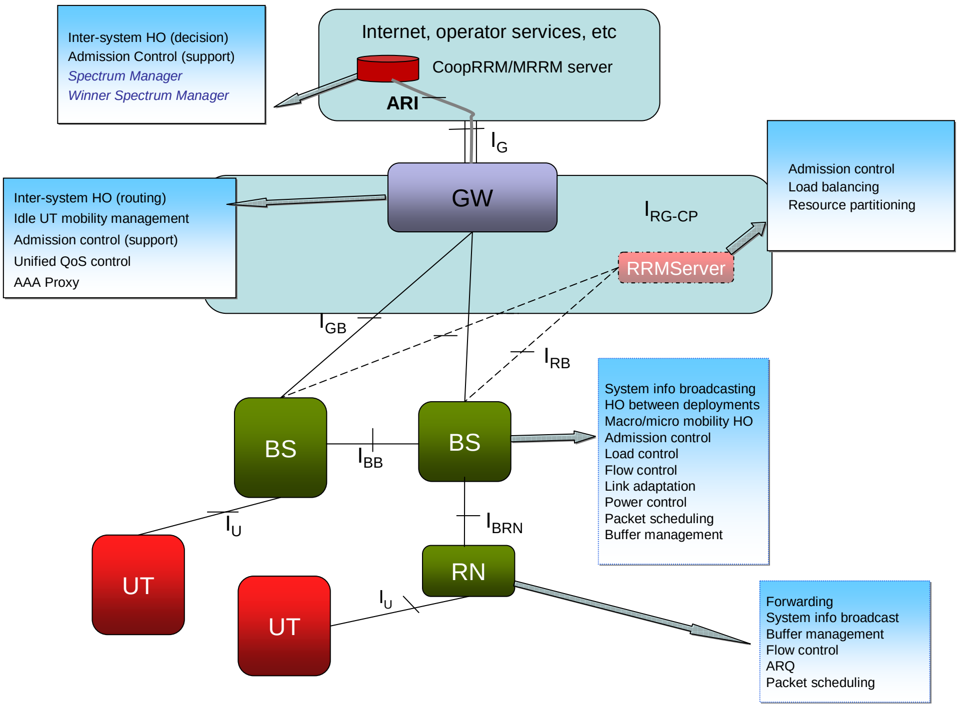

Figure 6-3: Cooperation Architecture.

In Figure 6-3, entities as CoopRRM Server/MRRM Server (Multi Radio Resource Management Server), Gateway (GW), Base Station (BS), Relay Node (RN), User Terminal (UT) and RRM Server are listed, where the CoopRRM server is a logical entity dealing with inter-system cooperation. All the interfaces between the corresponding entities are also depicted, as have been defined in Chapter 3.

6.3.1 Inter WINNER-Legacy System Interworking

The I_G interface defines the interworking between WINNER and legacy system. To be compliant with the Ambient Networks Project ([AN2006]) the existing AN interface ARI is defined to be a sub-signalling interface of the I_G interface. The GW control plane has to support the necessary request, measurement report, confirmation through the I_G interface.

6.3.2 Cooperative relaying

Cooperative relaying has been identified as a suitable technology to improve the user throughput both in the WA and the MA CG scenario. The next common RN or BS in the upper layers of the tree structure coordinates the cooperative transmission of multiple RAPs, i.e. the BS in a two hop relay network. Thus, no direct communication between two RNs is required for RN-RN cooperation.

In the following we will describe the cooperative relaying scheme for a two hop relay network. Please note that the role of the BS in the following description can be replaced by the next common RN in the upper layers of the tree structure. In order to coordinate cooperative transmissions the BS performs resource allocation for cooperatively served UTs in the same way as for UTs served by the BS. The BS sends together with the data, the resource allocation for the cooperatively served UTs in the next frame to the RN. An efficient signalling scheme has been developed for the WINNER system to signal the resource allocation to the UTs. The same number of bits will be required as additional overhead to signal the resource allocation for cooperatively served UTs to the cooperating RN(s). However, as the link quality of the BS-RN link will be much higher than the average link quality to the UTs located in the cell, it will require much less resources. Therefore the signalling load will not be a problem. One of the cooperating RAPs is the main serving RAP of the UT. It signals the allocation table to the UT and is also responsible for HARQ retransmissions.

One of the big challenges in the design of an OFDMA system with frequency-adaptive scheduling is to keep the signalling load low, while still capturing most of the potential scheduling gain. In the case of cooperative relaying this is even more crucial because if a RN is the main serving RAP, all the feedback signalling has to be forwarded by the RN to the BS. Thus, one option would be to restrict cooperatively served UTs with a RN as main serving RAP to frequency non-adaptive flows. In this case only limited feedback is needed, e.g., one channel quality indicator (CQI) value for the whole bandwidth and the signalling load could be kept low. However, this would limit the performance of cooperative relaying schemes that utilize frequency-adaptive scheduling and much of the performance might be lost. Therefore, the WINNER cooperative relaying scheme also support frequency-adaptive scheduling for cooperatively served UT. However, if a RN is the main serving RAP, the feedback of the UT has to be forwarded to the BS and the resource allocation has to be sent back to the RN before the resource allocation is broadcasted to the RN. The resulting additional delay will restrict the applicability of frequency-adaptive scheduling and it will probably only be beneficial for lower speed UTs compared to the non-cooperative case.

MIMO cooperative relaying as add-on

Two-hop relaying protocols are divided into two phases: in the first phase the BS transmits data to the RN and in the second phase the RN forwards this data. To gain on large-scale spatial diversity, most cooperative relaying protocols further benefit from a combination of the first phase and second phase transmissions. In a multiple-antenna based system this implies that dedicated MIMO algorithms cannot be applied (for instance beamforming and other SDMA algorithms), since *one* stream is only optimized to *one* destination. Furthermore, as the position of a RN can be well chosen to obtain good channel conditions, the data rate on the relay link is likely to exceed the data rate on the RN-UT links. This inhibits cooperation on physical layer because both links use different modulations as well as information combining and the UT would be unable to decode the BS transmission. One solution might be to decrease the data rate on the relay link and to forbear from using MIMO algorithms, which could lead to significant performance degradation. Thus, we propose to a cooperative relaying scheme that does not rely on the combination of first and second hop transmissions.

Secondly, the cooperative relaying scheme should be compatible with the spatial temporal processing schemes utilized in the WINNER system. Hence we use the same spatial temporal processing schemes, such as SMMSE multi-user precoding. The BS simply treats the RN antennas as additional antennas. In the case of multi-user precoding techniques the RN has to forward also the channel state information to the BS. The additional overhead will restrict the usefulness of e.g. the SMMSE multi-user precoding scheme to lower speed UT and very good BS-RN links.

6.3.3 Inter-entity cooperation

Optional RRM server

The RRM server has potential gains provided by the centralised joint radio resource control through the interfaces between the RRM server and the BSs. The system capacity gain obtained from the deployment of RRM server is in principle the enlargement of the number of operational servers from the queuing model viewpoint, which therefore results in a higher trunking gain. On the other hand, by alternatively allocating the resource to call units among the interworking coexisting base stations, the load balancing effect among the radio networks is realized. In an interference limited network, such effect is very significant.

The more servers available, the less the call blocking/dropping rate will be obtained, meaning higher trunking gain thanks to the admission control over multiple available cells and modes. In addition, for HCS (Hierarchy Cellular Structure) situation with overlapping e.g. WA BS and MA BS, RRM server may allow traffic splitting over the involved BSs. It may result in an even less network response time due to integration of radio resources [Luo05]. This gain can be termed as multiplexing gain.

However, deployment of RRM server will potentially increase the complexity and CAPEX. WINNER architecture design supports an optional RRM server in order to support scalable network solutions.

Strategically, the necessary functions are not allocated in the optional entity such as the RRM server. If operator is able to deploy the RRM server, the potential performance gains given by RRM server can be obtained. This solution is applicable for hotspots and high end UT concentrating service zones. Another option is to switch off RRM server, when the direct interworking between BS or coordinated by GW still allow a running network.

Inter Service Server – WRAN interworking

Service server is expected to interact with the RAN through the GW functions. The control plane GW deals with functions as policy enhancement, packet detailed inspections and charging control. The user plane GW deals with IP layer convergence including the functions as header compression and ciphering mechanism. The WINNER GW communicates with service server to provide radio context dependent appropriate service according to user QoS demand.

QoS is reflected from the SLA, which is an official negotiated agreement between a number of parties, e.g., mobile users, network operators. The SLA can also be inter-operators. It is contracts the level of QoS being provided by one provider to its customer. It records the common understanding about services, priorities, responsibilities, guarantee, etc. with the main purpose to agree on the level of service. For example, it may specify the levels of availability, serviceability, performance, operation or other attributes of the service like billing and even penalties in the case of violation of the SLA.

SLA mainly refers to the contractual agreement between subscriber and the operator. Based on the previous contracted agreements between these two parties, the operator should provide optimised QoS to the end user.

To support the proposed service optimization concepts, participating entities need to communicate and negotiate QoS parameters for their communication. This section puts focus on appropriate signalling within the IP Multimedia Subsystem (IMS). The IMS utilizes the PS domain to transport multimedia signalling and bearer traffic [3GPP04]. The IMS enables operators to offer their subscribers multimedia services based on Internet applications, services and protocols. Transport services for multimedia signalling and user traffic are preformed by the PS domain in the WRAN. In terms of the ISO/OSI reference model, IMS is responsible for providing session layer functionality.

6.3.4 Inter-Function Cooperation

All functions cooperate with each other to provide an optimized WINNER solution. For instance, functions such as Admission Control and Load Control are listed under the category of congestion avoidance control. To better explain the inter-relationships, Figure 6-4 and Figure 6-5 show the interworking among some of the RRM-related WINNER functions. Since congestion control itself is not an explicit function which evokes explicit WINNER air interface updates, it is classified to the network function and listed on the right hand side. The numbered functions are the priorities of execution.

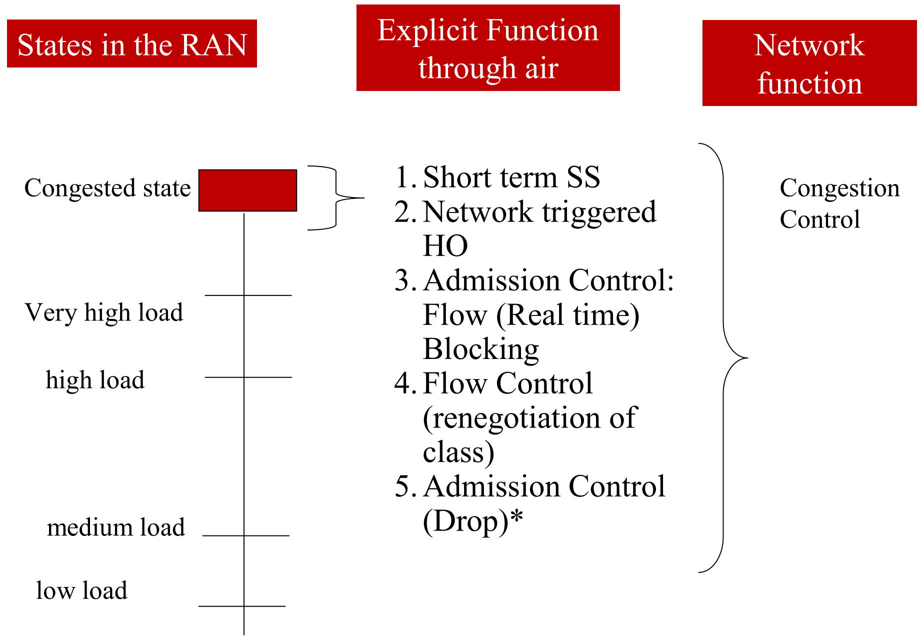

Figure 6-4: Interworking of selected inter-RRM function at congested State.

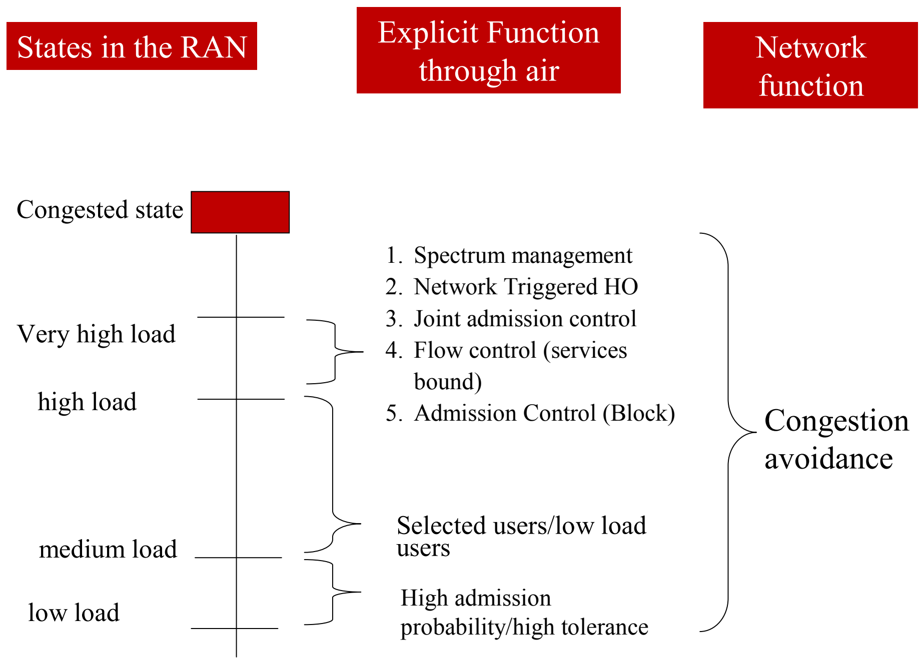

Figure 6-5: Interworking of selected inter-RRM function at high and low congested State.

The term inter-mode handover and intra-mode handover are basically two types of intra-Winner system handover, where the modes are the WINNER deployment modes including wide area, metropolitan area and local area.

In the high load and medium load situation, the congestion avoidance function may trigger (network triggered) a handover by shifting some selected users to another cell/mode or frequency in order to avoid the congested state.
Under the scope of RRM, the functions need explicit RRC messages that are termed as 'Explicit function through the air' in Figure 6-4 and Figure 6-5. The rest are rather 'implicit' functions. The sequences of functions when a congested state or very high load situation occurs are numbered. The choice of execution priorities for some of the involving functions still need further study and therefore is marked by '*'. Figure 6-4 and Figure 6-5 show the function interworking congested state and states before congestion respectively.

6.4 Mobility management

The focus of section on mobility management has been on the impact of the new flat architecture for the handover process, especially on distributed, centralised and hybrid approaches, and also on the integration of the macro mobility (L3) with micro mobility.(L2) Macro mobility is triggered when the UT changes of serving GW.

6.4.1 Micro mobility

Micro mobility is usually separated into two different mobility schemes, namely active mode mobility and idle mode mobility [3GPPTS25331]. This section will deal with these two schemes in more detail.

The purpose of active mode mobility (henceforth, handover will be used synonymously to active mode mobility) is to maintain connectivity and to support QoS in a radio resource efficient manner for mobile terminals. A handover from one RAP to another RAP (i.e. the UT will alter its point of attachment to the RAN from one RAP to another RAP) may be triggered for a number of reasons. The most essential trigger for handover is the signal strength that may vary due to mobility or changes in the radio environment, however, also the load situation within a cell, the interference experienced by the UT, as well as the location of the UT may trigger a handover. In a 2-hop (or multi-hop) scenario information reflecting the entire candidate path should be considered when deciding which RAP to handover to.

Developing a handover scheme for a relay based system is more demanding than in traditional cellular systems since new scenarios and options emerge.

The former one relates to that seven different HO cases need to be considered in a two-hop scenario:

- i) BSi-to-BSj,
- ii) BSi-to- BSi RNx
- iii) BSi RNx-to-BSi
- iv) BSi RNx-to- BSi RNy
- v) BSi-to- BSj RNx
- vi) BSi RNx-to-BSj
- vii) BSi RNx-to- BSj RNy.

where BSi RNx denotes the *x*:th RN associated to the *i*:th BS.

In a scenario where more than two hops are allowed there is also the special case where target RN is located downstream (upstream) of the source RN. Moreover, the induced protocol overhead needs to be considered to a larger extent than in conventional cellular networks as this overhead will be transmitted over the radio interface on every hop from the source to the destination. This will be further considered below.

In WINNER the BS takes the handover decision and updates the relevant routing tables in the REC, as well as the GW_C.

To obtain quality figures on signal strength in neighbouring cells one may consider three different measurement options:

- i) pure UL based handover based on UT pilots,
- ii) pure DL based handover based on RAP pilots and UT measurement reports
- iii) decision based on both DL and UL measurements.

The decision is that case ii) will be employed. Details and pros and cons with the different proposals are given in [WIN2D351].

Moreover, the number of RAPs/cells that the UT should measure on (as given in neighbour cell lists broadcasted by the BSs) should be kept as small as possible since: i) if the measurement duration is fixed, the measurement accuracy decreases with increasing number of RAPs/cells, and ii) if the number of samples necessary for obtaining the average values is fixed, the measurement duration will increase with increasing number of cells/RAPs. In relay based systems, more signalling needs to be conveyed over the radio interface and this will evidently increase the signalling overhead and thereby further exaggerates the need to minimize the number of RAP/cells that a UT should measure and report on as compared to conventional cellular networks. This problem increases with an increasing number of RAPs within a REC (at least if the coverage of a RAP is limited). Section B.1 further outlines some proposals on how uplink time alignment may be enhanced as compared to using the random access channel for this purpose.

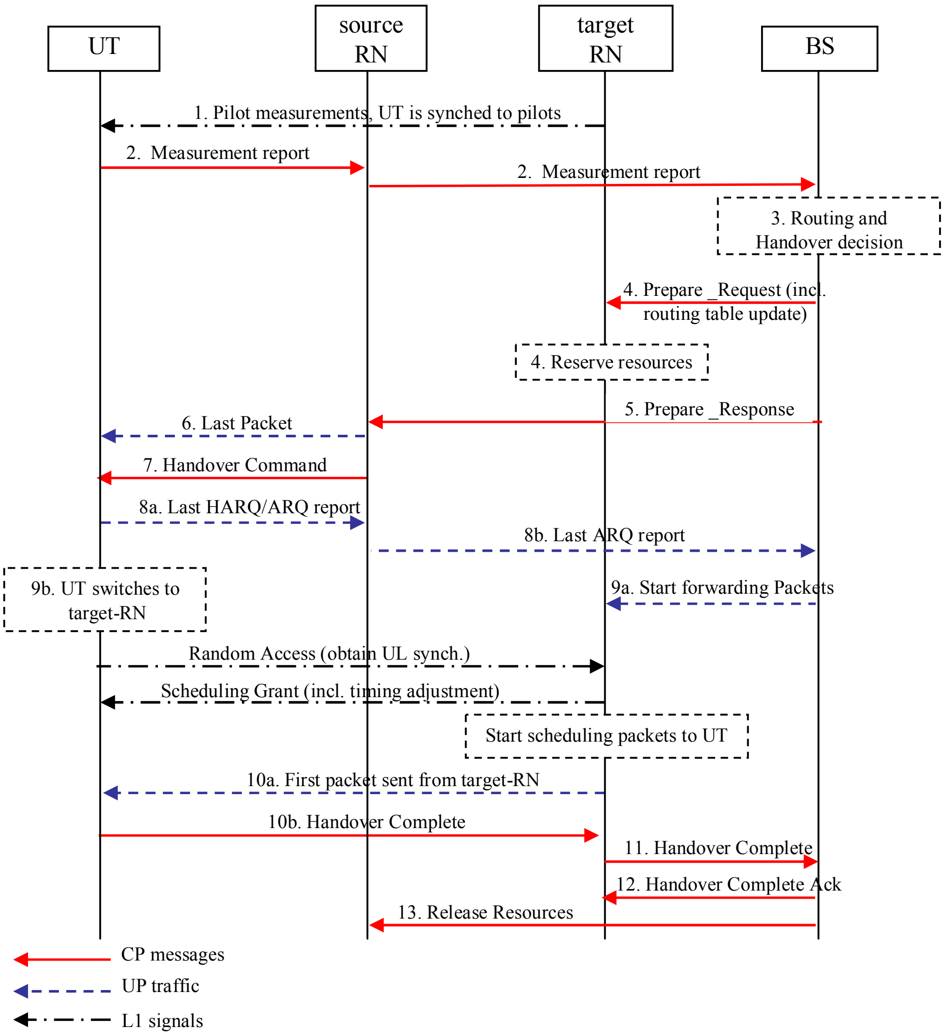

Figure 6-6: Message chart of the handover procedure in active-mode in a two-hop case (GW_C and GW_{IPA} are omitted in this case since they are not affected by this handover).

Figure 6-6 outlines an example flow chart of an intra-REC handover process (N.B. this is the simplest form of handover in a relay based system). As outlined above the UT (and RN) performs measurements on the neighbouring RAPs/cells that are included in the neighbour RAP/cell list . Based on these measurements and the associated measurement reports, as well as other higher layer triggers as outlined above, the BS may take a routing and handover decision. Subsequently, the BS conveys this information to affected RAPs and the UT. If necessary, the UT then commences with the procedure of obtaining uplink synchronization and when this is achieved the data communication will continue. For more details on (and optimizations of) the signals and procedures outlined in the figure, as well as example flow charts for the other handover scenarios, the reader is referred to section B.1.

The mechanism of handover may be enhanced with the introduction of additional centralised system entities (e.g. Hybrid Information System). In the centralised approach all the signalling and all the decisions go through a central RRM entity, which is the RRM Server. For more information about these potential enhancement see Annex B.2.4.

The purpose of **idle mode mobility** is to, in a resource and power efficient manner, keep track of the idle mode terminals and thereby enable network initiated connection setups, as well as to support efficient UT power saving. An overview of idle mode mobility will be given in the remainder of this section, but details are omitted to B.3.

To enable network initiated connection setups in a resource efficient manner, the mobile terminals are not tracked on a per cell basis. Instead, in idle mode the UT location is only known within a larger area here denoted the paging-area. At network initiated connection setups (i.e. the UT is moved from idle to active mode) paging messages are transmitted in all cells within the paging area to locate the UT.
To enable efficient power saving, a UT in idle mode will only occasionally wake up and listen to the paging channel to see if it is being paged. Moreover, during this active period it will also listen to the broadcast channel (for system information) and the common pilot channel to evaluate if it should perform a cell re-selection. If so, and if the cell belongs to a different paging area, the UT informs the network via a paging area update message.

6.4.2 Macro mobility

The mobility management mechanism in WINNER also covers macro mobility. Macro mobility is mobility over a large area. Macro mobility involves moving between two domains representing two different sub-networks, i.e. macro mobility/IP handover takes place when the UT changes the gateway it is associated with and thus enters a different IP domain with a new IP address. These domains might belong to one or several network operators. Thus, macro mobility is also named inter-domain mobility.

Macro mobility includes mobility support and associated address registration but also security and context transfer are important aspects. Mobile-IP can be seen as a means to provide macro mobility.

The micro mobility would exist between BS of the same pool of GWs and the macro mobility would exist in intersystem handovers (between WINNER and legacy RANs that are not tightly or very tightly coupled, i.e., the two systems do not share the same GWs) and in handovers between BS that belong to different pools of GWs, or when potentially load balancing between GWs takes place. According to WINNER architecture, the two layers of handover will be executed independently, that is when moving from one overlapping pool area to another, the time for the macro mobility handover is not dependent on the time for the micro mobility handover. This is also one of the arguments for having overlapping GW pool areas.

The benefit of independent execution of micro and macro mobility is that if the dual handover process should instead be completed in one step then there is a large amount of signalling that should be exchanged in a very short time. This follows from that the UT does not only have to change BS, but the network also has to execute a macro mobility handover to change the IP address and the GW associated with the UT. This process is complicated and could fail in a high network load situation.

7 Reference design and example physical nodes

7.1 Introduction

The WINNER system concept consists of the logical node architecture (described in section 3), protocol and service architectures (section 4), and the functionalities interconnecting these described in cross layer interactions (section 5), and radio resource management mechanisms (section 6) sections. The topics and items covered in these sections are such that they need to be standardized or specified when a new system is defined. In order to develop a working system, these are not enough, but a large amount of schemes and algorithms, filling the gaps and missing details in the general specification are required. As these solutions are only implicitly defined by the specification, they are vendor or manufacturer specific.

This section mainly focuses on these solutions, schemes, and algorithms that are not covered by a standard (in this case a hypothetical standard based on WINNER system concept), but which will need to be developed by the equipment vendors and manufacturers. Most of the presented topics are such that their conceptual description can be found from the earlier chapters (especially from chapters 3 to 6), and examples of their preferred implementations are described in this section. In some special cases (e.g. where the solutions details would depend on a specific system deployment) only a general description is given in the “system concept” description, and the design solution, which actually would need to be standardized, is given in this section. These exceptions were made to enhance the readability of the text.

Topics that are covered in this section include

- Channel coding; preferred encoding and decoding methods for QC-BLDPC, modifying the block length of these codes, channel coding for control channels
- Combined link adaptation and scheduling; overview of the multiple access, scheduling, spatial adaptation
- Synchronization; link level and network synchronization
- Channel estimation
- Utilization of cooperative relaying
- Interference mitigation schemes; with and without spatial processing
- UT location determination
- Flow handling mechanisms; congestion, load, and admission control
- Spectrum technologies; spectrum assignment negotiation, long- and short-term spectrum assignment algorithms
- Required measurements.

At the end of this section there are examples of physical deployments of WINNER systems. The four examples given are single-cell-layer wide area deployment, multi-cell-layer metropolitan area deployment, home base station deployment, and office / shopping mall deployment.

7.2 Channel Coding

7.2.1 Quasi-Cyclic Block LDPC Codes

After the global description of the nature of Low Density Parity Check codes chosen in WINNER (see section 4.8.3), we give hereafter more in-depth description of specific encoding/decoding methods well suited for such codes, together with an interesting technique enabling the increase of maximum codeword length.

7.2.1.1 Efficient Encoding Methods

There exist two main families of encoding methods suitable for encoding our particular structure of QC-BLDPC Codes. Those two encoding procedures share in common the structure promoted by Richardson and Urbanke [RSU01], and illustrated below on Figure 7-1. Full details can be found in the deliverable [WIN2D223], section 2.

$$\begin{array}{c} \quad\quad \begin{array}{ccc} N-M & g & M-g \end{array} \\ \mathbf{H}=\begin{bmatrix} \mathbf{A} & \mathbf{B} & \mathbf{T} \\ \mathbf{C} & \mathbf{D} & \mathbf{E}\end{bmatrix} \begin{array}{l} M-g \\ g=\gamma\cdot p \end{array} \quad M=m\cdot p \\ N=n\cdot p \end{array}$$

Figure 7-1: Richardson and Urbanke form of Parity-check matrix.

7.2.1.2 Decoding Algorithm

Thanks to the in-depth analysis and comparison evaluations carried out in WINNER (see [WIN1D210]), the decoding algorithm promoted for the reference design of WINNER system is a simple but efficient Normalized Min-Sum algorithm (MSA*), with a Scaling Factor of order 0.8 (the value can be tuned by post-processing). In addition to this, due to the particular structure of the QC-BLDPC Codes, it has been demonstrated, that adopting a Group-Layered Shuffling (Horizontal scheduling) could halve the average number of iterations needed to reach convergence in the decoding process, and thus end up with doubling the decoder throughput when compared to conventional Flooding Scheduling of LDPC decoder.

Moreover this not only optimizes the parallelism degree of the decoding architecture whilst still avoiding memory access conflicts, and optimizing the node processors activities, but can lead to enhance performance for high SNR regions by reducing the number of trapping-set (robustness to pseudo-codewords).

7.2.1.3 Lifting Property

Important part of the reference design solution for channel coding is the ease of adaptation of the block length of the channel code to fit to the required block sizes. A special requirement for WINNER system concept is the need to support very large block lengths, e.g. above 27000 bits. As demonstrated in [MY05] and [MYK05], applying a process called ‘Lifting’ to existing LDPC codes, enables one to increase the maximum allowable codeword length, whilst keeping same performance for previous range of codeword lengths (backward compatibility). In order to ensure consistent performance, the block length of BLDPC Codes developed in WINNER (R_c=1/3, 1/2, 2/3, 3/4), will be adjusted by applying ‘Lifting’ to our original parity-check matrices.

Our current constraints are the following:

- N_b=48 ➔ codeword length is multiple of 48 (cf. dimension of base-model matrix)
  - Maximum Expansion Factor = Zf_{max} = 576
  - Maximum Codeword Length = 27648 = 576 * 48
- Modulo Lifting procedure
  - With notation introduced in [MY05], this means the resulting Exponents $E(H_k)$of the parity check matrix H_k corresponding to Expansion factor L_k is given by:

$$E(\mathbf{H}_k) = E(\mathbf{H})\,\mathrm{mod}(L_k)$$

By applying step by step the Modulo-Lifting procedure described in [MY05], we have produced new parity-check matrices from the original parity check matrices (stemming from WINNER phase I) for the above mentioned code rates: R_c=1/3, 1/2, 2/3, and 3/4.The full details of thee lifted parity-check matrices, and the performance assessment of codes obtained with them can be found from Annex (A.6) of [WIN2D223].

7.2.2 Low-Rate Convolutional Codes

As mentioned in section 4.8.3, the channel coding candidate for short packets Transmission, such as control signalling, will make use of an Optimum Distance Spectrum (ODS) convolutional code as proposed in [FOO+98] , with coding rate Rc=1/4.

According to [FOO+98], an optimum distance spectrum convolutional code is a code generated by a feedforward encoder with a superior *distance spectrum* compared to all other like encoders with the same rate *R* and constraint length *L*. The superior distance spectrum is defined as follows:

> *A feedforward encoder with error weights* c_d *, giving a code free distance* d_f *has superior distance spectrum to encoder with error weights* $\tilde{c}_d$ *, giving a code free distance* $\tilde{d}_f$ *, if one of the following conditions is fulfilled:*
>
> *1)* $d_f > \tilde{d}_f$
>
> *2)* $d_f = \tilde{d}_f$ *and there exists an integer* $l \geq 0$ *such that:*
>
> *a)* $c_d = \tilde{c}_d$ *for* $d = d_f, d_f + 1, \ldots, d_f + l - 1$
>
> *b)* $c_d < \tilde{c}_d$ *for* $d = d_f + l$.

This means that for the same code rate and constraint length an ODS code has the same free distance as a Maximum Free Distance (MFD) code, but a lower or equal information error weight spectrum. Performance results and complexity investigations can be found in section 2.3 of [WIN2D223].

7.3 Combined Link adaptation and Scheduling

The task of selecting the packets to be transmitted and allocating the transmission resources within each frame is solved by the scheduler at the MAC layer. Its main characteristics have been described in Section 4.7.2. We will here discuss particular solutions and parameterizations that form parts of the WINNER reference design. In particular, we present solutions for

- The characteristics of the scheduler
- The characteristics of time-frequency resources that are allocated (the multiple access scheme)
- The selection of spatial transmit scheme, which can be adjusted to the requirements of each flow class and to changing channel properties (spatial adaptation).

7.3.1 Scheduling

Structure of the scheduler

In WINNER phase 1, the scheduler was assumed to be composed of a set of components, see Section 3.1.5 and Appendix C of [WIN1D210]. This main functional separation has in general been assumed within WINNER II. It distinguishes functions that work at two different time scales, namely super-frame or slower time scale, or slot/frame time scale.

- A ***constraint processor***, which combines constraints on the use of chunks and chunk layers. These arise from interference between user terminals, interference avoidance scheduling with neighbouring cells, resource partitioning between RNs an UTs in relay-enhanced cells and spectrum sharing between operators. The output of the constraint processor is in the form of a set of frame-specific chunk masks that define the restricted use of the chunks within a super-frame. The restrictions may either be hard or soft, in the form of transmit power constraints.These constraints are assumed to be prespecified one super-frame in advance.
- A ***spatial scheme controller***, which performs the spatial adaptation, i.e. adjustment of the spatial scheme used for each flow class and user. This adjustment is performed on a time-scale slower than the super-frame.
- A resource partitioning function now denoted ***resource division*** function. This function divides the resources into sets to be used for frequency-adaptive transmission and non-frequency-adaptive transmission, respectively. This partitioning is adjusted to the aggregate traffic need for these two types of transmission, but is precalculated and held fixed on a super-frame time-scale. The resource division may differ between the frames of the super-frame.
- A ***resource scheduler***, which performs the scheduling and link adaptation for each slot.

The constraint processor, spatial scheme processor and resource division work on a time scale slower than that of the resource scheduler. The adaptation performed by the resource scheduler, on a slot time scale, can therefore proceed with the parameters that change at a slower time scale acting as semi-fixed constraints.

Algorithms used by the resource scheduler

The resource scheduler has the over-all aim of allocating the transmission resources on a frame-by-frame basis, to satisfy the QoS constraints of all flow classes. By channel-aware scheduling, it should also allocate transmission resources that are advantageous for each transmission and perform joint link adaptation and scheduling.

The satisfied user criterion has been used as an over-all evaluation criterion in the WINNER project. It maximizes the number of users within the system among whom a high fixed percentage can be served with fulfilled QoS constraints. Specifically, the spectral efficiency has been measured under the constraint that 95% of the users within the system obtain a promised QoS level, which has in general been stated as a guaranteed bit rate [WIN2D6137]. This criterion therefore maximizes spectral efficiency and network capacity, with user-individual QoS requirements used as constraints.

Another key aim of the WINNER projects has been a low delay over the air interface. This enables adaptivity with respect to fast channel variations, it facilitates high-throughput TCP/IP traffic and it enables fast link retransmission, which is of advantage for the performance perceived at high layers. To attain this goal, a constraint on the computational complexity of the scheduling has been used. It has been assumed in Section 5.2.3 that the delay from the last information needed for scheduling a slot until the scheduling is finalized is 0.1 ms at maximum.

For these reasons, the ideal scheduling algorithm for the WINNER system concept would be an algorithm that directly optimizes the satisfied user criterion and that executes in at most 0.1 ms. Such an algorithm has not been yet found. However, simplified algorithms exist that provide reasonable performance and have very low computational complexity, see [WIN2D61311] for some evaluations. Examples that take the channel quality into account to maximize throughput-related performance measures are the Score-based scheduler, the proportional fair scheduler and various variants of the weighted sum rate scheduler [SSO+07].

The *score-based scheduler* [Bon04] has been used as default option in the WINNER baseline design [WIN2D6137]. This algorithm is an improved variant of the proportional fair scheduler [Ben00]. It allocates each channel resource to the user whose predicted channel quality is highest, relative to the predicted quality for this user in the other resources under evaluation. This evaluation is performed in a way that makes the allocation insensitive to variations in the statistics of the channel variability among users. The proportional fair algorithm, which has been used in some of the evaluations, tends to favour users who have a high relative variability of the channel quality.

However, while working well, the utilized algorithms take neither the satisfied user criterion nor a large variety of QoS constraints into account explicitly in their designs. The scheduling algorithms and most performance evaluation assumptions have instead been based on a full buffer assumption, i.e. there is always data to send. A useful performance evaluation with non-full buffers can be found in Section 4.7 of [WIN2D61310].

There is a large potential for improved performance in constructing improved scheduling algorithms. The construction of improved scheduling algorithms is an outstanding open problem at the end of the WINNER II project, with intriguing potentials for future performance gains.

Link adaptation algorithms

For both frequency-adaptive and non-frequency-adaptive transmission, the link adaptation algorithm proposed for the WINNER reference design is the MI-ACM algorithm (Stiglmayrs algorithms), which is explained in detail in Chapter 3 of [WIN2D223]. This algorithm provides a significant improvement over the state of the art link adaptation algorithms.

In case of frequency-adaptive transmission, the MI-ACM scheme uses fixed transmit power in all chunk layers onto which a code block is mapped. The modulation is adjusted individually for each chunk layer while the coding is not adjusted individually: an average code rate is used over the whole code block. Very little performance could be gained by in addition also adjusting the transmit power and/or the code rate individually per chunk layer, but the computational complexity and control overhead would become much higher.

7.3.2 Multiple Access

7.3.2.1 Frequency-adaptive transmission

The multiple access scheme for frequency-adaptive uplink and downlink is based on chunk-wise frequency-adaptive **TDMA/OFDMA** in both the FDD and the TDD modes. The same scheme is used for both uplink and downlink.

Chunk-based TDMA/OFDMA means that flows to/from one UT are mapped onto individual chunk layers[14]. The mapping is exclusive within the cells, i.e. each chunk layer carries data belonging to only one UT[15]. Individual link adaptation is used within each chunk layer, based on predicted SINR that will be perceived within that particular chunk layer for that particular user, at the time instance when the transmission will occur. One set of link adaptation parameters is used within the whole chunk.

Fast control loops for FDD and TDD enable reliable prediction and CQI collection up to vehicular speeds. Simulation results in [WIN2D461] prove that without channel prediction, the impact of UT's speed on the spectrum efficiency is quite severe: 33% performance loss from 3 km/h to 50 km/h. In [WIN1D24] it was demonstrated that the use of channel quality prediction enables frequency-adaptive transmission up to a limiting velocity (up to 70 km/h at 5 GHz carrier frequency) that depends on the average user SINR. Non-frequency-adaptive transmission has to be used outside of this feasible terminal velocity range and SINR, which has been shown to depend on the Doppler spectrum, see [WIN2D223] and references therein.

As outlined in Section 5.2.3, the uplink control is based on an initial request for transmission in frame *j-2*. If granted, the transmission is scheduled and prepared during frame *j-1* and is then performed in the uplink slot of frame *j*. The corresponding out-band frame control information is transmitted in the preceding downlink slot of frame *j*. The downlink transmission in frame *j*, is scheduled during frame *j-1*. The data transmission and the associated out-band control signalling then follows in the downlink slot of frame *j*.

This tight feedback loop enables the frequency-adaptive transmission to utilize the fine-grained channel variations, by relying on the prediction of the SINR within each chunk (layer). These channel variations will differ between channels to different user terminals. We therefore obtain significant multi-user scheduling gains when flows can be allocated to chunks that provide the best channel gains and interference conditions. These gains are quantified in e.g. [WIN1D24] and for multi-antenna schemes in [WIN2D341].

To realize frequency-adaptive transmission, a significant but still reasonable control and feedback overhead is required, as discussed in Section A.1. The downlink control information used for controlling downlinks as well as uplinks has to be transmitted with adequately low error rate at a reasonable coding overhead. For this reason, the use of frequency-adaptive transmission is thus suggested to be used above a certain limiting SINR.

In order to take advantage of both small-grained channel adaptation and strong FEC coding, a FEC coding concept is developed that support large FEC blocks that span multiple chunks with individually calculated link adaptation parameters, see [WIN2D223] for further details of the FEC schemes.

Frequency-adaptive transmission should typically be combined with multi-antenna transmit schemes that preserve the channel variability. This is due to the well known effect of transmit diversity to reduce the channel variations in frequency. In this case, the amount of channel variations that can be exploited by the scheduler is lower, which reduced the scheduling gain.

7.3.2.2 Non-frequency-adaptive transmission

With non-frequency-adaptive allocation, bits from each flow are allocated onto sets of chunks that are dispersed in frequency and/or space. Forward Error Correction (FEC) coding and interleaving are used to combat the small-scale frequency selective fading, aided by a Hybrid-ARQ scheme [WIN2D223]. No fast feedback from receiver to transmitter is used. Still, link adaptation is performed with respect to shadow fading, but not with respect to frequency-selective fading, i.e. the same modulation and coding is used in all frequency resources allocated to the flow in a frame slot (i.e. chunks and chunk layers), but the resource scheduling can be as fast for non-frequency-adaptive transmission as for frequency-adaptive transmission.

Non-frequency-adaptive transmission mode is required for control signaling and is also used for point-to-multipoint transmission in multicasting and broadcasting, which is useful for MBMS services. Due to the chunk-wise link adaptation and also to the multi-user diversity gains, the achievable spectral efficiency with frequency-adaptive transmission is higher whenever reliable CQI can be obtained. However, adaptive transmission becomes less reliable below a certain SINR threshold and above a certain user

[14] A chunk comprises a certain time-frequency resource consisting of a few OFDM subcarriers in a few consecutive OFDM symbols. A chunk can be spatially reused, and the spatial reuse dimension is denoted a chunk layer.

[15] In WINNER II we support MAC multiplexing of RLC PDUs, thus many flows to/from a UT can be multiplexed into the same chunk layer, compare with the flow concept in WINNER Phase 1 [WIN1D210].

terminal velocity. Thus, the non-frequency-adaptive mode offers a robust option for scheduled flows in situations reliable CQI cannot be obtained, and it thus serves as a fallback solution for frequency-adaptive scheduling. Non-frequency-adaptive transmission could also be used in an initial startup phase for adaptive flows before channel prediction is reliable enough, although recent results reported in [WIN2D461] chapter A3.1.3 (briefly mentioned in its Section 5.6.2 as well) indicate that adaptive transmission could be competitive also when working on channel estimates only.

The multiple access schemes for non-frequency-adaptive uplink and downlink for the WINNER system concept were introduced in [WIN2D461]. They are called Block Interleaved Frequency Division Multiple Access (B-IFDMA) and Block Equidistant Frequency Division Multiple Access (B-EFDMA) respectively. The resource allocation for B-IFDMA and B-EFDMA are the same and is illustrated in Figure 7-2 below. The difference between the schemes is that in B-IFDMA a common DFT precoding step is performed over the allocated blocks, see [SFF+07] for a detailed signal definition of B-IFDMA. A DFT-precoding is not seen as useful in the downlink since the RAP has to transmit a sum of frequency-adaptive and non-frequency-adaptive flows over the entire system bandwidth.

