Mobile and wireless communications Enablers for the Twenty-twenty Information Society-II

Deliverable D2.4
Final Overall 5G RAN Design

Version: v1.0

2017-06-30

The 5G Infrastructure Public Private Partnership

http://www.5g-ppp.eu/

Deliverable/Report D2.4
Final Overall 5G RAN Design

Grant Agreement Number:	671680
Project Name:	Mobile and wireless communications Enablers for the Twenty-twenty Information Society-II
Project Acronym:	METIS-II
Document Number:	METIS-II/D2.4
Document Title:	Final Overall 5G RAN Design
Version:	v1.0
Delivery Date:	2017-06-30
Editor(s):	Hans-Peter Mayer, Nokia Bell Labs Ömer Bulakci, Huawei Technologies GRC Gerd Zimmermann, Deutsche Telekom Mårten Ericson, Ericsson
Authors:	Paul Arnold, Nico Bayer, Jakob Belschner, Thomas Rosowski, Gerd Zimmermann, **Deutsche Telekom** Mårten Ericson, Olav Queseth, **Ericsson** Ömer Bulakci, Emmanouil Pateromichelakis, Malte Schellmann, Panagiotis Spapis, **Huawei Technologies GRC** Yang Yang, Ingolf Karls, **Intel** Shubhranshu Singh, **ITRI** Jens Gebert, Tommi Jokela, Patrick Marsch, Hans-Peter Mayer, Athul Prasad, Mikko Säily, **Nokia Bell Labs** Salaheddine El Ayoubi, **Orange** David Gutierrez Estevez, Tomasz Mach, Mehrdad Shariat, **Samsung** Sergio Barberis, Giorgio Calochira, Alessandro Trogolo, **Telecom Italia** Ji Lianghai, Hans Schotten, **University of Kaiserslautern** Daniel Calabuig, Jose F. Monserrat, **Universitat Politècnica de València**
Keywords:	5G, RAN design, requirements, CP/UP design, air interface landscape, harmonization, frequency bands, bandwidth demand, spectrum management, CN/RAN interface, functional design, network slicing
Status:	final
Dissemination level:	Public

Executive Summary

This deliverable presents the consolidated results of the METIS-II project on the 5th generation (5G) Radio Access Network (RAN) design.

The 5G RAN shall be designed to efficiently support a **wide range of services** and use cases spanning from extreme Mobile BroadBand (eMBB /xMBB) through massive Machine Type Communications (mMTC) to Ultra-Reliable Low Latency Machine Type Communications (URLLC/uMTC) ultimately using the entire **range of available spectrum** for 5G.

We summarize the results, findings and innovations of METIS-II at project end, the main ones being:

- The design rules underlying the **5G air interface** (AI) and the definition of a **framework for** the **harmonization and integration** of the different AI Variants (AIVs) which were developed for the different services. LTE(-A) is included here as a variant. This integration happens at RAN level, it allows to re-use most of the network functions and it foresees a common RAN-Core Network (CN) interface.
- The logical split between **RAN and CN** together with the **interfacing options**, and functions like mobility and paging that may be shifted from CN to RAN to better support new 5G services.
- A framework for **agile resource management** from functional, protocol and deployment perspectives.
- An analysis of the **split options** that exist in the RAN for both the control and the user plane enabling deployment topologies between the extreme cases of fully centralized and fully distributed deployment topologies.
- An architectural solution implemented into the Management & Orchestration (MANO) framework to **dynamically manage all spectrum** classes (licensed, license-exempt, etc.) in space and time.
- **Functional design** extensions like the introduction of a **new radio resource control (RRC) state** offering a reduced control plane latency for an improved battery efficient support of machine-type devices, the introduction of a **make-before-break handover** for enhanced reliability or **RAN based paging** allowing the tracking of devices on cell level. Further, there is the introduction of a new, **more efficient initial access** following the rules of lean design, and the integration of **device-to-device** communication and **self-backhaul** as integral parts of the RAN design.

In this deliverable, we link the design options and proposed solutions to the key performance indicator (**KPI) framework** established at the beginning of the project and the **performance analysis** used to assess the design choices.

Contents

List of Abbreviations and Acronyms

3GPP	3rd Generation Partnership Project
4G	4th Generation of mobile networks
5G	5th Generation of mobile networks
5G PPP	5th Generation Infrastructure Public Private Partnership
5G-PSS	5G Primary Synchronisation Signal
5GS	5G System
5G-SSS	5G Secondary Synchronisation Channel
AC	Admission Control
AI	Air Interface
AIV	AI Variant
AM	Acknowledged Mode
AMF	Access and Mobility Management Function
AN-I	Access Network – Inner layer
AN-O	Access Network – Outer layer
AP	Access Point
API	Application Programming Interface
ARP	Allocation and Retention Priority
ARQ	Automatic Repeat Request
AS	Access Stratum
AR	Augmented Reality
AUSF	Authentication Server Function
BF	BeamForming
BH	BackHauling
BLER	BLock Error-Rate
BS	Base Station
BSS	Business Support Systems
CA	Carrier Aggregation

C-RAN	Centralized/Cloud-RAN
CDF	Cumulative Distribution Function
CM	Configuration Management
CMP	Control-Management Plane
cmW	Centimeter Wave
CN	Core Network
CoMP	Coordinated MultiPoint
CP	Control Plane
CP-OFDM	Cyclic Prefix-OFDM
CPF	Control Plane Function
CPRI	Common Public Radio Interface
CRS	Cell-specific Reference Signal
CSI	Channel State Information
CSI-RS	Channel State Information Reference Signal
CU	Centralized Unit
D2D	Device-to-Device
DC	Dual-Connectivity
DL	Downlink
DM	Domain Management
DMRS	DeModulation Reference Signal
DU	Distributed Unit
DRX	Discontinuous Reception
DTS	Dynamic Traffic Steering
DTX	Discontinuous Transmission
E2E	End-to-End
EM	Element Management
EMBB	Enhanced Mobile Broadband
EPC	Evolved Packet Core
ETH	Ethernet
ETSI	European Telecommunications Standards Institute
FBMC	FilterBank Multi-Carrier

FBMC-OQAM	FBMC-Offset Quadrature Amplitude Modulation
FBMC-QAM	FBMC-Quadrature Amplitude Modulation
FEC	Forward Error Correction
FFT	Fast Fourier Transform
FP	Frame Programme
FQAM	Frequency Shift Keying and Quadrature Amplitude Modulation
FTP	File Transfer Protocol
GHz	Giga Hertz
GRE	Generic Routing Encapsulation
GTP	Generic Tunnelling Protocol
HARQ	Hybrid Automatic Repeat reQuest
HW	HardWare
I/Q	In-phase/Quadrature
ICIC	Inter-Cell Interference Coordination
IFFT	Inverse FFT
IM	Interference Management
IMT	International Mobile Telecommunications
IMT-2020	IMT for year 2020 and beyond
IoT	Internet of Things
IP	Internet Protocol
IRP	Integration Reference Point
ISG	Industry Specification Group
Itf-N	Northbound Interface
ITS	Intelligent Transport Systems
ITU	International Telecommunication Union
ITU-R	ITU – Radiocommunication Sector
JT	Joint Transmission
JT-DC	Joint Transmission with Dummy Symbols
KPI	Key Performance Indicator
LAA	Licensed Assisted Access
LDPC	Low-Density Parity Check