These schemes aim to maximize frequency diversity, accommodate small as well as large packets per frame, enable micro-sleep within chunks, support high speed trains, keep low addressing overhead and to simultaneously enable low envelope variations of transmitted uplink signals. The similarity of the uplinks and downlinks further simplify the system.

Note that in B-IFDMA and B-EFDMA, we do not define the resource unit per chunk duration (FDD or TDD slot) based on the chunk size, but rather as a function of the block size, repetition distance and number of blocks in a slot, e.g. in Figure 7-2 there are 48 channel symbols allocated for each of the users in the slot. The appropriate repetition distance depends on the coherence bandwidth of the channel to maximize diversity, and of the bandwidth allocated for non-frequency-adaptive transmission. The chunk concept is however still useful; it defines the common resource unit for resource allocation between adaptive and non-frequency-adaptive transmission and resource partitioning between access points in case effective frequency reuse distance larger than one is used. Furthermore, it is a common entity for channel estimation due to its property of being essentially flat in time and frequency.

A basic parameterization of B-IFDMA and B-EFDMA is a block size of 4 subcarriers x 3 OFDM symbols in both FDD and TDD, as illustrated in Figure 7-3. This block size fits the chunk size in both the FDD and the TDD frame structure as defined in Section 4.7.1. These small blocks can be used for resource allocation for small packets, encoded by convolutional coding. A frequency-hopping component is not built-in to the non-frequency-adaptive multiple access schemes, but could be included using short blocks in deployments with severe other system interference. For larger code block sizes, it is advantageous to combine the 4x3 blocks into larger units, which also improves the channel estimation performance [WIN2D233]. The appropriate block sizes for different scenarios are discussed in Annex C.

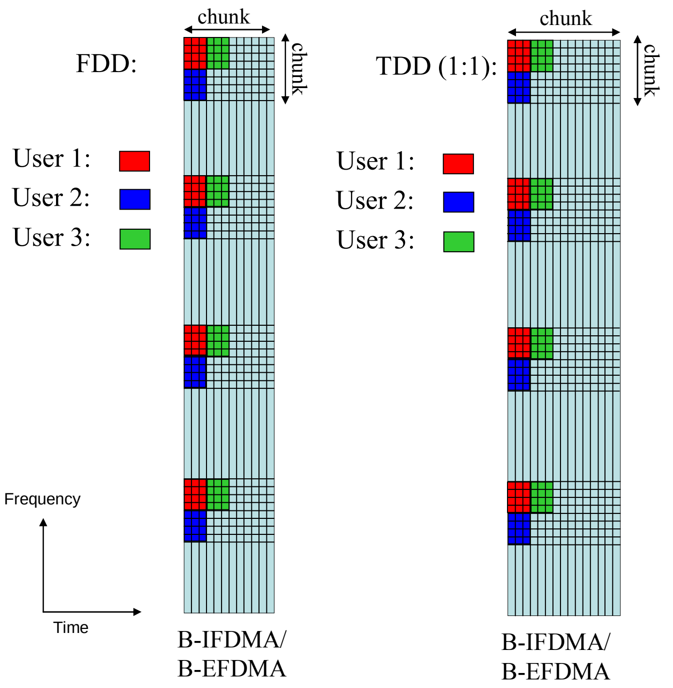

Figure 7-2: Illustration of B-IFDMA and B-EFDMA resource allocation in FDD and TDD with the smallest Block size, to be used for transmission of small packets using maximal frequency-diversity.

7.3.2.3 Co-existence of multiple access schemes

Frequency-adaptive and non-frequency-adaptive transmission both benefit from a large resource pool with independent fading statistics. Non-frequency-adaptive transmission needs the independent resources to combat the small-scale fading by diversity combining techniques. Frequency-adaptive transmission benefit from independent resources to minimize the service outage probability for semi-static users and to enable large multi-user diversity gains. The default assumption is to do frequency multiplexing of the resources for frequency-adaptive and non-frequency-adaptive transmission, but for deployment with low system bandwidth and low carrier frequency, time multiplexing of resources for frequency-adaptive and non-frequency-adaptive transmission could be preferred.

7.3.2.4 Switching between multiple access schemes

As shown in Figure 7-3, within a frame different sets of chunks are pre-allocated for each of the two transmission modes, frequency-adaptive and non-frequency-adaptive transmission. These sets are fixed within the whole super-frame, but may be changed between super-frames. The selection of multiple access scheme is made at the flow set-up, but switching between multiple access schemes can be made on an ongoing flow at a superframe time scale if needed[16]. Indeed, the typical large cell size and high number of users per cell make the WA scenario very composite. Below, we present criteria and mechanisms for switching of multiple access schemes within a cell. During handover to other cells, there are cases when switching is needed. This re-negotiation has to be made as part of the handover process.

The initial selection of MA scheme for a flow is based on two sets of criteria, preselection criteria and switching criteria. The *preselection criteria* are rather static. They are determined by the capability and the configuration of the RAPs (BS and RNs), the spectrum constraints and the UT capabilities and their distribution in the cell. They do not rely on measurements. The *switching criteria* are instead more related to parameters that change dynamically within the cell and for the flow. They are monitored and later used for possible switching of multiple access schemes for already established flows.

[16] It may be advantageous to let flows switch between these transmission types after the flow setup. For example, the first packet of a newly setup "frequency-adaptive" flow may be transmitted by non-frequency-adaptive transmission, since the channel predictors used in the frequency-adaptive transmission need a certain time to produce predictions with close to steady-state accuracy, and its accuracy is not yet adequate for the required prediction horizon for this particular UT velocity. In such a case, the described switching would reduce the flow setup delay, as compared to just waiting for the predictors to reach the required accuracy.

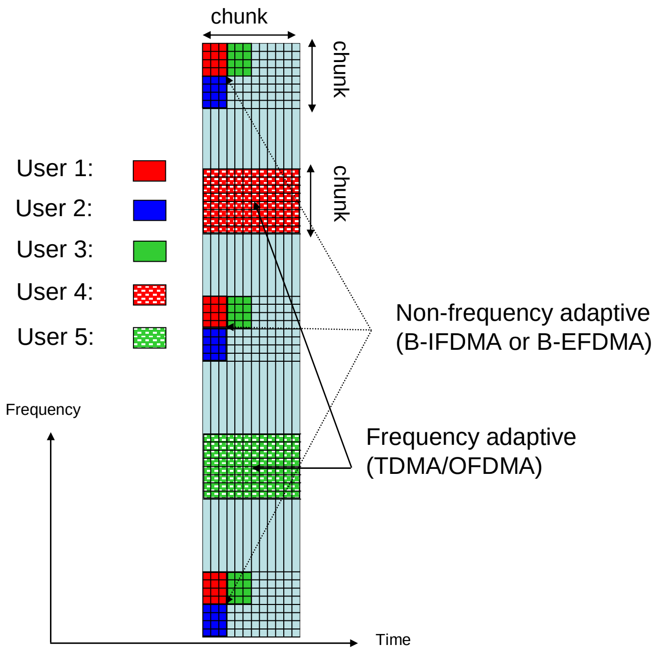

Figure 7-3: Frequency multiplexing of resources for frequency-adaptive and non-frequency-adaptive transmission.

The identified set of preselection criteria is:

- Flow class
- Next hop node type (BS, RN or UT)
- UT capability (from e.g. UT ID database)
- Chunk resources and their current constraints
- Cell load.

For example, relay links are characterized by good semi-static channel conditions, possible antenna directivity gains and large continuous data transfer rate due aggregation of many user flows served by the RN. This calls for selecting frequency-adaptive transmission, while keeping the CQI update rate and hence the feedback and pilot overhead rather low. The use of some of the criteria for the initial selection of frequency-adaptive transmission or non-frequency-adaptive transmission for a newly set-up scheduled flow is exemplified in Section 5.2.1 of [WIN2D461].

The identified set of switching criteria is:

- CSI/CQI quality, as measured by the estimation/prediction Normalized Mean Square Error (NMSE), see [WIN2D223] for a definition of prediction NMSE
- Terminal velocity (e.g. based on Doppler spread or satellite aided measurements)
- Average SINR
- Downlink SINR, which determines the reliability of downlink control information and the associated required coding overhead of this control information
- The number of recipients of a multicast transmission.

An important enabler for integrated efficient switching is the approach for hop-by-hop ARQ described in [WIN2D223] chapter 4. It is based on a flexible incremental redundancy Hybrid-ARQ scheme that relies on a soft-bit interface. The soft bit interface enables a seamless switching of transmission method also for packets subject to retransmission.

The pilot patterns and pilot schemes that have been selected for the WINNER reference design (see Section 7.4) do affect the possibilities to use frequency-adaptive transmission and the preferable combinations of adaptive transmission and multi-antenna schemes. These aspects are summarized as follows:

- For FDD downlink, a fixed grid of beams is used, with common pilots per beam. These pilots are present in each chunk, and they thereby support frequency-adaptive transmission at vehicular velocities. The prediction schemes and feedback loop designs that were proposed in Section 3.1

of [WIN1D24] can be used for this purpose. As outlined in Section 3.1.4 of [WIN1D24] and [EO07], the feedback signalling can be reduced to acceptable rates, around 0.25 bits per chunk layer at 50 km/h, by using its time and frequency correlation to compress the feedback.

- For frequency-adaptive transmission in FDD uplinks, the uplink pilots used for channel prediction are assumed to be transmitted only once per super-frame. This choice has been made to limit the uplink pilot overhead. This reduces the accuracy of channel prediction so it may not be possible to use frequency-adaptive FDD uplink at vehicular velocities (as was the case for the designs investigated in Section 3.1 in [WIN1D24]).
- For TDD systems, frequency-adaptive transmission in downlink would be integrated with one of several possible multi-user MIMO-OFDM schemes (cf. section 3.2.8 in [WIN2D233]). For downlink transmissions that use the SMMSE (successive minimum mean square error) multi-user MIMO scheme with short term CSI at the transmitter, the appropriate pilots to use would be uplink pilots transmitted in the super-frame preamble from all user terminals that take part in the competition for the set of frequency resources. This SMMSE transmission scheme is limited to users below 10 km/h and the super-frame preamble pilots allow frequency-adaptive transmission to be used at these velocities. Spatial multiplexing with per antenna rate control is the preferred scheme at velocities 10-50 km/h in metropolitan area deployments. In such cases, unweighted pilots would be transmitted from each antenna in each downlink slot. The UTs can generate CQI estimates on all chunks where these downlink pilots are transmitted. These CQI estimates are compressed as described in [WIN1D24] and transmitted to the BS/RN over the uplink. This enables the use of frequency-adaptive transmission in both downlinks and uplinks, due to the TDD channel reciprocity, up to velocities determined by the vehicle velocity and the Doppler spectrum properties of each channel.
- For non-frequency-adaptive B-IFDMA transmission in uplinks, the pilot scheme uses one pilot symbol per 4x3 block. See [WIN2D233] for investigations of the resulting channel estimation errors, with and without iterative channel estimation schemes.

7.3.2.5 Power control

The reference design for power control in *downlink* is that the RAP transmits with a constant power spectral density (psd), i.e. each chunk transmit power is the same over the allocated unconstrained chunks within the assigned spectrum. However, the constraint processor can assign other power levels due to e.g. spectrum sharing and interference avoidance schemes.

No reference design of an *uplink* power control scheme has been defined for the WINNER system concept, but there is ongoing research on this topic. A candidate scheme is likely to take into account:

- Requirements on acceptable intercarrier-interference levels in the RAP receiver due to the frequency-selective channel gains over the different uplink channels for adjacent subcarriers, the Doppler spread and the imperfections at the transmitter side causing spectral leakage.
- Maximum UT RF power capability. At a given path loss, this implies a restriction on the number of concurrent resources (chunks or blocks) that can be allocated to the UT in order to satisfy a minimum received power spectral density. Thus, the UT capability and path loss act as input to Resource Scheduler on maximum number concurrent resources that can be assigned to a UT at given time (i.e. it controls the actual signal bandwidth in terms of number of used subcarriers). In a low path loss scenario, an upper limit might be needed on the received power spectral density. Note that the maximum UT RF power can depend on the resource allocation as well, since the power amplifier backoff may have to to be adjusted for different uplink signals.
- No power loading is assumed in the link adaptation schemes [WIN2D223], thus the power control should try to maintain a receiver SINR variability in the assigned resources (SINR per chunk layer in frequency-adaptive transmission and SINR average over the utilized frequency subband for non-frequency-adaptive transmission), that is matched to the SINR range used by the MCSs in the link adaptation scheme [WIN2D223]. I.e. also excess receive power spectral density above the MCS range should be avoided, since the power is wasted for that link and causes unnecessary interference in other cells.

Thus, a slow power control scheme that adjusts the transmit power for uplink data channels seems to be needed, and it should adapt on a time scale of the long-term shadow fading and path loss. The users might be allowed to chose their own transmit power levels as long as the received power spectral density is within a received power window. Outside of that window, the RAP directed slow power control can take over.

7.3.3 Spatial adaptation

The process of selecting the spatial scheme is called spatial adaptation. The spatial adaptation consists of selecting the appropriate spatial scheme for transmission or reception in such a way that spatial scheme is adapted continuously to the spatial properties of the channel. Spatial adaptation aims at improving the signal reception at the receiver and to manage the downlink or uplink intra-cell interference.

In order to be able to choose a capacity achieving scheme the spatial adaptation involves grouping the users into various spatial groups. The grouping is followed by selecting the adequate user(s). User grouping coordinates transport block transmissions in such a way that the experienced intra-cell interference at the user terminals during downlink transmission or at the base station during uplink transmission is reduced.

In the following, a description of a generic user selection algorithm is given in section 7.3.3.1, while section 7.3.3.2 gives an overview of the spatial schemes that have been identified as most suitable in different scenarios.

7.3.3.1 Spatial User Selection/Grouping

Consider a downlink transmission with M transmit antennas and K users with N antennas each. Let S, T and P be respectively the set of possible user allocations, the set of possible transceiver strategies and the set of power allocations. The optimum (capacity achieving) scheme can be found by solving the following expression

$$\max_{\bar{S}\in S,\bar{T}\in T,\bar{P}\in P} \sum_{k=1}^{K} \alpha_k \log\left(1+\gamma_k\left(\bar{S},\bar{T},\bar{P}\right)\right)$$

where $\gamma_k\left(\bar{S},\bar{T},\bar{P}\right)$ is the signal to noise-plus-interference ratio (SINR) of the k-th user for a given choice of $\bar{S}$, $\bar{T}$ and $\bar{P}$, and α_k is the quality of service (QoS) weight given by the scheduler.

The capacity achieving scheme is the so-called *dirty paper coding* (DPC) scheme [CS03,VT03,YC04,WSS06] where the signal of different users are successively encoded at the transmitter side. The optimum receive and transmit coefficients are calculated using the *duality* concept [VT03]. The user selection stage is embedded in the standard power control technique applied to the dual channel [BTC06].

Suboptimum linear precoding (also known as beamforming) techniques have also been proposed in which the transmitted signal is a linear combination of the users' data signals. With respect to DPC, linear techniques guarantee lower performance at a reduced computational complexity. Moreover, due to the fact that there is no cancellation of the interference at the transmitter side, linear precoding schemes can be extended to the case of imperfect channel state information at the transmitter side. Differently from DPC, the user selection stage is not embedded in the precoder design stage.

In the following we present a general framework for spatial user selection schemes under linear precoding. For simplicity, we first treat the case where one stream is transmitted for each user. Afterward, we present the multi-stream case.

Single Stream User

Let $R(\bar{S})$ the achievable rate for a given set of users $\bar{S}$

$$R(\bar{S}) = \max_{\bar{T}\in T,\bar{P}\in P} \sum_{k=1}^{K} \alpha_k \log\left(1+\gamma_k\left(\bar{S},\bar{T},\bar{P}\right)\right).$$

Assuming K the total number of users, we need to consider all possible sets of users with cardinality $\left|\bar{S}\right| \leq \min(M,K)$.

The user selection problem can be written as

$$R(S^*) = \max_{\bar{S}\in S} R(\bar{S})$$

where S^* is the optimum set of users.

Finding S^* requires a brute force search over all possible sets of $1,2,\ldots,\min(K,M)$ users, whose number is $\sum_{j=1}^{\min(M,K)} \frac{K!}{(K-j)!\,j!}$. So the complexity of search becomes unacceptably high for large K.

Different suboptimum greedy user selection schemes have been proposed in order to lower the computational complexity. The general idea is that a new user is added to the set of allocated users at each iteration of the algorithm, in order to maximize a given metric.

Let $S^{*(i-1)}$ be the set of users allocated at the i-1th iteration of the algorithm. At the ith iteration, the set of selected users is updated as follows

1. A new set of candidate users $C^{(i)} \in \{1,\ldots,K\}$ is generated such that $C^{(i)} \cap S^{*(i-1)} = \Phi$.
2. The best candidate is chosen in order to maximize a given metric ξ

$$c^* = \arg\max_{c \in C^{(i)}} \xi(S^{*(i-1)} \cup c).$$

For example in [DS05] or [FDGH05a] the previous user selection approach is used with a zero-forcing beamforming technique. In [DS05] the set of available users is simply $C^{(i)} = \{1,\ldots,K\} \setminus S^{*(i-1)}$ and the metric is the sum-rate $\xi(S^{*(i-1)} \cup c) = R(S^{*(i-1)} \cup c)$. To reduce the complexity, approaches such as ProSched ([FDGH05a] [FDGH05b]) may be applied which approximate the metric using orthogonally projected channels of the users. An implementation guide is available in [WIN2D341]. A similar approach has been taken in [YG05] where the set of candidate users is chosen in order to have semi-orthogonal channels with respect to the users' channels chosen in the previous iterations. Such approaches can also be successfully applied to non-ZF techniques.

Multi-Stream user

Let us assume that multiple independent streams can be sent to each user. The described approach can be extended to the case of multiple streams per user, under the assumption of a block diagonalization transceiver design [BH07]. The achievable rate can be written for a given selection of users and eigenmodes $\overline{E}$. The problem is then reduced to find the optimum set of eigenmodes, E^*. The optimization is preformed for a given user and eigenmode selection. A generalization of the greedy allocation algorithm can be used instead of the brute-force allocation in order to avoid the high computational complexity.

Let $E^{*(i-1)}$ the set of eigenmodes allocated at the i-1th iteration of the algorithm. At the ith iteration, the set of selected eigenmodes is updated as follows

1. A new set of candidate eigenmodes $C_E{}^{(i)}$ is generated such that $C_E{}^{(i)} \cap E^{*(i-1)} = \Phi$.
2. The best candidate is chosen in order to maximize a given metric ξ

$$c^* = \arg\max_{c \in C_E{}^{(i)}} \xi(E^{*(i-1)} \cup c).$$

7.3.3.2 Spatial Schemes

The generic transmitter, see Section 4.8.2, can be configured to realize a vast number of different spatial schemes. The particular scheme to be used in a certain situation depends on several parameters, e.g. physical antenna configurations, properties of the radio propagation channel, type of deployment, availability of and quality of channel knowledge, system load, etc. To investigate and recommend a particular spatial scheme for all different situations that can be imagined would be a very challenging and time consuming task, and is not feasible. Instead, we will here give preferred schemes for three basic deployment scenarios:

- Macro-cell deployment with above rooftop antennas
- Urban micro-cell deployment with below rooftop antennas
- Isolated cell indoor deployment with distributed antenna arrays

Macro-cell deployment with above rooftop antennas

The wide-area channels typical for macro-cell deployment are characterized by their small angular spread (assuming BS antennas above rooftop), and the antenna correlations are significant, which results in a low rank channel matrix [WIN2D112]. This makes transmit beamforming schemes suitable. Furthermore, short-term CSI will not be available at the transmitter as user mobility is non-negligible and the feedback bandwidth and delay is limited. On the other hand, for low to medium user velocities scalar CQI information can be used on a fast fading granularity.

Together, this suggests that fixed beamforming with a grid of fixed beams (GoB) is a suiable scheme for the downlink, since it exploits small angular spreads which makes it possible to focus a large portion of the transmitted energy to a restricted angular zone. In addition, SDMA can be implemented on top of it, i.e. several users can be scheduled on the same resources in the same cell but in different beams. The suggested antenna configuration is a 4-element uniform linear array (ULA) forming 8 beams, while the UTs have two receive antenna elements in the form of one cross-polarized antenna element and employs IRC receive processing. The number of spatial streams per user is one, i.e. no spatial multiplexing is used, and the number of simultaneous users is limited to four in order to keep the inter-beam interference at an acceptable level.

For the uplink, it is recommended to implement SDMA based on IRC/SIC processing in the BS receiver.

Further details and performance of these schemes are given in [WIN2D341] and [WIN2D610], while the pilot design required to support the schemes are described in sections below

Urban micro-cell deployment with below rooftop antennas

In the urban micro-cell scenario the propagation conditions are challenging due to high-rise buildings and the shadowing they cause, which are typical for metropolitan areas. The user mobility is however limited, and it can be expected that a majority of the users are stationary or slowly moving (pedestrian), while a fraction is moving with vehicular velocities up to 50 km/h. The throughput requirements are high, which suggests that spatial multiplexing is important to achieve.

For the majority of users that are stationary or slowly moving (up to 10 km/h), availability of high quality short-term CSI at the transmitter (CSIT) can be assumed. Hence, for these users it is recommended to use MU MIMO precoding techniques on the downlink, which exploits the CSIT in order to achieve a flexible combination of SDMA and spatial multiplexing providing very high system capacity. It is recommded to use 4 cross-polarized antenna elements at the BS, i.e. the number of transmit antennas is 8, and to allow multiplexing of up to 4 spatial streams. For higher mobility, the quality of the CSIT is degraded, and it is instead recommended to switch between spatial multiplexing (e.g. in the form of PARC) and a linear dispersion code such as the well-known Alamouti scheme. Typically the spatial multiplexing mode (with up to 2 streams) is used for high SNRs, while Alamouti is used at low SNRs. Adaptation mechanisms for switching between these schemes are outlined in [WIN2D611]. At the UT, one cross-polarized antenna element and IRC receive processing is recommended.

For the uplink, as for the macro-cell deployment with above rooftop antennas, it is recommended to implement SDMA based on IRC/SIC processing in the BS receiver. The low mobility users employing MU MIMO precoding on the downlink, should use antenna hopping at the uplink in order to ensure that the full transmit covariance matrix is made available at the BS.

Further details and performance of these schemes are given in [WIN2D341] and [WIN2D611], while the pilot designs required to provide the schemes with necessary CSI and CQI are described in the section below.

Isolated cell indoor deployment with distributed antenna arrays

In the isolated cell indoor scenario, very high data rates are targeted. The mobility of the users is expected to be rather low, i.e., the terminal velocity will not exceed pedestrian walking speeds significantly. The distributed antenna array is another particular feature that chracterizes this local area scenario.

All this suggest that advanced MU MIMO precoding techniques are suitable for downlink transmission. As described for the urban micro-cell scenario above, they can explot the high quality CSIT in order to spatally multiplex a large number of streams. The use of the distributed antenna array even further boost this property. It is recommended to employ 32 transmit antennas in the form of 4 cross-polarized 4-element arrays, which allows multiplexing of up to approximately 30 streams. At the UTs, as for the two other scenarios it is recommended to have one cross-polarized antenna element and IRC receive processing.

For the uplink, many of the benefits used for downlink transmission also apply. The large number of antennas can be used efficiently, if they are deployed as distributed antenna arrays. This allows simulataneous transmissions from several users, and MU MIMO decoding is used at the BS to separate them. It is wise to assume that the UTs do not have any CSI. Acquiring it would be very expensive both in terms of pilot overhead (due to the large number of antennas) as well as in complexity. For the preprocessing of the data at the terminals there are two proposals: As a non-frequency-adaptive version we propose to use an open-loop linear dispersion code, such as Alamouti. Its advantages are that it is suitable for two antennas, easy to implement at the UTs and still enables us to extract the antenna diversity. As a frequency-adaptive version, the BS can pre-compute the optimum beamforming weights for the UTs and signal them to the UTs.

Further details and performance are given in [WIN2D341] and [WIN2D612], while the pilot design is required to provide the MU MIMO precoding schemes with CSI is described in Section 7.4.3 below.

7.4 Reference pilot design

The reference pilot design is closely tied to the multiple access and spatial processing schemes summarized in Section 7.3. Moreover, the combination of spatial schemes also implies challenges to the pilot design specified in Section 4.8.4 and Annex A.3, due to the inter-dependencies of CSI/CQI required for adaptive spatio-temporal processing and the provision of reliable estimates of the effective channel at the receiver. Due to these mentioned inter-dependencies, we will here consider the same three deployment scenarios that were considered for the spatial schemes above:

- Macro-cell deployment with above rooftop antennas described in Section 7.4.1
- Urban micro-cell deployment with below rooftop antennas described in Section 7.4.2
- Isolated cell indoor deployment with distributed antenna arrays described in Section 7.4.3

7.4.1 Pilot design for macro-cell deployment with above rooftop antennas

The pilot grids for the FDD mode (shown in Figure A-9, Annex A.3.2.2) are applicable for wide area (WA) deployment [WIN2D61310]. Furthermore an uplink super-frame pilot preamble (see Annex A.3.3) is foreseen to enable adaptive transmission on FDD uplinks for slowly moving users, as described in Section 7.4.1.2.

For the reference design the BS is equipped with a 4-element ULA, i.e. $n_T = 4$ transmit antennas, while the UTs exhibits $n_R = 2$ receive antennas in the form of one cross-polarized element [WIN2D61310].

7.4.1.1 Downlink

The reference design employs a GoB with 4 antennas forming 8 beams, see section 7.3.3.2. Common pilots per beam (CPB) are inserted. With a spatial pilot reuse of 2 and $P_n = 4$ orthogonal pilot sets up to 8 beams can be supported. CPB allow for interpolation over the whole frequency band, the pilots benefit the same beamforming gain as the data symbols, and CQI estimates at the UT can be determined. As the pilot overhead of $\Omega_p = 16.7\%$ is within acceptable limits (see Table A-2), CPB offer an attractive choice for the WA reference design.

On chunks where data is broadcast without beamforming (e.g. control information or MBMS services) common pilots per antenna (CPA) are used. We note that on a certain chunk CPA are never inserted together with CPB, rather CPA *or* CPB are picked dependent on the chosen spatial transmit processing scheme.

7.4.1.2 Uplink

In WA uplinks dedicated pilots per antenna (DPA) are used. For OFDMA the pilot pattern specified in Figure A-9 are used, while for B-IFDMA the pattern shown in Figure A-11 are used. For non-frequency-adaptive transmission no other pilots are necessary, i.e. users do not transmit uplink pilots in the super-frame preamble.

Slowly moving users that are scheduled for frequency-adaptive transmission need to transmit uplink pilots in the super-frame preamble to enable the BS receiver to scan for those chunks with the best CQI over the entire band. The uplink super-frame pilot preamble contains unweighted dedicated pilots per antenna over the full band (DPA-FB). With a pilot spacing of $D_f = 8$ up to $N_u = 8$ users can share the preamble. In case, $N_u > 8$ competition bands are introduced, which means that a certain user may access only a fraction of the available bandwidth. Therefore, DPA-FB pilots only need to be inserted in the corresponding competition band.

7.4.2 Pilot design for urban micro-cell deployment

The micro-cell deployment in metropolitan area (MA) utilizes the TDD mode [WIN2D61311]. Hence, the pilot grids for the TDD mode apply, shown in Figure A-10, and specified in Annex A.3.2.3 for the downlink and Annex A.3.2.4 for the uplink. Furthermore an uplink super-frame pilot preamble (see Figure A-12) as described in Annex A.3.3 is foreseen to enable MU-MIMO precoding with CSIT on the downlink for slowly moving users.

For the reference design the BS is equipped with 4 cross-polarized antenna elements, giving $n_T = 8$ transmit antennas, while the UTs exhibits $n_R = 2$ receive antennas in the form of one cross-polarized antenna element [WIN2D61311].

7.4.2.1 Downlink

For the reference design the BS is equipped with $n_T = 8$ transmit antennas, and up to 4 spatial streams per chunk can be simultaneously transmitted. In MA downlinks a number of MIMO schemes need to be supported, see or [WIN2D61311]:

- MU MIMO precoding with short term CSI at the transmitter (CSIT) with up to 4 spatial streams per chunk. Targets slowly moving users (velocity ≤ 10 km/h) with high median SNR.
    - *In-band dedicated downlink pilots*: beamforming is applied to pilots in the same way as to payload data. If the beams are sufficiently spatially separated, i.e. the cross-talk between beams is below a certain threshold, an orthogonal pilot set can be reused.
    - *Uplink pilots for CSI transfer*: In case a user transmits and receives data on both uplink and downlink, the necessary CSI for spatial precoding at the BS is readily available by the in-band uplink pilots. Unfortunately, users that receive on the downlink typically do not transmit data on the uplink. In this case, the UTs insert 2 pilots per UT antenna on the last OFDM symbol of uplink slots, on those chunks where this UT is receiving data on the downlink. These pilots facilitate transfer of CSI to the BS transmitter and are transmitted unweighted.
    - *Uplink super-frame pilot preamble*: the preamble contains unweighted dedicated pilots per antenna over the full band (DPA-FB). With a pilot spacing of $D_f = 8$ up to $N_u = 8$ users can share the preamble. In case, $N_u > 8$ competition bands are introduced, which means that a certain user may access only a fraction of the available bandwidth. Therefore, DPA-FB pilots only need to be inserted in the corresponding competition band.
- Spatial multiplexing (SMUX), e.g. with per antenna rate control (PARC), with up to 2 spatial streams per chunk. Targets users with medium velocity (less than 50 km/h) with high SNR. No spatial transmit processing, i.e. the payload data is transmitted unweighted.
    - *In-band downlink pilots*: Like the payload data, pilots are transmitted unweighted. On these chunks the pilot grid in Figure A-10 is then transformed to common pilots per antenna with a regular pilot spacing of $D_f = 4$ and $D_t = 12$ in frequency and time. For each spatial stream one orthogonal pilot set is used.
    - *CQI feedback*: The UTs can generate CQI estimates on all chunks where unweighted downlink pilots are transmitted. These CQI estimates are compressed as described in [WIN1D210] and transmitted to the BS. This enables frequency-adaptive transmission on both uplink and downlink, due to the TDD channel reciprocity.
    - *Uplink super-frame pilot preamble*: Due to the high velocities of users associated to SMUX, no uplink pilots are inserted in the preamble.
- Linear dispersion codes by two antenna transmit diversity, the well known Alamouti scheme, applicable to low SNR and/or high velocities.
    - *In-band dedicated downlink pilots*: Pilots are transmitted unweighted, and for each active transmit antenna one orthogonal pilot set is used. In analogy to SMUX the pilot grid in Figure A-10 is then transformed to common pilots per antenna with a regular pilot spacing of $D_f = 4$ and $D_t = 12$ in frequency and time.
    - *Uplink super-frame pilot preamble*: No uplink pilots are inserted in the preamble for users associated to LDC.

The MA downlink is characterised by a variety of spatial schemes. The fact that all these spatial schemes are to be flexibly combined imposes great challenges for the reference design. Figure 7-4 illustrates how the reference pilot design facilitates the coexistance of all considered flavours of spatial processing.

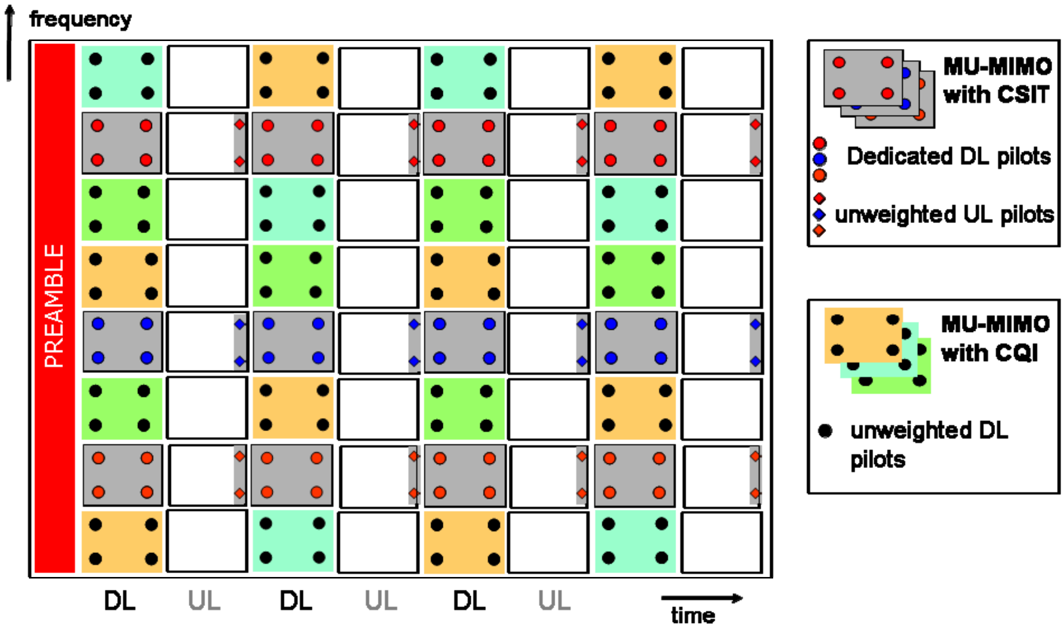

Figure 7-4: Pilot design to enable spatial processing in metropolitan area.

Figure 7-4 illustrates the chunk allocation and the coexistence of the considered spatial schemes. The chunk allocation works as follows:

- Downlink users scheduled for MU-MIMO precoding with CSIT transmit uplink pilots in the super-frame preamble.
- The BS selects up to 4 users that transmit simultaneously in one chunk. The most appropriate chunks for these users are assigned, and the spatial precoding matrix is computed.
  ➔ The chunk allocation for MU-MIMO persists until the next super-frame pilot preamble is transmitted. This is reasonable since the low mobility suggests that the channel conditions do not significantly change during one super-frame. More importantly, now the user specific uplink pilots for CSI transfer only need to be transmitted on the allocated downlink chunks, and not over the full competition band. *This is a key requisite to keep the pilot overhead for CSI transfer on the uplink at an acceptable level*, while maintaining a regular update of the spatial precoding matrix on a frame-by-frame basis.
- The remaining chunks are then assigned to spatial processing without CSIT, but with CQI at the transmitter (short-term or long-term).
  ➔ Since these users transmit unweighted downlink pilots, UTs can measure the CQI on all those remaining chunks that are not reserved for MU-MIMO precoding with CSIT. In other words, unbiased CQI is available at the UTs as no user-specific spatial precoding is applied to the pilots, which is a key requirement for multi-user scheduling based on short-term CQI on a frame-by-frame basis.
- CQI is reported to the BS on the uplink as encoded data packets, by DCT (discrete cosine transform) based data compression [WIN1D24], [WIN1D210].
  ➔ This is a key enabler for frequency-adaptive transmission up to mobile velocities. We note that no uplink pilots for CSI transfer are needed for high velocity users, avoiding prohibitive overheads for feedback on the uplink.

To conclude, flexible operation and coexistence of various flavours of spatial processing is established on the downlink, the overall burden on the uplink for feedback of CQI and CSI is kept remarkably low.

7.4.2.2 Uplink

In MA uplink dedicated pilots per antenna (DPA) are used. For OFDMA, the pilot pattern specified in Figure A-10 is used, while for B-IFDMA the pattern shown in Figure A-11 is used. The reference design for downlink and uplink is designed such that the spatial processing schemes chosen for downlink and uplink have the same requirements in terms of signalling. We therefore benefit from the same pilot design for both downlink and uplink reference design [WIN2D341].

As the MU-MIMO precoding schemes with CSIT on the downlink require uplink pilots to update the spatial precoding matrix, the last OFDM symbol is reserved for pilots of those users which require short-term CSI at the BS on that chunk. As this applies only a subset of users, one bit of signalling per chunk per super-frame is necessary, so to inform the UT whether the last OFDM symbol of a particular chunk is reserved for CSI feedback or not.

7.4.3 Pilot design for isolated cell indoor deployment with distributed antenna arrays

Indoor deployment in local area (LA) utilizes the TDD mode [WIN2D61312]. Hence, the pilot grids for the TDD mode apply, specified in Annex A.3.2.3 for the downlink and Annex A.3.2.4 for the uplink. Furthermore an uplink super-frame pilot preamble (see Figure A-12) as described in Annex A.3.3 is foreseen to enable MU-MIMO precoding with CSIT on the downlink. As in indoor deployment all users can be considered slowly moving, the coherence time of all users' channels are expected to exceed the duration of a super-frame. On the other hand, a distributed antenna array with up to n_T=32 BS antennas, together with the most advanced MIMO schemes that are able to serve up to 30 users simultaneously on one chunk, are considered [WIN2D341], [WIN2D61312]. To this end, one pilot in time direction may suffice, i.e. the pilots corresponding to pilot index $\ell_p = 2$ in Figure A-10 may be omitted, which cuts the pilot overhead by a factor of 2.

7.4.3.1 Downlink

Unlike the MA reference design, the spatial processing in LA deployment is always MU-MIMO precoding with CSIT. The challenge for LA is the potentially very large number of simultaneously served users per chunk. With the 4 orthogonal pilot sets, the spatial reuse of pilots in LA can be up to 8. Since, distributed antenna arrays are considered which decorrelate users and alleviate the near far problem, even such a high spatial reuse of pilots is feasible.

The general rules for the pilot design follow the MA section 7.4.2. The uplink pilot super-frame preamble provides the BS with the necessary CSI of all active users, which is used for chunk allocation and user grouping. Due to the low velocities and the resulting long channel coherence times, users may not need to send pilots on each super-frame preamble. Together with the provision of competition bands, this keeps the pilot overhead for signalling within acceptable limits. Likewise the spatial precoding matrix may not necessarily be updated every frame, so the feedback rate of pilots for CSI transfer on the uplink may also be reduced.

7.4.3.2 Uplink

The pilot design for the uplink closely follows the downlink case described above. Instead of downlink dedicated pilots per beam, dedicated pilots per antenna are used in the uplink. Similar to the downlink, the super-frame preamble provides the BS with the CSI of the uplink users. After chunk allocation and user grouping based on the super-frame preamble are carried out, the in-band uplink pilots provide the short-term CSI for the spatial receiver processing at the BS.

7.5 Synchronization

7.5.1 Link Level Synchronisation

7.5.1.1 Preamble Description

Coarse and fine time and frequency synchronisation algorithms utilize first OFDM symbol of *DL Synch* time slot, named *T-Pilot* whose time domain structure is illustrated in Figure 7-5.

CP	c(0)A	c(1)A	c(2)A	c(3)A	c(4)A	c(5)A	c(6)A	c(7)A

Figure 7-5: Time domain structure of T-Pilot design.

The samples $c(m)$ modify the signs of the A symbols, and their vector form is defined as follows

$$\mathbf{c} = [c(0), c(1), c(2), c(3), c(4), c(5), c(6), c(7)] = [-1, 1, 1, 1, 1, 1, 1, 1]$$

The *T-Pilot* consists of eight equal subsymbols A, each of length $0.125 N_{\mathrm{FFT}}$, where A is generated from the IFFT of a Gold sequence. To ease timing synchronisation the sign of the first subsymbol is negated.

In order to create above such time domain structure, pilot tones should be transmitted on every eighth subcarrier, named here active subcarriers, among all used subcarrier. Moreover some active subcarriers are excluded to ease integer frequency offset estimation.

The preamble *A* should not only have good correlation properties but also low Power-to-Average-Power Ratio (PAPR). The lowest PAPR of the preamble was achieved using Gold sequence of degree 9 and shift registers which states were initiated by 247_{oct} and 503_{oct}. Indices of excluded subcarriers were found using the PAPR reduction algorithm [SS06]. The algorithm finds pilot tones that have the greatest influence on the PAPR. Those pilots are excluded from the pattern and not predistorted as in the algorithm from [SS06]. The following set of excluded pilot tones indices minimising PAPR was found

$$40, 384, 392, 408, 568, 1480, 1640, 1664, 2008 \tag{7.1}$$

for indoor and micro-cellular scenarios, and

$$88, 96, 264, 464, 488, 1560, 1584, 1784, 1952, 1960 \tag{7.2}$$

for urban scenario. The resulting PAPR of the preamble becomes 5.98 dB.

7.5.1.2 Algorithm Description

The coarse time synchronisation is performed during the first half of the *T-Pilot*. During that time also fractional frequency offset is estimated. After detecting and receiving the transmitted symbol the integer frequency offset and fine timing synchronisation are performed.

In order to perform the coarse timing synchronisation only four out of eight *A* symbols are used. The remaining symbols are used for fine frequency synchronisation in time domain and fine timing synchronisation in frequency domain. Timing metric used to perform coarse timing synchronization uses an averaged value of three cross-correlations computed between four consecutive sample blocks of length $0.125N_{FFT}$. Maximum value of such timing metric is detected, completing coarse time synchronization process.

To estimate the frequency offset a similar procedure as to perform coarse timing estimation is employed. This time the argument of the correlation between two subsequent pilot symbols determines the frequency offset.

To perform integer frequency offset estimation it is possible to exploit a specific subcarrier pattern in the frequency domain created by exclusion of some extra subcarriers in the OFDM pilot symbol, as described in Section 7.5.1.1. Knowing this pattern the receiver is able to correlate the instantaneous powers of the received signal samples in the frequency domain with the known pattern and select that set of bins of the FFT demodulator which minimizes a specific cost function [DW02].

The fine symbol timing synchronisation is realized by an energy detection (ED) algorithm using pilot symbols multiplexed into OFDM symbol on certain subcarriers [YLCC02].

Exact procedures for coarse timing synchronization, fractional frequency offset estimation, integer frequency offset and fine timing synchronization are described in [WIN2D233].