LE	Logical Entity
LEG	Logical Entity Group
LSA	Licensed Shared Access
LTE (-A)	Long Term Evolution (-Advanced)
MAC	Medium Access Control
MANO	Management and Orchestration
MBB	Mobile BroadBand
MC	Multi-Connectivity
MCP	Measurement Configuration Profile
MHz	Mega Hertz
MIB	Master Information Block
MIMO	Multiple-Input Multiple-Output
MME	Mobility Management Entity
mMTC	Massive MTC
mmW	Millimeter Wave
MNO	Mobile Network Operator
MRS	Multi-port multi-beam Reference Signal
MT	Mobile Terminated
MTC	Machine Type Communications
NAS	Non-Access Stratum
NE	Network Element
NF	Network Function
NFV	Network Function Virtualization
NGMN	Next Generation Mobile Networks
NIDD	Non-IP Data Delivery
NE	Network Element
NM	Network Management
NN	Network Node
NR	New Radio
OAM	Operations, Administration and Maintenance
OFDM	Orthogonal Frequency Division Multiplex
ORI	Open Radio equipment Interface
OSS	Operations Support System
PBCH	Physical Broadcast Channel

PCF	Policy Control Function
PDCCH	Physical Downlink Control Channel
PDCP	Packet Data Convergence Protocol
PDU	Protocol Data Unit
PGIA	Pre-emptive Geometrical Interference Analysis
PGW	Packet Gateway
PHY	Physical layer
PLMN	Public Land Mobile Network
PNF	Physical Network Function
P-OFDM	Pulse-shaped-OFDM
PPDR	Public Protection and Disaster Relief
PRB	Physical Resource Block
PVNO	Private Virtual Network Operator
QCI	Quality of service Class Identifier
QoE	Quality of Experience
QoS	Quality of Service
RACH	Random Access CHannel
RAN	Radio Access Network
RCM	RAN Configuration Mode
RF	Radio Frequency
RLC	Radio Link Control
RM	Resource Management
RMF	Reconfiguration Management Functionality
RNTI	Radio Network Temporary Identifier
RRC	Radio Resource Control
RRM	Radio RM
RRU	Remote Radio Unit
RS	Reference Signal/Symbol
RSC	Regulator Spectrum Coordination/ Resource Sharing Cluster
RSPG	Radio Spectrum Policy Group
RSRP	Reference Signal Received Power

RTA	RAN Tracking Area
RU	Radio Unit
SAC	Spectrum Assignment Coordination
SC-FDMA	Single Carrier Frequency Division Multiple Access
SC-SACF	Self-Configuration Spectrum Assignment Function
SCaaS	Small Cell as a Service
SCS	Self-Configuration Supervisor
SDAP	Service Data Adaptation Protocol
SDN	Software Defined Networking
SDR	Software Defined Radio
SDU	Service Data Unit
SE	Spectral Efficiency
SFN	Single Frequency Network
SGW	Serving Gateway
SIB	System Information Block
SINR	Signal-to-Interference and Noise Ratio
SL	SideLink
SLA	Service Level Agreement
SMF	Session Management Function
SMS	Spectrum Management System
SN	Sequence Number
SON	Self-Organizing Networks
SSS	Secondary Synchronisation Signal
SW	SoftWare
TAU	Tracking Area Update
TeC	Technology Component
TDD	Time-Division Duplex
TTI	Transmit Time Interval
UC	Use Case
UDN	Ultra-Dense Network
UDP	User Datagram Protocol
UE	User Equipment
UL	Uplink
UM	Unacknowledged Mode

uMTC	Ultra-reliable MTC
UP	User Plane
UPF	User Plane Function
URLLC	Ultra-Reliable Low Latency communications
V2X	Vehicle-to-Anything
VNF	Virtual Network Function
VR	Virtual Reality
VRM	Virtualized Resource Management
WG	Working Group
WLAN	Wireless Local Area Network
W-OFDM	Windowed-OFDM
WP	Work Package, White Paper
WRC	World Radiocommunication Conference
XaaS	Anything as a Service
xMBB	Extreme Mobile Broadband
x-haul	Backhaul / Midhaul / Fronthaul
ZT-DFTs-OFDM	Zero Tail DFT-spread-OFDM

1 Introduction

1.1 Motivation and Scope of this Deliverable

The METIS-II project aims at a consolidated overall design for the Radio Access Network (RAN) of 5G, responding to the service requirements and Key Performance Indicators (KPIs) that have been set up for 5G systems while relying on the key RAN design paradigms.

The project has started from Technology Components (TeCs) originating from related 5G PPP projects (e.g., [FANTASTIC-5G], [mmMagic], [5GNORMA]), and earlier EU FP7 projects [METIS] and [5GNOW]). METIS-II has then assessed these components and integrated the needed ones into a comprehensive overall functionality framework for the 5G RAN. To make the integration process viable, a number of additionally required TeCs and enablers had to be developed and proposed inside METIS-II. This was done within the METIS-II project.

This overall process, which has been pursued within individual Work Packages (WPs) in METIS-II as well as in the overall RAN design team in WP2, is depicted in Figure 1-1.

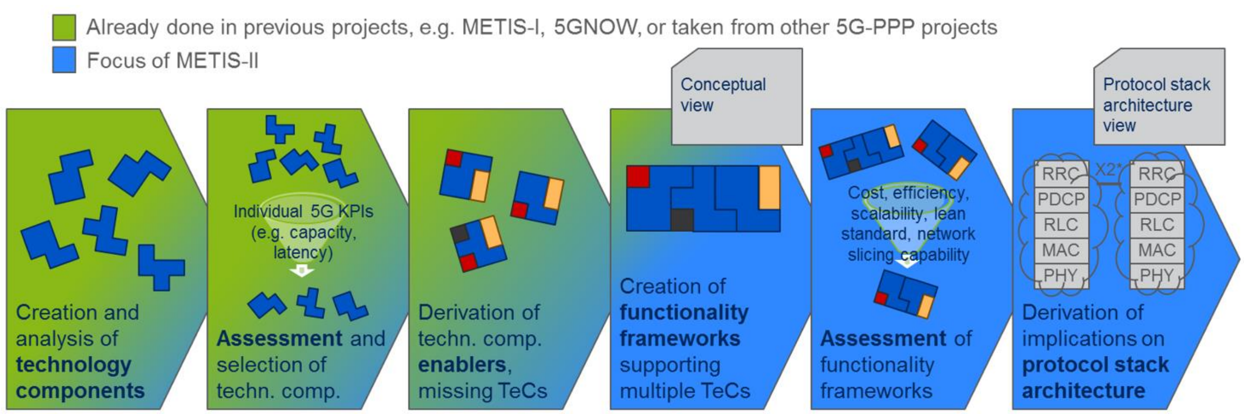


Figure 1-1: 5G RAN design process pursued in the technical WPs in METIS-II.

This deliverable presents an overview of the METIS-II 5G RAN design at project end, describing the current view and high level of consensus reached.

During the runtime of the project, 3GPP has been introducing its 5G terminology [3GPP-38912], which is partially different from the one used in the METIS-II project. For simpler reading, we show both terminologies, using the notion of METIS-II terminology/new 3GPP terminology.

The purpose is to provide a compact and comprehensive summary of the project's results while staying easily readable; however, this deliverable relies on various other documents published by METIS-II. In this document, we are recapitulating the points necessary to understand the context, however, for more detailed information please refer to the earlier documents which are available on the METIS-II web page [METIS-II]. In Figure 1-2, we illustrate the relation between the recent deliverables.

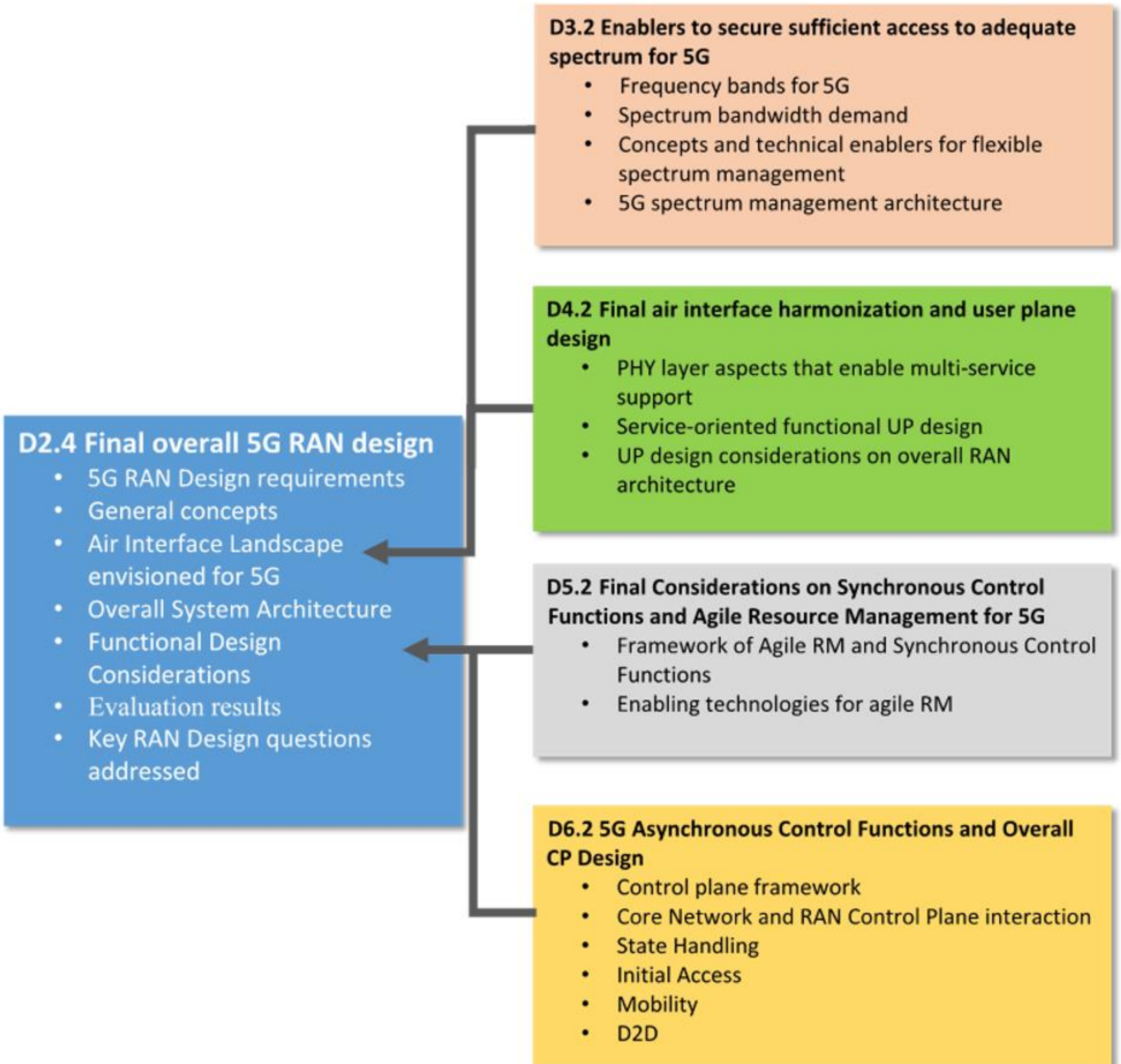


Figure 1-2: Relation of this deliverable to other deliverables with more technical depth.

At the end of the METIS-II project, we have now the performance evaluations available for the concepts considered. A part of these results have lately been presented in the Deliverable D2.3 [MII-D23] not listed in Figure 1-2. In D2.4 (this document), we include the most relevant performance results with the description of each concept, either directly in each paragraph or with a reference to the annex or the corresponding deliverable.

1.2 Structure of this Deliverable

This deliverable is structured as follows: **Chapter 2** starts with a short overview of the 5G service vision that has driven the work through METIS-II with the impact on the RAN ecosystem, and the innovation pillars required to achieve the transition from existing networks to 5G.

Chapter 3 then presents the RAN design requirements that result from the diverse service requirements in 5G, the introduction of network slicing, the integration of multiple Air Interface Variants (AIVs), and the support of functional splits in the RAN. Further considerations aim at fulfilling system targets on latency, spectrum usage, and energy efficiency.

In **Chapter 4**, the air interface aspects are treated starting from bandwidth and spectrum requirements which are considered in particular for the “new” 5G Use Cases (UCs). Then the different modes how the users can access and share spectrum are described, followed by a set of design principles and evaluation criteria for the 5G AI. A brief description of the main AIVs considered in the project is then reported, with the main focus on a framework allowing the integration of different AIVs.

The overall system architecture is discussed in **Chapter 5**, for both the User Plane (UP) and the Control Plane (CP). Here the possible RAN-Core Network (CN) constellations and interfaces are being described for 5G-only systems as well as for the interworking between 4G and 5G including the protocol options for the interface between RAN and CN. For the RAN, the possible functional splits between central and distributed entities are analysed. Furthermore, the elements required for assuring the interworking of different AIVs and for the support of network slicing are presented. The chapter concludes with the description of a high-level framework for an agile spectrum handling and its implementation into the network Management & Orchestration (MANO) framework.

Chapter 6 then presents the functional design aspects describing a service-tailored network according to the “lean design” paradigm that shall assure forward-compatibility and energy efficiency. This is followed by a description of the control functions needed to achieve energy efficient operation. Functions to handle extended topology like native relaying, self-backhauling or Side Links (SLs) are discussed in the following paragraphs, and functions for agile traffic steering are introduced, targeting the integration of multiple AIVs and network slicing. The functions for the initial access, paging mechanisms and mobility management are analysed at the end of this chapter with a particular focus on the newly introduced Machine Type Communication (MTC) services.

While in Chapters 3 to 6 we describe the most relevant evaluation results to motivate the corresponding design elements and decisions, in **Chapter 7** the results of an overall KPI analysis.

Chapter 8 summarizes the key RAN design questions that were set up at project start in a table together with the corresponding findings and **Chapter 9** concludes with a summary and outlook.

2 The METIS-II Vision on 5G

2.1 Envisioned 5G Service Landscape

There is a broad consensus that 5G will not just be a "business-as-usual" evolution of 4G networks with new spectrum bands, higher Spectral Efficiency (SE) and higher peak throughput, but will also target new services and business models. The main 5G service types are (see Figure 2-1):

- **Enhanced Mobile BroadBand (eMBB)**: Human-centric use cases for enhanced access to multi-media content, services and data with improved performance and increasingly seamless user experience. This usage scenario covers a range of cases with different requirements, e.g., the hotspot case with high user density, very high traffic capacity and low user mobility, as well as the wide area coverage case with seamless radio coverage providing strongly-improved user data rate when compared to existing systems with medium to high user mobility. This service is also referred to as **xMBB**, for extreme Mobile BroadBand. We will use the two terms, eMBB and xMBB, interchangeably.
- **Ultra-Reliable Low Latency Communications (URLLC)**: Stringent requirements for capabilities such as throughput, latency, reliability, and availability. Examples: wireless control of industrial manufacturing or production processes, remote medical surgery, distribution automation in a smart grid, transportation safety, etc. This service is also called **uMTC**, for ultra-reliable MTC. The two terms, uMTC and URLLC, are used interchangeably.
- **Massive Machine Type Communications (mMTC)**: Characterized by a very large number of connected devices typically transmitting a relatively low volume of non-delay-sensitive data. Devices like simple sensors are required to be low cost, and have a very long battery life.