7.5.2 Network synchronisation

Network synchronisation is defined as aligning all internal time references within the network, so that all nodes agree on the start and end of a super-frame. To do so in a self-organised manner, similar rules to the ones used in nature by fireflies are applied: each node maintains a time reference, referred to as the phase function, which is updated upon reception of a pulse from other nodes. The update rule is extremely important, and it was shown by Mirollo and Strogatz in their seminal paper [MS90] that the phase increment, which is a function of the internal phase, needs to be always strictly positive.

The issue of frequency synchronisation from a network perspective is not addressed here. The running oscillator of each base station is considered sufficiently precise, so that there is no need for inter-cell frequency synchronisation. Thus, under this assumption, only UT frequency synchronisation is needed, and is performed by the link-level synchronisation unit.

In this section, we first review the super-frame preamble structure that is currently defined and used for synchronisation. In Section 7.5.2.2 a self-organised network synchronisation scheme derived from the firefly synchronisation principle is extended to fit into the super-frame preamble. Thanks to the coarse misalignment mode, cells that are initially completely misaligned are always able to synchronise.

7.5.2.1 Preamble Structure and Constraints

Within the first phase of the WINNER project, network synchronisation was integrated into the system. The most noticeable contribution is the inclusion of two synchronisation words within the preamble of the WINNER super-frame. The super-frame preamble of the WINNER system consists of five consecutive time slots, which are shown in Figure 7-6.

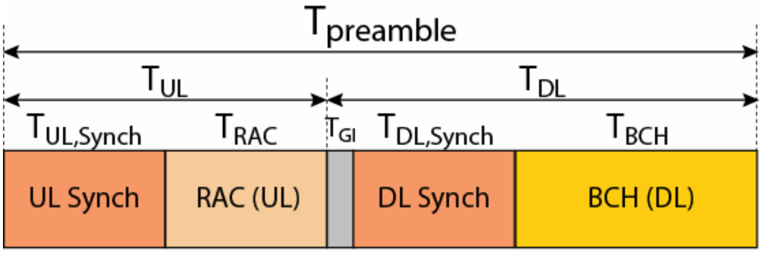

Figure 7-6: WINNER Super-frame Preamble.

The sum of these time slots corresponds to the length of the preamble:

$$T_{\text{preamble}} \quad T_{\text{UL,Synch}} \; T_{\text{RAC}} \; T_{\text{GI}} \; T_{\text{DL,Synch}} \; T_{\text{BCH}}$$

The network synchronisation procedure uses the two synchronisation words as follows. The first sync word, labelled “UL Synch”, is used by UTs to transmit a synchronisation word that is received by BSs to adjust their internal clock and synchronise. In a similar fashion, the second sync word, labelled “DL Synch”, is used by BSs to broadcast a synchronisation word that is used by UTs to synchronise. The necessity for having two distinct transmission slots is justified by the fact that nodes cannot receive and transmit simultaneously. Hence one group relies on the other to synchronise.

To reproduce the behaviour based on pulses as in the Mirollo and Strogatz scheme, the network synchronisation unit is placed close to the link synchronisation unit. The link synchronisation unit uses the UL and DL synchronisation words to estimate the start of a super-frame. This information is then used by the network synchronisation unit to identify the start of transmission from other nodes.

7.5.2.2 Synchronisation Rules and Process

Initially when a UT accesses the network, it needs to synchronise with its base station by following its timing reference, so that it does not disturb ongoing transmissions. This Master-Slave type of synchronisation is fine for intra-cell synchronisation, and is currently deployed in GSM and UMTS networks.

The problem of inter-cell synchronisation consists of aligning the timing reference of all base stations. As cooperation for network synchronisation among all network operators is not mandatory [WIN1D71], a self-organised synchronisation process that relies on UTs and BSs is presented.

Given the super-frame structure, two groups need to form: one formed by BSs and the other by UTs. To enforce this formation, the phase function of BSs is adjusted when detecting a transmission from UTs, and vice versa. Hence two distinct synchronisation sequences “UL Synch” and “DL Synch” are used. Figure 7-7 shows the two state machines defined for BSs and UTs as well as the super-frame preamble structure, when nodes are synchronised.

Based on the two state machines of Figure 7-7, interactions occur between the two groups (BSs and UTs) when a node transmits and nodes from the other group detect this transmission. Detection of the distinct synchronisation words is done by the link level synchronisation procedures. This allows for robust detection and avoids too much additional processing at the receiver.

Separation of nodes into two predefined groups is done in two parts:

- Coupling at Base Stations: if at instant τ_j, a BS node *i* is in ’Listen’ state, where its phase function ϕ_i linearly increments over time, and a UT node *j*, which can communicate with *i*, started transmitting $T_{\text{UL,Synch}} + T_{\text{DL,dec}}$ before, then the receiving BS node *i* increments its current phase ϕ_i :

$$\phi_i(\tau_j) \rightarrow \phi_i(\tau_j) + \Delta\phi_{\text{BS}}(\phi_i(\tau_j)) \text{ where } \phi + \Delta\phi_{\text{BS}}(\phi) = \alpha_{\text{BS}} \cdot \phi + \beta_{\text{BS}}$$

- Coupling at User Terminals: if at instant τ_i, a UT node *j* is in 'Listen' state, where its phase function ϕ_j linearly increments over time, and a BS node *i*, which can communicate with *j*, started transmitting $T_{\mathrm{DL,Synch}} + T_{\mathrm{UL,dec}}$ before, then the receiving UT node *j* increments its current phase ϕ_j:

$\phi_j(\tau_i) \to \phi_j(\tau_i) + \Delta\phi_{\mathrm{UT}}(\phi_j(\tau_i))$ where $\phi + \Delta\phi_{\mathrm{UT}}(\phi) = \alpha_{\mathrm{UT}} \cdot \phi + \beta_{\mathrm{UT}}$

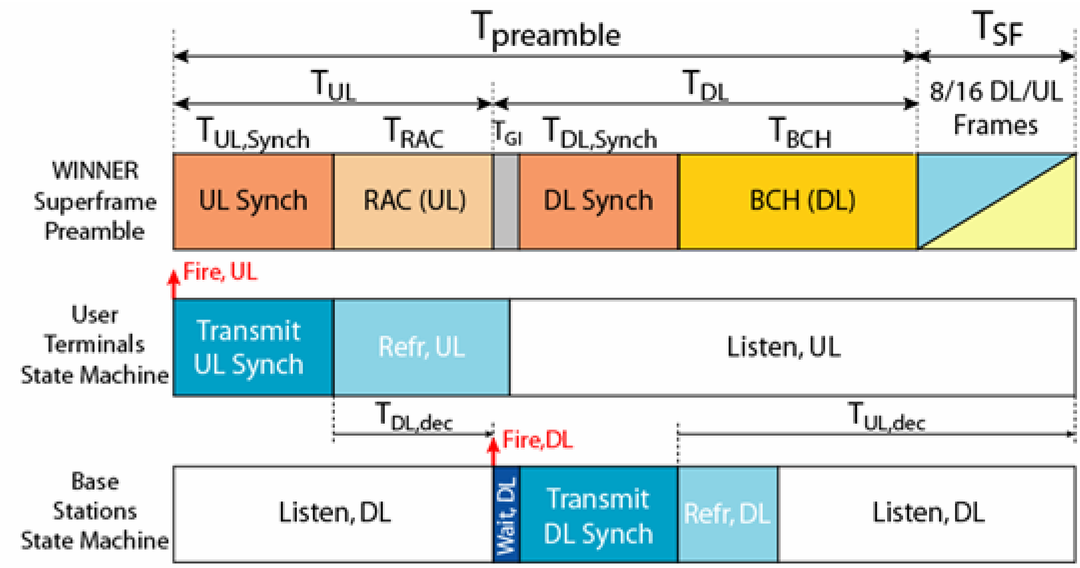

Figure 7-7: State Machines of Network Synchronisation units for coarse misalignments.

Thanks to this strategy, the formation of two groups is controlled: starting from a random initial state, where all nodes fire randomly, after following the simple coupling rules, UTs and BSs separate over time into two groups, all BSs firing T_{UL} after UTs and all UTs firing T_{DL} after BSs. This state corresponds to a synchronised state.

Regarding scalability, the time needed by the algorithm to synchronise starting from any misalignment grows rapidly as the size of the network increases. To reduce the convergence time and to prevent instabilities in the network it is necessary to impose a global reference onto the network synchronisation algorithm. This can be done if a few nodes in the system have access to a Primary Reference Clock, which can be obtained for example from the Global Positioning System (GPS). These reference nodes have a different behaviour from the algorithm described in Figure 7-7, and a complete description of their behaviour is given in [WIN2D233].

7.5.2.3 Compensating Propagation Delays: Timing Alignment

In the Wide Area scenario, given the large propagation delays, the main concern for the network synchronisation scheme is the achieved accuracy. More precisely the misalignment in time between neighbouring base stations cannot be larger than the guard interval duration of an OFDM symbol, which is equal to $3.2\mu s$ in the Wide Area case.

A common procedure for compensating the propagation delay is for terminals to advance their transmission by the propagation delay. Thus the firing instant of user terminals $\tau_{\mathrm{UT},i}$ is advanced by the propagation delay with its own base station $\theta_{\mathrm{UT},i,\mathrm{BS(i)}}$, so that uplink transmissions are effectively performed according to the new timing reference instant:

$$\tau_{\mathrm{UT},i} \to \tau_{\mathrm{UT},i} - \theta_{\mathrm{UT},i,BS(i)} .$$

Thanks to this re-alignment, it was shown through simulations, that base stations synchronise with a maximum misalignment of $1\mu s$. Thus the achieved accuracy fulfils the requirement of being smaller than the OFDM guard interval.

7.6 Channel estimation

In the WINNER system, channel estimation (CE) is aided by the transmission of pilots, frequency-multiplexed with data. Pilots create extra overhead, and it is therefore important to use channel estimation techniques which make most efficient use of pilots. Channel estimation is essential for satisfactory performance of equalizers, smart antennas, multi-user scheduling, transmitter array processing and precoding, optimum or near-optimum detection and decoding, and signal quality estimation. Channel estimation for generalized multi-carrier signals (GMC) produces explicit estimates of the frequency response (channel transfer function) between transmitters (with or without spatial processing) and receivers. These estimates are then used in the algorithms that perform the abovementioned functions. In some cases, channel estimation is implicit, producing estimates of receiver parameters indirectly; an example of this is the least squares adaptation for suppressing low level out of cell interference.

The channel estimation (CE) reference design utilizes a scattered pilot grid as summarized in Section 4.8.4 and detailed in Annex A.3. Interpolation between pilots over time and frequency by pilot aided channel estimation (PACE) provides initial channel estimates for the entire frame. Interpolation over time and frequency is separated and realized by two one dimensional FIR filters, referred to as 2x1D PACE. As demonstrated in [WIN1D21], [WIN1D23], [WIN1D210], 2x1D PACE experiences only marginal performance degradation with respect to optimal 2D PACE (i.e. PACE implemented by a two dimensional FIR filter), but significantly reduces the computational cost.

The FIR interpolation filters are implemented by a Wiener interpolation filter (WIF) with model mismatch. The filter coefficients of a WIF with model mismatch are generated with the following prior knowledge about channel statistics: the maximum *delay of the channel* τ_{max}, *Doppler frequency* $f_{D,max}$ and *average SINR* are assumed to be known; however, no further knowledge of the 2nd order statistics (i.e. the channel covariance matrices in time and frequency) are assumed. If the required measurements of τ_{max}, $f_{D,max}$ and average SINR are unavailable at the receiver, the *worst case* design of the WIF is adopted: the maximum delay of the channel is set equal to the CP-length, the maximum expected velocity is set with respect to a certain deployment scenario (LA: 3km/h, MA: 70km/h, WA: 250km/h), and the highest expected SINR set to 30 dB. We note that the worst case design of the WIF will have significantly poorer performance, so it is recommended to implement and utilize means to measure τ_{max} and $f_{D,max}$.

As the WINNER pilot design in Section 4.8.4 and Annex A.3 requires in some cases dedicated pilots as well as allows for spatial reuse of pilot symbols, the attainable performance of PACE may be insufficient to meet the ambitious targets of the WINNER system. Two enhancements for the CE reference design are foreseen:

- **Channel estimation over multiple frames** exploits the correlation in time, in the way that pilot symbols from previous symbols provide significantly improved channel estimates. A Kalman filter is an efficient means to exploit the correlation of pilots over multiple frames [WIN2D233].
- **Iterative channel estimation (ICE):** can be efficiently implemented in terminals that contain turbo receiver consisting of an inner and outer receiver that exchange soft information in the form of log-likelihood ratios (LLRs) [WIN1D210], [WIN2D233]. In this case the channel estimation unit is included in the turbo loop and additional computational complexity is acceptable. PACE estimates are used as initial estimates for ICE and feedback needed for ICE is derived from *a posteriori* information, as shown in Figure 7-8. Application of ICE leads to significant performance improvements over PACE.

Although channel estimation over multiple frames and ICE are both optional extensions of PACE, for the reference CE preferably both extensions should be implemented, so to achieve the ambitious spectral efficiency targets of the WINNER system.

In order to improve performance in a multi-user MIMO scenario where channel estimation is more challenging, ICE can be extended by genetic algorithm (GA) aided joint ICE and multi-user detector (MUD) at the expense of additional computational complexity [WIN2D233].

For DFT-precoded OFDM decision-directed iterative technique, which suppress in-cell and out of cell interference with the aid of least squares adaptation is applied. To adaptively suppress out-of-cell interference, without explicitly estimating it, a least-squares processing over several successive FFT blocks can be applied. In order that least-squares processing functions properly, it is required that the channels do not change much over observed blocks [WIN2D233].

PACE, iterative techniques for OFDM which supplement PACE; and decision-directed iterative techniques for DFT-precoded OFDM, which suppress in-cell and out of cell interference with the aid of least squares adaptation are described in [WIN2D233].

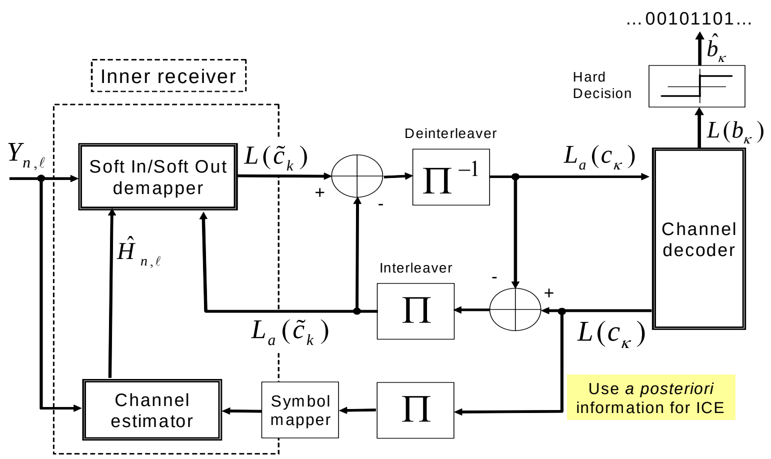

Figure 7-8: Iterative receiver structure. *A posteriori* information is used for ICE

7.7 MIMO cooperative relaying

Cooperative relaying is typically seen as a macro diversity transmission scheme, where the base station and the relay node are transmitting (or receiving in uplink) simultaneously the same information and the user terminals receiving both of these transmissions are diversity combining these signals. Cooperative relaying methods have been extended to distributed MIMO systems in WINNER as described in [WIN2D352] and [WIN2D61311], where extensive performance assessment results are also shown. With multi-antenna transceivers the spatial transmission schemes described in sections 4.8 and 7.3.3 are utilised also with cooperative relaying.

The performance assessment results of the distributed MIMO schemes in [WIN2D353] and [WIN2D61311] do not include a quantitative overhead analysis or a consideration of imperfections such as imperfect channel state information. Hence, the results shown there rather serve as upper limit and for a qualitative evaluation. The following overhead categories can be identified (exemplified using the DL):

- Before BS and RN can cooperatively transmit, the BS needs to transmit the data for the cooperative transmission to the RN. This overhead has been taken into account in the cooperative relaying concept and its' performance assessment, and it is the same overhead as in a single-path relaying system with the difference that the data must be transmitted twice by the BS whereas in a single-path relaying protocol the BS is only forced to transmit it once. On the other hand one can assume a BS-RN link to be of high quality and hence only few resources are necessary for the additional transmission. Furthermore, the cooperative transmission of BS and RN offers significant improvements in terms of intra-cell interference management especially in those areas where BS-UT and RN-UT are of similar quality where in single-path relaying the performance is anyway limited due to the BS interference.

- In a cooperative transmission the RN must additionally transmit the CSI for its RN-UT links to the BS as well as vice versa such that a distributed precoding can be performed. The additional CSI which needs to be exchanged is increased by a factor of $A*(M_{BS} + M_{RN})*N_{UT}$. The pre-factor A highly depends on the quantization quality and technique of the CSI as well as on the robustness of the precoding. Furthermore, as distributed MIMO is only applied to slowly moving users we might assume a high coherence time and bandwidth which further decreases the necessary overhead.

- Finally, we must consider the additional distributed precoding operation which has to be performed at the RN and BS. As this step is a standard procedure used in usual MIMO system, this overhead can be neglected.

Obviously, only the second bulletpoint represents serious additional overhead which has not been considered in the performance assessment but should be included while comparing single-path and cooperative relaying. Nonetheless, the question for a quantitative measure of this overhead highly depends on the quantization quality as well as time and frequency coherence of the channel and can not be answered in general.

Performance-wise the main impact of cooperative relaying, shown in [WIN2D353] and [WIN2D61311], is that it significantly improves the 5%-ile of the user throughput (the target level for the Satisfied User Criterion, used to measure the cell-edge user performance in WINNER project), in addition to improving significantly the average cell throughput, when compared to ordinary single-path relaying. Therefore, cooperative relaying can serve as an important tool for WINNER systems, especially for boosting up the performance for cell edge users.

7.8 Interference mitigation

Inter-cell interference mitigation is directly coupled with achievable re-use of the available spectrum. The ultimate goal is to achieve high system capacity using an optimal effective frequency re-use factor. Three main types of inter-cell interference mitigation techniques were investigated in the WINNER II framework as listed below:

- Inter-cell interference averaging [WIN2D471]
- Inter-cell interference avoidance [WIN2D472]
- Inter-cell interference mitigation using smart antennas [WIN2D473]

Interference mitigation by resource partitioning for relay enhanced cells is addressed in [WIN2D351], [WIN2D352] and [WIN2D353].

7.8.1 Without spatial processing – resource partitioning

In general inter-cell-interference avoidance techniques are based on restrictions of resources available for allocation towards the UTs. They usually apply in a coordinated way between cells so that inter-cell interference avoidance is often mentioned together with inter-cell interference co-ordination. Resource partitioning is based on restrictions on the transmit power in pre-defined parts of the spectrum.

From a cell point of view, Flexible Re-use Partitioning (FRP) is a method which implements different re-use zones per cell i.e. which divides the cell coverage area into concentric zones with different frequency re-use factors. The simplest yet promising form of implementing FRP is to divide the cell into inner (re-use 1) and outer (e.g. re-use 3) zones.

The performance of resource partitioning schemes is tightly coupled to the used scheduling policy. In [WIN2D472] scheduling policies are considered that prioritise the cell-edge UTs into sub-bands (or cell-edge sub-bands) that are protected from inter-cell interference. UTs are classified as “cell-edge” if their SINR averaged over the whole allowed bandwidth is below a given threshold.

Unlike BSs, relay nodes should be cost efficient, they are not equipped with as many antennas, and resource partitioning has been identified as an effective technique to coordinate the interference in relay enhanced cells [WIN2D352], [WIN2D353].

7.8.1.1 Fractional frequency reuse

Fractional Frequency Re-use (FFR) typically involves a sub-band commonly used by all cells (i.e. with a frequency re-use of 1) that is exempt from any transmit power restriction, while the allocation of the remaining sub-bands is coordinated among the neighbouring cells, so to create for each cell one sub-band with a lower inter-cell interference level. FFR means that RAPs and UTs are not allowed to transmit on the cell-edge sub-band assigned to adjacent cells, which suppresses inter-cell interference from surrounding cells within the cell-edge sub-bands, at the expense of a reduced amount of available resources.

7.8.1.2 Soft frequency reuse

In the case of Soft Frequency Reuse (SFR) power masks are assigned to each RAP. For SFR all cells have access to the entire band, however, the transmit powers on specific sub-bands is reduced. At the cell-edge, UTs are assigned the sub-band with reduced power of their strongest interfering base station. This scheme allows keeping the inter-cell interference at a low level, while still using all available frequency resources for the transmission.

7.8.1.3 Resource partitioning in relay enhanced cells

7.8.1.3.1 Static load based resource partitioning

A static load-based resource partitioning scheme has been presented and assessed in [WIN2D352], [WIN2D353] and [WIN2D61310]. The target of the proposed scheme is to provide a low-cost solution with low hardware requirements put on the relay nodes. The envisaged deployment is therefore characterized by (a) single-antenna (omnidirectional) Relay Nodes and (b) single-transceiver Relay Nodes (Relaying/Forwarding is performed in the time domain).

Simulation results in the WA CG scenario [WIN2D61310] indicate that RECs that utilize static load-based resource partitioning can sustain up to 20% higher downlink traffic and substantially improved UL cell coverage compared to a BS only deployment.

The basic concept of static load-based resource partitioning applies a fragmentation of the resources in the frequency domain. The resources are assigned to so-called groups of RAPs, where - in the extreme case - each RAP node belongs to a distinct group while – in the other extreme – all RAPs can belong to one group. The groups are used by the partitioning algorithm for intracell frequency planning – RAPs belonging to the same group may reuse the same resources. In order to exploit the resources within a REC as efficiently as possible, the static resource partitioning tries to identify RAPs within the REC that are sufficiently well separated from each other to enable re-using the same resources. In the case of centralized resource partitioning, the groups may also be used for intercell frequency planning.

The optimal fragmentation of resources within the REC also highly depends on the distribution of the users, i.e. of the offered traffic within the REC. Therefore, a dynamic algorithm has to be able to take this distribution into account when assigning resources to the different groups of RAPs within one REC.

7.8.1.3.2 Soft resource partitioning

A resource partitioning scheme based on SFR has been studied in the MA CG scenario [WIN2D352] and [WIN2D353]. By utilizing soft resource partitioning the throughput in a relay enhanced cell could be increased by 28% compared to a BS only scenario.

Phased Scheduler

In order to gain from soft frequency reuse, a suitable scheduling scheme is required. Therefore, we developed the phased scheduler, a new scheduling algorithm for soft frequency reuse. It aims to increase the throughput of low throughput users while keeping an at least similar cell throughput than the proportional fair scheduling algorithm. The scheduling algorithm exploits SINR information of each user in the different power mask areas. It uses this information to schedule the lowest SINR users in power mask areas where they have their highest SINR. A detailed description of the scheduling algorithm as well as a detailed discussion of the simulation results can be found in annex B2.2.2 of [WIN2D352].

Adaptive soft frequency reuse

A REC can have multiple RNs, each providing different traffic load to the network due to cell size and user density variations. Therefore a static soft-frequency reuse pattern will not make efficient use of the available radio resources. Thus, a local power mask adaptation is desirable and a mechanism for adapting the power mask based on the traffic load situation within the REC is presented in [WIN2D351]. The simulation results illustrate that the power mask adaptation improves the throughput especially for low throughput users.

7.8.2 Interference mitigation with spatial processing

7.8.2.1 Interference mitigation by transmit beamforming with a Grid of Fixed Beams

Using fixed beams at the base stations of a cellular system is a relatively simple but yet efficient spatial processing technique [WIN1D27], [WIN2D341] and is the reference design for macro-cell deployments, see section 7.3.3.2. Beamforming with fixed beams exploits small angular spreads, which makes it possible to focus a large portion of the transmitted energy to a restricted angular zone. Small angular spreads are typically observed in wide-area channels (assuming BS antennas above rooftop). The obvious advantages are enhanced SNR with reduced intercell-interference. Moreover, multiple users/terminals can be served simultaneously with controllable interference by forming grids of fixed beams (GoB). Clearly, the amount of interference depends on the applied beam pattern, the number of beams used simultaneously, and the number of transmit antennas.

Whereas adaptive beamforming aims at directing one or more beams to preselected user(s), the fixed beam approach works to some extent reversely. Here, the beam or grid of beams (GoB) – to be applied on a certain time-frequency unit, i.e., chunk – is selected first and then the users are assigned to those beams typically based on some fed back channel quality indicator (CQI). The advantage of such a scheme compared to adaptive spatial processing techniques is a higher robustness since misadjustment of the antenna weights due to defective channel state information is excluded.

Interference mitigation in relay enhanced cells

Two strategies for dynamic spatial resource sharing within a REC have been presented and assessed in [WIN2D352] and [WIN2D353]. By utilizing dynamic spatial resource sharing, the peak spectral efficiency can be increased by 25% compared to a BS only deployment. Depending on the number of users in the cell, the spectral efficiency is up to 60% higher than for the BS only deployment.

The basic idea is that the users, instead of being statically grouped according to their positions only, are dynamically assigned to logical beams, i.e. groups of users served by a single RAP and which are unable to share the same resources, according to their mutual spatial correlation. It should be noted that the beams represent groups of users. They can be seen as a dynamic version of the sectors in a RAP. Although particularly fitting when GoB is used, this procedure can be applied to any spatial processing scheme, and no direct relationship exists with the physical transmission beams. The resource partitioning procedure will then assign resources to these groups, allowing spatially compatible ones to share the same chunks. Three main steps can be identified: the creation of the beams, their grouping and the actual resource partitioning. They will be described in the remaining of this section.

The beams creation procedure follows a tree-based approach, which is derived from the one proposed in [FDH05]. In the original algorithm, however, compatible users are grouped together, while in our implementation the outputs are groups of incompatible users. Different beams creation metrics could be used, such as the maximization of the average inter-user correlation within a beam, or the maximization of the minimum inter-user correlation. Furthermore, under reasonable assumptions the correlation can be considered as a function of the angular spread between the users.

In order to select appropriate groups of logical beams that share the same resources, a simplified estimate of the transmission rate of a beam is used. The computation of this term makes use of estimates of the inter-beam interference, i.e. to measure the interference experienced by the users of a beam when a transmission is occurring on another one. The actual value can be calculated in many ways, e.g. through the maximum or average inter-user interference experienced by the users of the beam. It should be noted that the beam rate only represents a long-term, average channel quality estimate for the users of the beam. The actual transmission rate for each user will depend on the short-term scheduling performed by each RAP, which is unknown during the resource partitioning. No time or frequency selectivity is therefore assumed, and the same value is used in the whole bandwidth and within the whole superframe.

The beam rate provides a metric for evaluating the quality of an allocation of concurrent beams on the same time-frequency resources. For example, we propose a simple beam grouping procedure which starts with an empty group, and at each step adds the beam allowing the highest overall throughput increase. It stops when adding another beam would decrease the throughput due to excessive interference.

Furthermore, the beams rate estimation can also be used to evaluate the amount of resources which need to be assigned to each beam in order to achieve a certain data rate. Both the grouping quality and the estimation of the required resources are used as inputs for the resource partitioning algorithm.

Two resource partitioning algorithms are described in [WIN2D352] chunk-by-chunk balancing of the first and second hop (CCB) that focuses on achieving the maximum aggregate cell throughput and iterative independent balancing of the first and second hop (IIB) that tries to achieve a fair allocation between the users.

7.8.2.2 Interference mitigation by spatial processing at the receiver

The use of multi antenna receivers at both base stations and user terminals has been identified to be efficient means to mitigate interference. This allows implementation of receive diversity combining schemes such as Maximum Ratio Combining (MRC) and spatial interference suppression schemes such as Interference Rejection Combining (IRC) [Vau88]. IRC allows to cancel out multi-user interference originating from SDMA in the own cell [WIN2D341], as well as inter-cell interference from other cells [WIN2D473].

7.8.3 Combination of interference mitigation schemes

7.8.3.1 Interference mitigation by resource partitioning

7.8.3.1.1 Hexagonal cell deployment scenario

Among the studied schemes, Fractional Frequency Re-use (FFR) appears the preferred choice for practical implementation. Unfortunately, the significant cell-edge user throughput enhancement is traded off with a degradation in sector throughput. In order to ensure a fair allocation of resources to cell-edge users, FFR is considered to be an integral part of the reference design for single antenna transmitters. Figure 7-9 shows an example for the bandwidth allocation of (FFR).

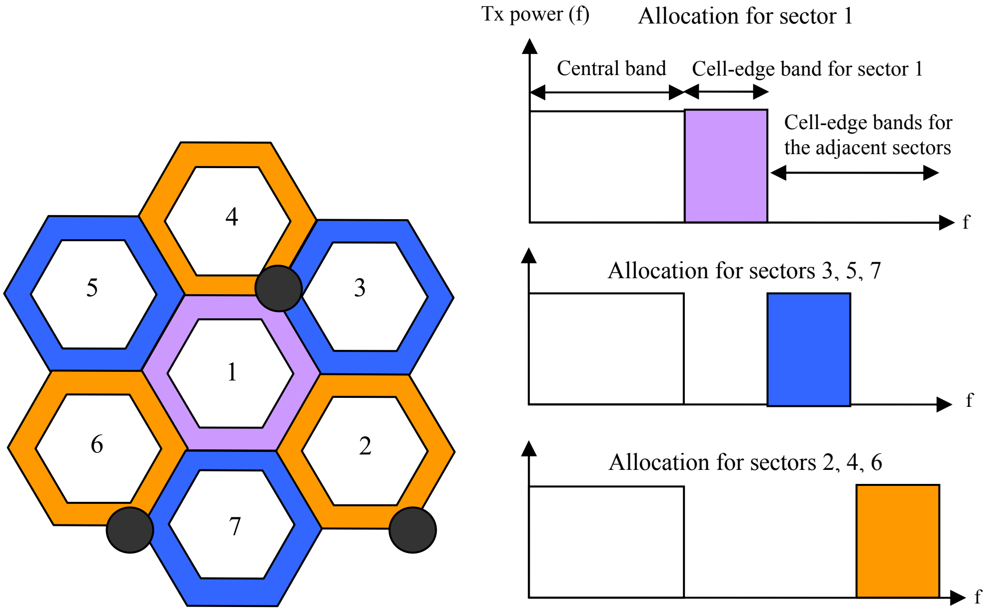

Figure 7-9: Fractional Frequency Re-use (FFR) with reuse 3 zones at the cell edge.[17]

7.8.3.1.2 Manhattan Grid deployment scenario

Resource partitioning with an inner ring with frequency re-use of 1, and an outer ring with frequency reuse of 3 can also be applied to Manhattan grid deployment. The corresponding reuse 3 pattern for Manhattan grid deployment is illustrated in Figure 7-10. With this reuse pattern LoS interference from adjacent BS is completely avoided.

7.8.3.2 Downlink interference mitigation with multiple antennas

7.8.3.2.1 Urban macro cells with hexagonal cell deployment

Figure 7-11 depicts a summary of the studied interference mitigation schemes in wide area. While Figure 7-11 summarizes the results for non-frequency-adaptive transmission, similar results are obtained for frequency-adaptive transmission, so the conclusions and recommendations apply to the FDD mode in general. In the downlink interference mitigation with multiple antennas comprise transmit beamforming at the BSs, as well as receive diversity and interference suppression techniques based on multiple antenna reception in the UTs. It is demonstrated in [WIN2D473] that both transmit beamforming and UT multiple antenna reception are efficient means to mitigate inter-cell interference. Moreover, the combination of transmit beamforming and spatial receiver processing provide additive gains, making both schemes attractive choices for interference mitigation. The gains provided by multiple antenna techniques are generally lower in the frequency-adaptive mode than in frequency non-adaptive mode, but still remain very attractive. The reason is that the interfering beams, and thus the interference environment the cell edge users are particularly sensitive to, change at each frame in an unpredictable manner for the UT. The benefits of frequency adaptivity are consequently reduced, since this mode relies on the channel quality predictability.

[17] Note the geographical representation of the cell-edge bands allocation in the left-hand side figure is symbolic only, since low-SINR UTs are not necessarily at the cell border.

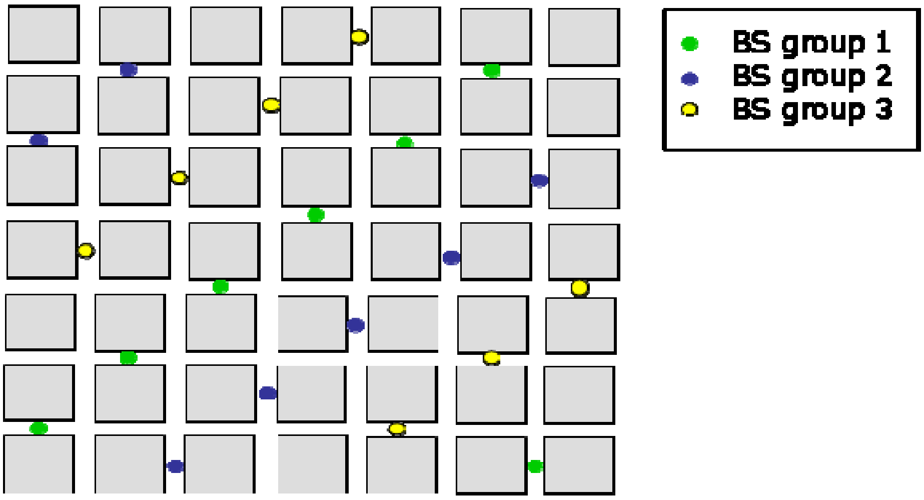

Figure 7-10: Resource Partitioning for the Manhattan Grid deployment scenario.

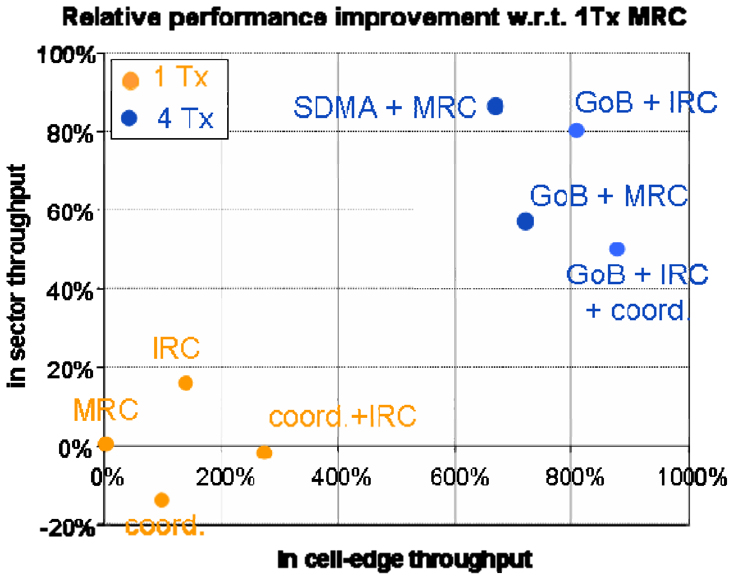

Figure 7-11: Performance improvement for non-frequency-adaptive downlink of various interference mitigation schemes relative to a single antenna transmitter with a 2 Rx antenna MRC receiver.

SDMA on top of the transmit beamforming, as suggested as spatial processing reference design cf. section 7.3.3.2, affects the inter-cell interference mitigation properties. Results presented in [WIN2D473] indicate that an SDMA based grid of beams (GoB) outperforms the single stream GoB in terms of sector throughput; however the cell-edge user throughput of the SDMA based GoB is slightly poorer with respect to the single stream GoB. This difference of performance at cell edge between both techniques can be explained by the fact that the SDMA based GoB is optimised considering mostly the interference situation inside the cell. However, inter-cell interference is more spread by simultaneous transmission in several beams. This degrades the situation at the cell border, due to the partly lost directivity of the interference compared to single stream beamforming. However, the degradation is minor, as seen in Figure 7-11.

Selection criteria for the number of simultaneous beams are provided in [WIN2D341]. Dependent on the SINR it is proposed to switch between SDMA and single stream GoB (or switch between multiple SDMA modes for higher number of antennas at the BS). As the achievable throughput of the SDMA based GoB is upper bounded for increasing SINR, it is recommended that for high SINR there is a switch from SDMA to single stream GoB. The optimum switching points depend on the number of users in the sector, the SINR (including inter-cell/sector interference), the user distribution in the cell, and is elaborated in detail in [WIN2D341].

Generally, the optimal switching behaves as follows: For very small SINR, use $M = n_T$ beams. The higher the SINR, the less beams should be used until for high SINR single stream GoB is best. The SINR switching points (in dB) as a function of the number of symmetrically distributed users in Rayleigh fading are [WIN2D341]:

# users	2	3	4	5	6	7	8	9	10
SINR [dB]	-5	0	3.7	6.75	9.2	11.1	12.7	14.1	15.3

The detailed assessment of the above beam selection scheme on inter-cell interference mitigation is for further study. It is expected however, that beam selection, which essentially will result in cell-center users transmitting with fewer number of beams than cell-edge users, boosts the cell edge performance. The reason is that the interference from cell-centre users becomes more directional, which should have a positive impact on the interference statistics of the cell-edge users. In any case, an SDMA based GoB provides significant performance gains in average sector throughput and is therefore the preferred choice for the reference design, cf. section 7.3.3.2.

The combination of FFR with multiple-antenna interference mitigation is assessed in [WIN2D472]. FFR benefits from increased performances when used in conjunction with IRC. However it hardly improves the cell-edge throughput when used in combination with the GoB, and even degrades for frequency-adaptive transmission. This behaviour can be explained by the fact that in FFR schemes, the interference reduction in the cell-edge bands is achieved at the expense of a transmit power reduction in other resources. The GoB, because of the directivity of the interference, dynamically creates almost interference-exempt resources, when the UT is not hit by an interfering beam. As a consequence, the GoB has very little to gain from sub-bands with lower interference levels (the actual gain is likely to depend on the probability to be interfered by a beam versus the probability to be scheduled in resources with reduced interference, as well as the effective interference reduction in the latters). On the contrary, the reduced power on other sub-bands affects the useful signal power received on these sub-bands, which effectively reduces the efficiency of the GoB compared to frequency re-use 1.

It is important to note that all result presented in [WIN2D472], [WIN2D473] (summarized in Figure 7-11) assumed perfect link adaptation and neglected channel estimation errors. This may affect in particular the interference mitigation schemes using multiple antennas. As common pilots per beam are assumed for FDD downlinks, the pilots potentially observe different interference compared to the beamformed data symbols. As this may severely affect the attainable sector throughput of a frequency reuse 1 system, resource partitioning cannot be completely ruled out for the reference design with multiple transmit antennas, despite the clear indication in favour of reuse 1 for idealized channel knowledge in Figure 7-11.

7.8.3.2.2 Urban micro cells with Manhattan grid deployment

While for wide area a grid of beam (GoB) was shown to be very effective in mitigating inter-cell interference [WIN2D473], for Metropolitan area a MU MIMO precoding scheme, that relies on high median SNR and instantaneous CSI at the transmitter, is instead a cornerstone in the spatial processing solution, as described in section 7.3.3.2 and [WIN2D341]. Hence, inter-cell interference avoidance by resource partitioning is an effective means to ensure sufficiently high SINRs required for multi-stream spatial processing.

The street canyons of the Manhattan grid deployment generate only a small number of dominant interferers. Consequently, spatial receiver processing and in particular interference rejection combining (IRC) as described in [WIN2D473], provides larger gains in MA compared to WA. This enables the UT to cancel out much of the interference through IRC spatial processing, which should particularly support cell-edge users.

7.8.3.3 Uplink interference mitigation with multiple antennas

7.8.3.3.1 Urban macro cells with hexagonal cell deployment

Receive diversity in the form of MRC or IRC at the BS receiver provide significant gains and effectively enabling a frequency reuse 1 deployment. In the uplink, the gains of IRC with respect to MRC are rather modest, due to the high number of interfering sources which tends to whiten the interference. In any case, as IRC is the choice for BS receiver spatial processing it is the preferred interference mitigation scheme for BS receivers with multiple receive antennas. In the uplink more interferers are encountered than in the downlink, which consequently reduces the IRC gain.

7.8.3.3.2 Urban micro cells with Manhattan grid deployment

The street canyons of the Manhattan grid deployment generate only a small number of dominant interferers. Consequently, spatial receiver processing and in particular interference rejection combining (IRC) as described in [WIN2D473], is likely to provide larger gains in MA compared to WA. As IRC is the preferred for uplink spatial processing for a BS equipped with multiple receive antennas, it is the natural choice for interference mitigation on the uplink.

7.9 UT location determination

WINNER provides location based services, where the service is a generalised term that covers radio resource management operations such as handover. Location determination can be obtained both in band and out of band. In [WIN2D482] approaches to obtain Time Difference of Arrival (TDOA) timing information by OFDM synchronization algorithms were presented. Furthermore, the integration of Global Navigation Satellite System (GNSS) based positioning techniques was described for the static case.

7.9.1 TDOA based location determination

Approach is the use of the pre-amble for downlink synchronization to perform timing measurements with at least three BSs. Generally, these measurements are processed afterwards in a hybrid data fusion (HDF) entity to solve the navigation equation and get a mapping from the TDOAs to the UT position in x and y coordinates.

The BSs are organized in a – not necessarily cellular – network with cell radius R according to Figure 7-12. Clearly, also a BS distribution according to Manhattan grid is possible.

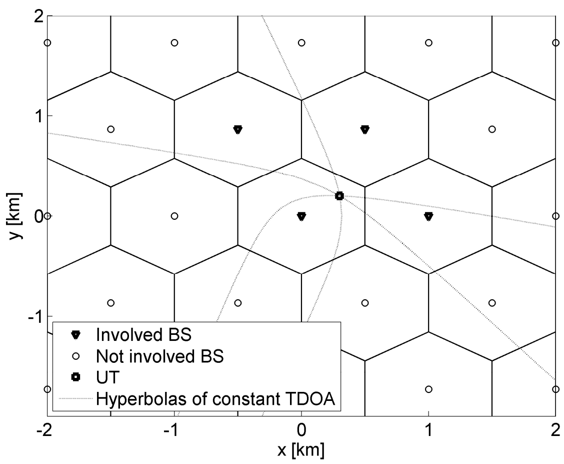

Figure 7-12: TDOA positioning in cellular networks with four involved BSs.