At the beginning of METIS-II, 21 UCs were identified in total, with the five main ones:

- the "**Dense urban information society**" with the connectivity required at any place and at any time by humans in dense urban environments,
- the "**Virtual reality office**" use case is related to the evolution of today's tele-presence services into high-resolution 3D versions.
- the "**Broadband access everywhere**" use case is related to the constant increase of the demand for very high data rate.
- the "**Massive distribution of sensors and actuators**" use case covers the massive deployment of low cost and low energy consumption connected devices.

- the “**Connected cars**” use case addresses information exchange among vehicles and with the road-side infrastructure to enable the provision of safety hints to the driver or warnings about the road status. It also addresses xMBB services on-board of cars.

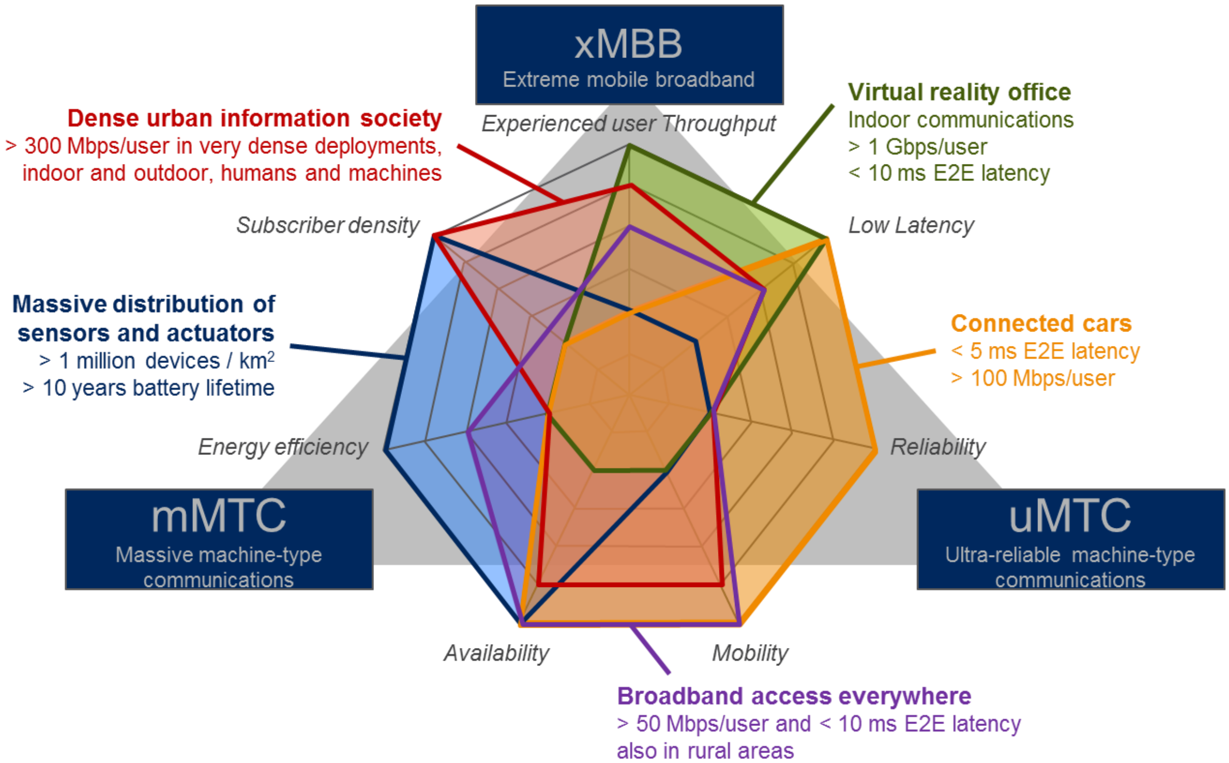


Figure 2-1: Main 5G service types considered and representative use cases [MII-D11].

2.2 RAN ecosystem evolutions with 5G

The new service landscape described above, with stronger focus on MTC and vertical industries in the future, in addition to the generalization of virtualization and the massive deployments of small cells required to fulfill the capacity demand for xMBB, are expected to induce major changes into the ecosystem of the RAN in the 2020-2030-time horizon. These evolutions have been described in [MII-D11], and a discussion around the techno-economic aspects can be found in [MII-D12]. In particular, new value chains are expected to emerge, such as:

- the **Small Cell as a Service (SCaaS) value chain**, where one or several players invest in the deployment of small cells in a particularly crowded place. These small cells may be deployed and managed by a single operator, a joint venture between operators, an urban street furniture owner, a manager of dedicated facility (e.g. a shopping mall), etc.

- the **partner service provider value chain**, where a manufacturer would sell his device or smart object to the end-user. This product would include a service relying on a connectivity that could be provided either directly by contracting a Mobile Network Operator (MNO), or indirectly through a partnership between the device manufacturer and a connectivity provider.
- the **Private Virtual Network Operator (PVNO) value chain**, where a utility provider would decide to rely on frequencies and RAN infrastructures of a commercial MNO, but would still own and operate all or part of the elements of the CN. This configuration is facilitated by the “network slicing” concept developed later in this document.

These new evolutions will introduce major changes in the value chain of MNOs, with new actors entering and others changing their roles or taking new roles. For example, the increased heterogeneity and the virtualization of networks are expected to diversify the list of suppliers of MNOs, where IT companies will be able to provide processing servers and virtual network SoftWare (SW). Verticals will be omnipresent in the evolved value chain, partly buying services from the operators, partly running own networks, and, thus, responding to new needs for wireless connectivity in the society.

2.3 Summary of the Innovation Pillars

METIS-II has developed key innovation pillars [MII-D22] as frameworks of essential concepts and functional designs to enable 5G services and their requirements. While the key innovation pillars are described and analysed in their respective deliverables in detail, herein, we provide the highlights only and refer to the associated sections within this deliverable.

- The **Holistic spectrum management architecture** defined by METIS-II enables flexible spectrum management and multi-operator collaboration in 5G, by integrating numerous frequency bands within a wide spectrum range with possibly different spectrum access schemes, and coping with the versatile spectrum requirements from different user groups. It is based on the enhanced concept for spectrum management and sharing, briefly introduced in Section 4.1.3. The architecture concept embraces the regulatory domain covered by a Spectrum Management System (SMS), and the operator domain which consists of a central Spectrum Assignment Coordination (SAC) entity supported by a number of further functional blocks (see Section 5.5.2). The SAC is going to be integrated into the 5G network MANO framework as briefly outlined in Section 5.5.3. More details can be found in [MII-D32].
- **Holistic air interface harmonization framework** is a new approach to common overarching 5G AI design based of various AIVs, including novel and legacy ones such as Long Term Evolution-Advanced (LTE-A). The objective of the framework is to address 5G use cases and KPIs, minimize the implementation cost and complexity (e.g. in a multi-waveform transceiver implementation) without significantly sacrificing the performance of

individual AIVs. This could be achieved by introducing a degree of common or parameterized generic functions across AIVs in different layers of the protocol stack including Physical layer (PHY), Medium Access Control (MAC) or higher. Some examples include harmonized common frame structures, parameterized PHY numerologies or generic Hybrid Automatic Repeat Request (HARQ) concept tailored to different service requirements or radio channel characteristics. While the holistic AI harmonization framework was studied in detail in [MII-D41] and [MII-D42], the main 5G AI design recommendations are highlighted in Section 4.2 of this deliverable.

- **Agile Resource Management (RM) framework** provides holistic RM solutions that consider and exploit the novel aspects of 5G systems, such as, very diverse service requirements, existence of multiple AIVs in the overall 5G AI, dynamic topologies (e.g., based on vehicular Nomadic Nodes (NNs) as part of moving networks), and novel communication modes (e.g., Device-to-Device, D2D). Within the context of agile RM, METIS-II extends the notion of a resource beyond conventional radio RM (RRM), and aims to attain the optimum mapping of 5G services to any available resources when and where needed within this extended realm of resources. The framework comprises paradigm shifts in terms of the re-design of functions, e.g. operation of control functions on a faster time scale as compared to legacy, as well as new functional considerations for the emerging enablers of a 5G system, such as, multi-slice RM. Another peculiarity of this framework is the AIV-overarching RM, where the developed hierarchical CP design enables a multi-AIV operation including novel 5G AIVs and legacy AIVs. Consequently, design recommendations are derived as the main outcome of this work. While the agile RM framework is detailed in [MII-D52], various highlights are captured in the following Sections 5.3.3, 0 and 6.2.
- The **cross-layer and cross-air-interface system access and mobility framework** enables system access and mobility functions to be used by different AIVs, regardless of the use case. To improve the reliability and resource usage of the system, a Multi-Connectivity (MC) solution is developed. This enables the User Equipment (UE) to be connected to one or more Base Stations (BSs) with same or different AIV simultaneously. A single flexible system access solution is developed that can serve different 5G services with a wide range of requirements. The cross-layer and cross-air-interface system access and mobility framework is covered in detail in [MII-D62], and a summary is given in Section 6.3.
- A **common framework for control and user plane** which consists of both the synchronous and asynchronous CP and UP functions. The synchronous functions require frame/slot/sub-frame or any time-domain level synchronization between a set of functions (for instance related to dynamic scheduling and power control), which are captured in [MII-D52]. The asynchronous functions do not require frame/slot/sub-frame or any time-domain level synchronization (for instance mobility and initial access functions), which are handled in [MII-D62]. This resultant innovation pillar is seen as essential to enable fast data flow routing, CP/UP diversity, throughput aggregation, reliable mobility management, etc. in

order to fulfil the diverse and stringent 5G requirements. A summary of the results are presented in the context of describing the envisioned overall 5G RAN architecture in Chapter 5.

3 Key 5G RAN Design Requirements

In order to support the design of the 5G RAN, METIS-II has identified several requirements [MII-D22] which are needed to meet the diverse service requirements stated in Section 2.

- The **5G RAN should be highly scalable** with respect to parameters like throughput, the number of devices or the number of connections.
- One enabler for the system to handle the diverse service requirements stated before is that **the overall network (both RAN and CN) should be software-configurable**. This means, for instance, that it is configurable which sets of logical and physical entities are to be traversed by CP and UP packets.
- The 5G RAN must be **designed to operate in a wide spectrum range** with a diverse range of characteristics such as bandwidths and propagation conditions. For higher frequency bands such as millimetre wave (mmW), Beam Forming (BF) will become essential.
- The 5G RAN should enable a tight interworking between **LTE-A evolution and novel 5G radio technology on RAN level**.
- The **5G RAN should natively and efficiently support MC,** i.e. the case when the UE is connected to more than one radio node (inter-node, i.e. not co-located) and / or more than one AI (which may be co-located or not).
- **The 5G RAN should natively support network-controlled D2D** (i.e., point-to-point, multicast and broadcast), including the option that some **5G devices could flexibly act as if they were infrastructure nodes**, one example being self-backhauled, possibly nomadic access nodes.
- The 5G RAN should be designed such that it can maximally **leverage from centralized processing of radio layers**, but also operate well in the case of distributed BSs with imperfect x-haul (backhaul / midhaul / fronthaul) infrastructure.
- The 5G RAN design must be **energy efficient**. This means that permanently active network functions or signals have to be avoided.
- **The 5G RAN design must be future proof**, i.e., it should enable an efficient introduction of new features and services (e.g., by minimizing the spreading of signals over radio resources and facilitating the introduction of new physical channels) and guarantee backward-compatibility of devices in future releases.

In the following sub-sections, we will present more detailed requirements for specific design aspects: Treatment of diverse services, integration of AIVs, support of functional split in the RAN, and control functions.

Solutions that respond to the design requirements will be described in Sections 4 to 6. An evaluation with respect to the KPIs will be given in Section 7, which adds to the performance results in [MII-D23].

3.1 Design Requirements specifically related to Diverse Services and Network Slicing

The envisioned **set of 5G services** described in Section 2.1 **and their diverse and partially conflicting requirements** will pose the following requirements on the 5G RAN design:

- **Traffic differentiation**: The 5G RAN should support more sophisticated mechanisms for traffic differentiation than legacy systems in order to be able to treat heterogeneous services differently and fulfil more stringent Quality of Service (QoS) requirements, e.g. raised by uMTC/URLLC services. Potential solutions are described in Section 6.2.
- **Resource reuse**: 5G networks should support a strong reuse of resources (e.g., radio, functional, and infrastructure resources; see the extended notion of a resource in [MII-D52]) to enable an economically viable solution for emerging 5G services.

An efficient joint utilization of infrastructure resources by multiple services and differentiated service treatment also prepare the grounds for the introduction of network slicing[1] in 5G RAN. Beyond these aspects, some additional requirements have been identified that are specific to network slicing:

Slice-aware RAN: Slices should be visible to the 5G RAN to enable a treatment of all service flows (or bearers) within one or across slice instances according to related KPIs by applying resources in a joint (shared) or dedicated (separated) way.

- **Slice selection and association**: The 5G RAN functionalities should support the UEs with dedicated selection and association procedures to appropriate slice instances. Simultaneous associations to more than one instance of different network slice types should be feasible for a UE.
- **Slice protection**: The 5G RAN should offer slice isolation and protection mechanisms so that critical fault- or security-related events within one slice instance do not have a negative impact on another one.
- **Slice management**: The 5G RAN should support efficient mechanisms for life-cycle management of slice instances on the common infrastructure.
- **Slice-specific network management**: The 5G RAN should allow offering slice-specific network management functions as a service.

[1] A “network slice” supports the communication service of a particular connection type with a specific way of handling the CP and UP for the services included in the created slice instance throughout core, transport, and (radio) access network, and is seen from a customer perspective as a separated logical network [NGMN15, MII-WP].

3.2 Design Requirements specifically related to Air Interface Integration in 5G

Requirements for the 5G AI design and integration have been elaborated in detail in Deliverable D4.1 [MII-D41]. In line with discussions and agreements ongoing in 3GPP by the time of writing D4.1, analysis of the KPIs defined for 5G enabled the derivation of the following requirements and/or assumptions for the design:

- The main 5G services **xMBB, uMTC, and mMTC** with their adverse requirements **should efficiently co-exist**. It is envisioned that this will drive solutions such as flexible frame structures for the PHY layer design. For the low-data rate case (mainly mMTC services) a massive amount of devices will likely drive the amount of signalling. The design shall also support very high-data rates for xMBB and some uMTC applications such as remote control of infrastructure with high quality video.
- **Efficient usage of spectrum** shall also be supported, as spectrum is the most valuable and scarce resource for radio communication. This would require the design of RRM solutions that support mMTC, uMTC, and xMBB service multiplexing on a time scale and a frequency granularity capable to capture the dynamics of the traffic. Also, dedicated spectrum should be supported in extreme scenarios. Further, means for interference estimation and mitigation facilitate the efficient use of spectrum.
- The design of **reference signals (RS) should allow for a high level of configurability**, possibly exploiting UE-specific reference signals.
- **Low UP latency** for the radio access should be supported, being in the order of 1ms in selected scenarios.
- **Ultra-high reliability within tight latency limits** should be supported for selected services. This will drive more flexible frame structures. Current systems have been designed for delay-tolerant services: With LTE, targeted Block Error Rate (BLER) for the first transmission is in the order of 10%, as it is assumed that consecutive retransmissions based on HARQ process can compensate for information losses and finally achieve the desired reliability by extending the transmissions over time. However, this paradigm needs to be reconsidered for the context of ultra-high reliability with latency constraints.
- **D2D mechanisms should be defined and efficiently exploited** to solve coverage issues as well as to enable the availability and retainability level (which ensures the reliable operation of a service) required for ultra-reliable 5G services like those from the Vehicular to Anything (V2X) context.
- **The UP design should support** sensors or other **low cost devices** with strong demands for low complexity and energy efficiency. Reducing the use of control channels and control signals as well as reducing PHY latency can be considered suitable means to address these demands.