When the network is synchronized e.g. by utilizing the methods presented in Section 7.5.2, the TDOA measurements of the BSs is straight-forward. For non-synchronized BSs location measurement units (LMUs) might be used for estimating the possible time offset of the involved BSs. The LMUs are associated to the BSs and compensate the missing synchronization of the BS network. The UT is located at $\mathbf{x} = [x, y]^{\mathrm{T}}$ and only the N_{BS} nearest BSs at $\mathbf{x}_\mu, \mu \in \{1, 2, \ldots, N_{\mathrm{BS}}\}$ are used for positioning [3GPP04]. The distance between the BSs and the UT is given by

$$r_\mu(\mathbf{x}) = \|\mathbf{x}_\mu - \mathbf{x}\|^2 = \sqrt{(x_\mu - x)^2 + (y_\mu - y)^2}. \tag{7.3}$$

This equation can also be seen as a result of Time-of Arrival (TOA) measurements. With TOA the absolute time for a signal travelling from the BSs to the UT is measured. It is not even required that all BSs are synchronized with each other (directly or by LMUs), synchronized time knowledge, i.e., the time of transmission, is necessary at the UT as well. In case that no exact time knowledge is available (time offset of the UT), an additional BS is necessary to estimate this offset. The propagation time from the BSs to the UT is proportional to the distance. Hence, we get the distance between the UT and all involved BSs. From a geometrical point of view, the UT lies on circles around the BSs. The intersection gives the position of the terminal [WIN2D481].

The problem of processing TOA measurements is the fact, that the UT is usually not synchronized to the BSs in the manner that the time of transmission is not known to the UT. To avoid this problem, with

TDOA the time difference of signals received from various BSs is measured directly, i.e., the unknown time offset of the UT w.r.t. the synchronized BSs is not relevant for TDOA processing. In the geometrical interpretation, the UT lies on hyperbolas with foci at the two related BSs. The intersection gives the position of the UT. Note that TDOAs are defined w.r.t. an arbitrary chosen reference source which in case of the WINNER system should be the serving BS, e.g., the TDOAs for BS $\nu \in \{2,3,\ldots,N_{\mathrm{BS}}\}$ w.r.t. BS 1 can be written as

$$d_{\nu,1}(\mathbf{x}) = r_{\nu}(\mathbf{x}) - r_1(\mathbf{x}), \tag{7.4}$$

where – without loss of generality – we use BS 1 as the reference BS.

Extracting timing measurements from the communications signals directly results in a synchronization problem where normally the arrival time of the signal transmitted from the involved BSs has to be measured at the UT. We make use of the downlink synchronization symbols included in the preamble of the super-frames. Aim of the symbol-timing synchronization is to find the start of these OFDM symbols. This information will afterwards be used to calculate the TDOA measurements. For the timing measurements, same methods as for the link level synchronization, presented in Section 7.5.1, can be used.

Finally, when link-level synchronization has been performed with at least three BSs at the same time the navigation equation has to be solved in the HDF unit, i.e., the position information has to be extracted from the resulting TDOA measurements which yield a non-linear estimation problem.

Performance assessment of the location determination methods with WINNER system parametrisation, and channel models is presented in [WIN2D482]. The results show that a reasonable location accuracy is achievable with these methods in most of the cases and environments. For further improvement we can include tracking algorithms for the solution of the navigation equation in the dynamic case. Usually, the user is moving around a certain track in different scenarios. Clearly, there are certain correlations between the positions of the UT over time. This information can be integrated in the overall position estimation process and can help to improve the estimates in average.

The Kalman filter (KF) is a flexible tool for providing positioning estimates in the context of tracking applications. However, the standard KF just performs optimal if several criterions on, e.g., linearity or Gaussianity, are fulfilled which is usually not the case in practical applications. Nevertheless, even if these conditions are not fulfilled completely, the KF gives reliable and robust estimates.

7.9.2 Out-band location determination using GNSS

Besides system-internal location determination also global navigation satellite system (GNSS) based support from GPS and/or Galileo can be included if appropriate capabilities of the UT are available.

For WINNER-internal location determination time-difference of arrival (TDOA) timing information obtained by OFDM synchronization algorithms was used and assessed [WIN2D482]. This stand-alone location determination should be supported from all UTs to satisfy the WINNER system requirements [WIN2D6114], e.g., for the localization of emergency callers. However, for more advanced location based services which can exploit available location information, e.g., location based handover and location based radio resource management, a higher accuracy is required. Hence, we have combined stand-alone location information with GNSS based positioning techniques. These satellite based systems provide high accuracy and reliable position information under optimum conditions, i.e., in situations where line-of-sight (LOS) access to several satellites is given. However, especially in critical urban scenarios it can happen quite often that only less than the required four GNSS satellites are visible. In these situations the WINNER TDOA measurements can compensate the lack of satellites and improve the overall accuracy. Additionally, for further improvement tracking algorithms based on an extended Kalman filter (EKF) for the solution of the navigation equation in the dynamic case were applied which helps to improve the estimates in average. Further details and performance assessment results on this hybrid locationing scheme can be found from [WIN2D483].

7.10 Flow handling mechanisms

Flow handling mechanisms mainly includes congestion control, load control and admission control. These functions are very tightly interworking together. Flow handling admits, reconfigures flow QoS classes, rejects and triggers other RRM functions to maximize the flow admission rate, minimize the flow rejection rate that turns out to be an optimised system capacity. Some of the flow control mechanisms use explicit RRC message, as explained in the logical node architecture part.

7.10.1 Congestion Control

Congestion situations increase interference, loss of packets, low bandwidth availability and from the user's point of view they cause decreased quality in the service reception which leads to the user's disappointment. In very high congestion situations, the network could collapse if a congestion control scheme is not adopted. The congestion situations can be avoided applying a set of techniques belonging to "Congestion Avoidance Control" network mechanism, which has the important task of controlling the load of the network by restricting the admission of new user sessions and resolving unwanted overload situations. Selected from the overall relationships between Congestion Avoidance Control to other RRM functions, besides network triggered handover, spectrum management, congestion Avoidance Control mechanism relies mainly on two system functions that can be seen in the next figure:

- Admission Control (AC), which is the mechanism that receives the requests for new sessions (whether they come from a new user or from ongoing users) and checks if the users are authenticated to the network and if the network has sufficient resources based on the requested resources by the new session.
- Load Control (LC), which is the mechanism that controls the load of the network by continuously monitoring the load and acting if the load exceeds some pre-defined thresholds. It can be triggered due to false, incomplete, insufficient and imperfect Admission Control design.

These two mechanisms are tightly coupled to and complete each other and their combination into the Congestion Avoidance control network function can lead to excellent network performance control.

As it can be seen in the next figure, there is a clear interaction between the Admission and the Load Control mechanisms. The AC mechanism is the one that admits the users and has as a result a load increase; the Load Control mechanism monitors the load increase and tries to decrease it when it comes above a predefined threshold. Load Control sends messages to AC to reject any (not emergency) user when there is a congestion situation and sends also the message of normal operation when the congestion has been resolved. The AC sends messages to LC to perform some actions when needs to admit a high priority user to the network and gets the responses of these messages and the actions of the Load Control (if it has gained the needed resources to admit the new user).

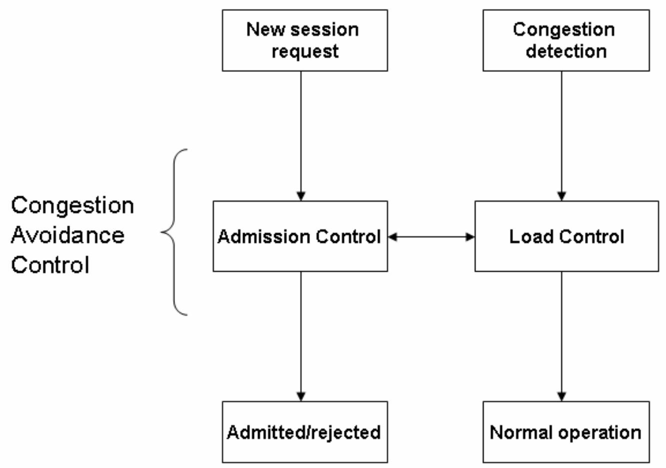

Figure 7-13: Congestion avoidance control.

7.10.2 Load control

Admission Control algorithms are designed to make decisions about the new session requests, in order to maintain the load of the networks under a certain threshold. Though, if for different reasons the load of the network exceeds that threshold, then the network experiences an overload/congestion situation. Network congestion control is a very critical issue and has high priority, especially given the growing size, demand, and speed (bandwidth) of the increasingly integrated services networks. Designing effective congestion control strategies for future wireless networks is known to be difficult because of the complexity of the structure of the networks, nature of the services supported, and the variety of the dynamic parameters involved.

According to the International Telecommunication Union (ITU) definition (ITU-T: Rec. I.371) congestion is defined as a state of network elements in which the network is not able to meet the negotiated network performance objectives for the already established connections and/or for the new connection requests. Congestion control is a set of actions that are taken by the network in order to minimize the intensity, spread, and duration of congestion. It is a protocol of the network that detects and resolves congestion situations.

Congestion situations can be caused by saturation of network resources such as communication links, channels, throughput, etc. For example, if a communication link delivers packets to a queue at a higher rate than the service rate of the queue, then the size of the queue will grow. If the queue space is finite then in addition to the delay experienced by the packets until service, losses will also occur. Networks need to serve all users requests, which may be unpredictable and bursty in their behaviour (starting time, bit rate, and duration). However, network resources are limited, and must be managed for sharing among the competing users. Congestion will occur, if the resources are not managed effectively. The optimal control of networks of queues is a well-known, much studied, and notoriously difficult problem, even for the simplest of cases.

The basic result of a congested network is the degradation of the network performance. The users are experiencing long delays in the delivery of the packets, perhaps with heavy losses caused by buffer overflows. The jitter and delays values are then very high and the network available throughput can be close to zero. The degradation of the network performance, the delays and the packet losses result to retransmissions of the lost packets, which in turn results to a waste of the available throughput, which is consumed to retransmit the lost packets.

An effective Load Control mechanism should be first of all preventive in order to avoid congestion situations. Since it is not easy to avoid network overloading, the Load Control mechanism should be designed to react very fast and minimize the duration of the congestion and the amount of the load. The Load Control mechanism tries to settle the problems with network overloading. Its function seems to be very close to the function of Admission Control. The difference is that Load Control takes place after overload occurs. Admission control tries to prevent overload situation. If the network is well planned and Admission Control and packet scheduling algorithms work sufficiently then congestion situations will be exceptional. But an always effective AC is a very complicated task. In addition the interference level in the cell is not static and can vary very much over the time and fast moving users cause more interference than stationary or slow moving users. If congestion occurs, the Load Control must decrease the load to the limits defined by network planning.

In the context of the WINNER network, the congestion control mechanism is monitoring the network and if an overload situation occcurs it tries to decrease the load of the network performing several actions and especially trying to make the networks cooperate. The basic idea is that when there is congestion detected in the WINNER network then at first the Load Control mechanism tries to solve the problem locally within the WINNER network and if it fails then it escalates the problem for global resolution, including the legacy cooperating networks.

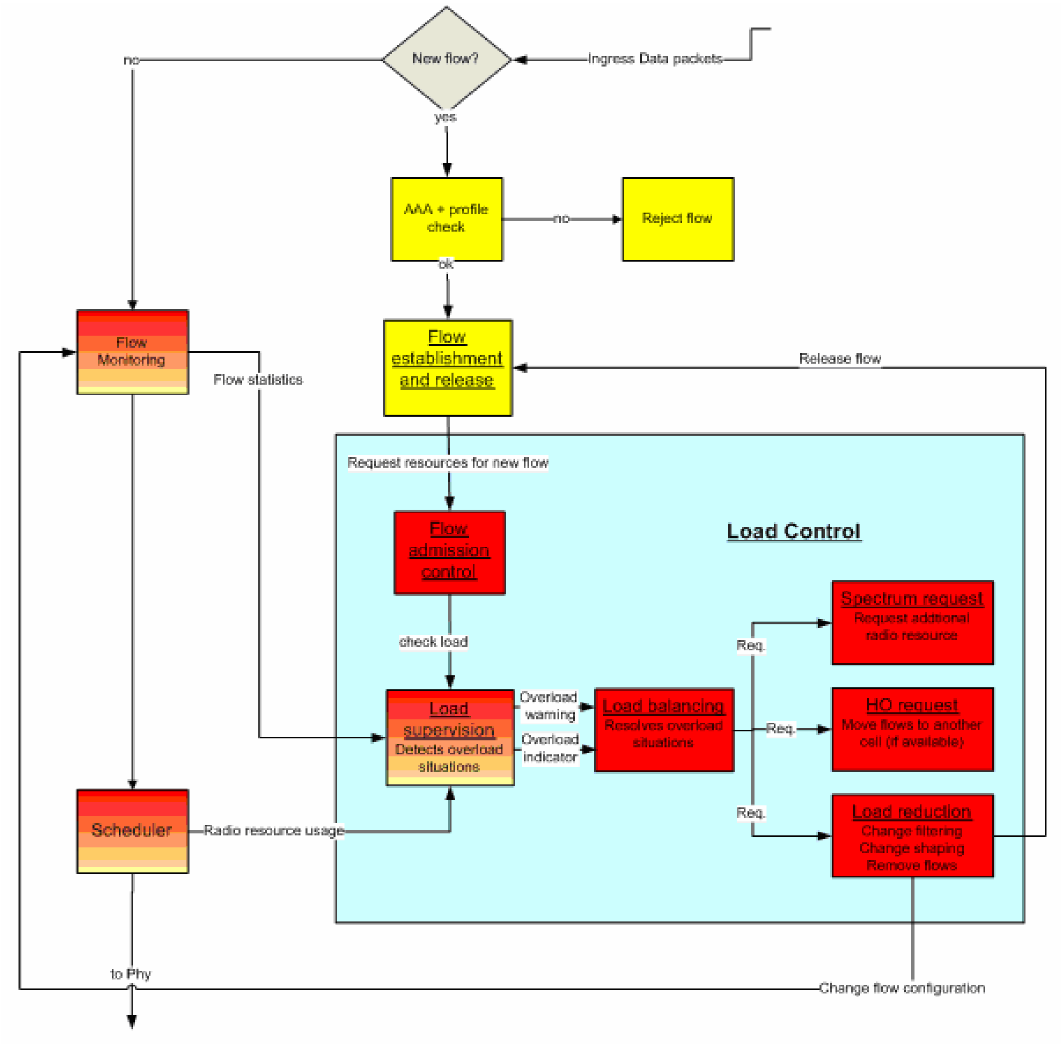

Figure 7-14: Load Control.

In the previous figure it is presented an overall description of the Load Control procedure. As it can be seen in the figure there are many functions involved in the Load Control procedure. The Load Control procedure can be split in three phases.

- **Load monitoring or congestion detection phase**. The congestion control algorithm continuously monitors the network and periodically checks the load in order to detect a congestion situation in any of the cells of each deployment mode. It is considered that a cell of a deployment mode or a network is overloaded if the load factor is over a certain pre-defined threshold during a certain amount of time, i.e. if $n \geq n_{thC}$ for an amount of time ΔT_C, then we can say that the cell is overloaded and the congestion control algorithm is triggered.
- **Congestion resolution phase**. This is the phase in which the algorithm is trying to resolve the problem that causes the overload situation. At first, the Load Control must interact with the Admission Control to order it to reject all the incoming sessions (new and handoff) as they will increase more the load of the network. The only exception is for emergency sessions that should be handled in any case. Then there are other functions that can help decrease the load of the network and are related with spectrum management, handovers, resource renegotiation and dropping of ongoing flows.
- **Congestion recovery phase.** After performing the actions described in the congestion resolution phase, the algorithm enters the congestion recovery phase, where it tries to restore the QoS (the datarate) of the sessions, which rate was decreased in the congestion resolution phase. A congestion recovery algorithm is necessary, because these sessions should not continue to receive low QoS that violates the service agreements. The problem here is that if we restore the sessions' bitrate incautiously, the network could fall once again in a congestion situation and that's why we must be very careful to that. For all the sessions that we have decreased their

bitrate, we check (separately) if we can restore their transmission rate without causing congestion again, by computing the amount of load needed in order to restore the transmission rate.

There are many ways to detect or sense a congestion situation in the network:

- Packet loss sensed by the queue as an overflow, by destination (through sequence numbers) and acknowledged to a user or by sender due to a lack of acknowledgment (timeout mechanism) to indicate loss.
- Packet delay, which can be inferred by the queue size, observed by the destination and acknowledged to a user (e.g. using time stamps in the packet headers), or observed by the sender, for example by a packet probe to measure Round Trip Time (RTT).
- Loss of throughput, which can be observed by the sender queue size (waiting time in queue).
- Other events, like increased network queue length and its growth or queue inflow and its effect on future queue behavior.

Since these phases are not easy to be discriminated in the previous figure, it is presented the next figure, with the distinction of the phases, the functions that reside in each phase and their interactions. In the Congestion Resolution phase five different steps for decongesting the network are presented. These steps are numbered from 1 to 5. These numbers show a kind of prioritization, in terms of which step is applied first by the network. The first action of course would be to reject or delay any non – emergency new requests. In the mean time, the load control entity communicates with the scheduler to decrease priorities of some elastic traffic classes, through e.g. weight re-assignement to the scheduler. Then the network will do some spectrum management techniques in order to gain more resources in the congested area. Then other techniques like handovers, resource renegotiation and dropping high loaded flows are applied. These priorities are assigned to these steps taking into account their effectiveness on the load reduction and on the user's perception. The prioritization is not fixed. Which function should be triggered first is determined by the context of the communications. For instance, if short term Spectrum Sharing (SS) can be easily performed without dropping the newly arrived calls, it can be placed with higher priority.

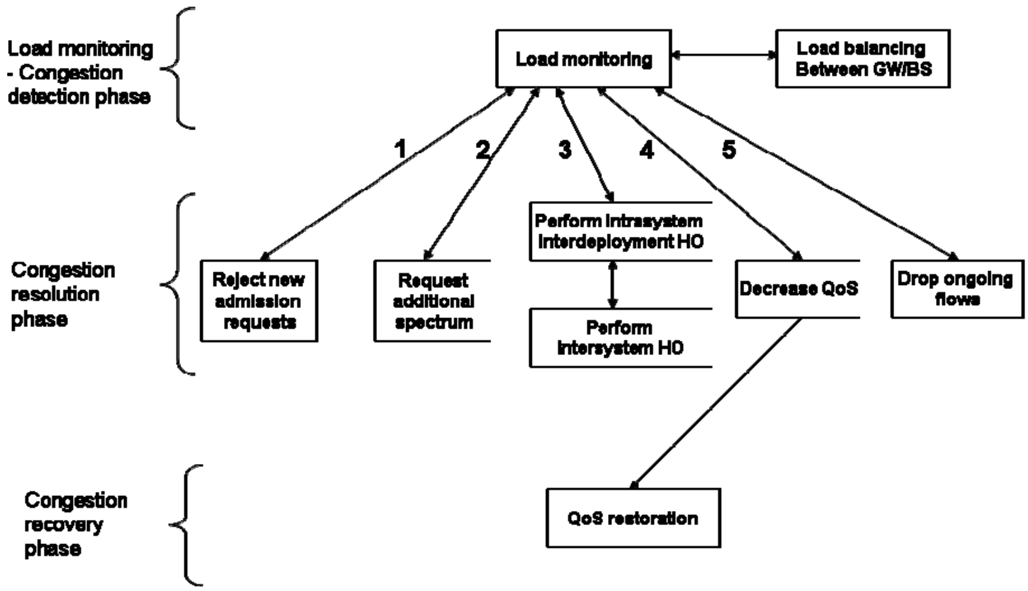

Figure 7-15: Load Control functions interactions.

7.10.3 Admission control

Admission Control (AC) is one of the key RRM mechanisms that ensure the good operation of a network, by admitting or rejecting new user requests based on criteria such as the load of the network. In WINNER AC is applied at data flow level. The basic goal of an AC algorithm in cellular networks is to control the admission of new or handover sessions within the network with the goal of maintaining the load of the network within some boundaries. The main function of an efficient AC algorithm for heterogeneous networks is to decide at a specific point of time if there is a network that has the available resources to serve (to satisfy the QoS requirements of) a new session. The decisions of the AC algorithm must be made very carefully in order that the following two situations don't happen at all, or are minimized:

- Bad rejections, which occur when the algorithm rejects a session, although there is actually a network that can meet the session's requirements (there is enough capacity to allocate the session). In this case, capacity is wasted and because of that the operator's revenue is not optimized.
- Bad admissions, which occur whenever the algorithm accepts a session although there isn't a network that has the available capacity for it. In this case, QoS guarantees are not provided and user's satisfaction is degraded.

There are several criteria that the Admission Control algorithms use for accepting or rejecting a flow. Very basic algorithms can be power-based, which means that they use periodic measurements of the transmitted power, computing the interference at the user's receiver and based on the statistic evaluation of the SINR and the QoS target, make the decision of admitting or rejecting the user. In throughput-based algorithms, the throughput that can be delivered by the system is determined according to some dimensioning calculations, assuming some conditions in the system. There are also algorithms that use the equivalent capacity of aggregated traffic, which is an estimation of the arrival rate of a class of traffic. Other algorithms check the load of the system (it must not exceed a pre-defined threshold) and the bandwidth and delay constraints of each flow, according to the current system's data.

In mobile environments it is not adequate to admit a new call only based on the status of the current cell, where the call is being generated, because when the user attempts to move from the original cell to a next cell, there may not be sufficient resources in the destination cell for accepting the handoff. This may result in dropping the call and increasing the call dropping rate. This requests an efficient Admission Control algorithm, which should check also the available resources in the adjacent cells, except in the current cell. However, to avoid call blocking and dropping rate, the RRC may trigger other RRM functions such as inter-mode handover, resource partitioning or even spectrum sharing with other systems.

Moreover, not all sessions have the same characteristics and requirements. For example a voice session (telephony) has very low delay and bandwidth requirements, while e-mail delivery is very tolerant to delay and it might need more bandwidth. That's why in radio access networks there are defined different service classes for the users according to the services the networks will offer and to the different characteristics of each service. Future wireless networks will have to consider many service classes to meet future user requirements for best quality of service offers, because they want to offer the best quality of service to the users and they have to meet the future users' requirements. The service classes considered have different delay, throughput and BER requirements. The different service classes of the RANs will be acknowledged by the AC algorithm, in terms of resource allocation and especially in terms of prioritization. Different service classes will have different priorities in the algorithm. The criteria for each class's priority should be based on the characteristics and requirements of each class, i.e. the delay sensitivity, the bandwidth requirements, etc. A class with high priority should be checked (admitted or rejected) before a low priority class, although the low priority class arrives first, or a high priority class could be admitted although a low priority class will be rejected. Due to the nature of the future service classes there could be services that should have higher priority irrespective to is the call new or from handover. For example if a service class is defined in a network for "emergency calls" (i.e. calls for police or ambulances during a car accident), these calls (which are new calls) should have the highest priority of all calls.

In current legacy systems the Admission Control only examines the requests for entering the network in a specific cell. In the WINNER network, the algorithm will not only examine the admission of a user into a specific cell, but also it should select the most suitable WINNER mode for the service that the user requests. If the WINNER system is overloaded, it should find the most suitable legacy system to serve the user. In future wireless systems like WINNER, where multiple overlapping deployment modes exist together with overlapping legacy systems, the approaches of distributed, centralised and scalable or hybrid AC have been investigated.

Distributed AC means that the AC decisions are only taken by the BS that the user is trying to connect. This is very similar to the existing AC procedure for independently operated systems. If the user is not admitted to that BS he will try to be admitted to other BS in the same area. This approach is very fast and does not need a lot of signalling exchange between the network entities. However, it may encounter a non proper admission decision.

Centralised AC can only be done when a central entity (like the RRM server in WINNER) exists in the network. The RRM server receives periodically RTTMs (Real Time Traffic Measurements) from the BSs and the GWs, so it has permanently the knowledge about the load of the BSs and the GWs. When a user tries to change his state from idle to active, the AC is handled by the RRM server, which, based on the load information, can find the most suitable (and not overloaded) mode and BS to serve the user.

To solve the problems in the two approaches described above, it is proposed a combination of them to obtain a hybrid and more scalable approach. In this approach the AC is being handled by the BS, but it can request assistance from the RRM server to take a better decision. The hybrid/scalable AC solution is usually fast since the BS takes the decision and does not need so much signalling between the entities, since the information exchanges between the BSs and the RRM server could be performed only on-demand and not all the time.

In WINNER system, by default, the entity responsible for the admission decision is the base station that the user terminal is connected. This base station when it receives the service request it tries to admit the user but if it fails then it communicates with the other neighbour base stations to try to admit the service in one of them. So it sends messages to request RTTMs from the BSs and then calculates them and decides to which BS the UT should be admitted and communicates with that BS to finalize the admission process. The signalling process for the distributed Admission Control can be seen in the next figure. In this figure the involvement of the GW is also presented, as it is the entity that communicates with the AAA server to request the authorization, authentication and accounting information for the user, namely to check if the user is subscribed to the network. Also the GW has the profile information of the user. The green messages are the AC – CP messages, the orange are the RTTM messages and the blue is the UP signalling.

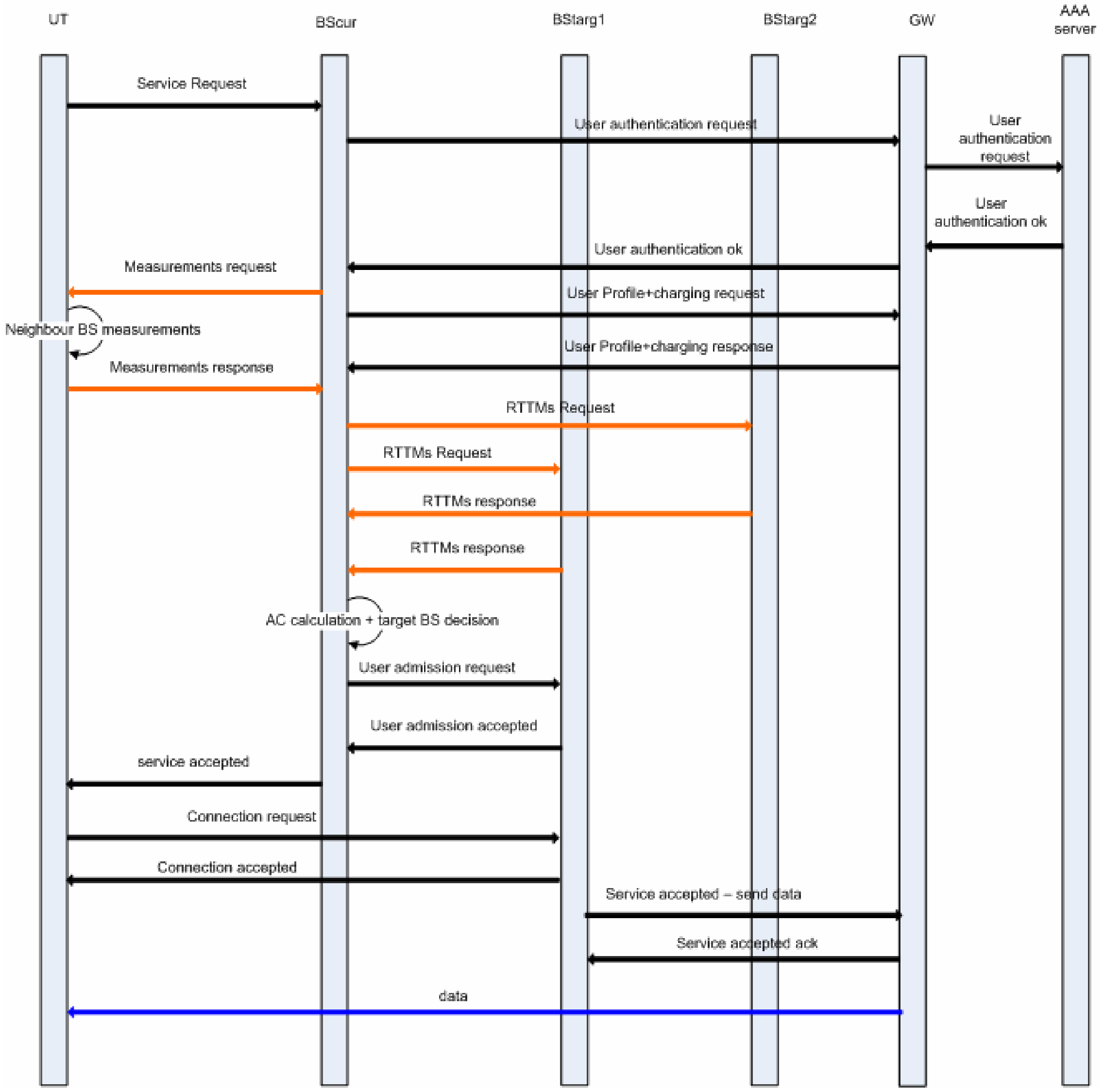

Figure 7-16: Distributed Admission Control signalling process.

In the hybrid approach, the current BS is the one that communicates with the neighbour BS in order to take the decision about admitting the new user request. The current BS also communicates with the RRM Server in order to get advice for the target BS, so as it can take the best decision for the user. The BS can also complete the AC procedure without asking the RRM Server for advice and this can happen in low load situations. The signalling process for the scalable/hybrid Admission Control is presented in the following figure. Same as in the previous figure, the dashed lines mean that the signalling is optional, when the RRM Server is to be taken into account.

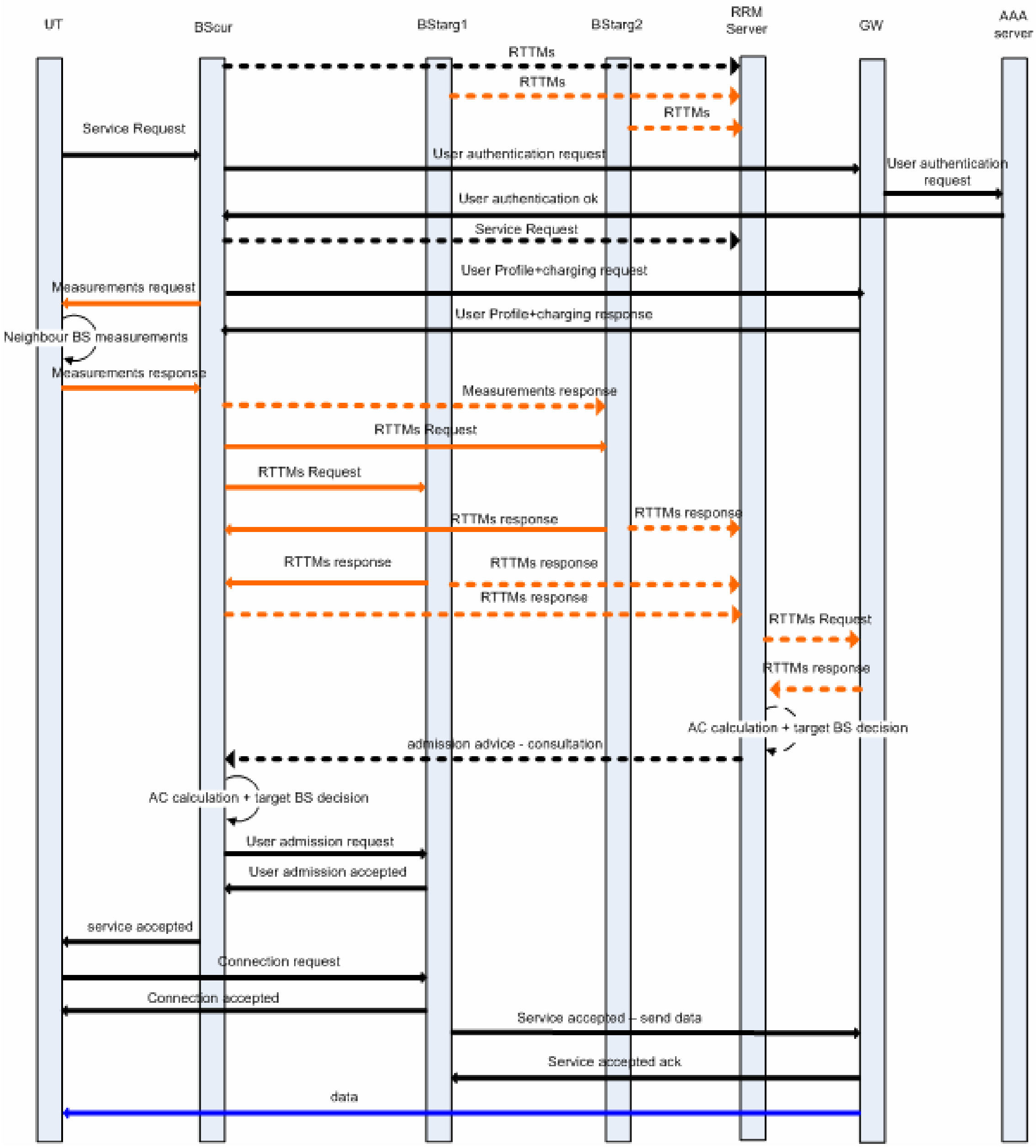

Figure 7-17: Hybrid / scalable Admission Control signalling process.

7.11 Spectrum technologies

7.11.1 Spectrum Assignment Negotiation

A possible conceptual structure for Long Term (LT) spectrum assignment was developed in WINNER I [WIN1D63] and is illustrated in Figure 7-18. From the figure, one can see that LT spectrum assignment is a periodical, or scheduled, function. One can also notice that although inter-RAN communications occur in several component functions, recovery from possible communication failure (not resolved by normal communication protocols) between the RANs is carried out in the resource update function.

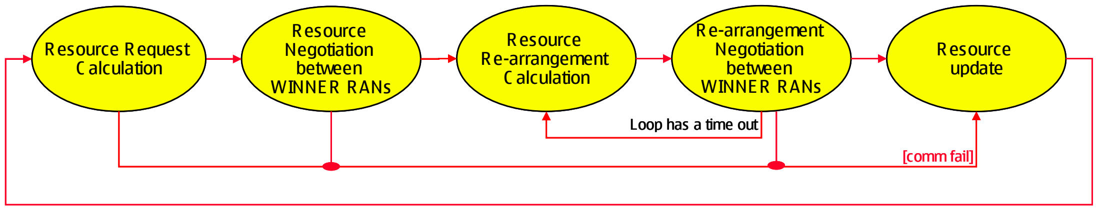

Figure 7-18: Main components of long-term spectrum assignment.

The **Resource request calculation** decides the requested spectral resources for the next assignment period based on the inputs from load prediction, MAC control feedback, and spectrum sharing, Due to the slow adaptation rate of LT spectrum assignment relatively long-term estimates on the network load are required. Hence the load predication is not only based on the current network load, but also on longer term traffic patterns such as daily or weekly averages. **Resource Negotiations** between WINNER RANs are carried out based on the resource requests and by utilizing appropriate fairness or cost metrics. The total amount of spectral resources assigned for each RAN is fixed during the resource negotiations with tentative explicit resource assignments. After the negotiations, the spectrum assignment function looks for the possibilities to improve the tentative spectral resource assignments (e.g., by minimising inter-RAN guard bands) during the **Resource Re-arrangement Calculation**. It is followed by the ***re-arrangement negotiation between WINNER RANs*** during which the exchange requests are sent and negotiated with other RANs. The purpose of this second negotiation phase is to remove any inefficient solutions that might have been established during the negotiation phase. During the **Resource Update** phase the negotiated resources are update to the system settings and error checking is done [WIN1D63]. During the resource update the following processes are performed:

1) Necessary spectral resource and constraint information messages are compiled for load control and resource partitioning located at the base stations (BSs), with timing signal for next assignment period.
2) If necessary, short-term spectrum assignment may be triggered with requests.
3) Logs on the spectrum use, fairness metrics, etc. are updated.
4) At certain time intervals, log updates are sent to the WINNER spectrum manager as well as updated parameters are downloaded from the database.
5) Recovery from any possible communication failures between the WINNER RANs is performed. During the recovery, networks fall back to the use of predefined basic spectrum assignments.

From this high level functional overview a further functional break-down was made. The main flowchart showing this process is shown in Figure 7-19. This figure is a flowchart of the spectrum assignment subfunction. The function takes several input signals depending on the type of spectrum assignment; these signals include traffic load, information from the spectrum register, the spectrum manager, and sharing and coexistence schemes. Based on these inputs the logical node determines its spectral resource needs, both in terms of current need and projected future needs. In order to incorporate the future needs in the analysis the node may add a margin to its prediction. This margin can be a fixed percentage, e.g. 10 or 20 % of the current load depending of the amount of conservatism build into the mechanism. It is one of the conclusions from WINNER I to limit the amount of conservatism since it makes less resources available for sharing and reduces its gains [WIN1D63].

After having determined the current and future spectral needs, this prediction is mapped on the available spectral resources. The available resources can consist of resources dedicated to the network as well as resources that have already been exchanged between networks.

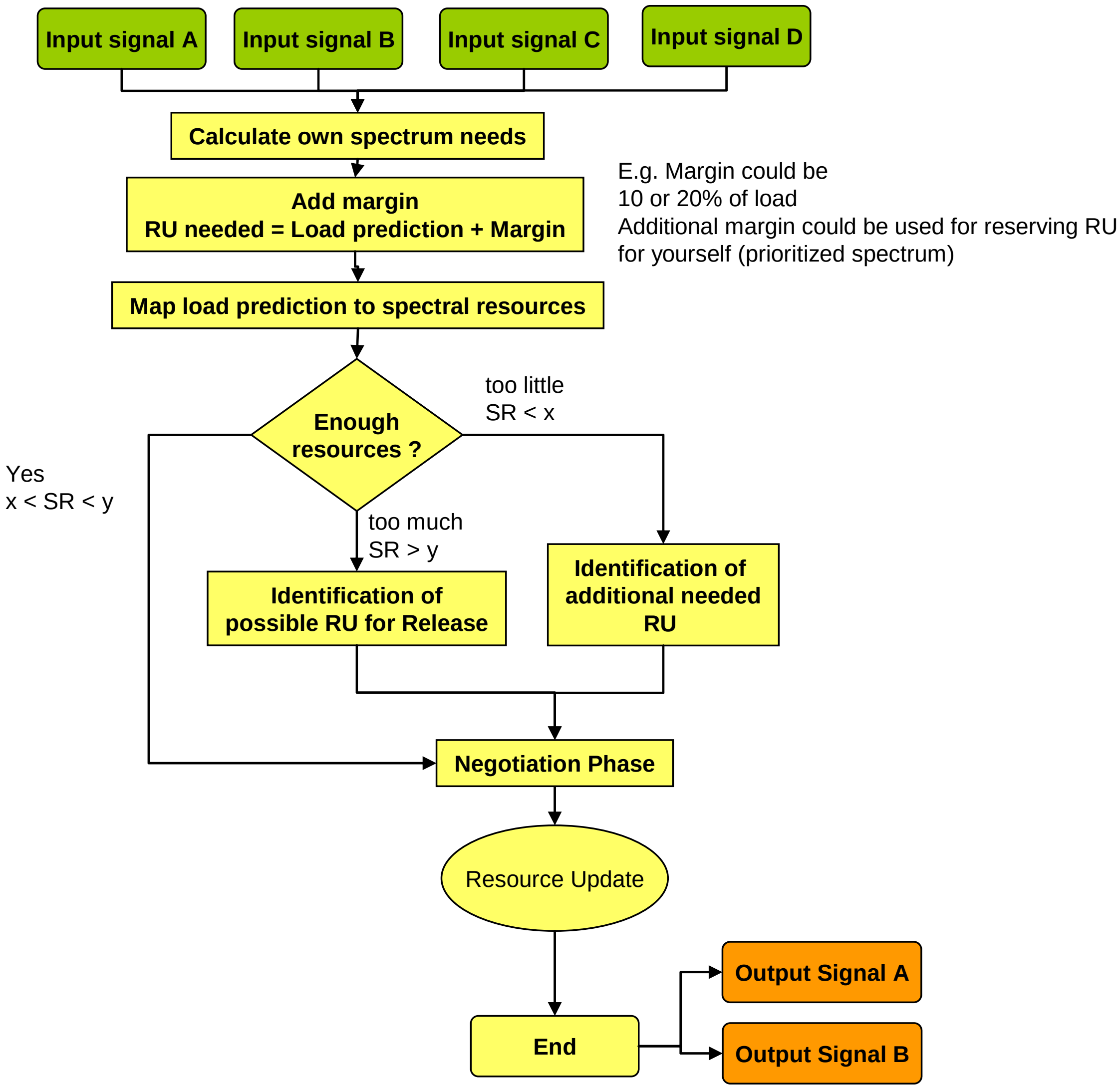

Figure 7-19: Spectrum Sharing Sub function.

If the current spectral resources are more than sufficient to meet the predicted load the wireless node may identify resource units for release (if it is currently borrowing spectrum), or it could decide to lent out some of its current resources. If the current amount of resources is insufficient, then a procedure should be started to obtain additional resources, this includes obtaining back resources that were previously lent to other networks. After either one of these two tracks, or in case of a matching amount of resources in which case no spectrum changes are indicated, the nodes may engage in the negotiation phase. During the negotiation phase two parallel processes take place. Firstly, the nodes evaluate their own situation and check if their resources are available; this information is collected in a Resource Unit MAP (RU MAP). It can be indicated in the RU MAP if resources are for example occupied (in use), reserved (part of the margin), or considered (available for spectrum trading). Secondly (at the same time), the nodes are collecting incoming requests from other nodes and analyze these. Based on this analysis a RU MAP can be constructed with the request for resources. When all these requests are received the node checks the possible combinations and their technical feasibility. From the set of feasible solutions a sub set of solutions that are most interesting from a performance point-of-view is selected. These options are input to an auctioning and billing process after which the best option from both a technical and an economical point-of-view is selected [WIN2D591].