3.3 Design Requirements related to inherent Support of Functional Split

One important requirement for the 5G RAN is to provide sufficient flexibility for the placement of Network Functions (NFs). Furthermore, the number of interfaces between the NFs, clustered on a horizontal layer structure according to the radio protocol stack, should be as small as possible. This is important in order to keep the standardization and testing effort lean, as these interfaces may be standardized and in any case have to be tested together with all alternative combinations before going into operation. It may be even more stressed by multi-vendor implementations and inter-operability testing.

The key rationale behind any choice of function split is to obtain the largest possible extent of centralization that a specific deployment architecture supports. A large extent of centralization of functionalities allows to exploit pooling gains related to, e.g., centralized Joint Transmission (JT), centralized scheduling, centralized flow control etc., but following aspects have to be considered [MII-D22]:

- The **data rate required on the resulting x-haul interfaces**, for instance between a Remote Radio Unit (RRU) at the antenna site and a BaseBand Unit (BBU) hosting the full radio protocol stack or upper parts of it in a decentralized or centralized way (cloud/centralized-RAN, C-RAN).
- In addition, the **latency aspect** is a critical issue for the selection of suitable splits, for instance limiting the implementation of certain functionalities (e.g., Coordinated Multi-Point (CoMP) processing) in the case of some deployment scenarios. A key consideration here is to design 5G RAN functions in a way to avoid strict timing relations between the protocol layers, and to have a clearer split between time-synchronous and time-asynchronous functions, as we will discuss later.
- Finally, **the level of complexity** and **maintenance** requirements on RRU side, on the one hand, versus delivering x-haul interface requirements in **forward compatible** manner, on the other hand, can be other key factors in selection of the splits.

3.4 Design Requirements specifically related to AIV-agnostic and AIV-specific Control Functions

It is envisioned that in 5G, the overall AI comprises different AIVs that are optimized, e.g., for the specific frequency bands of operation (below 6 GHz, mmW, etc.) and for one or more target use cases [MII-D22]. Furthermore, **the AI landscape includes both novel 5G AIVs and legacy AIVs**, where, for example, a RAN level tight interworking is aimed for evolved LTE integration. Accordingly, the 5G RAN CP design needs to factor in not only the diverse service requirements but also the diversity introduced by the existence of the peculiarities of different AIVs. To this end, the CP functions can be categorized under AIV-agnostic and AIV-specific functions. AIV-agnostic functions operate over multiple AIVs and enable addition of new AIVs. On the other hand, AIV-specific functions are tailored towards the characteristics of the target AIV, e.g., PHY layer design. In what follows, design requirements for the AIV-agnostic and AIV-specific control functions are outlined.

- **RM framework shall include both AIV-agnostic and AIV-specific CP functions.** While there can be different AIVs, the RM framework should be agile to operate in an AIV-overarching manner [MII-D51] [MII-D52]. That is, AIV-overarching RM functionality framework shall remain agnostic to the design of the PHY layer of the AIVs that are involved, and, thus, can also operate over a newly introduced AIV. For example, traffic steering should be able to route the service flows to the right AIV based on instantaneous radio link conditions of the AIVs. At the same time, AIV-specific RM mechanisms, e.g., interference management and dynamic resource scheduling, shall operate over the radio frame design of the AIV, such as, duration of time slots and subcarrier spacing.
- **AIV-specific context information shall be made available for AIV-agnostic CP functions.** For efficient operation of the AIV-agnostic CP functions, real-time radio link feedback is needed. The radio link feedback can be utilized to characterize the AIVs, e.g., in terms of mapping mission-critical service flows or activation of packet duplication. The frequency of the radio link feedback depends on the carrier frequencies of AIVs and the associated radio channel characteristics as well as the mobility of the UEs. It can be expected that for a high-speed UE connected to an AIV above 6 GHz, more frequent radio link feedback is required.
- **AIV-agnostic CP functions shall exploit MC**. It is envisioned, that a UE connecting to multiple AIVs will be an essential component of the 5G RAN design [MII-D52] [MII-D62]. This can improve not only the achievable data rates but also the overall link reliability. For example, make-before-break handover relies on momentary MC and enables high-reliability mobility management.
- Beside constraints due to the physical deployment scenarios, AIV-agnostic CP functions potentially reside in a central location, while AIV-specific functions can be implemented

near the antenna sites. Thus, **the functional split shall consider both AIV-agnostic and AIV-specific CP functions.**

- AIV adaptation can be applied to cope with the semi-dynamically changing network conditions, e.g., during stadium events, an AIV tailored for mMTC devices can have extended bandwidth to serve a large number of wearables [MII-D52] [MII-D62]. On this basis, **AIV-agnostic CP functions shall be able to adapt to AIV reconfiguration**.
- **AIV-specific CP functions shall enable service prioritization.** One of the vital aspects of the 5G RAN design is the fulfilment of wide range of service requirements. Particularly, for the mission-critical services a fast access to the network is vital. Initial access schemes [MII-D62], hence, shall take into account the delay requirements of the services and prioritize the service on such needs.

4 5G Air Interface

4.1 5G Spectrum Aspects

This section contains a brief summary of some important results on the spectrum work in METIS-II. Covered aspects are frequency bands, bandwidth demand, and spectrum management concept. For outcome on further issues, e.g. rationale for spectrum above 6 GHz (trials, technologies, wave propagation, coverage, co-existence), or technical enablers, please refer to the respective publications [MII-R31], [MII-D31], and [MII-D32].

4.1.1 Frequency bands for 5G

Spectrum bands under consideration for 5G

Frequency spectrum under consideration for 5G span from 600/700 MHz up to 86 GHz. In addition to the bands already allocated and in use for mobile communications, the bands illustrated in Figure 4-1 are under study for identification for 5G/IMT-2020 at the World Radiocommunication Conference in 2019 (WRC-19).

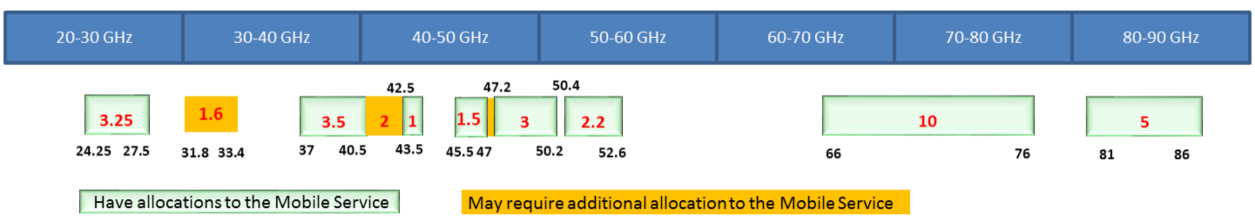


Figure 4-1: Frequency bands under study for identification for 5G/IMT-2020.

In Europe, the Commission has launched an initiative to accelerate the deployment of 5G by 2020. This action plan [EU16-COM588] sets a clear roadmap for investment in 5G infrastructure in the European Union. Furthermore, the Radio Spectrum Policy Group (RSPG) recommended bands at 700 MHz, 3.6 GHz and 26 GHz as pioneer bands for the introduction of 5G based services in Europe [RSPG16-032]. Similar activities are under way in other regions and countries.