7.11.2 Long-term spectrum Assignment Algorithms

Gradual spectral deployment of WINNER systems following the gradual increase in the WINNER subscribers has been investigated [WIN1D63]. Spectral resource requests representing the spectrum demands of the networks were assumed to be similar on the average and to show either large random variations or no variations at all, from day to day. Based on the results, one carrier was sufficient for the considered 3 networks when the maximum average spectral resource requests per network, normalized with the number of chunks per carrier, were below 20%. Similarly, two carriers were found sufficient for the normalized resource requests per network were below 40%. In the final case of fully evolved WINNER system with 3 carriers, the gains from FSU depended on the amount of required guard chunks, and on the differences between the networks on the spectral resource request. Based on the results, it can be concluded that the FSU facilitates a smooth gradual spectral deployment / evolution for WINNER system, achieving higher intensity in the spectrum use during the evolution than the conventional fixed spectrum assignment.

One of the earlier conclusions on FSU was that it did not provide any gain if the spectrum demands were similar on the involved networks. It was also noted that if the requests are random with relatively high variance and uncorrelated between the networks, the FSU can provide gain in terms of higher achievable operating point [WIN1D63][WIN2D593]. These variations can be exploited using Short-term assignment schemes.

7.11.3 Short term Assignment Algorithms

For each RAN we define *greedy cells* and *generous cells* depending on the additional or insufficient resources. When ($N_a > N_d$) then these cells are categorised as generous cells where extra resources are available for negotiation. If ($N_a < N_d$) these cells are categorised as greedy cells since they are starving for more resources. N_a is the number of resources allocated to each cell during LT assignment and where as N_d is the actual number of resources needed from the traffic demand curves. During the ST assignment period only generous and greedy cells negotiate with each other. Also to avoid conflicts between more than one cell pairs entering into the ST negotiation within the analysis we have limited number of cell pairs involved into negotiation as only one [WIN2D593].

The selection of cell pair is based on the following two algorithms:

1. "the least satisfaction criteria"-Algorithm
2. "maximum flexibility criteria"-Algorithm

The first algorithm is based on the selection of cell pair with least satisfaction. In this case the generous cell satisfies the minimum requirement of the greedy cell. The second algorithm is based on cell pair with the most flexibility in allocating chunks. In this case generous cell provides the maximum satisfaction to the greedy cells or much more than the required resources from the greedy cell. Having more resources than the required amount allows for flexibility in resource allocation for the greedy cell thus avoiding interference to neighbouring cells of the generous cell [WIN2D591][WIN2D593].

In the case of the least satisfaction criteria the amount of offered frequency chunks in the generous cell is similar to the number of extra frequency chunks required by the greedy cell. In the second case the cell pair selection is based on maximum flexibility criteria. In the latter case the generous cell offers much more than required by the greedy cell. For successful ST assignment performance it is necessary to reduce extra interference caused by offered frequency chunks to the neighbouring cells of the generous cell. With plentiful resources the greedy cell has the flexibility of deploying chunks minimising interference to neighbouring cells in the 1st tier of the generous cell [WIN2D593].

It has been shown that in both cases the amount of negotiation chunks or gain in terms of frequency units is higher with higher variation, independent of cell selection algorithms. Therefore more bursty traffic patterns at cell level contribute towards positive performance gains in ST assignment.

Long-term spectrum assignment allocates the resources based on longer-term averages and provides a good mechanism for sharing resources over time. Short-term assignment exploits local variations in load and traffic patterns to further optimise the performance.

7.12 Measurements

In order to operate, WINNER system requires a lot of various kinds of measurements. Actually, it will be more dependent of this data as current wireless systems, due to its' highly adaptive nature starting from the physical layer link adaptation to flexible utilization of available spectrum. This section collects the most important measurments the WINNER system needs to carry out.

Table 7-1: Summary of most important measurements that WINNER system needs to carry out.

Name	**Short description / comments**	**Example used of measurement**	**Need for signalling**
Channel Quality Indicators (CQI)	Superordinate term for various scalar channel related measurements (no unique definition), e.g., SINR of received streams before or after receive processing		
Received SINR (after receive processing)	Received SINR per chunk	Flow and resource allocation, Adaptive Modulation and Coding, Scheduling, Multi-antenna techniques, Handover.	Signalling is required, and is urgent for fully adaptive scheduling (FDD operation)
Received SINR accuracy	MSE of prediction		
Average / Mean SINR	Can be derived from Received SINR		
Channel State Information (CSI)	channel state information (frequency response of the radio transmission channel) - at transmitter (CSIT) - at receiver (CSIR)	CSIT: assignment of flows and radio resources to subcarriers/chunks; selection of modulation and coding scheme per subcarier/chunk(i.e., prediction of received SINR possible in certain techniques); selection of SDMA scheme; computation of precoding weights in TDD mode CSIR: computation of received beam CQI in fixed beam SDMA techniques; equalization of MIMO channel without transmit precoding (e.g., in FDD uplink)	TDD mode: no signalling when reciprocity can be assumed (calibration needed) FDD mode: CSIT through (quantized) feedback of downlink measurements, only for hybrid schemes adapting spatial codebook
CSI accuracy	Measure of the quality of the available CSI	Selection of frequency-adaptive or non-frequency-adaptive transmission mode	
Effective Channel State Information (ECSI)	Including effect of spatial transmit processing (at receiver)	Equalization of effective MIMO channel acting on data (aka combining)	
Doppler Spectrum	related to CSI accuracy		
Transmit Power		Estimation of expected interference: interference mitigation or avoidance schemes; power sharing algorithms	
Noise plus Interference Power (NIP) at RX		MCS selection, precoding, decoding, equalization, scheduling	
Interference Power (IP)	Frequency dependent	Required if frequency-adaptive interference coordination is performed	

Frequency Synchronisation accuracy	Accuracy of carrier frequency in BS and UT	Frequency synchronisation mechanism	
Cell Synchronisation accuracy	Accuracy of the superframe time synchronisation with respect to neighbouring cells	Cell synchronisation mechanism	
Propagation delay	Propagation delay between BS and UT	Timing Advance on the uplink	
ToA TDoA	Time of Arrival and Time Difference of Arrival	Location Determination, UT speed, intermode handover	Non-urgent from UT back to BS, and from BS to location nodes
UT position		Handover trigger. UT velocity (derived from UT position) could be an inter-mode handover trigger.	
Cell load	Used chunks/Available chunks	Load Balance between modes, congestion control	BS -UT and BS -BS
Dominant Inter-cell Interferers identification	Typically only the dominant or the 2 dominant interferers need to be identified. In the dowlink this can be done from the average power received from the neighbouring RAPs.	Complexity reduction of inter-cell interference cancellation.	
Correlation Matrix of the received signal	Correlation matrix of the received signal, including inter-cell interference.	IRC weights computation	
Inter-cell Received Signal Strength (RSS)		Handover	
Inter-RAN RSS / SINR		Handover between different RAN deployments	
Inter-cell Interference per chunk	Absolute measure of inter-cell interference per chunk	Minimum Interference Dynamic Channel Allocation	

7.13 Evaluated physical WINNER deployments

In the final subsection of the reference design chapter, concrete examples for deployments of the WINNER concept in specific scenarios are given that have been used for evaluation. The physical deployments include both concrete examples for the implementation of logical nodes in physical nodes and examples for application of WINNER technologies, concept and reference algorithms in a specific scenario. The scenarios are chosen to cover a broad range of deployments with respect to propagation conditions, coverage area, capacity and available transport technology. By that, it is visualized how the WINNER concept can flexibly adapt to specific scenarios and how it can be scaled to different traffic requirements. Furthermore, scenarios are chosen in order to highlight the application of WINNER key technologies.

Note however, that the WINNER system is very flexible and allows for other parameterizations. Especially it should be emphasized that depending on the available bandwidth and whether the available spectrum is paired or unpaired, other spectrum configurations than those used in the evaluated scenarios are suitable.

7.13.1 Single – cell – layer wide area deployment

The single-cell wide area deployment scenario reflects a classical operator deployment situation who owns a licensed frequency band and wants to provide base coverage layer, i.e. provide a large region with relatively low number of users with broadband service according to WINNER requirements and satisfied user criterion. Cell sizes are kept as large as possible and base stations are deployed as macro cells for high coverage. Addtionally, cell area is extended by usage of relay nodes.

The chosen physical nodes and interfaces are depicted in Figure 7-20. Base stations are connected with GW_IPA_{LN} and GW_C_{LN} by using transport infrastructure as leased lines or microwave. Number of gateways and size of pool area is chosen according to user density and can be scaled up flexibly by adding more gateways. Logical interfaces I_{BB} between basestations are established for handover and use the desribed transport infrastructure. Spectrum server enables spectrum sharing with other systems and flexible spectrum usage with WINNER deployments of other operators.

While applying frequency – adaptive transmission for low – to medium velocity users, high velocity can be supplied with stable data rates by non-frequency-adaptive transmission. Multi-antenna arrays allow usage of grid-of-beam technology for extension of cell – range cell edge capacity. Additionally, multi- and single user multi-stream transmission (SDMA) increases cell capacity. Combination of adaptive scheduling with grid - of – beam antennas in a propagation environment with low angular spread provides an effective technology for inter-cell-interference mitigation. Furthermore, the antenna array can be exploited by using interference rejection combining at the basestation, improving system capacity and number of satisfied users, only by adding more signal processing capacity to the basestation.

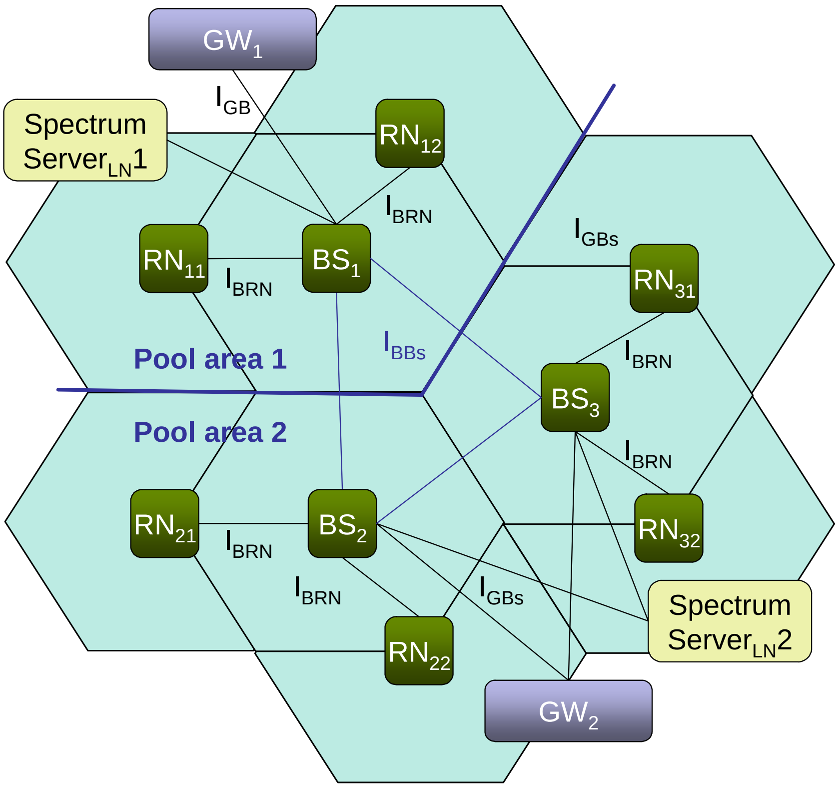

Figure 7-20: Exemplary single cell layer deployment.

A possible parametrization of the WINNER concept that was chosen for the performance evaluation (refer to WIN2D61310) is given in Table 7-2. It should be emphasized that the pool areas generally overlap (which is not visible in the figure above).

Table 7-2: Example WINNER wide area deployment parameters

general	duplexing (asymmetry)	FDD (1:1)	OFDM parameter	Subcarrier distance Δf	39062.5 Hz
	carrier frequency f_c	3.95 GHz DL / 3.7 GHz UL		Useful symbol duration T_N	25.6 µs
	channel bandwidth	2 x 50 MHz		Guard interval T_G	3.2 µs
	Deployment	cellular, hexagonal layout		Total symbol duration	28.8 µs
base station	location/height	Above rooftop		used subcarriers	**[-576:576]** subcarrier 0 unused
	number of sectors per BS	3		Signal bandwidth	**2 x 45 MHz**

	number of antennas per sector	4		System bandwidth	2 x 50 MHz
	antenna configuration (per sector)	Linear array | | | |		FFT bandwidth, sampling rate	80.0 MHz

7.13.2 Multi – cell – layer metropolitan area deployment

The multi-cell-layer metropolitan area deployment scenario corresponds to micro cellular deployment situation where the operator targets to provide high system capacity to areas having high user density, like densely populated urban areas and city centers. The operator for the metropolitan area deployment owns a licensed frequency band for the deployment. The operator (or a collaborating operator) can have the wide area deployment coverage available for the same area realizing the multiple cell layers. Due to the high traffic demand, and typically challenging radio propagation environment, the achievable cell size is rather small on average. Micro cellular (below rooftop level) deployment of base stations is selected to obtain reasonable cell sizes. Relaying is integral part of the system deployment, as the coverage can be efficiently expanded with them.

The chosen physical nodes and interfaces are depicted in Figure 7-21. Base stations are connected with GW_IPA$_{LN}$ and GW_C$_{LN}$ by using transport infrastructure as leased lines or microwave. Number of gateways and size of pool area is chosen according to user density and can be scaled up flexibly by adding more gateways. In the same area there are at least two layers of deployments and base stations; one layer for the base coverage layer, and one for the metropolitan area capacity enhancement layer. Logical interfaces I$_{BB}$ between basestations on different layers are established for handover and use the desribed transport infrastructure. Spectrum server enables spectrum sharing with other systems and flexible spectrum usage with WINNER deployments of other operators.

As the majority of the users have low mobility in the metropolitan area deployments, frequency-adaptive transmissions will be heavily used. Multi-user MIMO schemes are effective means for transmission for most of the users present in the coverage area of this deployment. Usage of relaying is very beneficial for mitigating the challenging shadowing conditions that frequently prevail.

A possible parametrization of the WINNER concept that was chosen for the performance evaluation (refer to WIN2D61311) is given in Table 7-3.

Table 7-3 Example WINNER metropolitan area deployment parameters

general	duplexing (asymmetry)	TDD (1:1)	OFDM parameter	Subcarrier distance Δf	48828.125 Hz
	carrier frequency f_c	3.95 GHz DL		Useful symbol duration T_N	20.48 µs
	channel bandwidth	100 MHz		Guard interval T_G	2 µs
	Deployment	cellular, Manhattan grid layout		Total symbol duration	22.48 µs
base station	location/height	Below rooftop		used subcarriers	**[-920:920]** subcarrier 0 unused
	number of sectors per BS	2		Signal bandwidth	**89.84 MHz**
	number of antennas per sector	8		System bandwidth	100 MHz
	antenna configuration (per sector)	Cross-polarized Linear array X X X X		FFT bandwidth, sampling rate	100.0 MHz

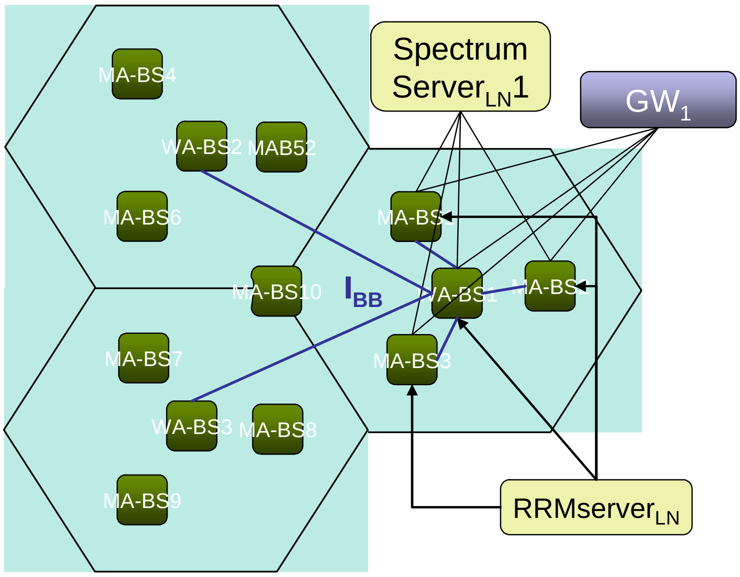

Figure 7-21: Example multi-cell-layer deployment (interfaces only shown for BS1-3)

7.13.3 Home-basestation deployment

The home base station deployment scenario consists of the wide area network complemented with small home base stations providing indoor coverage in customers own premises. The home base station is connected directely to Internet with e.g. ADSL, or cable modem. There is an efficient handover between the wide area deployment and the home base station. The home base station utilizes different band than the wide area deployment, e.g. unlicensed bands.

The chosen physical nodes and interfaces are depicted in Figure 7-22. The home base stations have their own GW_IPA$_{LN}$ providing local breakout to internet. Number of gateways and size of pool area is chosen according to user density and can be scaled up flexibly by adding more gateways.

Important technologies in this deployment include spectrum technologies, and seamless handovers. Spectrum technologies facilitates the flexible usage and sharing of the available frequency resources between the the home base stations of different operators. Smooth and seamless handovers between the home base station and wide area deployment is a prerequisite for the success of this deployment.

A possible parametrization of the WINNER concept that was chosen for the performance evaluation (refer to WIN2D61312) is given in Table 7-4

Table 7-4 Example WINNER home base station deployment parameters

general	duplexing (asymmetry)	TDD (1:1)	OFDM parameter	Subcarrier distance Δf	48828.125 Hz
	carrier frequency f_c	5 GHz DL		Useful symbol duration T_N	20.48 µs
	channel bandwidth	100 MHz		Guard interval T_G	2 µs
	Deployment	Isolated cell, indoor		Total symbol duration	22.48 µs
base station	location/height	3 m		used subcarriers	**[-920:920]** subcarrier 0 unused
	number of sectors per BS	4 arrays operated as remote radio heads		Signal bandwidth	**89.84 MHz**
	number of antennas per sector	8 elements per array		System bandwidth	100 MHz
	antenna configuration (per sector)	Cross-polarized Linear array X X X X		FFT bandwidth, sampling rate	100.0 MHz

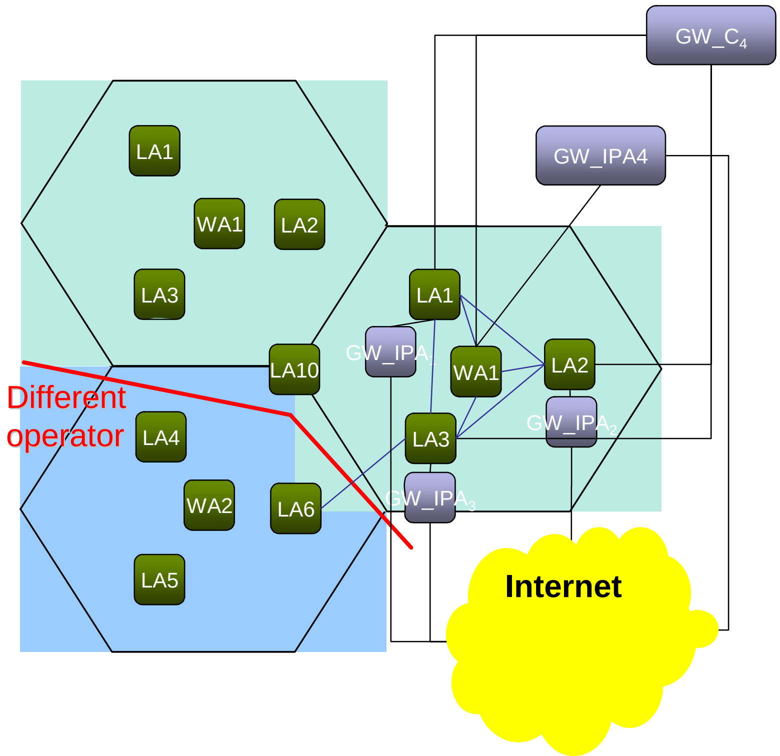

Figure 7-22: Example home BS deployment with WA overlay and logical interfaces (not all interfaces showwn)

7.13.4 Office/shopping mall deployment

The office/shopping mall deployment is targeted for large office spaces or shopping malls for providing service for large and dense indoor user population. The basic coverage in this deployment is provided by the wide area and metropolitan area deployments. The main high capacity service is provided by local area base stations which are interconnected with a local infrastructure, e.g. ethernet. This interconnection facilitates the usage of different advanced techniques like interference coordination, and application of distributed antennas.

The chosen physical nodes and interfaces are depicted in Figure 7-23. The local area base stations are connected to a local GW_IPA_{LN} and to an associated GW_C_{LN} pool. Number of gateways and size of pool area is chosen according to user density and can be scaled up flexibly by adding more gateways.

Due to high traffic demand, and low mobility of users, MU-MIMO pre-coding using distributed antennas for providing very high system throughput is one of the key facilitating technologies for this deployment. The above mentioned interference coordination is also one enabling technology for reaching the challenging targets. Flexible spectrum usage between the local area base stations and the metropolitan area network is also seen as a significant enabler. As with the home base station, seamless handovers between the indoor and outdoor deployments are prerequisites for successful operation for this deployment scenario.

A possible parametrization of the WINNER concept that was chosen for the performance evaluation (refer to WIN2D61312) is given in Table 7-4

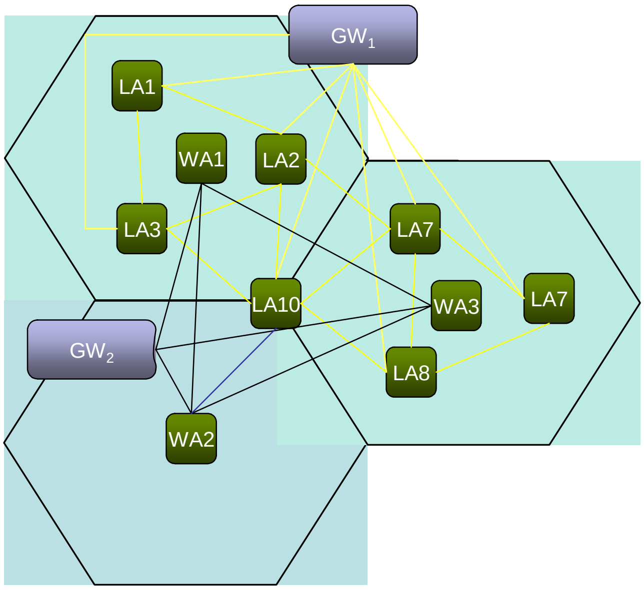

Figure 7-23: Example shopping mall deployment with fast interconnections using local infrastructure.

7.13.5 Summary

The brief examples described in sections 7.14.1-7.14.4 attempt to highlight the flexibility and scalability of the WINNER system concept, and high number of different possible ways of deployments in many scenarios and spectrum configurations, see Figure 7-24. A direct implication of the huge flexibility and scalability is that there can be a very large variety of base stations and access points, tuned to different deployments while keeping the number of different terminal classes (highlighting the differences in capabilities of the terminals) reasonable.

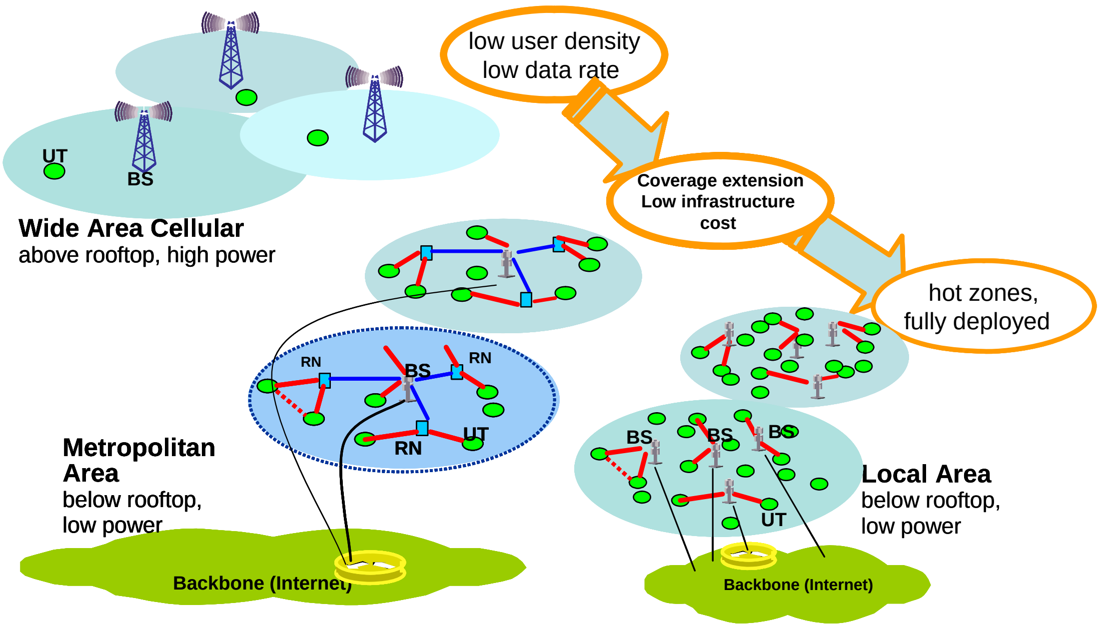

Figure 7-24: Scalability of WINNER deployment.

8 Conclusion

This deliverable shows the top-down framework for logical node architecture, protocol and service architecture and cooperation architecture in the WINNER project. These architectures form a common abstract description of functional relationships for all possible designs optimised for a certain scenarios.

Furthermore, the specific algorithms and protocols developed in WINNER can readily be mapped to logical nodes or interfaces, respectively, and their impact to the system as a whole can be understood.

The logical node architecture efficiently supports also single frequency networks and MBMS services.
The interfaces between the logical nodes are described and some basic functional interactions are provided.

The air interface protocol architecture descriptions are given layer-wise in some detail so that a basic understanding of the layer functionality is provided. Cross-layer interactions and optimizations are also described in some detail.

These architectures are designed to accommodate the WINNER requirements over a large span of scenarios including Wide Area, Local Area and Metropolitan Area deployments with or without relaying.
Furthermore, the architectures are optimised with respect to key enabling technologies such as relaying, spectrum management and key services such as packet oriented traffic and MBMS.

Reference designs optimized for these environments are given as well as exemplary algorithms for e.g. scheduling and link adaptation.

This report also serves as an introduction to many of the other more detailed WINNER deliverables.

9 References

[3GPP04] 3GPP "IP Multimedia Subsytem (IMS) Stage 2", Technical Specification, TS 23.228 V5.13.0 (2004-12) (Release 5), 2004

[3GPP07a] 3GPP RAN1 Tdoc R1-073227, "Way Forward for DL Resource Allocation Mapping," 3GPP TSG RAN1 #49bis, Orlando, USA, June, 2007. (available at: http://www.3gpp.org/ftp/tsg_ran/WG1_RL1/TSGR1_49b/Docs/R1-073227.zip).

[3GPP07b] 3GPP TSG RAN WG2 #58, Tdoc R2-071905, ETRI "Optimization of downlink persistent scheduling," Kobe, Japan, May 2007.

[3GPP07c] 3GPP TSG RAN WG2 #58, Tdoc R2-072461, ETRI "Remaining Issues on persistent scheduling," Orlando, USA, June 2007.

[3GPP08] TS36.300 Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRAN); Overall description; Stage 2

[3GPP R2-052057] 3GPP, R2-052057, "Comparison between single and two-layer ARQ for LTE", August 2005, Ericsson

[3GPP R4-060867] 3GPP TSG-RAN WG4 R4-060867, "Uplink Power Control for E-UTRA", Ericsson, Sep 2006.

[AN2006] "Ambient Networks System Description", Ambient Networks Phase 2 (EU-project IST-2004-027662), Public Deliverable D7-A.2, December 2006.

[Bajzik07] Bajzik, Lajos; Horvath, Peter; Korossy, Laszlo; Vulkan, Csaba; "Impact of Intra-LTE Handover with Forwarding on the User Connections", Mobile and Wireless Communications Summit, 2007. 16th IST, 1-5 July 2007

[BH07] F. Boccardi, H. Huang, "A Near-optimum technique using linear precoding for the MIMO broadcast channel," IEEE ICASSP, Honolulu, Hawaii, US, April 2007.

[BJD01] C. Berrou, M. Jézéquel, C. Douillard, S. Kerouédan, L. Conde Canencia, "Duo-binary turbo codes associated with high-order modulation", *ESA TTC'2001*, Noordwijk, The Netherlands, Oct. 2001.

[BKOL+07] M. Bennis, J. –P. Kermoal, P. Ojanen, J. Lara, S. Abedi, R. Pintenet, S. Thilakawardana and R. Tafazolli, "Advanced spectrum functionalities for 4G WINNER radio network", *Wireless Personal Communication Journal*, Springer 2007 (Accepted for publication).

[Ben00] P. Bender et.al., "CDMA/HDR: A bandwidth-efficient high-speed wireless data service for nomadic users,"IEEE Communication Magazine, vol. 38, pp. 70-77, July 2000.

[Bon04] T. Bonald, "A score-based opportunistic scheduler for fading radio channels," European Wireless Conference, Barcelona, February 2004.

[BTC06] F. Boccardi, F. Tosato and G. Caire, "Precoding schemes for the MIMO GBC," in*IEEE International Zurich Seminar on Communications*, (Zurich, Switzerland), Feb. 2006.

[CS03] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1691–1706, July 2003.

[DW02] Z. Długaszewski, K. Wesołowski, "Simple Coarse Frequency Offset Estimation Schemes for OFDM Burst Transmission", in *Proc. PIMRC 02, Lisbon, Portugal*, pp. 567-571, Sept. 2002

[Dor07] J-B. Dore, "Optimisation conjointe de codes LDPC et de leurs architecture de decodage et mise en oeuvre sur FPGA", Ph.D Thesis (French), October 2007.

[DS05] G. Dimic and N. D. Sidiropoulos, “On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm,” *IEEE Trans. Commun.*, vol. 53, pp. 3857–3868, Oct. 2005.

[EO07] T. Eriksson and T. Ottosson, “Feedback and adaptive modulation and scheduling in wireless systems,” ,” *Proceding of the IEEE*, Special Issue on Adaptive Transmission, Vol. 95, no. 12, December 2007.

[ETSI05] ETSI EN 301 790 V1.4.1, “Digital Video Broadcasting (DVB); Interaction channel for satellite distribution systems”, September 2005.

[ETSI02] ETSI EN 301 958 V1.1.1, “Digital Video Broadcasting (DVB); Interaction channel for Digital Terrestrial Television (RCT) incorporating Multiple Access OFDM”, March 2002.

[FDGH05a] M. Fuchs, G. Del Galdo, and M. Haardt. A novel tree-based scheduling algorithm for the downlink of multi-user MIMO systems with ZF beamforming. In Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing(ICASSP), volume 3, pages 1121–1124, Philadelphia, PA, March 2005.

[FDGH05b] M. Fuchs, G. Del Galdo, and M. Haardt, ”Low complexity space-time-frequency scheduling for MIMO systems with SDMA”, Submitted to IEEE Transactions on Vehicular Technology, September 2005.

[FOO+98] P. Frenger, P. Orten, T. Ottosson, and A. Svensson, “Multi-rate Convolutional Codes”, Communications Systems Group, Department of Signals and Systems, Chalmers University, Technical report no. 21, ISSN 02083-1260, Göteborg, April 1998.

[Fos04] M. P. C. Fossorier, “Quasi-Cyclic low density parity-check codes from circulant permutation matrices”, in *IEEE Trans. Inform. Theory*, Vol. 50, pp. 1788-1794, Aug. 2004

[FDH05] M. Fuchs, G. Del Galdo, and M. Haardt, “A Novel Tree-Based Scheduling Algorithm for the Downlink of Multi-User MIMO Systems with ZF Beamforming,” in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP ’05)*, Philadelphia, USA, vol. 5, pp 21-24, Mar. 2005.

[Gol61] M. J. E. Golay, “Complementary Series”, IRE Transactions on Information Theory, vol. IT-7 , pp. 82-87, April 1961.

[HeWiMe+05] Henning Wiemann, Michael Meyer, Reiner Ludwig, Chang Pae O, “A Multi-Hop ARQ Concept”, VTC Spring, 2005

[HH02] B. Hassibi and B. M. Hochwald „High-Rate Codes That Are Linear in Space and Time,” *IEEE Transactions on Information Theory*, Vol. 48, No. 7, July 2002.

[HLPS+05] K. Hooli, J. Lara, S. Pfletschinger, M. Sternad and S. Thilakawardana, “Radio resource management architecture for spectrum sharing in B3G systems”, *Wireless World Research Forum (WWRF) meeting 15*, Paris, Dec. 2005.

[Hua05] X. Huang, “Simple Implementations of Mutually Orthogonal Complementary Sets of Sequences”, in *Proc. International Symposium on Intelligent Signal Processing and Communication Systems*, pp. 369-372, December 2005.

[JWI06] Jang, Won and Im, "Cyclic Prefixed Single Carrier Transmission with SFBC over Mobile Wireless Channels". IEEE Signal Processing Letters, May 2006

[IE16e04] IEEE Std 802.16-2004, "Part 16: Air Interface for Fixed Broadband Wireless Access Systems”, IEEE standard institution, NewYork, USA, 2004.

[Lar73] K. J. Larsen, “Short convolutional codes with maximal free distance for rates ½, 1/3 and 1/4”, *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 371-372, May 1973.

[LR+06] T. Lestable, and M. Ran, et al. “ Error Control Coding Options for Next Generation Wireless Systems”, White Paper WG4, WWRF#17, 2006.

[Luo05] J. Luo, et al., Affecting Factors for Joint Radio Resource Management and a Realization in a Reconfigurable Radio System, WWRF 13th Meeting, Jeju Island, Korea, 02-03, March, 2005.

[LZ04] T. Lestable and E. Zimmerman, “LDPC Options for Next Generation Wireless Systems”, WWRF#15, Paris

[MBL03] H. Minn, V. K. Bhargava, and K. B. Letaief, “A Robust Timing and Frequency Synchronization for OFDM Systems”, IEEE Transactions on Wireless Communications, vol. 2, no. 4, July 2003.

[MP06] C. Mensing and S. Plass, “Positioning Algorithms for Cellular Networks Using TDoA”, Proc. ICASSP, May 2006.

[MS90] R.E. Mirollo and S.H. Strogatz, “Synchronization of pulse-coupled biological oscillators”, *SIAM J. APPL. MATH*, 50(6):1645–1662, Dec. 1990.

[MY05] S. Myung and K. Yang, “Extension of quasi-cyclic LDPC codes by lifting”, in Proc. IEEE Int. Symp. Information Theory 2005, pp. 2305-2309.

[MYK05] S. Myung, K. Yang, and J. Kim, “Quasi-Cyclic LDPC codes for fast encoding”, IEEE Trans. Inform. Theory, vol. 51, pp. 2894-2901, Aug. 2005.

[Racz07] Racz, Andras; Temesvary, Andras; Reider, Norbert; “Handover Performance in 3GPP Long Term Evolution (LTE) Systems”, Mobile and Wireless Communications Summit, 2007. 16th IST, 1-5 July 2007

[RSU01]

[Sachs06] Sachs, J.; Khurana, B.S.; Mahonen, P.; “Evaluation of Handover Performance for TCP Traffic Based on Generic Link Layer Context Transfer”, Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium on, Sept. 2006

[Simonsson05] Simonsson, A.; Peisa, J.; Pettersson, J.; “Analytic Study of TCP Performance over a Soft Rate Switching WCDMA Bearer”, Personal, Indoor and Mobile Radio Communications, 2005. PIMRC 2005. IEEE 16th International Symposium on, Volume 1, 11-14 Sept. 2005 Page(s):326 – 330

[SC97] T. M. Schmidl and D. C. Cox, “Robust Frequency and Timing Synchronization for OFDM”, IEEE Transactions on Communications, vol. 45, no. 12, December 1997.

[SFF+07] T. Svensson, T. Frank, D. Falconer, M. Sternad, E. Costa and A. Klein, ”B-IFDMA - A Power Efficient Multiple Access Scheme for Non-frequency-adaptive Transmission.”, *IST Mobile and Vehicular Summit* 2007, Budapest, July 2007.

[SS06] S. Sezginer, H. Sari, “OFDM Peak Power Reduction with simple amplitude predistortion”, *IEEE Comm.Letters*, vol. 10, issue 2, pp. 65-67, February 2006

[SSO+07] M. Sternad, T. Svensson, T. Ottosson, A. Ahlén, A. Svensson and A. Brunstrom, ”Towards systems beyond 3G based on adaptive OFDMA transmission,” *Proceding of the IEEE*, Special Issue on Adaptive Transmission, Vol. 95, no. 12, December 2007.

[STK05] A. H. Sayed, A. Tarighat, and N. Khajehnouri, “Network-Based Wireless Location”, IEEE Signal Processing Magazine, Vol. 22, No. 4, July 2005.

[TS36300] TS 36.300 V8.0.1, Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2, July 2007

[VT03] P. Viswanath and D. Tse, “Sum capacity of the vector Gaussian channel and uplink-downlink duality,” *IEEE Trans. Inform. Theory*, vol. 49, pp. 1912–1921, Aug. 2003.

[Vau88] R. G. Vaughan, "On optimum combining at the mobile," in *IEEE Trans. Vehicular Technol.*, vol. 37, no. 4, Nov. 1988.

[WIN1D21] IST-2003-507581 WINNER, “D2.1Identification of radio-link technologies”, June, 2004.

[WIN1D24] IST-2003-507581 WINNER, “D2.4 Assessment of adaptive transmission technologies”, February, 2005.

[WIN1D26] IST-2003-507581 WINNER, “D2.6 Assessment of Multiple Access Technologies”, October 2004.

[WIN1D27] IST-2003-507581 WINNER, "D2.7 Assessment of advanced beamforming and MIMO technologies", Ed. M. Döttling, Feb. 2005.

[WIN1D35] WINNER I Deliverable D3.5 "Proposal of the best suited deployment concepts for the identified scenarios and related RAN protocols", December 2005

[WIN1D29] IST-2003-507581 WINNER, "D2.9 Implementation Impact of Candidate Key Technologies", June 2005

[WIN1D24] IST-2003-507581 WINNER, "D2.4 Assessment of adaptive transmission technologies", February, 2005.

[WIN1D210] IST-2003-507581 WINNER, "Final Report on identified RI key technologies, System Concept, and their assessment", 2005

[WIN1D63] WINNER I, "D6.3: WINNER Spectrum Aspects: Assessment reports", 2005.

[WIN1D71] IST-2003-507581 WINNER, "D7.1 System Requirements", July 2004

[WIN1D76] IST-2003-507581 WINNER, "D7.6, WINNER System Concept Description", Oct. 2005.

[WIN2D112] IST-4-027756 WINNER II., D1.1.2 "WINNER II Channel Models, Part I Channel Models", September, 2007.

[WIN2D221] IST-4-027756 WINNER II "D2.2.1 Joint Modulation and Coding Procedures", October 2006.

[WIN2D223] IST-4-027756 WINNER II, D2.2.3, "Modulation and Coding Schemes for WINNER-II System", November 2007

[WIN2D233] IST-4-027756 WINNER II., D2.3.3, „ Link Level Procedures for the WINNER System", November 2007.

[WIN2D341] IST-4-027756 WINNER II, D3.4.1, "The WINNER II Air Interface: Refined Spatial-Temporal Processing Solutions," November 2006.

[WIN2D351] WINNER II Deliverable D3.5.1 "Relaying concepts and supporting actions in the context of CGs", October 2006

[WIN2D352] WINNER II Deliverable D3.5.2 "Assessment of relay based deployment concepts and detailed description of multi-hop capable RAN protocols as input for the concept group work", June 2007

[WIN2D353] WINNER II Deliverable D3.5.3 "Final assessment of relaying concepts for all CGs scenarios under consideration of related WINNER L1 and L2 protocol functions", September 2007

[WIN2D461] IST-4-027756 WINNER II "D4.6.1 The WINNER II Air Inteface: Refined multiple access concepts", November 2006.

[WIN2D471] IST-4-027756 WINNER II deliverable D4.7.1 "Interference averaging concepts", June 2007.

[WIN2D472] IST-4-027756 WINNER II deliverable D4.7.2 "Interference avoidance concepts", June 2007.

[WIN2D473] IST-4-027756 WINNER II deliverable D4.7.3 "Smart antenna based interference mitigation", June 2007.

[WIN2D481] WINNER II, "D4.8.1: WINNER II Intramode and Intermode Cooperation Schemes Definition", Deliverable, IST-4-027756 WINNER II, June 2006.

[WIN2D482] WINNER II, "D4.8.2: Cooperation Schemes Validation", Deliverable, IST-4-027756 WINNER II, June 2007.

[WIN2D483] WINNER II, "D4.8.3: Integration of cooperation on WINNER II System Concept", Deliverable, IST-4-027756, Nov. 2007

[WIN2D591] IST-4-027756 WINNER II, "Flexible Spectrum Use Functionalities for WINNER II", 2007

[WIN2D592] IST-4-027756 WINNER II, “Implementation of efficient spectrum sharing and coexistence capabilities for WINNER II”, 2007

[WIN2D593] IST-4-027756 WINNER II, “WINNER II spectrum efficiency and verifications”, 2007

[WIN2D6114] IST-4-027756 WINNER II, D6.11.4 v1.0, Final WINNER II System Requirements

[WIN2D6134] IST-4-027756 WINNER II “D6.13.4, “Intermediate concept proposal (metropolitan area) and evaluation”, November 2006

[WIN2D6135] IST-4-027756 WINNER II “D6.13.5 Intermediate concept proposal (local area) and evaluation”, November 2006

[WIN2D6137] WINNER II, “D6.13.7: Test Scenarios and Calibration Cases Issues 2”, Deliverable, IST-4-027756, December 2006.