Spectrum bands suitable for 5G use case families

As already outlined in [MII-D31], following general conclusions on the suitability of spectrum bands for the three 5G use case families can be drawn:

- For xMBB service types, a mixture of frequency spectrum is required comprising lower bands for coverage and low traffic, and higher bands with large contiguous bandwidth to cope with extreme traffic demand, including wireless backhaul solutions. Exclusive licensed spectrum is essential to guarantee coverage and service quality, supplemented by spectrum access with other licensing regimes (e.g. Licensed Shared Access (LSA) or license-exempt access) to increase the overall spectrum availability.

- For most mMTC service types, frequency spectrum below 6 GHz is more suitable and spectrum below 1 GHz is needed in particular when large coverage areas and outdoor to indoor penetration are needed. Exclusive licensed spectrum is the preferred option. However, other licensing regimes might be considered depending on the specific application requirements.
- For uMTC services, licensed spectrum is considered as most appropriate. For automotive traffic efficiency and safety communications, the frequency band 5875-5925 MHz harmonized for Intelligent Transport Systems (ITS) is an option. For high-speed applications and rural environments, spectrum below 1 GHz is particularly suitable.

4.1.2 Bandwidth demand for 5G use cases

Bandwidth demand estimations for xMBB and uMTC

The spectrum bandwidth demand for 5G services depends on a number of factors, including the use case, the applications used, the deployment scenario, the frequency band, user density and spectrum efficiency. For example, with specific assumptions a total bandwidth demand of 2.4 - 7.1 GHz has been estimated for the xMBB use case “dense urban information society” [MII-D32]. An analysis for the uMTC use case “connected cars – traffic efficiency and safety” showed a bandwidth demand estimate of about 400-800 MHz for communication ranges of 500-1000 m [MII-D32].

5G vertical use cases leading to additional demand

There are important vertical use cases (e.g., broadcast, automation, and public transportation) whose details are only now getting defined. Thus, additional demand for spectrum and bandwidth for 5G is foreseeable and should be taken into account already now.

While [MII-D32] and other spectrum demand studies are evaluating the broad range of existing or foreseeable use cases where the role of 5G is already understood, the quickly evolving digitalization of society and industry will in the future result in many new use cases for 5G that are not yet fully defined or whose feasibility is not yet predictable. For these use cases only initial estimates of the traffic demand are available. The resulting spectrum demand has to be evaluated once details of the use cases and the concepts for 5G support are defined in detail. Many of the use cases are resulting in a traffic demand that is tied to specific locations, vehicles, scenarios, or other side conditions. Thus, the figures cannot simply be compared to the overall traffic and spectrum demand studies described in [MII-D32]. Some examples are:

- Replacing **TV broadcast** services will result in a broadcast traffic load of 200-400 Mbps with the option to use Single Frequency Networks (SFN). Here, various projects [IRT] have studied implementation options using multicast concepts as already defined for LTE. It is, however, expected that even more optimized solutions will help to reduce the spectrum demand significantly.

- **Mobile broadband** support for **public transportation**: High speed trains, for example, might require several Gbps per train. It is expected that due to the competition of bus services (that can much easier support broadband access for their customers) and the growing expectations of customers, train operators will in some years have to support these data rates.
- **Automation** of various **operation processes in public transportation** (e.g., up to driverless trains and the railway station operation) will result in a broad range of new broadband and, in particular, URLLC services.
- Increasing the efficiency of maintenance workers using **Augmented Reality (AR)** will require support of weakly compressed or even uncompressed (if compression latency is too high) HD video for one or several workers in factories but also outside (process automation). Similar to Virtual Reality (VR) office use cases, that would require several Gbps, however here with seamless mobility.
- The **digitalization in industry** is expected to dramatically change industrial processes and services. Ubiquitous and resilient connectivity will be one key enabler; in many cases, today's solutions are not able to meet the requirements and, thus, 5G is an important building block of most factory of the future visions. Details of the many industrial use cases depending on vertical and horizontal connectivity are not yet fully understood. However, it is obvious already today that the use of AR, VR, and video inspection in low latency applications will sometimes result in very high bandwidth needs. Motion control, remote control of robots, communication of autonomous shop floor vehicles etc. will result in URLLC traffic demand. Support of process automation with its thousands of sensors and actuators will result in demand for resilient and secure URLLC services. Spectrum usage options are investigated in [KOI]. Respective use cases and their requirements are currently collected in [3GPP-SP170169], complementing the use cases already analyzed in [3GPP-22862] and [3GPP-22261].

4.1.3 Enhanced Concept for Spectrum Management and Sharing

Radio spectrum is generally authorized in two ways, first by "individual authorization" in the form of license granting, and second by "general authorization" also referred to as license-exempt or unlicensed. The concept developed in [MET-D54] has been enhanced. In METIS-II, four different user modes have been defined under which 5G radio systems are expected to operate: "service dedicated user mode", "exclusive user mode", "LSA user mode" and "unlicensed user mode". The relationship between these user modes and the authorization schemes is visible in the upper part of Figure 4-2, named "regulatory framework domain".

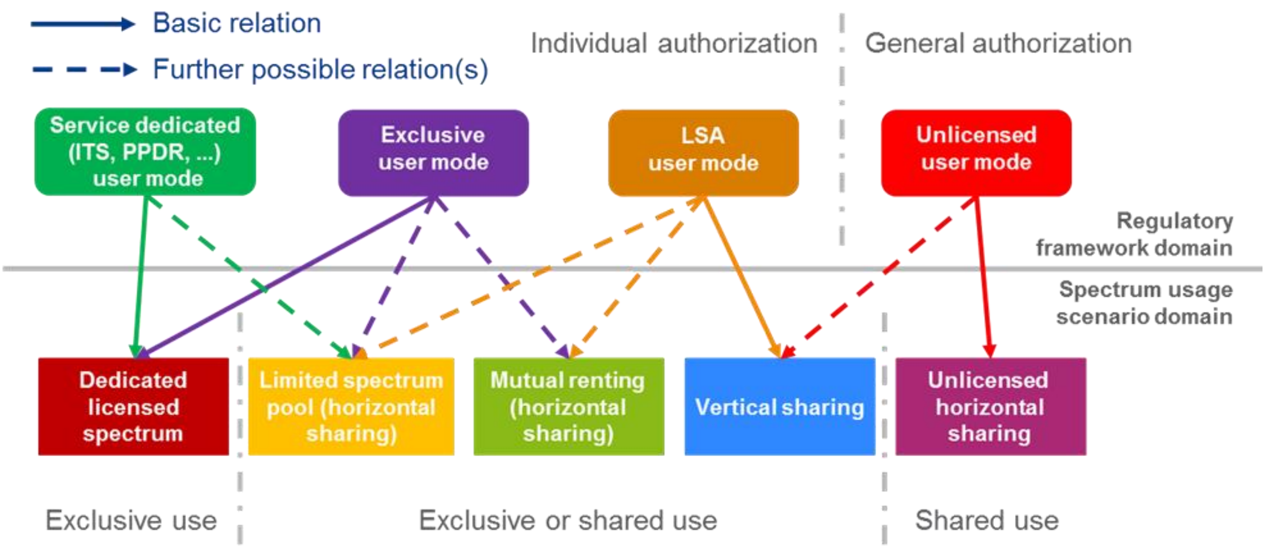


Figure 4-2: Enhanced concept for spectrum management and spectrum sharing for 5G mobile networks.