[WIN2D6138] WINNER II Deliverable D6.13.8 “Intermediate WINNER II concept”, November 2006

[WIN2D61310] IST-4-027756 WINNER II, "D6.13.10 Final CG "wide area" description for integration into overall System Concept and assessment of key technologies," November 2007

[WIN2D61311] IST-4-027756 WINNER II, "D6.13.11 Final CG "metropolitan area" description for integration into overall System Concept and assessment of key technologies," November 2007

[WIN2D61312] IST-4-027756 WINNER II, "D6.13.12 Final CG "local area" description for integration into overall System Concept and assessment of key technologies," November 2007

[WSS06] H. Weingarten, Y. Steinberg and S. Shamai (Shitz), “The capacity region of the Gaussian multiple-input multiple-output broadcast channel,” *IEEE Trans. Inform. Theory*, vol. 52, pp. 3936–3964, Sept. 2006.

[YC04] W. Yu and J. Cioffi, “Sum capacity of Gaussian vector broadcast channels,” *IEEE Trans. Inform. Theory*, vol. 50, pp. 1875–1892, Sept. 2004.

[YG05] T. Yoo and A. Goldsmith, “Optimality of zero-forcing beamforming with multiuser diversity”, in Proc. Of IEEE ICC, May 2005.

[YJG075] T. Yoo, N. Jindal, and A. Goldsmith, Multi-Antenna Broadcast Channels with Limited Feedback and User Selection, To appear: *IEEE Journal Sel. Areas in Communications*, 2007.

[YLCC02] B. Yang, K.B. Letaief, R.S. Cheng and Z. Cao, “Timing Recovery for OFDM Transmission”, *IEEE Journal on Selected Areas in Comm.* Vol. 18, no. 11, pp. 2278-2291, Nov. 2002

A Annex: Detailed technical descriptions of layer functionalities

A.1 Frame control: Signalling and overhead.

The downlink out-band control signalling that is needed to control each frame contributes a significant part of the total control overhead. This Section describes a systematic framework for reducing and controlling this overhead. The essence is that only a small parts of the total information is broadcast to all users, since broadcasts are expensive in terms of transmission resources: Their modulation and code rate must be adjusted to the worst potential user SINR. Sets of allocation tables are then transmitted to individual users or groups of users, with modulation and code rates adjusted to the SINRs of these users/groups.

A.1.1 Adaptive Control Signalling Concept

A particular advantage of OFDMA-based wireless systems is that opportunistic scheduling can be done in several dimensions, such as time, frequency, and space. Small portions of the overall radio resources, so-called resource elements (RE) or chunk layers can be individually and flexibly allocated to different users.

This allows fostering significant scheduling gains, but on the other hand requires informing the peer entity (e.g. the user equipment (UT) in case of downlink) about which chunk layers are allocated to him. Thus either excessive control signaling to express which *A* resource elements (RE) out of *N* total resource elements are assigned to a particular user will result or the flexibility of the adaptive scheduling must be reduced by limiting the possible assignments. The latter would, however, also restrict the available gain due to opportunistic scheduling.

An straight-forward bitmap encoding of all possible allocations of $A \in \{1,\ldots, N\}$ chunk layers out of *N* would require *N* bits to be sent to each user, i.e. the total information overhead could be up to N^2 bits (if each resource element is allocated to a different user).

Different approaches to limit control overhead are known, like bitmaps, combinatorial techniques, subband or tree approaches, as summarised in [3GPP07a]. All methods are evaluated for fewer chunks than available in WINNER. Furthermore, they reduce overhead by limiting the number of possible allocations in different ways. This reduces the achievable scheduling gain. Furthermore, these algorithms are based on fixed length of control information elements and of the total control information channel allocation. They therefore do not adapt and scale with largely varying operational scenarios with respect to number of users, user data rate, etc. – which is of utmost importance for WINNER.

Depending on the current operation point of a cell, the WINNER system needs to maintain low control signaling overhead for few high-rate users, as well as for many low-rate users. Due to the increased spectral efficiency and bandwidth, the number of users that a scheduler can immediately assign resources (called active users in the following) will increase considerably. Therefore also the design of the control channel needs to scale with these different operation conditions and maintain low overhead and at the same time high flexibility of resource assignment.

Most beneficial encoding of resource map depends on the number of users and the distribution of chunks to them. Therefore the WINNER control channel design is based on optimised flexible-length control information, which reduces overhead by a combination of slow and fast control signalling, of broadcast and multicast signalling, as well as of individual and table-based allocation information.

The approach

- allows flexible configuration of the control information that allows to adopt and optimise overhead to a large range of operation conditions (e.g. ranging from few high-rate users to many low-rate users, full load, low load),
- supports switching between individual and table-based indication of allocated chunk layers depending on the operation conditions,
- saves overhead due to mapping of users into different *control groups* based on their average SINR, which maintains the advantages of multicasting information, but uses different

modulation and coding of control information, therefore avoiding that the total overhead is dominated by the strongly-encoded information sent to users in bad conditions,
- minimises the information that needs to be decodable by all user terminals, while keeping full flexibility of resource element (chunk) allocation,
- applies adaptive length of control information that allows re-use of left-over physical resources for data transmission, i.e. no waste of resources due to pre-determined overall length of the control information,
- allows efficient channel coding of the resource allocation and transport format information.

The control signalling is organized in several parts:
- "a priori" knowledge on basic information **BI** and constants,
- a slow broadcast configuration table **CT,**
- an optional broadcast control message length indicator **LI,**
- multicast allocation tables **AT** per control group, and
- multicast transport format tables **TFT** per control group.

The following basic information **BI** is known a priori, e.g. by information at cell association, other signalling with slow update rate, or by a fixed pre-determined rule:

- number of control groups N_{CG}, i.e. groups that receive multicast control information with different modulation and coding MCS_{CG}
- maximum number of users per control group $N_{max,i}$, i = 1, …, N_{CG}. Unicast control signalling can be configured as a special case with many groups each containing only one connection
- maximum number of chunk layers allocated to one particular user in each control group $K_{max,i}$
- the size (i.e. number of bits) of the pointers defined further below
- transmission formats, physical resources and their sequence of usage for downlink control information. This includes the configuration of the control message, i.e., whether *case A* or *case B* described below is used in the cell. *Case A* refers to a per-user information, which chunk layers are allocated to him, whereas *case B* is a table-based approach using user indices indicating which chunk layers are allocated to different users. It will be shown, that depending on the operation conditions either *case A* or case *B* shall be used and only the combination of both possibilities allows small overhead in all conditions.
- mapping of the user to a control group (based on the users' average channel quality) and its particular index in this group, i.e. index *ij* which serves as a short user ID
- the length of control channel information elements with fixed length,
- information on the total number of resource elements R_i for each control group in the cell

The following configuration table **CT** is jointly encoded, protected (e.g. by a cyclic redundancy check CRC), and broadcast in the cell at a timescale comprising at least one, typically many resource allocation time steps (called slots in the following):

- the actual number of users in each control group N_i (requires ceil($\log_2(N_{max,i})$) information bits per entry), i.e. the number of users that can be scheduled in this control group
- optionally the actual number of maximum resource elements allocated to one particular user in each control group K_i (requires ceil($\log_2(K_{max,i})$) information bits per entry. This option allows to further reduce the size of the following tables, in particular by reducing the number of resource allocation possibilities that need to be signalled. If this option is not used, the following applies by setting $K_i = K_{max,i}$.
- Due to the fact that this table needs to be decodable for all users in the cell, it needs strong protection (i.e. high coding overhead). The information in table CT has been tailored such, that it provides minimum information length but allows high savings for the following tables.

The main purpose of the **CT** is to distribute basic information on the following control-group specific information with minimal number of information bits. In particular the actual number of users in each control group allows tailoring and reducing the size of the following allocation table **AT**.

An optional control message length indicator **LI** is broadcast every slot. It contains a pointer *ptr_dstart* to the first entry of the pre-defined physical resources for downlink control information, which is unused for control purposes. Starting from this entry these resources will then be used for transmission of data. In most of the cases the additional overhead due to broadcasting the **LI** will be less than the achievable savings due to re-use of left-over resource elements and the use of **LI** is therefore recommended.

However, whether **LI** is used can be part of the cell configuration. The **LI** needs also to be decodable for all users, i.e. it requires strong protection. As it needs to be sent every slot, especially this kind of information has been minimized. The major benefit is that **LI** enables flexible length control information and efficient use of the remaining radio resources for data. This allows adaptation to a wide range of operational scenarios.

For each of the configured control group an allocation table $\mathbf{AT}_i$ is jointly encoded and protected (e.g. by a CRC). Each $\mathbf{AT}_i$ uses its particular modulation and coding MCS_i that allows all users in control group i to decode the information. This allows to the system to maintain the efficiency of multicasting, while avoiding that strong coding (needed for the users with bad SINR) is required for all information. The encoded allocation tables $\mathbf{AT}_i$ are written sequentially in the pre-defined physical resources, the length of each table can be determined by any user from the information contained in the **CT**. Each user can therefore determine which part of these resources contains the **AT** for his control group.

The content of one $\mathbf{AT}_i$ is as follows:

- a pointer ptr_tft_i to the physical resource where $\mathbf{TFT}_i$ starts, thus allowing all users of one control group to retrieve the remaining control information without the need to be able to decode $\mathbf{AT}_k$ with $k \neq i$ (i.e. the **AT**s of the other control groups). As the total length of all **AT**s is known based on the information contained in the **CT** the position where $\mathbf{TFT}_1$ starts is also known and therefore for control group 1 this pointer needs not be be signalled explicitly. In a preferred implementation, therefore the pointer to the starting point of $\mathbf{TFT}_i$ is only used for $i > 1$. The pointer approach allows flexible length of control information and therefore allows adaptation to a wide range of total number of scheduled connections
- in *case A*, where individual resource mapping is used: N_i entries $k_{i,j}$ defining the number of RE allocated to user with index j in control group i. Each entry has ceil($\log_2(K_i)$) information bits in case K_i is signalled with CT, $K_{max,I}$ otherwise. $k_{i,j} = 0$ for users that are not scheduled in the particular slot

The **AT** contains not only information about which users are scheduled, but additionally, how many resources are allocated to a particular user. This allows efficient compression of the transport format information contained in the subsequent **TFT**. In particular the information *which* resources are allocated to a particular user and the adaptive modulation information per RE can be efficiently reduced as explained in the sequel.

The transport format tables $\mathbf{TFT}_i$ contain the necessary information for each user ij on:

- which RE are allocated,
- mapping of codeblocks to RE,
- transport format of the codeblocks,
- HARQ information (one HARQ channel may contain one or several codeblocks,
- etc.

Many particular implementation of the TFT format are possible. The key idea, however, is to benefit from the following a priori knowledge provided by the preceding tables, which can be done in the following ways

- Entries are only generated for scheduled users, i.e. $k_{i,j} > 0$,
- In *case A*, where individual resource mapping is used: The actual RE allocated can be signalled very efficiently for each users, since it is a priori known how many resources $k_{i,j}$ are allocated per user. This means that even for full flexibility of resource allocation only all possible combinations of $k_{i,j}$ out of R need to be signalled. Therefore the length of the resource allocation information by (ceil ($\log_2(k_{i,j,} R_i$)), where R_i the total number of RE that can be used in each control group. The length of the resource allocation information field will therefore be flexible. Case A is in particular relevant for relatively large RE allocations to few users

- In *case B*, where a table-based resource mapping is used, the length of each entry in the table can be further reduced by introducing shorter short-hand ids using an index of only the scheduled users, since this information is available to all users in one control group.

- In *case C*, which will be used to for control signalling of non-frequency-adaptive transmission, further compression is possible due to the inherent structure of B-I/EFDMA In non-frequency-adaptive transmission each RE is further divided into resource blocks (RB). B-I/EFDMA transmission allows to express the exact allocation of the RB for one user by the RB allocation within one RE (this can be a simple index out of the limited total amount of possibilities), the start index of the RE, the number of RE, and the interval between adjacent REs.

Figure A-1 shows the described resource mapping alternatives for an instructive example of 8 resource elements (RE) and 4 scheduled users. In case A, the AT contains information how many RE $k_{i,j}$ are allocated to user *j* in control group out of the resources R_i available in control group *i*. This is written sequentially for all user ids in the **AT.** In the example user 1 gets 2 resources, user 2 is not scheduled at all, user 3 will be allocated 1 resource, etc. This requires 8 x 3 bits = 24 bits (8 users each having a 3 bit entry, since there are 8 RE). The knowledge of the scheduled users within the control group, allows to establish a subindex, which only contains those. I.e. user 1 will get a short-hand id of 1, user 3 the short-hand id of 2, user 6 short-hand id 3, and user 8 the short-hand id 4. Therefore only 2 bit short-hand ids are required.

In the resource mapping in the **TFT**, the allocated resources are signaled for the 4 scheduled users in sequential fields. Each field contains an index into an ordered list of any possible combination to allocate $k_{i,j}$ resources out of R_i. The first scheduled user is allocated RE 1 and 3. As there are 28 possibilities for 2 out of 8, 5 bit would be required to signal the index into an ordered list of any possible combination of 2 out of 8. The same calculation is done for all users, resulting in a total length of the resource mapping of 19 bits. The total overhead related to resource mapping is thus 24 bits + 19 bits = 43 bits.

In case B, the AT does not contain overhead related to the resource mapping (pointer and CRC length are identical in both cases and not included in the resource mapping overhead comparison here). The TFT on the other hand consists of a sequence of entries, where each entry corresponds to on physical RE. The mapping of entry no. *x* to the physical RE is know to all users based on information contained in **BI** and **CT**. Each entry gives the user-id (i.e. 3 bits to distinguish 8 users), i.e. a total overhead of 8 x 3 bits = 24 bits is required.

Therefore in this particular example, case B would be preferable. However, as will be shown later, there are also situations, where case A results in significantly less overhead.

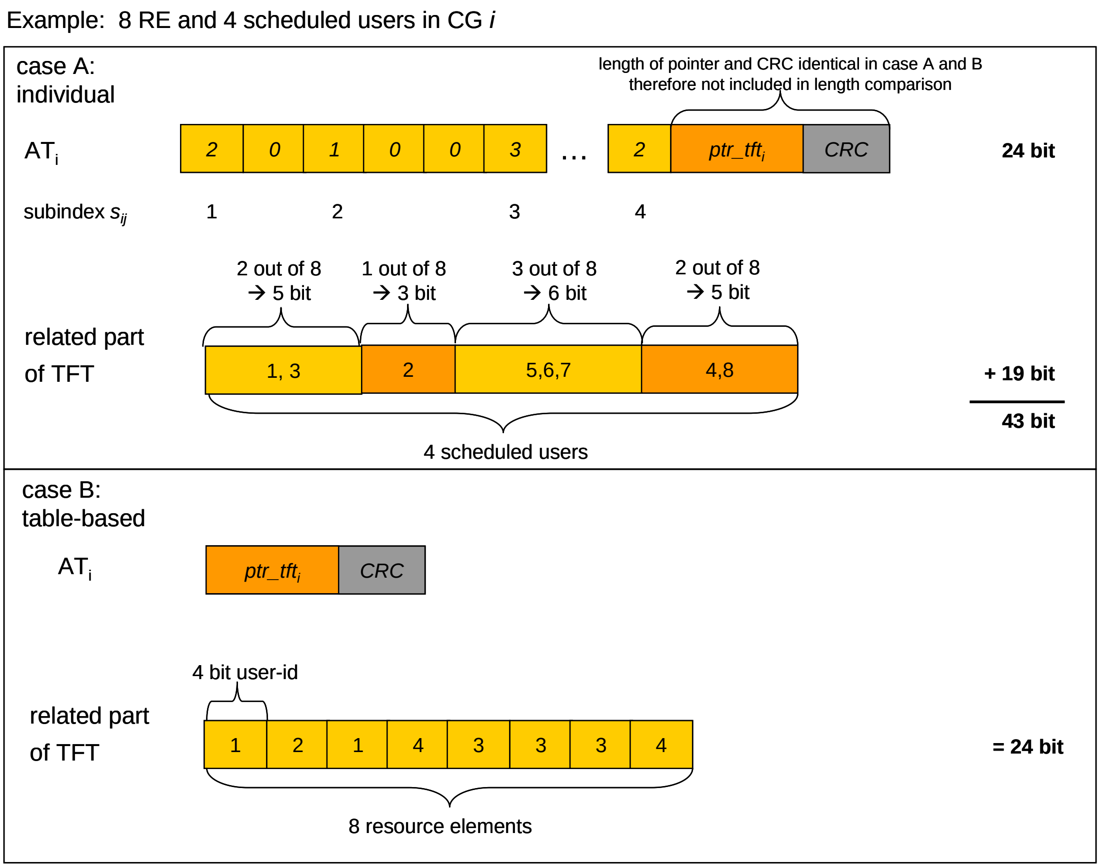

Figure A-1: Example to explain individual and table-based resource allocation table alternatives.

The modulation information for frequency-adaptive transmission might either be given explicitly per chunk layer (requiring 2 bit each) or based on a basic modulation, which is given once per codeword and then signalling of the difference in modulation for the individual chunk layers belonging to this codeword.

A sequential mapping of codeblocks to chunks is proposed, therefore it is sufficient to indicate the chunk where the current codeblock ends ("end chunk"). This end point can efficiently be signalled by using an index into the $k_{i,j}$ chunk layers allocated to user *ij*, i.e. it only requires ceil($\log_2(k_{i,j})$) information bits. If the "end chunk" field contains the number of the highest chunk allocated, the last codeblock of this user is reached and the subsequent information block belongs to the next scheduled user (otherwise information for the next codeblock of the current user follows). This implicit signalling of information block borders further reduces overhead. An alternative implementation might require that codeblocks of equal size are used only and therefore indicating the number of codeblocks is sufficient.

Figure A-2 provides a synopsis of the different elements of the control information and explains the use of the pointers. For the first transport format table no pointer ptr_tft_1 is required, since the starting point of this table can be determined by all users. *Ptr_dstart* can be read by all users. The entries readable for all users that are member of control group (CG) *i* are shown with bold boundaries.

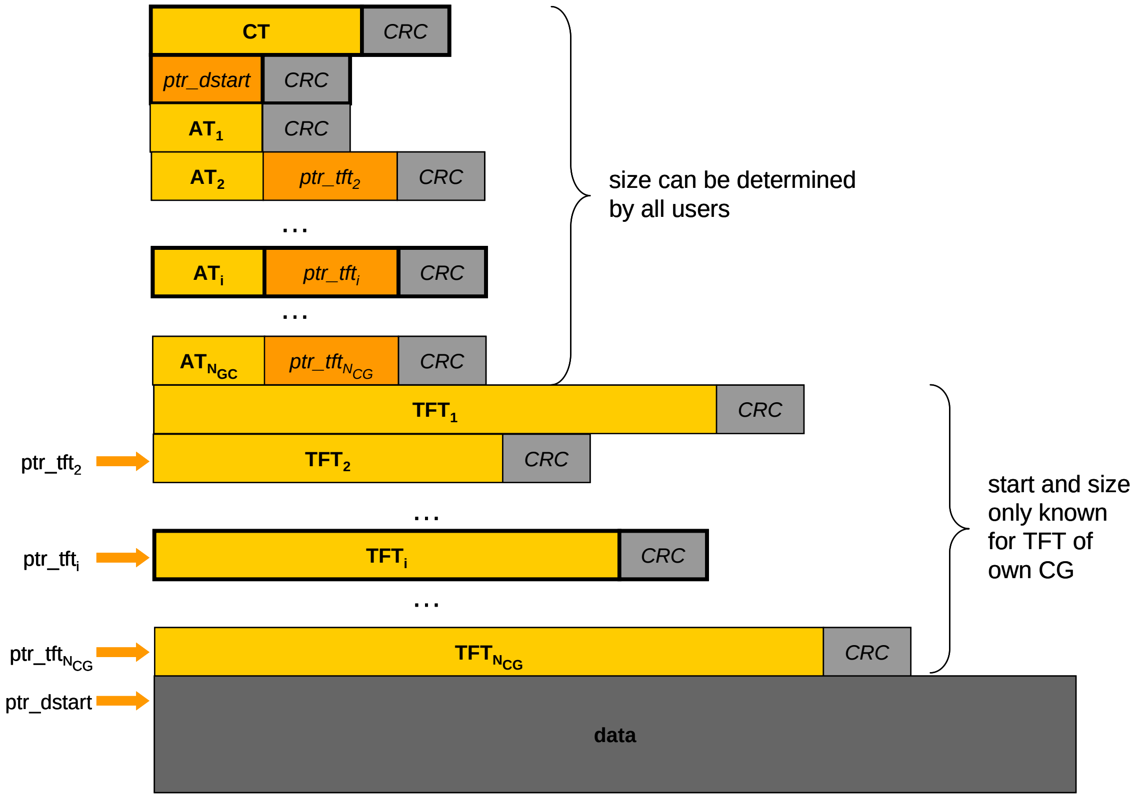

Figure A-2: Overview of the three-step control signaling using configuration table, allocation tables, and transport format tables.

A.1.2 Overhead Estimation of the Adaptive Control Signalling Concept

The benefits of this approach are demonstrated by the following assumptions in accordance with and derived from the frequency-adaptive transmission mode of the WINNER FDD and TDD physical layer mode, i.e. assuming R_{tot} = 144 RE per slot, 96 symbols per RE, 4 spatial layer per RE, 40 codewords (retransmission units with own HARQ-ID) per slot, and a CRC length = 12 bit.
Additionally some results are shown for the WINNER TDD physical layer mode, where R_{tot} = 230 RE and 120 symbols exist per RE.

The number of active users and control groups is varied in order to show that low control overhead is achieved in a wide range of operational scenarios.
The different tables described above have different requirements in terms of coding and temporal update rate. For the overhead calculation example the following assumptions (corresponding to an average operational case) were taken:

- **BI** overhead is signalled very rarely and therefore negligible,
- **CT** needs high protection as it needs to be decoded by all users, therefore 16 symbols/information bit are assumed, temporal update rate is assumed every 16^{th} slot (every 2 superframes in WINNER, i.e. still a high update rate every 11.5 ms), effective multiplication factor: 1 symbol/information bit/slot
- **LI** needs also high protection, but is sent every slot, effective multiplication factor: 16 symbols/information bit/slot
- **AT** and **TFT** information need to be sent every slot, however they have an optimized modulation and coding format within each control group. Averaged over the population of scheduled users it is assumed that the effective multiplication is 2 symbols/information bit/slot

The following implementation for the TFT format is investigated:

- asynchronous HARQ based on *n*-channel stop-and-wait protocol, supporting incremental redundancy and using
  - 2 bit for redundancy version
  - 1 bit for new data indicator
- 5 bit for code rate / transport block size
- 5 bit to describe the MIMO scheme used
- 2 bit description of basic modulation format of the code word and describes the actual modulation of the individual RE belonging to this codeword using differential encoding based on 3 states (up / same / down)
- stop index of chunk is given, i.e. variable size of codewords is supported

For the following configurable parameters an upper bound was used:

- $R_I = R_{tot}$, i.e. all resources can be used in all control groups (this means that full flexibility is kept, the concept of competition bands could be used additionally for further overhead reduction)
- a spatial scheme with 4-time re-use of RE in spatial domain is employed, thus introducing 4 spatial layers per RE
- modulation information is required for all spatial layers of each RE

For the variable-length information fields an average was assumed, in particular:

- $K_{max,I} = R_{tot} / N_{CG}$, i.e. the maximum number of RE allocated to one particular users is the total number of RE divided by the number of control groups
- the actually used number of RE k_{ij} is identical for all users, i.e. for a given number of RE per users, the total number of users is calculated by $N_u = \text{ceil}(R_{tot} / k_{ij})$
- it is assumed that the 40 codewords per slot are equally distributed amongst the users and the length of the HARQ-ID field is assumed to configured accordingly, i.e. $\text{ceil}(\log_2(40/N_u)$

The evaluations show different operation conditions for the WINNER FDD and TDD mode, ranging from only 8 active users to 1280 users that can be scheduled in the next slot and spans scenarios where many users get small allocations (few RE per users) up to allocations of a few high-rate users (many RE per users). For simplicity identical number of chunk layers are allocated to all users and the number of users is given implicitly by the quotient of available chunk layers and resource elements (= chunk layers per user).

Figure A-3 considers the overhead in a low load case, where only 8 users can be scheduled and 2 control groups are established. Figure A-3, left shows the total control information bits required for the resource allocation map, i.e. including all required information elements from **AT** and **TFT.** It can be seen that individual resource allocation is advantageous for few RE per user and for large allocations, whereas for intermediate size allocations, the table based approach results in minimum size of the resource allocation. Figure A-3, right shows that the total overhead is around 10% and is dominated by the **TFT.**

Figure A-4 considers a high load case with 320 active users organised in 5 control groups. In this case inidivual resource map signalling is beneficial in even more configurations, e.g. as soon more than 35 RE are allocated per user. When switching from the table-based approach to the individual signalling, the size of the allocation table is increased (cf. right hand side of Figure A-4), since information on the number of

RE allocated to each user is required additionally. However, this increase is more than compensated by the reduction of the required information in the following TFT. Total overhead is between 16% and 24% depending on the actual distribution of the resource allocation.

Figure A-5, finally shows some evaluation examples in a highly loaded TDD scenario where 640 or 1280 users can be scheduled immediately on all available resources. Even in this extreme scenario, overhead can be kept below 22%. In general the overhead fraction in TDD is lower, e.g. due to larger chunk sizes.

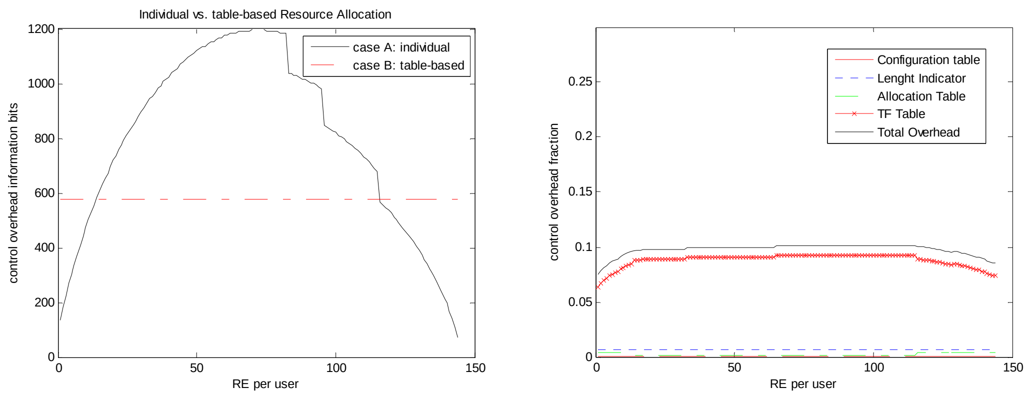

Figure A-3: FDD mode, 8 active users organized in 2 control groups: Comparison between overhead of individual and table-based resource map (left) and total overhead fraction (right)

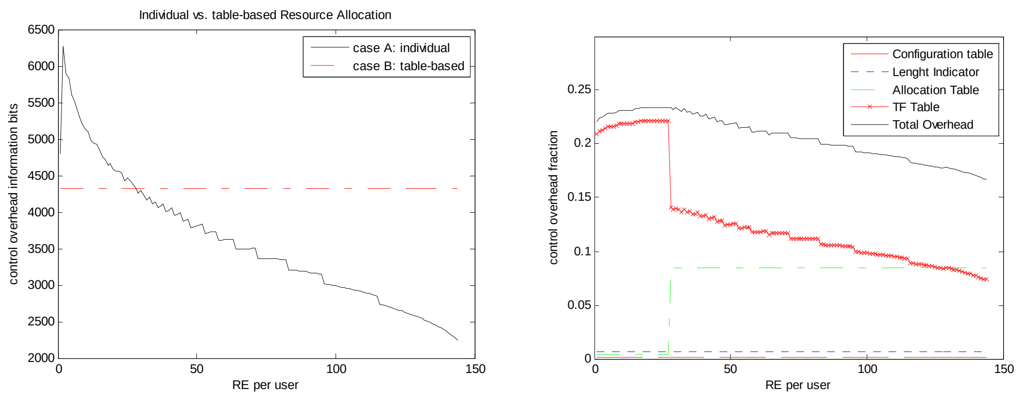

Figure A-4: FDD mode, 320 active users organized in 5 control groups: Comparison between overhead of individual and table-based resource map (left) and total overhead fraction (right)

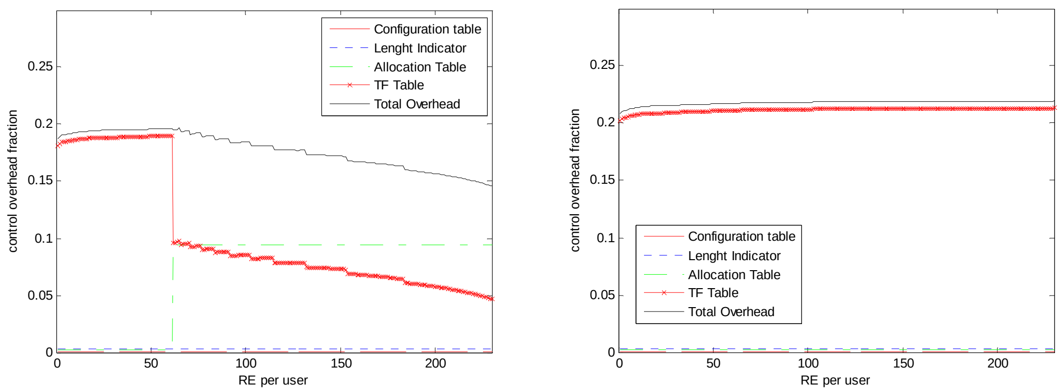

Figure A-5: TDD mode, Total overhead fraction for 640 active users (left) and 1280 (right) organized in 5 control groups.

These example show that control overhead an be kept low for both physical layer modes, from very few to more than 1000 users that can be scheduled, and for any configuration from many users with small allocations to few users (or a single user) with large allocations. This is achieved by using an optimised three-step signalling scheme with flexible length based on switching between table-based and individual resource allocation signalling. Such a control channel design is an important enabler for scalability and adaptivity of the WINNER system.

Further optimisations in the format of the transport format table might provide additional savings. Additional savings are possible by restricting the scheduler flexibility, e.g. using competition bands (restricting the allocation of a given transmission to a subset of the total resource pool).
While the adaptive control signalling concept optimises the overhead per control channel uses, the total number of required control channel usage can be kept low by exploiting coherence in time, frequency, and space [WIN1D29], as well as persistent scheduling techniques, where applicable. [3GPP07b, 3GPP07c, WIN2D61311].

These investigations focussed on the frequency-adaptive mode in WINNER, where modulation is adapted in each layer of the RE and irregularly dispersed allocation of the RE to one user is possible. In terms of control channel overhead, this mode can be considered worst-case. Overhead can significantly be reduced for the non-frequency-adaptive transmission, since there modulation information is only required once per codeword. Furthermore the use of regular RE allocation allows efficient encoding of the RE allocation. Despite of the partially deviating information required for non-frequency-adaptive transmission, the overall control channel design can be kept to a large degree similar.

A.2 Channel coding

A.2.1 Quasi-Cyclic Block LDPC Codes (Mandatory)

Among the increasing number of subsets of LDPC codes, only few of them are seen as serious candidates for next generation Wireless Systems ([LZ04], [LR+06]). Indeed for realistic future systems, many different constraints have to be taken simultaneously into account, such as e.g. performance, encoding and decoding complexity, decoder throughput (parallelism), resulting into what is called lately "Adequacy Algorithm Architecture" approach (**[Dor07]**).

Among those candidates, *Quasi-Cyclic Block Low-Density Parity Check Codes* (QC-BLDPCC[18]) are among the most promising ones ([Fos04]).

The full parity-check matrices for base-model matrices from agreed LDPC Codes can be found in Annex of [WIN2D223].

Structure and Scalability properties

The scalability property of LDPC Codes has been seen as a problematic, blocking feature for long time, compared with Turbo-Codes, till the notion of base-model matrix, or protograph was introduced. Indeed, thanks to a very simple but efficient 'Expansion' process, multiple codeword length can be generated by using only a unique base-model matrix, that can be seen as a 'mother' parity-check matrix (protograph).

These Block LDPC Codes are defined by sparse parity-check matrices of size $M \times N$ consisting of square submatrices (sub-blocks) of size $Z \times Z$ that are either zero or contain a cyclic-shifted identity matrix. *M* is the number of rows in the parity-check matrix, *N* is the code-length (number of columns) and the information size *K* is given by $K = N - M$.
These parity-check matrices are derived from the so-called base matrix $\mathbf{H_b}$ of size $m_b \times n_b$ and the expansion factor *Z*, which determines the sub-block size and hence the size of the derived code. I. e., from one base matrix different code lengths can be constructed using different expansion factors:

$$N = Z \cdot n_b \tag{A.1}$$

There is one base matrix specified per mother code rate:

$$R = K / N = 1 - m_b / n_b \tag{A.2}$$

The entries of the base matrix are integer values defining the content of the sub-blocks:

$$\mathbf{H_b} = (p_{ij})_{\substack{1 \le i \le m_b \\ 1 \le j \le n}} \tag{A.3}$$

In the *expansion process* each entry p_{ij} is replaced by a $Z \times Z$ square matrix that is:

- a zero matrix $\mathbf{0}_{Z \times Z}$ if $p_{ij} < 0$,
- or an identity matrix $\mathbf{I}_{Z \times Z}$ shifted to the right by $p_{ij} \bmod Z$, if $p_{ij} \ge 0$.

The base matrix always consist of a systematic part $\mathbf{H_s}$ and a parity part $\mathbf{H_p}$:

$$\mathbf{H_b} = \left[\mathbf{H_s} \mid \mathbf{H_p}\right] \tag{A.4}$$

Consequently a codeword c consists of a systematic part s and a parity part p:

$$\mathbf{c} = \left[\mathbf{s} \mid \mathbf{p}\right] = \left[s_1\, s_2 \ldots s_K \mid p_1\, p_2 \ldots p_M\right] \tag{A.5}$$

The parity part of the base matrix is in an *approximate lower-triangular* form:

[18] Alternatively abbreviated as BLDPCC or BLDPC codes in the following

$$\mathbf{H_p} = \begin{pmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & 1 & 1 & & 0 & & \\ & & 1 & \ddots & & & \\ & & & \ddots & \ddots & & \\ & & & & \ddots & 1 & \\ & & & & & 1 & 1 & \\ & & & & & & 1 & 1 \end{pmatrix}$$

Figure A-6: Parity part of the base matrix.

Further details concerning efficient encoding/decoding algorithms are given in the reference design section 7.

Rate Compatible Punctured LDPC Codes

Whilst dealing with advanced HARQ features, such as Type-II HARQ, also called Incremental Redundancy (IR), the key enabling characteristic from a channel coding point of view is to allow the use of single lower coding rate mother code, that will be punctured to obtain multiple higher coding rates. This code is then called ‘Rate Compatible’ Punctured code (RCPC).

This was also another key feature to be solved to enable LDPC Codes to become suitable candidates for Next Generation Wireless Systems. During this 2^{nd} phase of WINNER, we have developed very simple puncturing scheme resulting in an efficient QC-BLDPC Code, allowing many different coding rates with thin granularity, whilst still maintaining reasonable performance results.

Block LDPC codes are quasi-cyclic, i.e., a cyclic-shift by a number smaller than the sub-block size Z of a codeword by yields again a codeword. From the symmetry of the codes it follows that each bit within one sub-block is equally important for the decoder and hence, equally suitable for puncturing. It is therefore reasonable to define the puncturing pattern “sub-block-wise”.
For R = 1/2 base matrix given in Annex (A.3) from [WIN2D223] a set of puncturing patterns was optimized for the code-rates in region $\frac{24}{26} \leq R \leq \frac{24}{48}$. All these puncturing patterns are described by the *priority vector* P:

$$\begin{aligned}\mathbf{P} = [&1,2,3,4,5,6,7,10,11,12,14,15,16,17,18,19,20,21,23,24,30,34,38,41,42,...\\ &13,29,46,8,32,25,37,40,44,48,27,45,33,35,36,47,31,28,26,39,9,43,22\,]\end{aligned} \quad \text{(A.6)}$$

The priority vector P gives the order in which sub-blocks of the codeword should be sent in a HARQ process. It can be used to define an interleaver in order to implement arbitrary punctured code-rates elegantly.

Further details about performance results can be found in [WIN2D223].

A.2.2 Low-Rate Convolutional Code for BCH (Mandatory)

The modulation and coding requirements for control channel signalling are different than the ones for user data transmission. The information sent through the control channel is very important for proper functioning of the advanced protocols of the WINNER concept. Although the proposed BLDPCC and DBTC provide an excellent coding performance as shown in [WIN1D210], they can not be used for encoding the control information due to very short packet sizes being considered (25 information bits). Therefore low rate convolutional codes (CC), which can be used for encoding of such a short packets by choosing a *tail-biting* algorithm, are still considered for the WINNER reference design (CC were already proposed in Phase I, cf. [WIN1D210]).

Instead of the *maximum free distance* (MFD) convolutional code [Lar73] defined in the previous proposal for the WINNER reference design with $R = 1/3$ and $G_A = [575, 623, 727]_{oct}$, one of the *optimum distance spectrum* (ODS) convolutional codes [FOO+98] with $R = 1/4$ can be used, with the following generator polynomials: $G_B = [473, 513, 671, 765]_{oct}$.

Higher Coding rates are then obtained through usual puncturing of the mother code.
Full comparison and performance results details can be found in [WIN2D223].

A.2.3 Duo-Binary Turbo-Codes (Optional)

Whilst the overall evaluation process carried out during the second phase of WINNER, ended up with selecting the Quasi-Cyclic Block-LDPC (QC-BLDPC) Codes as primary and mandatory coding scheme for data transmission, the use of Duo-Binary Turbo-Codes (DBTC) can still a suitable candidate for medium block length and low coding rates, thus resulting in keeping this scheme as an optional alternative within WINNER Concept.

The block diagram in Figure A-7 shows a parallel concatenated convolutional encoder and the corresponding iterative decoder. The message u is encoded twice: directly by the encoder X_1 and a permuted version of the message by the encoder X_2. Both encoded bitstreams as well as the message itself are transmitted. At the receiver side, each coded bitstream is decoded separately by a *soft-in soft-out* decoder and the obtained information is used by the other decoder, which in turn returns new *extrinsic* information to the first decoder. After several iterations, the obtained L-values are mapped to an estimate of the message u.

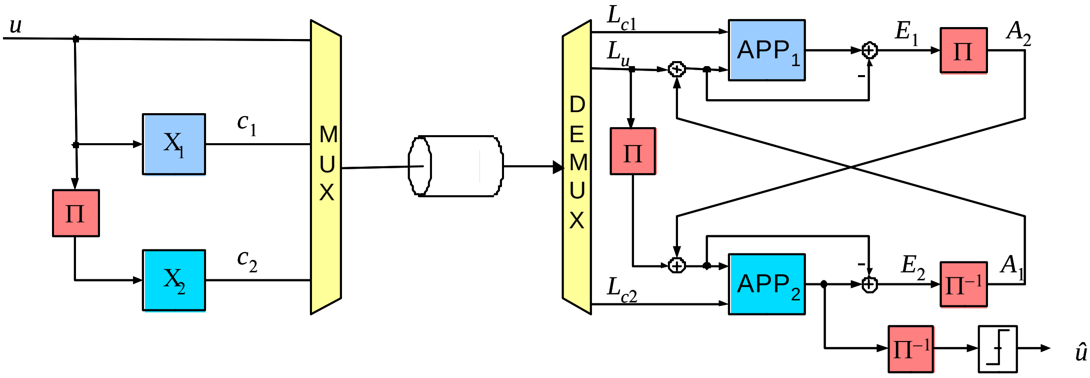

Figure A-7: Block diagram of a parallel concatenated turbo code.

Duo-binary turbo codes are used in several standards, e.g. [ETSI02], and have been found to offer very good performance in conjunction with higher-order modulation [BJD01].

The Component Code

The main enhancement from the original turbo codes to the DBTC lies in the component codes, which encode two bits at a time. As usual for parallel turbo codes, both component codes are identical. The term “duo-binary” is somewhat misleading since the component codes are still binary convolutional codes (only the number of input bits per transition is $k_0 = 2$) and all operations are carried out in the binary field GF(2).

The specific generator polynomial used within the 2nd phase are given hereafter (Eq. 4.7):

$$\mathbf{G}(D) = \begin{pmatrix} 1 & 0 & \dfrac{1+D^2+D^3}{1+D+D^3} & \dfrac{1+D^3}{1+D+D^3} \\ 0 & 1 & \dfrac{1+D+D^2+D^3}{1+D+D^3} & \dfrac{1+D^2}{1+D+D^3} \end{pmatrix} \tag{A.7}$$

As in all parallel turbo codes, the component codes are *recursive*. One of the features of turbo codes is that the component codes are relatively simple codes with low memory. This is also true here since the component codes defined by G(D) have only $S = 8$ states.

Puncturing

The DBTC are described in [WIN2D221] in detail. These codes have been defined and evaluated already in phase I. However, the DBTC given in [WIN2D221] are not rate-compatible and the lowest code rate is 1/2, which is considered as too high for some scenarios. Therefore, the code rates 1/3, 2/5 and 4/7 have been added and all puncturing patterns for the DBTC are now rate-compatible according to the puncturing matrices in Table A-1.

The puncturing patterns in Table A-1 are applied to both component encoders.

Table A-1: Puncturing patterns for rate-compatible DBTC.

Rate	Puncturing matrix
1/3	$\mathbf{P}=\begin{pmatrix}1&1&1&1&1&1&1&1&1&1&1&1\\1&1&1&1&1&1&1&1&1&1&1&1\end{pmatrix}$
2/5	$\mathbf{P}=\begin{pmatrix}1&1&1&1&1&1&1&1&1&1&1&1\\1&0&1&0&1&0&1&0&1&0&1&0\end{pmatrix}$
1/2	$\mathbf{P}=\begin{pmatrix}1&1&1&1&1&1&1&1&1&1&1&1\\0&0&0&0&0&0&0&0&0&0&0&0\end{pmatrix}$
4/7	$\mathbf{P}=\begin{pmatrix}1&1&1&0&1&1&1&0&1&1&1&0\\0&0&0&0&0&0&0&0&0&0&0&0\end{pmatrix}$
2/3	$\mathbf{P}=\begin{pmatrix}1&0&1&0&1&0&1&0&1&0&1&0\\0&0&0&0&0&0&0&0&0&0&0&0\end{pmatrix}$
3/4	$\mathbf{P}=\begin{pmatrix}1&0&0&0&1&0&1&0&1&0&0&0\\0&0&0&0&0&0&0&0&0&0&0&0\end{pmatrix}$
4/5	$\mathbf{P}=\begin{pmatrix}1&0&0&0&1&0&0&0&1&0&0&0\\0&0&0&0&0&0&0&0&0&0&0&0\end{pmatrix}$
6/7	$\mathbf{P}=\begin{pmatrix}1&0&0&0&1&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0&0&0&0&0\end{pmatrix}$

Note: the puncturing patterns in Table A-1 have been derived from the ones given in [WIN2D221] such that the condition for rate-compatibility is satisfied. No special effort has been made to find the best rate-compatible puncturing patterns. It is therefore probable that other patterns lead to better performance; however, no significant differences are expected.

Further performance evaluation, together with detailed implementation of encoding and decoding algorithms of such alternative can be found in [WIN2D223].

A.3 Pilot structure

Pilots are used for implementing certain physical layer support functions, e.g. connection setup, synchronisation, mobility support, power control, CQI measurements and most importantly channel estimation. On the other hand, pilots add overhead and consume transmission power. Thus, a proper pilot design should enable accurate and reliable channel estimation and on the other hand keep the induced spectral and power efficiency loses at an acceptable level. In WINNER phase I pilot aided channel estimation (PACE) for OFDM was extensively studied and serves as a basis for the WINNER pilot design. For channel estimation purposes, the following means for **multiplexing pilots** are foreseen [WIN1D21], [WIN1D23] and [WIN1D210]:

- For OFDM downlink and uplinks a **scattered pilot grid** is used for channel estimation and channel prediction. The pilot grids are specified in Annex A.3.2.
- For uplink (frequency domain generated) serial modulation, pilot patterns may be generated in the frequency domain equivalent to OFDM, enabling the use of a scattered pilot grid. This establishes a common framework for the pilot grid of generalized multi-carrier (GMC) signals.

The WINNER system concept heavily relies on provision of accurate channel state information (CSI) at both receiver and transmitter. To this end, two types of channel estimation must be distinguished:

- Channel estimation for data reception, where the receiver needs to measure the effective channel (ECSI), including the effect of the spatial processing at the transmitter, referred to as effective CSI (ECSI).
- Channel estimation (or more accurately channel prediction) for adaptive transmit processing based on CSI at the transmitter (CSIT), typically provided through return link feedback or measurements, where an additional extrapolation/prediction in time is required.

One key issue for pilot design is the spatial transmit processing, which is selected based on the available CSI or CQI at the transmitter [WIN2D341]. The pilot design for any specific embodiment of spatial processing is a challenging task on its own, and the optimum choices for the specific spatial processing schemes may be fundamentally different. Moreover, various flavours of multiple antenna transmission schemes are to be flexibly combined with opportunistic multi-user scheduling, and link adaptation, within the same air interface. To this end, a straightforward combination of the individual best choices would result in a combination of a large number of types of pilots (dedicated / common pilots per beam / antenna), which inevitably leads to prohibitive overheads. Hence, the objective of the WINNER pilot design is to reuse pilots for as many different functions as possible.

The WINNER pilot design is a modular concept consisting of basic building blocks defined on the chunk level. These building blocks are [WIN2D233]:

- The **pilot pattern** specifies the position of pilots on the chunk. The pilot positions is chosen such that a globally regular pilot pattern is obtained, i.e. a two dimensional (2D) grid with equidistantly spaced pilots by D_f and D_t in time and frequency, which is advantageous for channel estimation by interpolation.
- The **pilot type** specifies whether pilots include user specific transmit processing or not.
- The **orthogonal pilot set** specifies whether pilots associated to different spatial streams are orthogonally separated in time and/or frequency, or pilots are spatially reused, i.e. pilots of two spatial streams are placed on the same subcarriers.

This modular concept avoids that several pilot patterns corresponding to different pilot types are inserted within a frame. Instead only one pilot grid is inserted in the frame and the pilot type is determined by the chunk specific spatial transmit processing. Thus, a highly flexible and adaptive system concept can be supported with a modest pilot overhead.

The pilot grids for FDD and TDD modes as well as the super-frame pilot preamble are described in Annex A.3.2 and A.3.3. Furthermore, the realization of the WINNER pilot design for the reference design of wide area, metropolitan area and local area deployment are addressed in Sections **7.4.1**, 7.4.2 and 7.4.3.

A.3.1 Types of Pilots for Multi-Antenna Transmission

A scattered pilot grid with orthogonally spaced pilot symbols in time and frequency was proposed in [WIN1D21] and [WIN1D210]. This however, may result in prohibitive pilot overheads. For instance, in local area deployment a distributed antenna array with up to 32 antenna elements is foreseen. Fortunately, spatial precoding schemes forming beams that are spatially well separated allow to spatially reuse pilot symbols [WIN2D233].

Figure A-8 illustrates a scattered pilot grid. Pilots associated to beams that are spatially well separated are multiplexed. For instance pilots 1 and 3 as well as 2 and 4 do not spatially overlap, so they are located on the same subcarrier, yielding a pilot reuse of two. On the other hand, pilots associated to beams with significant spatial overlap (in Figure A-8 pilots 1 and 2, 2 and 3, as well as 3 and 4) need to be orthogonally separated in the time or frequency, i.e. placed on a different subcarrier.

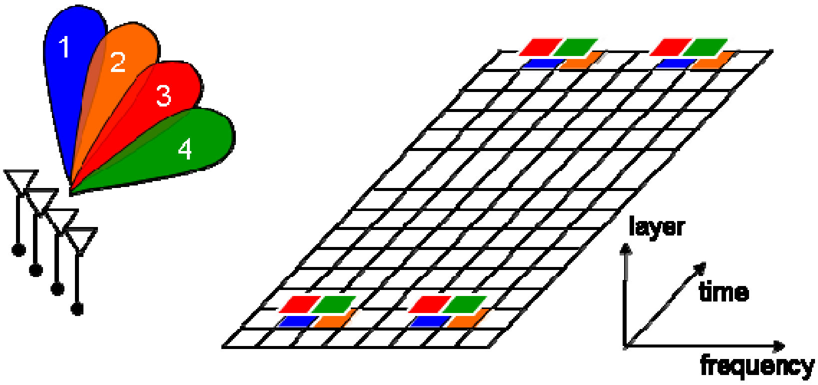

Figure A-8: Scattered pilot grid. Pilots are orthogonally separated in time and frequency. Furthermore, Pilots associated to beams that are well separated are spatially multiplexed.

Dependent on the transmit direction (uplink / downlink) and the kind of spatial processing being used, several types of pilots are being distinguished [WIN1D210]:

- ***Common pilots*** have the property not to include user-specific transmit processing and thus the interpolation in frequency is possible. In case of user-specific transmit processing, the amplitude and phase induced on common pilots by fading channel deviates from those of the data symbols which are induced by combination of user-specific transmit processing and fading channel, and therefore the receiver cannot detect those based on common pilots. Different variants of common pilots exist
  - ***Common pilots per antenna (CPA)*** are used to obtain the unweighted channel matrix **H** which describes the propagation channel between any combination of transmit and receive antennas in the MIMO case.
  - ***Common pilots per beam (CPB)*** are useful to estimate the effective channel (including the beamforming weights) and perform CQI measurements for the associated beam for fixed beamforming approaches. Note that measurements on such pilots in neighbouring beams could then be used for beam handover. Also, the common pilots per beam benefit from the beamforming gain, which reduces the transmit power required for a target channel estimation error and coverage area.
- ***Dedicated pilots*** may be required if user-specific transmit processing (i.e. a user-specific adaptation of amplitude and phase) is applied to the data symbols. These pilots are subject to the same transmit processing as the data symbols and therefore allow the receiver to estimate the effective channel. The use of dedicated pilots for other purposes, like CQI measurements, is limited, since they contain a user specific component, giving rise to biased measurements. Different types of dedicated pilots can be distinguished:
  - ***Dedicated pilots per antenna (DPA)***, are typically uplink pilots where the pilot symbols do not include the spatial processing, i.e. the beamforming weights.
  - ***Dedicated pilots per beam (DPB)*** are useful to estimate the effective channel (including the beamforming weights) and perform CQI measurements for the associated beam for adaptive beamforming approaches.
  - ***Dedicated pilots over the full bandwidth*** have identical weights for all chunks dedicated to a particular user. Therefore interpolation over these chunks is possible.

Both DPA-FB (full band) and DPB-FB exist, although the former is more common. DPA-FB and DPB-FB allow interpolation for all users, similar to common pilots. However, since pilots cannot be shared between users, the pilot overhead is typically about N_u times higher than the corresponding overhead for common pilots, where N_u is the number of users.

Due to the fact that common pilots can be used by several users, they are appealing for the downlink processing, since the overall energy to perform the associated functions has only to be spent once and the pilot symbols can be spread over all resources. Also they provide a basis for un-biased CQI measurements. However, certain user-specific spatial processing techniques require to estimate the effective channel (the channel including the spatial precoding), which is typically provided by dedicated pilots. To this end, when user-specific spatial processing is combined with multi-user OFDMA, even on the downlink an increasing amount of dedicated pilots is needed. On the other hand, dedicated pilots fail to deliver unweighted CSI and CQI estimates that are needed for adaptive transmission.

A.3.2 In-band pilot patterns

The pilot patterns for the FDD mode is illustrated in Figure A-9 and described in Section A.3.2.2.The pilot patterns for the TDD mode is illustrated in Figure A-10 and described in Section A.3.2.3 for the downlink and in Section A.3.2.4 for the uplink. Furthermore, the dedicated uplink pilot patterns for B-IFDMA are illustrated in Figure A-11 and described in Section A.3.2.5. The pilot spacings in frequency and time, D_f and D_t, as well as the corresponding overheads Ω_p of the WINNER pilot design are shown in Table A-2.

The pilot reference design for the FDD mode is described for the deployment wide area in Section 7.4.1. The pilot reference design for the TDD mode is described for the deployment metropolitan area and local area in Sections 7.4.2 and 7.4.3.

Table A-2: Pilot spacings and overheads for the WINNER pilot design. The overhead is given as a function of the number of orthogonal pilot sets P_n.

	FDD[19]	TDD[20]	B-IFDMA
D_f	4	4	4
D_t	10	12	3
Ω_p	$4.16\% \cdot P_n$, $(5.2\% \cdot P_n)$, P_n={1,2,3,4}	$3.33\% \cdot P_n$, $(1.67\% \cdot P_n)$, P_n={1,2,3,4}	$8.33\% \cdot P_n$, P_n={1,2}

Up to 4 orthogonal sets of pilots are allocated. Due to spatial reuse of pilots, the actual number of spatial streams can be significantly higher than 4. In many cases the number of spatial streams is below 4 and/or the associated beams are well separated; in this case, the number of orthogonal pilot sets, P_n, that are actually used may be smaller than 4. For instance, for SISO only one orthogonal pilot set is used, while for LDC with two antennas 2 orthogonal pilot sets are required.

In Table A-2 it is seen that the pilot spacing in frequency is always D_f = 4. This is due to the chunk dimension in frequency of 8 subcarriers. With D_f = 4 there are 2 pilots in frequency per chunk, and a globally regular pilot pattern is retained, where the pilot locations are independent relative to the start of the chunk. This is the key requirement for the WINNER pilot design, as it allows selecting the pilot type (common / dedicated pilot per beam / antenna) on the chunk level. A further advantage of having D_f an integer multiple of the chunk duration is for uplink transmission: if a user is allocated several adjacent chunks, interpolation over frequency on those chunks is possible in case of dedicated pilots per antenna. This is particularly beneficial for relay enhanced cells (REC), as a relay is forwarding data of several users on the uplink in a localized sub-band, so interpolation in frequency over that sub-band is possible.

[19] Pilot overhead in brackets correspond to chunks of high velocity users with speed exceeding 150 km/h, where additional pilots are inserted.

[20] Pilot overhead in brackets correspond to chunks of low velocity users with speed below 10 km/h.

Given the pilot indeces $n_p = \{1,\cdots,\lfloor N_c / D_f \rfloor\}$ and $\ell_p = \{1,2\}$ in frequency and time, the position of the pilot (subcarrier n and OFDM symbol ℓ) within the frame is determined by the following relation:

$$\begin{pmatrix} n \\ \ell \end{pmatrix} = \begin{pmatrix} D_\mathrm{f} & d_{0\mathrm{f}} \\ 0 & D_\mathrm{t} \end{pmatrix} \cdot \mathbf{d}_\mathrm{p} + \begin{pmatrix} d_{\mathrm{ort,f}} \\ d_{\mathrm{ort,t}} \end{pmatrix}, \quad \text{with} \quad \mathbf{d}_\mathrm{p} = \begin{pmatrix} n_p - 1 \\ \ell_p - 1 \end{pmatrix} \qquad \text{(A-1)}$$

where the pilot spacing in frequency is set to $D_f = 4$, while the pilot spacing in time is for FDD mode $D_t = 10$ and for TDD mode $D_t = 12$. In **(A-1)** the parameter $d_{0f} = \{-1,1\}$ specifies the shift in subcarriers between $\ell_p = 1$ and 2, while the vector $\mathbf{d}_{\mathrm{ort}} = [d_{\mathrm{ort,f}}, d_{\mathrm{ort,t}}]^T$ specifies the orthogonal separation of pilots associated to different transmit antennas or beams. The entries of $\mathbf{d}_{\mathrm{ort}}$ are within the range $\{d_{\mathrm{ort,f}}, d_{\mathrm{ort,t}}\} = \{1,2\}$. For the 4 orthogonal pilot sets the following parameters are chosen:

	Orthoghonal pilot set			
	1	2	3	4
d_{0f}	1	-1	1	-1
$d_{\mathrm{ort,f}}$	1	2	1	2
$d_{\mathrm{ort,t}}$	1	1	2	2

The pilot pattern from **(A-1)** allows pilots from multiple beams share the same orthogonal pilot set, which causes inter-beam interference. Moreover, interference from adjacent cells further corrupts the pilots. Hence, pilots originating from the various cells and beams that are spatially multiplexed are randomized through a cell *and* beam specific scrambling sequence, $\tilde{X}^{\mathrm{cell}}_{n,\ell}$ $\tilde{X}^{\mu}_{n,\ell}$, both of length $2 \cdot N_c / D_f$, where μ is the beam index, n and ℓ are the subcarrier and OFDM symbol position of the pilot from **(A-1)**. The scrambling does not remove the inter-beam and inter-cell interference, but decorrelates the pilots. In order to reduce the peaks inherent to multi-carrier signals, the pilot signal should preferably exhibit a uniform envelope and power spectrum. The recommendation for the pilot sequence is the DFT of a Chu sequence (or the DFT of a sequence with similar properties), as it produces a Chu sequence in the time domain.

A.3.2.1 Pilot boost

The effect of a pilot boost on the performance of a MIMO-OFDM system is investigated in [WIN2D233]. Only the link level was studied, so cellular interference is not taken into account. The optimum pilot boost is shown to be decreasing as the number of transmit antennas increase. In general, most of the attainable gains of a pilot boost are captured by setting the pilot boost to $S_p = 3$ dB.

In a cellular system the potential benefits of a pilot boost for the WINNER system appear limited. Only when in-cell pilots are guaranteed not to interfere with out-of-cell pilots a pilot boost was shown to provide substantial gains [WIN2D341]. Unfortunately, due to practical constraints it appears difficult to ensure that out-of-cell pilots will never interfere with in-cell pilots. For instance, dedicated pilots should be placed near the corners of a chunk, which greatly reduces the number of eligible positions to place pilots. In summary, the pilot boost should not exceed $S_p = 3$ dB in the WINNER system.

A.3.2.2 Pilots in FDD mode

On the downlink common pilots per antenna/beam (CPA and CPB) are used, while on the uplink dedicated pilots per antenna (DPA) are used. Since common pilots are not subject to user specific processing, interpolation in frequency is possible, and edge effects are less problematic. According to the WINNER pilot design, the selection of CPA or CPB is determined by the spatial processing within a particular chunk: for GoB we choose CPB, while for LDC we choose CPA. On the other hand, the pilot design is independent of the used multiple access scheme, i.e. whether OFDMA or B-EFDMA is used.

The pilot spacings in frequency and time are $D_f = 4$ and $D_t = 10$. The WINNER 2D grid for the FDD model is illustrated in Figure A-9 and the position for subcarrier n and OFDM symbol ℓ relative to the first symbol in the chunk at (1,1), denoting earliest OFDM symbol and lowest frequency within a chunk, are given in Table A-3. Note that in the case of LDC with 2 transmit antennas only two sets of orthogonal common pilots are required.

Table A-3: Pilot symbol location for the FDD mode relative to the first symbol in the chunk at (1,1). Up to 4 orthogonal pilot sets are supported[21]

	Orthoghonal pilot set							
	1		2		3		4	
$\begin{pmatrix} n \\ \ell \end{pmatrix}$	$\begin{pmatrix} 2 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 6 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 7 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 2 \end{pmatrix}$	$\begin{pmatrix} 6 \\ 2 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 2 \end{pmatrix}$	$\begin{pmatrix} 7 \\ 2 \end{pmatrix}$
	$\begin{pmatrix} 4 \\ 6 \end{pmatrix}$		$\begin{pmatrix} 5 \\ 6 \end{pmatrix}$		$\begin{pmatrix} 4 \\ 7 \end{pmatrix}$		$\begin{pmatrix} 5 \\ 7 \end{pmatrix}$	
	$\begin{pmatrix} 3 \\ 11 \end{pmatrix}$	$\begin{pmatrix} 7 \\ 11 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 11 \end{pmatrix}$	$\begin{pmatrix} 6 \\ 11 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 12 \end{pmatrix}$	$\begin{pmatrix} 7 \\ 12 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 12 \end{pmatrix}$	$\begin{pmatrix} 6 \\ 12 \end{pmatrix}$

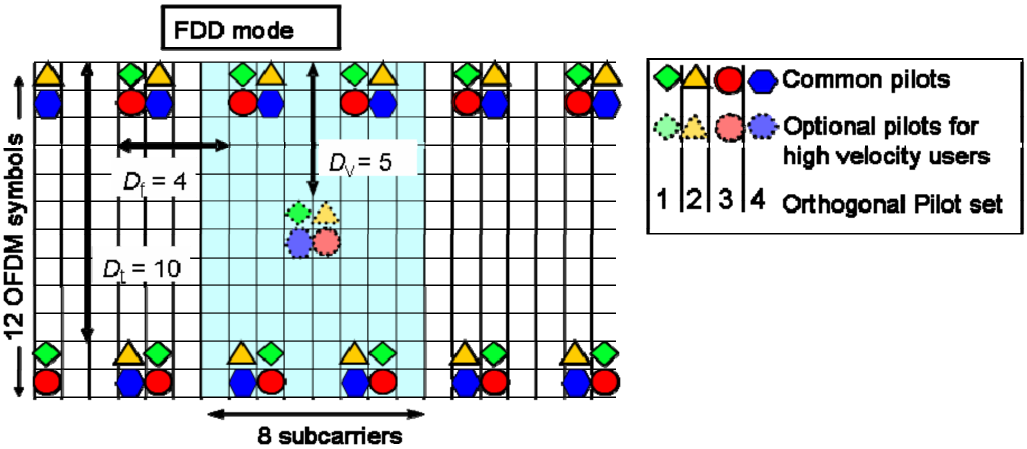

Figure A-9: Pilot grids for the FDD mode.

For chunks associated to high velocity users with speeds exceeding 150 km/h, one additional pilot per orthogonal pilot set is inserted to better track the high time variations on the channel response.

A.3.2.3 Downlink Pilots in TDD mode

For the TDD mode 4 pilots per chunk per orthogonal pilot set are arranged in a rhombus shape, as specified in and illustrated in Figure A-10. The spacing in the time direction D_t should be adjusted to the used uplink/downlink asymmetry ratio. With a link asymmetry ratio of 1:1, the pilot spacings become $D_f = 4$ and $D_t = 12$ in frequency and time. For low mobility users with velocities below 10 km/h, one pilot in time direction is sufficient, i.e. the pilots corresponding to pilot index $\ell_p = 2$ in Table A-4 may be omitted, which cuts the pilot overhead by a factor of 2.With spatial reuse of pilots, the $P_n = 4$ orthogonal pilot sets allow for a number of spatial layers of up to 32, with a modest pilot overhead of $\Omega_p = 13.3\%$ and 6.7% for mobile and pedestrian velocities.

Table A-4: Pilot symbol location for the TDD mode relative to the first symbol in the chunk at (1,1). Up to 4 orthogonal pilot sets are supported[22]

	Orthoghonal pilot set							
	1		2		3		4	
$\begin{pmatrix} n \\ \ell \end{pmatrix}$	$\begin{pmatrix} 2 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 6 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 7 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 2 \end{pmatrix}$	$\begin{pmatrix} 6 \\ 2 \end{pmatrix}$	$\begin{pmatrix} 3 \\ 2 \end{pmatrix}$	$\begin{pmatrix} 7 \\ 2 \end{pmatrix}$

[21] The pilots near the center of the chunk are optionally inserted for high mobility users with velocities > 150 km/h.

[22] For low mobility users with velocities ≤ 10 km/h, one pilot in time direction is sufficient, i.e. the pilots corresponding to pilot index $\ell_p = 2$ may be omitted

	$\begin{pmatrix}3\\13\end{pmatrix}$	$\begin{pmatrix}7\\13\end{pmatrix}$	$\begin{pmatrix}2\\13\end{pmatrix}$	$\begin{pmatrix}6\\13\end{pmatrix}$	$\begin{pmatrix}3\\14\end{pmatrix}$	$\begin{pmatrix}7\\14\end{pmatrix}$	$\begin{pmatrix}2\\14\end{pmatrix}$	$\begin{pmatrix}6\\14\end{pmatrix}$

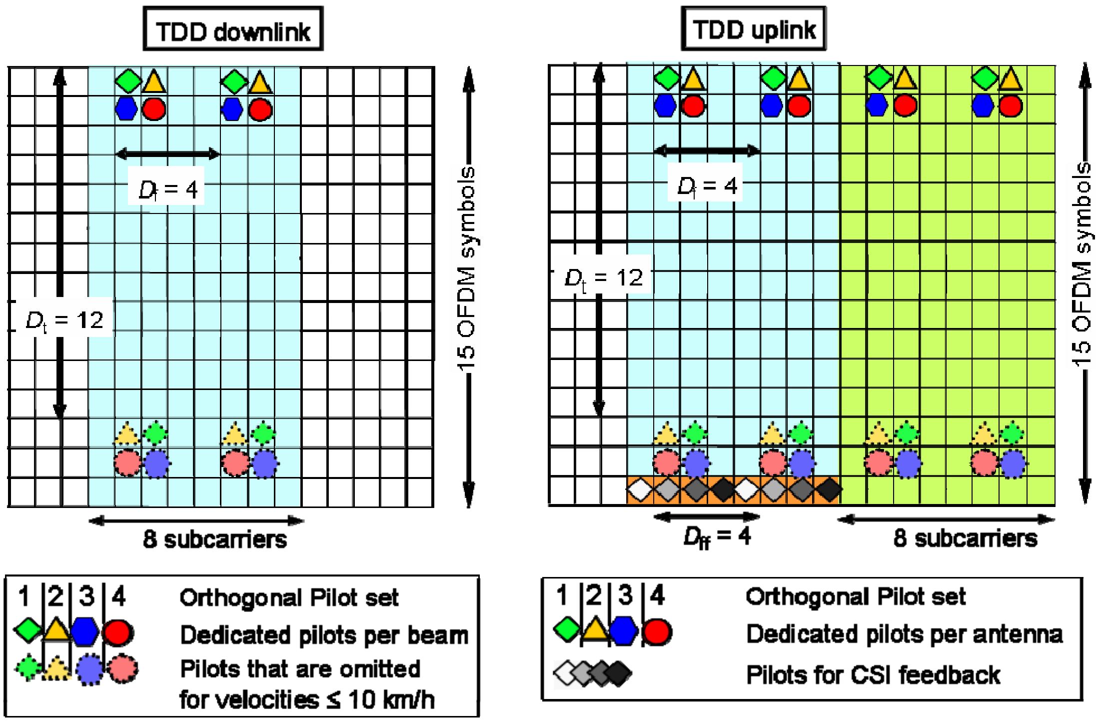

Figure A-10: Pilot grids for the TDD mode.

In case of spatial precoding at the BS the pilots are weighted with the same beamforming vector as the corresponding data symbols of that spatial layer, i.e. *dedicated pilots per beam* (DPB) are transmittted. However, for MIMO schemes without spatial precoding, the pilots are transmitted as *common pilots per antenna* (CPA). The receiver implicitly knows whether dedicated or common pilots are transmitted on a certain chunk, as it is uniquely determined by the spatial scheme selection.

A.3.2.4 Uplink Dedicated Pilots in TDD mode

It is apparent from Figure A-10 that the pilot pattern for TDD uplinks closely follows the TDD downlink. Dedicated pilots per antenna (DPA) are always used on the uplink.

The difference to the downlink is an additional set of pilots on the last OFDM symbol. For MU-MIMO schemes with spatial precoding based on short term CSI at the transmitter (CSIT) on the downlink, the last uplink OFDM symbol of each chunk is reserved for CSI transfer of downlink streams. As typically users that receive on the downlink typically do not transmit data on the uplink, the uplink pilots can in general not be used for updating the spatial precoding matrix at the BS. UTs insert 2 pilots with frequency spacing $D_{\mathrm{ff}} = 4$ per UT antenna on the last OFDM symbol of uplink slots, on those chunks where this UT is receiving data on the downlink. These pilots are transmitted unweighted, and their position in the uplink chunk is $(n,\ell)=\{(b_{\mathrm{ff}},15),(b_{\mathrm{ff}}+D_{\mathrm{ff}},15)\}$ in (frequency, time), where b_{ff} is the beam index of the corresponding downlink transmission. With 8 subcarriers per chunk up to 4 orthogonal pilot sets are available, which is sufficient to provide CSIT for the 4 spatial layers on the downlink. This means that users that are scheduled for CSIT spatial precoding on the downlink are reserved the last OFDM symbol of the corresponding uplink chunk to transmit pilots. These pilots provide the BS with the necessary CSI to update the spatial precoding matrix for the next downlink transmission.

A.3.2.5 Uplink dedicated pilots for B-IFDMA in both FDD and TDD modes

When using the B-EFDMA and B-IFDMA scheme for non-frequency-adaptive transmission [WIN2D461], one pilot symbol is included within each ***block,*** if possible located near the centre of the block. The assumed block size is 4 subcarriers by 3 OFDM symbols (abbreviated by 4x3 block). The resulting pilot pattern is depicted in Figure A-11. A larger number of pilots (one per 4x3 block per layer instead of four pilots per chunk layer) are thus required for the non-frequency-adaptive transmission, as compared to the frequency-adaptive transmission. With 8 blocks per chunk in FDD and 10 blocks/chunk in TDD, the pilot overhead becomes 8/96 and 10/120, respectively. With the specifications for B-IFDMA it is recommended that not more than 2 spatial layers should be used. The reasons are: first, high pilot overheads; and second, the dedicated pilots cannot be placed near the centre of a chunk, which severely affects the channel estimation accuracy.

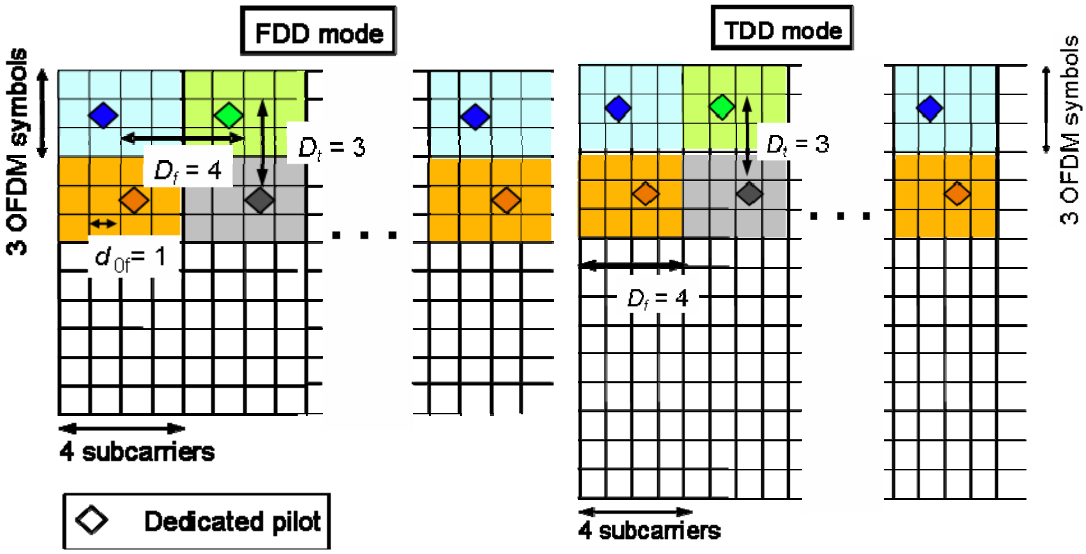

Figure A-11: Uplink pilot grids for non-frequency-adaptive transmission with B-IFDMA.

A.3.3 Uplink superframe pilot preamble

The dedicated pilot symbols per stream alone provide estimates of effective channel gains, i.e., channel gains affected by the spatial precoding scheme. To provide the BS with short-term CSI and CQI an uplink pilot preamble is inserted in the beginning of each superframe in both FDD and TDD modes, as shown in Figure A-12, so to obtain estimates of the unweighted channel matrix. As uplink pilots over the full-band are very expensive in terms of overhead and UT power consumption, these full band pilots are inserted at a lower rate (once per superframe). This essentially limits the maximum velocity for adaptive transmission to 10 km/h. Only users with sufficiently low velocities, and which are scheduled by the BS for adaptive transmission transmit pilots at the superframe preamble. Pilots are orthogonally multiplexed in frequency. With a pilot spacing of $D_f = 8$ up to 8 such users can be supported per competition band.

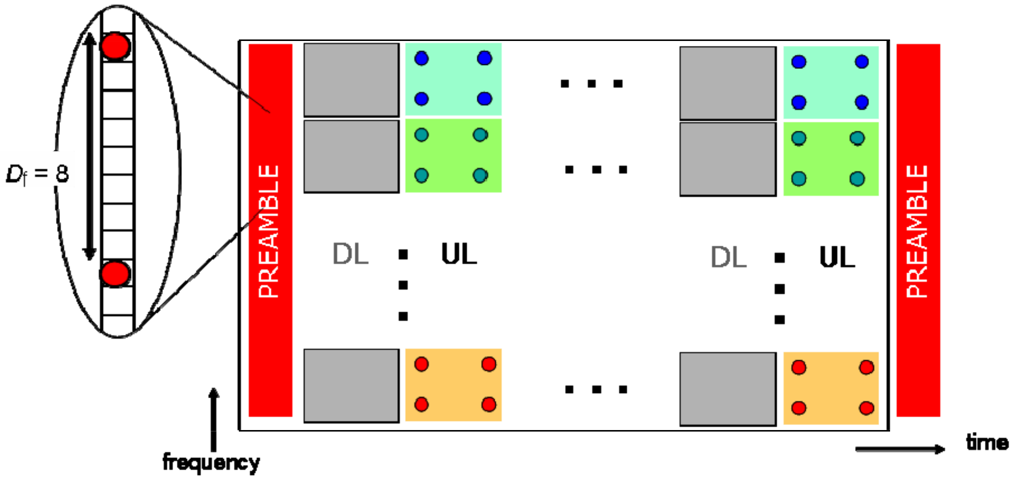

Figure A-12: Position of uplink full-band dedicated pilots in superframe preamble.

We note that frequency-adaptive transmission with mobile velocities up to 50km/h is possible through the downlink pilots in TDD mode. Hence, these mobile users do not insert pilots in the preamble. Pilots for channel estimation are frequency-multiplexed inside OFDM symbols. They are regularly spaced on a rectangular grid, as described in Annex A.3.2.

B ANNEX: RRM mechanisms

B.1 Resource partitioning within a REC

As indicated by simulation results in [WIN2D352] and [WIN2D353], a proper radio resource partitioning between BS and RN within a REC is essential to utilize the potential benefits of relay based deployments. Three types of links share the radio resources in a REC, BS-UT, RN-UT and BS-RN. If cooperative relaying is used in the REC, two additional link types have to be considered (BS, RN) – UT and (RN, RN) – UT. The WINNER system considers a "distributed" MAC [WIN2D351], i.e., the BS partitions the resources between itself and the RNs in the REC. The RNs can freely allocate these resources and thus frequency-adaptive transmissions and multi-antenna schemes for UTs served by RNs can be supported without forwarding all the required control signalling to the BS. Secondly, the complexity of the BS is reduced.

The resource partitioning for RECs has to work on a slower timescale than the inter-RAN load balancing (several seconds) and the short-term spectrum assignment (200ms-1s). It operates mainly on the same time-scale than the inter-cell resource partitioning (100ms). Additionally, a fast intra-REC resource partitioning, every superframe (~6ms), is supported.

Figure B-1: illustrates a resource partitioning, where the BS receives resource requests and measurement reports from the RN. Based on the received information the BS decides the new resource partitioning and sends it to the RNs within its REC. The resource partitioning and the associated signalling are performed in the RRC2 layer as presented in Figure 4-2, which covers RRC related functions that do not involve UTs.

The signalling is routed through an own RLC instance to ensure a reliable data transfer. The resource partitioning messages have to be forwarded to all the RNs in the REC before the new resource partitioning is active. To allow sufficient time for at least one retransmission, the message has to be sent at least 2 frames (~1.4ms) in advance for a two hop deployment. The decision on how to partition the resources between the RAPs should be based on most up to date data from the RAPs. This data includes interference measurements as well as resource request coming from the RAPs. For deployment scenarios with more than 2 hops, the additional delay for each hop will restrict the update frequency of the resource partitioning and a fast resource partitioning every superframe might not be feasible.

The BS in Figure B-1: can be replaced by a central RRM server that performs the resource partitioning, if such a logical node is part of the final WINNER system design.

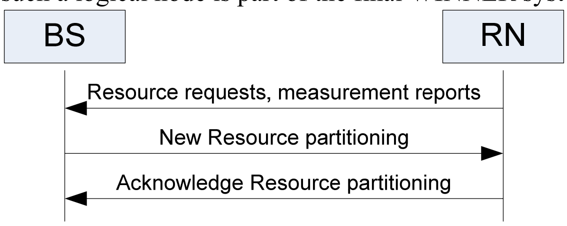

Figure B-1: Resource partitioning.

Examples of different centralized resource partitioning strategies and the content of the resource partitioning messages can be found in Chapter 7.

B.2 Active mode mobility

B.2.1 Introduction

The purpose of active mode mobility is to maintain connectivity and to support QoS in a radio resource efficient manner (seamless handover). A handover may be triggered due to mobility, changes in the radio environment and/or in the QoS requirements or to other reasons (e.g. load balancing).

B.2.2 Assumptions

The following assumptions and requirement are assumed:

- Single-mode case.
- Network controlled handovers
- L2 RNs, i.e. no repeaters.
- Within a REC all IDs are assigned and managed by the BS. Hence when a UT performs an intra-REC handover the ID of the UT does not need to be updated.
- RN (re-)broadcast preamble, i.e. cannot rely on direct signalling from BS to UT throughout the REC (derived from T5 working assumptions)
  - Further motivation for this assumption:
    - If the BS was to broadcast the preamble throughout the entire REC, the coverage may be impaired.
    - Larger cyclic prefix (CP) may be needed to cover time synchronization errors.
      - If the RN does not transmit any preamble, the UT synchronizes with the BS and is hence unaware of the RN. Since the distance to the RN is different, a timing error is introduced in addition to the multi-path delay spread. This may result in a large drop in capacity.
- Forwarding (routing) performed on the RLC layer. RN does not involve any security procedures - managed by GW/BS within a REC.
- Source BS and its RNs broadcasts the neighbour cell_list, their identity and resources used for control signalling.
- Neighbouring BSs exchange cell status information either periodically or on request
  - Since there is a many-to-many relationship between BSs and GWs, our working assumption is that this type of handover will be infrequent and does not need to be optimized for. Nevertheless, it may be desirable to avoid synchronous BS-to-BS handover and GW-to-GW handover, e.g. by assigning partially overlapping pool regions.
- The GW_C and GW_{IPA} functions are split into different entities with a standardized interface in between.
- The GW_C is not involved in the handover preparation signalling; instead the CP signalling is done between the BSs directly. The UP is handled by packet forwarding from source BS to target BS.
- The (target) BS performs admission control.
- The UT is not a trusted device, i.e. any action taken by the UT needs to be authorized.
- BS owns cell resources.
- Handover/routing is performed per user.[23]
- The same route is used in the uplink and downlink direction.

In the paragraphs below, it is assumed that the BSs take the handover decisions in a distributed manner i.e. no RRM server is present. Nevertheless, generalizations of the proposed schemes also to incorporate an RRM server is straight forward.

An important aspect of the handover procedure is how to get UL time alignment in the new cell. Here one may distinguish in between two different cases: i) the network is synchronized, and ii) the network is not synchronized.

For the first case UL time alignment may be accomplished in the following way. The UT estimates the (DL) propagation delay from both the source and target RAPs. By using the information on the UL time alignment towards the source RAP as well as the timing different in between the source and target RAPs the UT may estimate the UL time alignment to be used towards the target RAP.

For the second case, two further cases may be distinguished i) the UL timing difference is larger than the cyclic prefix, and ii) the UL timing difference is smaller than the cyclic prefix.

- For the first case UL time alignment may be accomplished in a number of ways:
  - Alt1: use of normal RACH procedure to gain UL time alignment (i.e. the UT transmits on the RACH, BS measures and reports back (e.g. in some scheduling grant message) the needed time alignment.
    - + Simple case, will have to be supported anyway.
    - - Potentially long handover delays.

[23] In a multimode scenario one may also envision the case where per flow handover/routing may be performed if the UT is equipped with more than one transceiver and that the two modes do not share resources .

  - Alt2: as for case 1, but the UT is assigned dedicated (RACH) resources in the target cell to transmit on.
    - + Lower handover delay than alt1 (collisions on the RACH (prolonging the handover procedure) are avoided).
    - - Requires that RACH resources may be reserved (and potentially unused) by the BS (implications unknown).
  - Alt3: the target RAP measures on UT transmissions (in the source cell) prior to handover.
    - + lower handover delay than the previous two cases
    - - Requires signalling of resource assignments over the BS-to-BS interface, however for intra cell RN to RN handovers this might be doable since they are both controlled by the same BS.
- For the second case the UT does not need to be updated on the time alignment in the target cell (this scenario may be very likely in case of small cells (especially areas covered by RNs may be envisioned here) - it is also a desirable case since this will lower the handover delays). The only problem in this case is to detect that the UT does not need UL time alignment in the target cell. This may be accomplished if i) it is time aligned to the source BS and ii) the same frame format structure is used in both cells, in the following way: the UT estimates the time difference between the two cells by measuring on the DL transmissions from BSs (same difference in the UL due to requirement i)) and reports this to the source BS. The source BS may then deduce whether the UT needs time alignment to any potential target BS or not (i.e. not needed if the time difference between the two BSs is shorter than the cyclic prefix in the DL). During this procedure the UT may receive an update of the time alignment from the source node to ensure time alignment to the source BS.

In the sequel, it is assumed that UTs use alternative 2 above, i.e., the UT is assigned dedicated resources on the RACH to acquire synchronization. Generalization to also incorporate other alternatives is straightforward. For more details, see [WIN2D351].

B.2.3 Proposed solution

This section will outline three handover/routing cases

- the BS-to-BS handover case (baseline case)
- the RN-to-RN handover within one REC case (as this is the simplest handover case between two RNs)
- the RN-to-RN handover in between two RECs case (as this is the most complex handover case in a two hop scenario).

B.2.3.1 BS-to-BS handover (single-hop case)

Figure B-2 shows the message sequence chart of the handover procedure, assuming an error-free, successful case. The figure shows both the control plane messages and the user plane data.

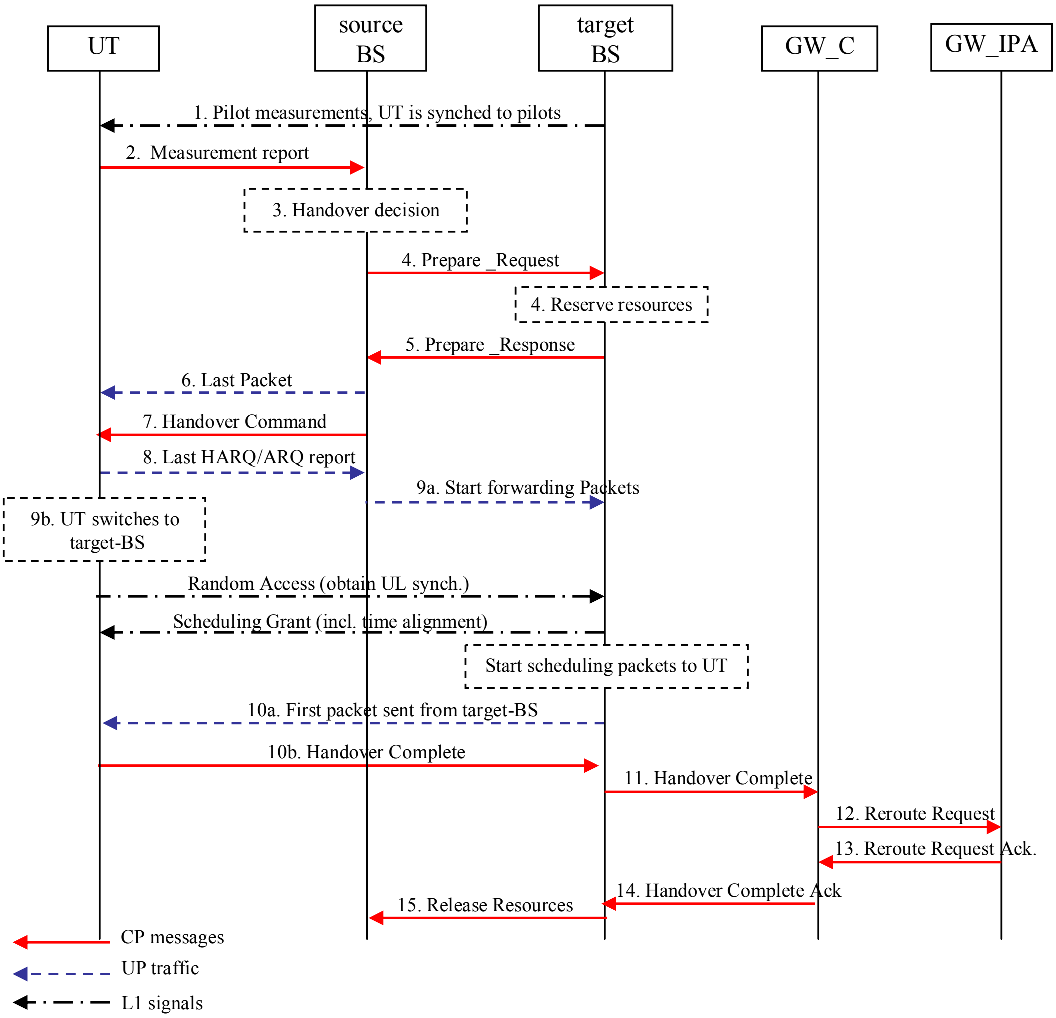

Figure B-2: Message sequence chart of the handover procedure in active-mode in a single-hop case.

1. The UT makes measurements on the pilot channels of the cells in the candidate neighbour set. While the UT performs measurements on some pilot channels of neighbouring cells it should also obtain downlink synchronization to the candidate cells as well as to the cell ID.

2-3. The UT sends the measurement report to the source BS. Based on these measurement reports (as well as other higher layer triggers as outlined above) the source BS decides to perform a handover. Prior to the handover decision, the source BS may poll the target BS on the cell status. However this may also be done in a periodic fashion.

4-5. The source BS transfers the "RRC context" of the UT to the target BS and asks for reservation of resources at the target BS (e.g., UT ID etc.). The transferred RRC context of the UT includes for instance, the configured measurements in the UT and the flow class parameters of the UT. The target BS performs admission control and if the UT is admitted the target BS sends back all information that will be necessary for the UT to initiate communication at the target BS (e.g., UT ID, information on where to find RACH (if applicable)) in the "Preparation Response" message. With this same message exchange the forwarding tunnel between the source BS and target BS can be also set up. If the UT is not admitted the source BS will select another cell and forward a new request to this new target BS.

6-7. At this moment the source BS can trigger the MAC layer to finish the ongoing HARQ processes. The source BS also sends the Handover Command to the UT.

8. The UT acknowledges the handover command message on the HARQ/ARQ layer and sends the last HARQ/ARQ report. After the handover command has been ACK:ed by the UT the source BS does not send any more data to the UT, since it assumes that the UT has started the handover execution.

9. The source BS starts forwarding packets to the target BS. At the same time the UT starts to obtain synchronization at the target BS. It sends a random access on RACH, in the response it gets timing alignment and scheduling grant assigned. 10. The target BS can start sending DL data to the UT. The UT sends a handover complete message to the target BS and can start sending UL data as well.

11-14. The target BS sends a handover complete message to the GW_C, which, in response, reconfigures the routing in the GW_IPA.
15. Finally, the target BS notifies the source BS that it can release the resources that were allocated to the UT (e.g., UT IDs, UT context).

There are a number of further issues that need to be considered with respect to the figures above since we have assumed that outer-ARQ is terminated in the BS, namely lossless delivery, duplication less delivery and in-order delivery.

Lossless delivery
Downlink packet forwarding guarantees that no packets will be lost at the source BS during the handover execution (packets may still get lost on the transport network). Packets that could not be sent out from the source BS prior to the handover and the rest of the incoming packets from the GW will be forwarded to the target BS. The forwarding begins with the first packet that has not been cumulatively acknowledged on the RLC layer. Note that forwarding is employed only for the downlink direction. In the uplink, the working assumption is that the UT resumes the transmission at the target BS by sending those packets that have not been acknowledged on the RLC layer at the source BS before the handover. This means that gaps may occur in the UL stream sent from the BS to the GW during handover and packets may arrive out of order.

Duplication-less delivery
There can be some packets that have been received at the receiver side and the corresponding packets delivered to the upper layers, while the corresponding HARQ/ARQ acknowledgements are not returned to the source side before the start of a handover execution. Such packets will be sent again to the target BS and will appear as duplicates on the receiving side. Such duplicates may occur either in the uplink or in the downlink directions but they are rare cases.

In the downlink two alternatives for duplicate detection are possible:

- IPCL in the UT detects and removes the duplicate packets
- RLC in the new BS uses the IPCL sequence number to detect duplicate packets

On the one hand, if alternative 1 is used and RLC in the BS performs duplicate detection, then RLC has to look into the IPCL sequence numbers, which violates the protocol layering. On the other hand, if alternative 2 is used and IPCL in the UT performs duplicate detection, then layering is preserved, but duplicates are transmitted all the way over the air to the UT.

In the uplink duplicates can be filtered out in the IPCL layer in the GW based on the corresponding IPCL sequence number.

In-order delivery
In the downlink the forwarded packets coming from the source BS and the rerouted packets coming from the GW may arrive out of order. The reordering can be done in the target BS based on the IPCL sequence numbers. Alternatively, the reordering can be done in the UT based on IPCL sequence numbers.

In the uplink, packets may arrive out of order at the GW due to that the source BS releases the received packets in a non-contiguous manner, i.e., gaps may occur, where the missing packet will arrive from the target BS after the handover. This reordering is performed at the IPCL layer in the GW.

Out of order packets may also occur due to delay differences in the transport network, if the delay on the source BS-GW path is significantly higher than the delay on the target BS-GW path. However, such delay differences should be rare.

B.2.3.2 Two-hop case

B.2.3.2.1 Intra-REC handover

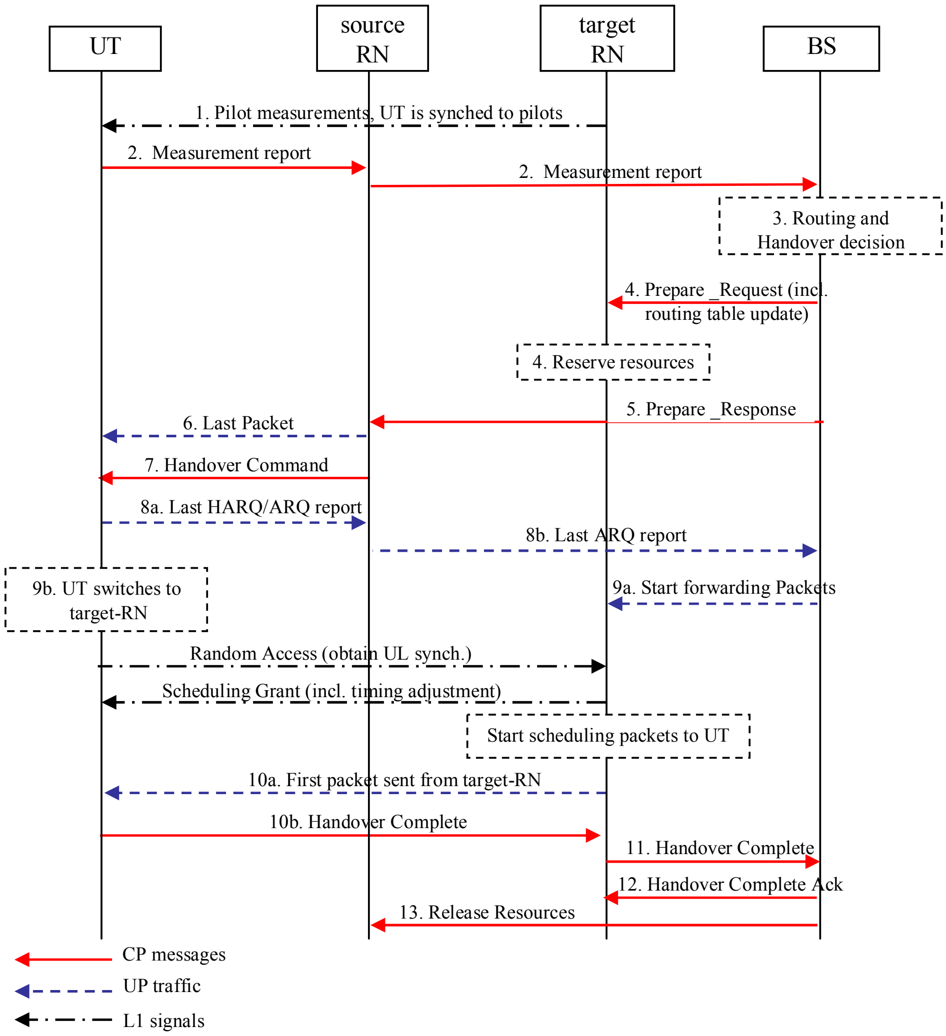

Figure B-3: Message chart of the handover procedure in active-mode in a two-hop case (GW_C and GW_IPA are omitted in this case since they are not affected by this handover).

Steps 6 and 8a may potentially also be omitted (in this case we heavily rely on outer-ARQ).

B.2.3.2.2 Inter-REC handover

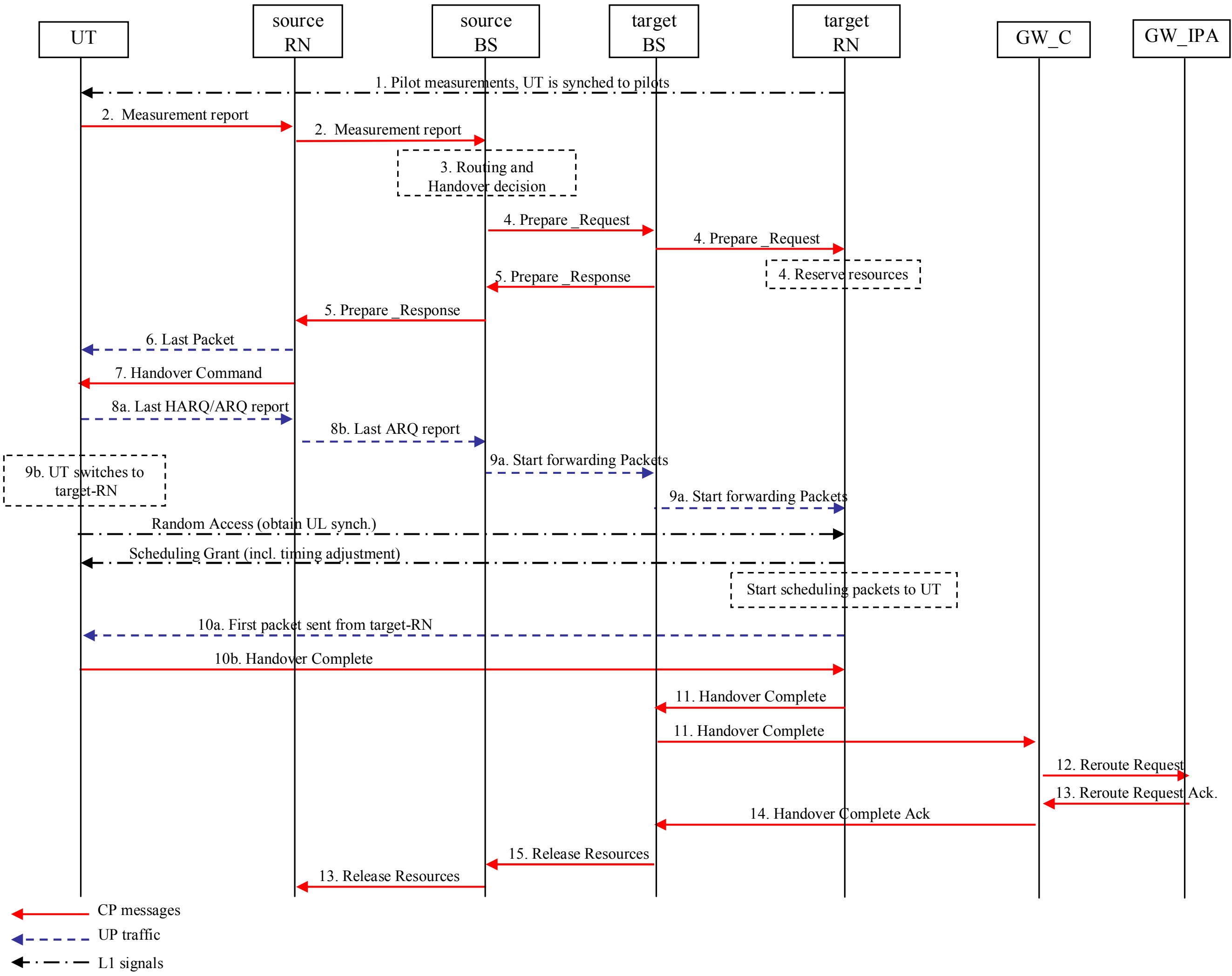

Figure B-4: Message chart of the handover procedure in active-mode in a two-hop case.

The RN along the route in the new REC has to obtain and record the ID of the UT.

B.2.3.3 Multi-hop case

May be deduced from Figure B-3 and Figure B-4 depending on whether the handover is intra- or inter-REC, respectively (the only difference is that the signalling in between the BS and the RAPs/UTs will traverse more hops).

B.2.4 Hybrid Information Systems enhancements

The WINNER system may interwork with an extra entity which provides location based services for the mobility management. That entity is termed as eHIS (enhanced Hybrid Information System).

The basic property of eHIS is to map incoming measurement reports to specific locations. Besides that, further parameters such as velocity, moving direction, current service consumption, etc., can be added. Such extended entries support personalized service provision. By evaluation of those data it is possible to set up user profiles. Additionally, it will be possible to predict user requirements.

A typical example for location based handover is given: each active UT reports about the current link condition. Together with the measurement report the location of the reporting UT is stored in the eHIS. The BS acquires the corresponding measurement report from the eHIS, depending on the current location of the UT and signals the handover decision (respectively related information that allows the UT to take the decision) to the UT. The UT can then perform the handover to the proper BS or the mode.

Predictive HO reflects the merit of employing eHIS: if for example a user moves with high speed along a road, it is very unlikely that it will spontaneously turn left or right. Such, an enhanced mobility management mechanism could exploit the eHIS information to prepare a planned handover to another serving BS/AP in the near future. Due to the interference maps, the current link condition in the future target system can be predicted so that it can be decided whether extra bandwidth needs to be reserved for the shortly to handover terminal. Moreover, if the geographic target is covered by several vertical systems, the eHIS may even trigger intersystem support for the terminal and support joint RRM in this way.

In this basic approach, measurements that are inherently available for each system are made available to heterogeneous systems as well to support the inter-working between heterogeneous systems. Depending on the new target system and the current location of the mobile, the mobile is supplied with state reports of the same system type (for horizontal handover, HHO) or a vertical system (vertical handover, VHO), and subsequently may perform the (V)HO, which is referred to as location-based VHO.

Obviously it makes not much sense if a mobile that is about to handover to another (vertical) system is provided with interference information being totally out-of–date. On the other hand, if system engineers want to get more information on areas with low coverage, they are not interested in present fading analyses. The concept of eHIS accommodates both needs by distinguishing between short-term, mid-term and long-term data. Short-term data is meant to support real time requests from the information clients. As soon as a feeding client provides new measurement reports, the essence is extracted and stored in the eHIS database. By such, short-term data reflects the latest entries in the data base. Respective information is used to serve as decision basis for short-dated handover triggers, linko adaptation and power control support and the like. Mid-term data instead is less time critical. It is based on short-term input but due to respective filtering and averaging time selective fading effects are equalized. Nonetheless, mid-term data is of interest for ongoing communication since it serves as set value especially for predictable actions. Especially in combination with prediction and profiling mid-term data is useful for planned handover triggering, (joint) RRM or Admission Control (CAC). The long-term data in eHIS system addresses either permanent impacts, e.g. to determine areas with ongoing insufficient link quality, or recurrent events such as the analysis support of daily occurring networking congestions during e.g. rush hours. Such long-term data comprises fairly static information. The period for long-term data is supposed to be longer than one day.

The eHIS is a concept facilitating seamless and proper handover and inter-system cooperation. Subject of the eHIS is to perform more accurate detection of complementary systems and to initiate more efficient handover execution by respective triggering.

The eHIS entails a decision unit that takes into account trigger origins as input and produces handover recommendations (=triggers) as output. The advantage is, that the eHIS is not restricted to local and system specific trigger origins. Besides incorporation of a multiple number of systems, eHIS supports load balancing and joint radio resource management. Further, specific user preferences may be requested from e.g. the home network provider and incorporated in any decision process.

B.3 Idle mode mobility

B.3.1 Introduction

The purpose of idle mode mobility is to, in a resource and power efficient manner, keep track of the idle mode terminals and thereby enable network initiated connection setups as well as to support efficient UT power saving.

Before discussing the necessary procedures in idle mode, it is necessary to define the different states a UT may be in. The Detached state is a “null”-state which the UT enters prior to attachment, i.e., when the UT power is switched on and the UT is not known to the network. Identified UTs will be in either idle mode or in active mode. In idle mode the UT location is known over a *paging-area* (incorporating a number of cells). Mobility in idle mode is handled by UT *cell (re-)selection*. The UT registers with a *paging area update* when it moves from one area to another. At network initiated connection setups (i.e. the UT is moved from idle to active mode) *paging* is used to locate the UT within the paging area. In active mode, the UT is known on a cell basis.

Moreover, the GW_C stores the context for idle mode UTs, which includes the location information of the UT on a paging area level, the security context and the context about the established flows of the UT.

Note that a UT in idle mode can only have flows that do not have reserved resources. There is no context stored in the BS about idle mode UTs.
The procedures in italic above will be further discussed below.

B.3.2 Assumptions and Requirements

This section relies on the same assumptions as for the active mode mobility as well as the following assumptions and requirements.

- A cell is defined by the geographical coverage area of its broadcast channel.
- The UT has an identity, which uniquely identifies the UT in a paging area;

B.3.3 Technology Options

B.3.3.1 Paging area and paging area updates

The location of a UT in idle mode is maintained on a paging area (PA) level. When a UT in idle mode moves into a cell that belongs to a PA different from the one it is currently registered with, it performs a PA update towards the GW_C. The GW_C selects a suitable PA for the UT to camp on and sends this information back to the UT in the PA Update Confirm message. The UE identifies the PA the given cell belongs to by reading the broadcast channel of the cell.
When configuring paging areas in the network there will always be tradeoffs between paging area size and paging load:

- Small paging areas leads to less paging load in the system, but to more paging area update messages.
- Large paging areas leads to more paging load in the system, but to fewer paging area update messages.

Therefore, it is desirable to define a paging area concept that allows some flexibility on the size of the area a given UT will be paged in. For instance a stationary UT can be assigned to a small paging area, while a UE that is moving through the system can be assigned to a larger paging area.
Moreover, it is desirable to define a paging area concept that avoids excessive PA Update signalling by UTs on paging area borders.
Two potential solutions fulfilling the two “wishes” above are given below:

- multiple PAs assigned to one UT, i.e., to allow that a UT can be registered at multiple PAs, or
- overlapping PAs, i.e., to allow that a cell belongs to more than one PA.

The “multiple PAs assigned to one UT” is very similar to existing Routing Area concepts in GERAN/UTRAN today. Each cell in the network broadcasts one paging area identity. When the terminal in idle state enters a new paging area, which it has not been assigned to, it will perform a paging area update procedure. As a result of this procedure the terminal will be assigned to the new paging area it enters (using the paging area update response message). In this concept the terminal may also be assigned additional paging areas as a result of the paging area update procedure. These additional paging areas are treated in the same way as in the single PA case, meaning that as long as the terminal moves within the paging areas it has been assigned to it will not perform any paging area update procedures except periodical updates. Subsequently, the terminal will be paged in all paging areas it has been assigned to. The network is however allowed to optimise the order in which it pages the terminal (e.g. it can start in last known cell or PA before it pages in all areas).
The “Overlapping PA concept” is the same as the UTRAN Registration Area concept that is used in RRC connected mode in UTRAN. Multiple paging area identities (PA_Id) may be broadcasted in each cell. As long as the terminal is moving in cells that belong to the same paging are that the terminal has been assigned to the terminal does not need to perform any paging are updates (except periodical updates). When the terminal enters a new cell not broadcasting the PA_id the terminal has been assigned, the terminal will perform a paging area update procedure. As a result of the paging area update procedure the terminal will be assigned a new paging area that is available in the cell.

B.3.3.1.1 Analysis/Reasoning

Both concepts meet the basic requirements on flexibility, and avoid frequent paging area updates. Nevertheless, the “multiple PAs assigned to one UT” concept might be slightly more flexible than the “overlapping PA concept” since the paging area for specific UTs can be configured dynamically, while in the “Overlapping PA concept” all paging areas are pre-allocated.

B.3.3.2 Paging

Paging is used for network-initiated connection setup. When the UT needs to be paged the GW_{IPA} will trigger the paging procedure at the GW_{IPA} and GW_C will page the UT in all cells of the current PA. The UT will respond to the page with a cell update and will go to active mode.

An efficient paging procedure should allow the UT to sleep with no receiver processing most of the time and to briefly wake up at predefined time intervals to monitor paging information from the network. Paging opportunities should be located at pre-defined resources and may be co-located with other frequent and periodically re-occurring control-signalling, e.g. resource assignments are transmitted in every frame at pre-allocated resources.

In order to support an efficient sleep mode procedure (i.e. minimize the information that needs to be decoded by the UTs at every paging instant) it is advisable to transmit as little information as possible on the pre-allocated resources. This may be achieved by grouping the UTs into paging groups and only indicate the paging group at the pre-defined resources and indicate (explicitly or implicitly) where the UT may find the additional information, i.e. only transmit a paging group indicator (similar to UMTS) at the pre-allocated resources and instead use resources which can be used dynamically also for traffic/other control channels to convey the paging information regarding which UTs within the group that where actually paged.

An example where the paging group indicator is co-located with the “normal” scheduling information is given below: If the UT detects scheduling information, through the presence of its own identity or a common paging group identity, the downlink transmission is received at the identified resources and further processed by higher layers (N.B. if a UT is identified at the paging group indicator no further resources are needed). Otherwise, the UT reverts to sleeping until the next paging occasion.

As has been the case for previous procedures presented in this working document, one may distinguish different cases depending on whether broadcast information (in this case paging messages) may be performed as a SFN (in case the propagation delay from some RAP in the cell to any UT in the sector/cell is shorter than the cyclic prefix one may potentially run the REC as a single frequency network) .

Four different options have been distinguished so far:

- All RAPs (in a given area, though larger than a REC) broadcasts paging information in a SFN manner
- All RAPs in a REC broadcasts paging information in a SFN manner
- All RNs in a REC broadcasts paging information in a SFN manner
- No SFN features

In case one, the paging information has to be made available to all RAPs prior to the paging instance in order to allow all RAPs to transmit the same information. In case two, the paging information has to be made available to the RNs prior to the paging instance in order to allow all RAPs in a REC to transmit the same information. This may be achieved by transmitting a control message (on resources which can be used dynamically also for traffic/other control channels) in a frame prior to the paging instance. In case three, the paging information transmitted in frame i may be re-transmitted by the RNs in a SFN manner in frame i+1 (or perhaps i+j where j>1, to enable the receiver to decode the information).

B.3.3.2.1 Analysis/Reasoning

Case one, two and three above both have the potential to lower the paging overhead as compared to case four as: i) fewer resources are consumed per paging message, however this gain largely depends on the average ratio of RNs and BSs and ii) diversity gains through combining.

B.3.3.3 Cell selection and cell reselection

When a terminal moves from detached mode to idle mode it performs cell selection in order to determine what cell to camp on and commence the cell reselection procedure. Cell reselection is primarily performed in idle mode to decide what cell to camp on.

No specific issues with respect to relaying have been identified so far.

C ANNEX: Configuration of multiple access schemes

In this section, we outline a recommendation of configuration of the non-frequency-adaptive multiple access schemes, but we start with some basic considerations regarding uplink power control and power efficiency for user terminals that are important for both frequency-adaptive and non-frequency-adaptive transmission.

C.1 Power efficiency in user terminals

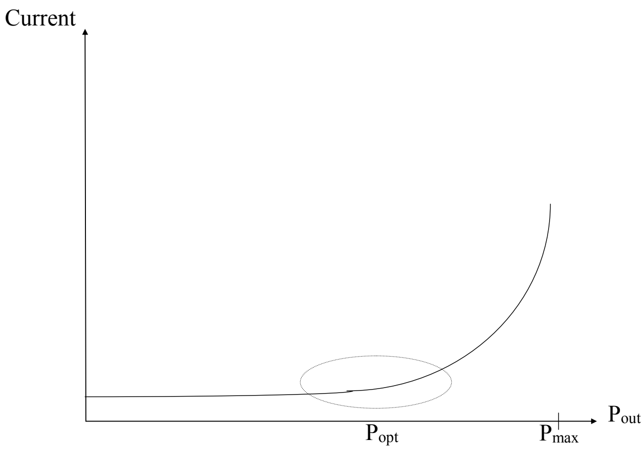

Figure C-1: Characteristics of a typical high power amplifier in a user terminal.

In Figure C-1, a typical characteristic of a high power amplifier (HPA) is depicted. To make a cost and power efficient user terminal, the requirement on instantaneous RF power P_{max} should be as small as possible and with such an HPA design, the normal operation point for the instantaneously generated RF power P_{out} of the HPA should be close to this maximum instantaneous transmit power level for maximum HPA efficiency.

This implies that, from a *user centric* point of view, to maximize the *energy efficiency* of the UT:

- At high data rates and/or bad channel conditions, the RF energy for uplink transmission should be generated over as long time as possible to maintain a low enough required instantaneous RF power P_{max} on the user terminal.
- At low data rates and/or good channel conditions, uplink transmission should be localized enough in time to enable the HPA working at the designed most efficient operation point P_{opt} as often as possible.

Apart from optimizing the operation point for the HPA, in cases when the transmission can be localized in time it also enables the user terminal to gain from micro-sleep, whenever the time slots with no transmission are long enough for turning off and on appropriate circuits in the UT, as discussed in [WIN2D461].

To highlight the potential of micro-sleep, assume the HPA is designed to be most efficient when transmitting at least with an instantaneous data rate of 2 Mbit/s, which corresponds to the value at the satisfied user criterion in WINNER. When using a low data rate service as e.g. VoIP, there is a large potential energy savings with micro-sleep. E.g., assume a G.723.1 codec that without header compression (i.e. before the IPCL layer) generates 33 kbit/s (with duplex traffic) with 66 packets/s. This corresponds to one 500 bit packet every 44th slot, and an instantaneous data rate of 1.45 Mbit/s in the used slots. After applying header compression in the IPCL and with a HARQ scheme, the instantaneous data rate is substantially smaller in the used slots, and the HPA would work more close to its optimal operation point

P_{opt} if the B-IFDMA/B-EFDMA blocks are shorter than a slot, thus increasing the instantaneous data rate within the used OFDM symbols.[24]

In addition to optimizing the RF transmit power operation point of the HPA, it is important to maintain a small enough envelope fluctuation. The DFT encoding applied in the non-frequency-adaptive uplink enables a lower required HPA backoff, which improves the HPA efficiency. The required backoff versus block size and number of blocks has been investigated for B-IFDMA in [WIN2D61310], [WIN2D233] and [WIN2D461]. The more subcarriers per block and the more parallel blocks allocated in frequency, the larger required power amplifier backoff. Thus, for medium to high data rates, longer blocks than the basic block size (4x3) are beneficial to maintain a low enough envelope variation and to enable large enough RF transmit energy per slot.

C.2 Uplink power control vs resource allocation

As discussed in Section 7.3.2.5, a power control algorithm is likely needed for both the non-frequency-adaptive and the frequency-adaptive uplink, with the goal to meet a certain minimum and maximum received power spectral density at the RAP.

Assuming a requirement on minimum received psd at the RAP, for frequency-adaptive transmission, this means that the UT energy efficiency would benefit from assignment of as many chunks as possible up to the maximum number dictated by P_{max} (or P_{opt}). This is beneficial also for accommodating strong FEC coding by using large code words that are terminated within the frame, as assumed in the WINNER system concept. But this assumes that enough chunks with good SINR can be found by the Resource Scheduler. Otherwise, there is a trade-off with the multi-user scheduling gains.

I.e. at a certain point, from a *system centric point* of view, the more user terminals that are transmitting concurrently, the larger multi-user scheduling gains are obtained with large total instantaneous RF power seen at the receiver side, which will improve the spectral efficiency in the uplink, but for some user terminals this is at the expense of operating *below* P_{opt}, since the few allocated chunks cannot benefit from more transmit power due to e.g. MCS limit and inter-cell interference control. I.e. at high load levels, the system might need to operative at a system centric optimum operation point that is suboptimum as seen from a user centric energy efficiency point of view, but that enables the system to fulfill the satisfied user criterion in terms of data rate and delay.

For non-frequency-adaptive transmission, the frequency-selective channel gains are not measured, and the power control adjusts to the path loss and shadow fading time scale only. Thus, the corresponding trade-off is how many B-IFDMA basic blocks (4x3 subcarriers x OFDM symbols) that should be assigned in the frequency direction to each user terminal for uplink transmission to meet targets on minimum and maximum expected average psd at the RAP in the assigned blocks. (The modulation and code rate is then adapted on the shadow fading time-scale to the received SINR in the slot.) However, not only the number of concurrent blocks is important, but also the position of them, since the envelope properties of the B-IFDMA signal, the diversity gains and channel estimation performance depends on these parameters.

For low SINR users, a requirement on a minimum expected receive psd per block directly translates to a limit on a maximum number of parallel basic blocks that can be allocated in the frequency direction. If the resources for non-frequency-adaptive transmission are frequency-multiplexed with the resources for frequency-adaptive transmission that are mainly serving high SINR users, the required minimum expected psd per block could be rather high to limit adjacent channel leakage caused by imperfections in the transmitters [3GPP R4-060867], as well as Doppler spread and receiver imperfections. This will limit the number of parallel blocks for low SINR users that can be allocated in the frequency direction and thus limit the frequency diversity gains. Multiple receive antennas at the RAP becomes important to compensate with spatial diversity gains, and also to lower the frequency-selective fading dips. In addition, allocation of multiple basic blocks in time within the slot becomes important to enable large enough code words and data rates.

[24] This also indicates that non-frequency-adaptive transmission could be more energy efficient in the uplink for the user terminal than frequency-adaptive transmission for low data rate real-time continuous services, especially since most slots are empty and channel tracking is of no use for these empty slots. But the trade-off with the full chunk based frequency-adaptive transmission using better channel knowledge is not investigated.

C.3 Configuration of non-frequency-adaptive multiple access

The above considerations motivate the introduction of the small basic blocks in B-IFDMA/B-EFDMA as a basic building block to enable adaptive block allocation in different scenarios. The optimization should take into account aspects of:

- UT energy efficiency (power amplifier operation point, HPA backoff requirement and sleep mode gains)
- Transmit robustness for small packets, especially when HARQ cannot be used as in the case of fast and broadcast control signals
- Amount of spatial diversity available
- Performance of channel estimation
- Allocated bandwidth for the non-frequency-adaptive transmission
- Employed power control algorithm

Below we outline some block allocations, which should be seen as recommendations based on the insights gained during the WINNER II project. The presented block allocations should be seen as the union of the block allocations used in different deployments. The number of different block allocations should probably be less in a given cell in order to keep the resource allocation complexity low, and should be optimized based on the deployment scenario and expected user characteristics.

Studies in [WIN2D461], [WIN2D233] and [WIN2D61310] have quantified the performance with different block allocations w.r.t. power amplifier backoff, frequency-diversity gains vs channel estimation loss, coding gains and performance of different receiver equalizers. The results of these studies are important input. However, further investigations are needed to quantitatively motivate the configurations below and such investigations would need some further studies as input regarding a suitable power control algorithm (taking both intra-cell interference due to transmitter and receiver imperfections and intercell interference into account), and quantify user terminal gains and losses in a reference transceiver design.

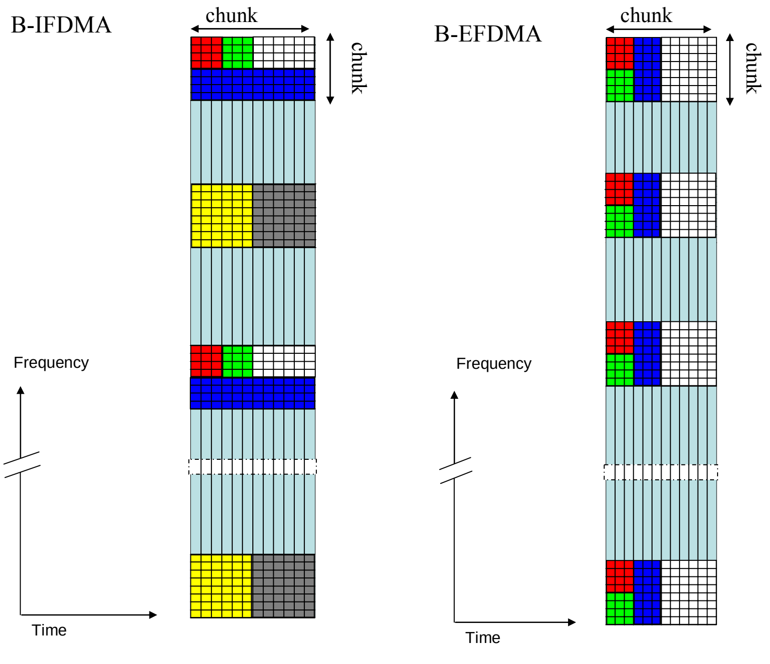

Figure C-2: Configuration of B-IFDMA and B-EFDMA in FDD Wide Area scenario.

C.3.1 FDD Uplink

Referring to Figure C-2 (left), in SISO deployment a frequency-diversity collected by up to 8-16 blocks in B-IFDMA seems useful for resource allocation units (RAUs) corresponding to 1-2 physical chunks, i.e. a block size equal to the basic block of 4x3 (subcarriers x OFDM symbols) is appropriate for small packets, with a repetition distance of 2-4 chunks in the frequency direction. The short 4x3 blocks are also needed for timely carrying the control signals for frequency-adaptive transmission, see [WIN1D26] for a discussion of the timing loops for channel prediction in frequency-adaptive transmission.

Larger RAUs should use larger blocks. The investigations in [WIN2D233] shows that a block size of 8x6 (subcarriers x OFDM symbols) is useful since it enables interpolation of channel estimates based on the dedicated pilots, and the channel estimation gain is larger than the loss in frequency-diversity. However, taking also uplink power control, power amplifier backoff and optimum working point for the power amplifier into account as discussed above, a resource block of 4x12 (subcarriers x OFDM symbols) seems most appropriate for bad SINR users, but has not yet been evaluated in the project.

With spatial diversity available, less frequency-diversity is needed, and with 1x2 or 2x2 spatial diversity, 4-8 blocks in the frequency direction per frame could be sufficient, which enables allocation of larger blocks per chunk. Taking the power efficiency and power control aspects into account, a block size of 4x6 (subcarriers x OFDM symbols) for small packets and 4x12 for large packets seems useful. The even number of OFDM symbols in the RAU enables use of Alamouti precoding (in the time direction, frequency-domain Alamouti cannot be combined with DFT precoding). With an odd number of OFDM symbols, a space-frequency diversity scheme can be applied instead, also with DFT precoding, [JWI06].

C.3.2 FDD Downlink

In the FDD downlink, the B-EFDMA scheme does not apply a DFT precoding step and the envelope property of the signal does not significantly depend on the block allocation. The reason to disregard the envelope properties of the downlink signal is that the RAP does not benefit from a low PAPR of each users signals, since multiple downlink signals are multiplexed including also signals for frequency-adaptive transmission. In addition, the RAP should be able to generate enough power to signal with a constant psd over the allocated system bandwidth (it essentially defines the cell size), and thus power shortage in the transmitter is not an issue as for the uplink.

Thus, localizing the transmission in time is of no disadvantage for the RAP and it enables the user terminal to benefit form micro-sleep within chunks. Common pilots are used for the downlink, so the user terminal could either apply a basic channel estimation algorithm and improve its performance by listening to pilots also in blocks not allocated for it, or use a more advanced channel estimator and benefit from micro-sleep in parts of the chunk.

Thus compared to the FDD uplink, in the FDD downlink a preference for short and wide blocks should be beneficial as shown in Figure C-2 (right), with block size 4x3, 8x3 or 8x6 depending on packet size and number of available antennas. Since no DFT precoding is applied, Alamouti precoding can be applied in the frequency direction for block allocation using an odd number of OFDM symbols.

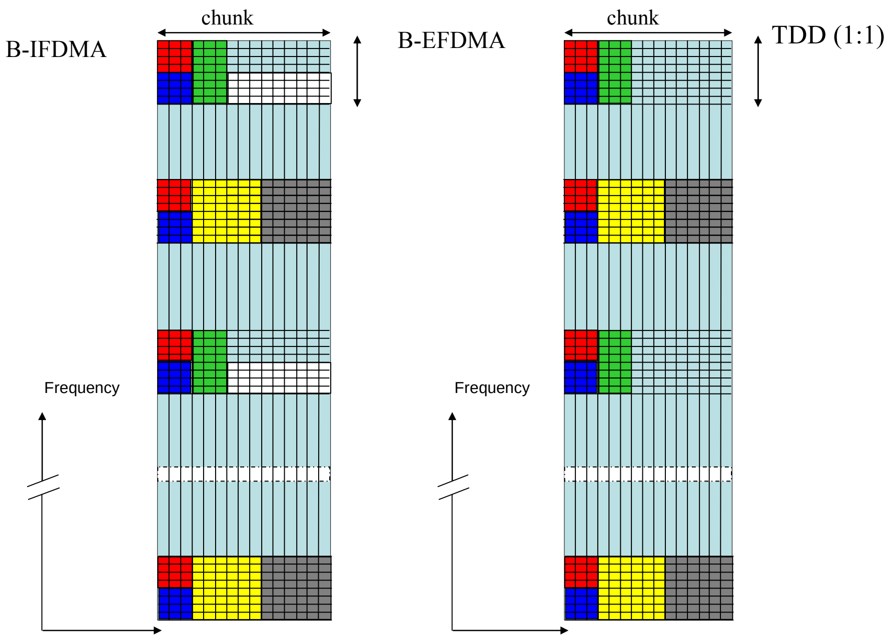

Figure C-3: Configuration of B-IFDMA and B-EFDMA in TDD.

C.3.3 TDD uplink and downlink

In local area deployments, most users will be served by frequency-adaptive transmission, but fast control channels would benefit from large diversity. Typically multiple antennas are used and in deployment in licensed bands, limited frequency-diversity might be needed. Thus referring to Figure C-3, a few rather large blocks could be used consisting of 8x6 (subcarriers x OFDM symbols) for payload data. Control signals for frequency-adaptive transmission still need short blocks for timing reasons, and the 8x3 blocks might be useful. In deployments in unlicensed bands with narrowband interferers, the smallest 4x3 blocks would be needed to maximize robustness towards the narrowband interferers.

In metropolitan area scenarios, there is typically a mix of users with good SINR that can benefit from frequency-adaptive transmission and users with bad SINR that cannot benefit from frequency-adaptive transmission. For the uplink, there are also users with good SINR but with a speed above pedestrian speed, which with the reference pilot design cannot use frequency-adaptive transmission. Thus the preferred block allocations are basically the same as for the FDD uplink and downlink.

In low load scenarios with low inter-cell interference levels, the user terminal could however be less power limited than in the FDD scenario due to the smaller cells. This would enable a preference for wider blocks 8x3, 8x6, keeping the same number of blocks in the frequency direction for the same frequency-diversity gains, thus localizing the transmission more in time and still operating with the power amplifier close to P_{opt} (but with somewhat larger heat dissipation due to larger backoff requirement of the power amplifier).

On the other hand, some users could be severely shadowed (indoor-to-outdoor) and they would be severely power limited under a power control scheme with a target psd per block. Thus, these users would benefit from a few long blocks (4x9 subcarriers x OFDM symbols in the example in Figure C-3) and large spatial diversity.