Spectrum usage rights granted by "individual authorization" are exclusive at a given location and/or time. The "service dedicated user mode" refers to spectrum designated to services other than public mobile communications, which are intended to be integrated into the 5G eco system (e.g. ITS or Public Protection and Disaster Relief (PPDR)). Spectrum designated to public mobile communications falls into the "exclusive user mode". In the "LSA user mode", a non- mobile communications license holder (e.g. fixed radio link service, military service) would share spectrum access rights with one (or more) mobile communications operator(s) under defined conditions subject to an individual agreement and permission by the relevant regulatory authority. These three user modes can occur in their basic form (continuous lines), or as evolution of current approaches in the form of "limited spectrum pool" or "mutual renting" (dashed lines) towards the respective spectrum usage scenario in the lower part of Figure 4-2.

In the "unlicensed user mode", spectrum access and usage rights are granted by general authorization, i.e. without an individual license but subject to certain technical restrictions or conditions like e.g. limited transmission power or functional features like duty cycle or listen-before-talk. In this user mode, no protection from interference caused by other users is guaranteed.

Section 5.5 will outline how to integrate the resulting spectrum management architecture into the network MANO framework of the 5G system. More details on the concept for spectrum management and spectrum sharing for 5G mobile networks can be found in [MII-D32].

4.2 Air Interface Design Considerations

METIS-II has proposed a design framework and suitability assessment process for 5G AI candidates. The proposed assessment methodology focuses on “harmonization KPIs” and how to measure them (qualitatively and quantitatively). It is expected that the elaborated evaluation criteria, which result from the wide consensus reached within METIS-II and have been aligned with 3GPP, will offer a long-term, integrated system view. They aim to impact research and standardization facing the technical and economic trade-offs to be taken into account when assessing new AI technologies.

4.2.1 5G AI Design Principles

A key question related to the 5G system is how the different AI candidate technologies, including LTE-A evolution, can be integrated into one overall 5G AI, such that

- this design supports the wide landscape of bands, cell types etc.,
- both the complexity of the standard and that of the implementation are minimized,

while the performance of individual technologies is not sacrificed. An adaptable and flexible 5G AI design is therefore required to address these issues while efficiently multiplexing multiple services.

In the following table, we summarize key METIS-II principles for the 5G AI design developed in the project [MII-D4.1] [MII-D22] and highlight their target and their relation to the AI design requirement as presented in Section 3.2.

Table 4-1: Key METIS-II design principles for the 5G AI

Design Principle	Target	Related to AI design requirement (Section 3.2)
Flexibility by design	Provide the required flexibility for multi-service support and non-traditional applications	Efficient coexistence of xMBB, uMTC, and mMTC
Forward-compatibility	Ensure future-proofness for upcoming variants of existing 5G services as well as potential new services	Configurable control and reference signals Low UP latency Ultra-high reliability
Easy interworking with evolution of LTE	Allow to integrate LTE-A evolution and novel 5G AIV on RAN level	Efficient coexistence of xMBB, uMTC, and mMTC
Lean Design	Minimize signaling overhead and unnecessary transmissions	Configurable control and reference signals
Beam-centric design of UP/CP signaling	Especially at high frequencies, signals will often be transmitted in beams to account for high path loss	Efficient usage of spectrum

Design Principle	Target	Related to AI design requirement (Section 3.2)
Application Programming Interfaces (APIs) to higher layers	Facilitates the implementation of network slicing, including logical aggregation of UP instances related to different bands	
Terminal complexity	The implementation of one widely harmonized AI is expected to decrease terminal complexity	UP design for low-cost devices

4.2.2 AI harmonization and evaluation criteria

Different proposals for the overall 5G AI design have been developed within METIS-II, but also within other 5G PPP projects, standardization bodies, and elsewhere. These different proposals contain different levels of harmonization. Some alternatives rely on the harmonization of the lower layers, while other solutions rely on the harmonization of the higher layer protocols (with a greater differentiation at lower layers). Each METIS-II proposal is a single framework comprised of multiple AI components selected to jointly fulfil the performance requirements of the different main service types and frequency bands, as depicted in Figure 4-3. Each of these harmonization alternatives could have several (potentially different) benefits. In general, benefits of harmonization include better utilization of available resources due to the flexibility even in short time scales, reduced complexity in the access nodes and the end devices, lower delay in case of switching between AIVs, less standardization and implementation effort and simpler upgrade of an existing system by implementing additional AIVs. In order to evaluate the degree of these benefits contained in different proposals, harmonization KPIs have been defined so that not only performance, but also other, equally important aspects (e.g., cost and complexity as well as switching delay) are taken into account when assessing the relative suitability of different proposals as 5G AI candidates. These harmonization KPIs are described in the following:

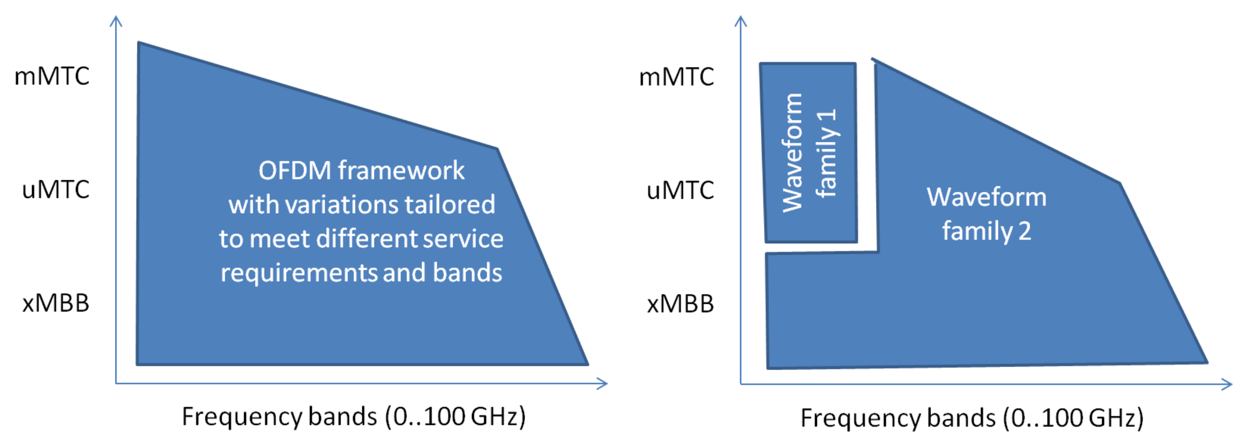


Figure 4-3: High-level examples of potential AI design proposals addressing the 5G landscape.

Ability to dynamically utilize radio resources

This KPI assesses in which time scale the proposed AI can utilize the frequency bands in a given location. The highest level is achieved when multiple services, possibly relying on the same numerology (e.g., frame structure) can be scheduled in the fastest possible time scale (i.e., on a Transmit Time Interval (TTI)-basis), in order to capture the dynamics of the traffic demands on these services and maximize the resource utilization. The lowest level is when a dedicated portion of the spectrum must be allocated in a large time scale (higher than minutes / hours) so that no other service can utilize that due to design reasons. In the case of multiple numerologies [MII-D42], one should assess the ability to schedule multiple shorter TTIs within longer TTI periods.

Support of UP aggregation

This KPI assesses the degree of ability to aggregate multiple AI components (i.e. functional blocks, for details see [MII-D41]) on different layers of the protocol stack to support UP aggregation. Aggregation on a certain protocol stack layer means that on and above that layer, there is only one single logical protocol stack instance, and hence the higher layers are agnostic to the existence of multiple protocol stack instances at the lower layers.

Ability to reuse software and hardware components among components of new AI

This KPI assesses the ability to reuse SW and hardware (HW) components by the different AI components / instantiations, for both the UE and the network equipment.

For networks with a heterogeneous set of AI components supported by the UEs and the network there will be variations in the number of devices using a particular component. This is caused by fluctuations of the number of users in the network, as well as a requirement to use AI components that are supported by both the network and the UE. Reusing components is good because it avoids implementation of multiple radio chains where only one is used at a time.

Standardization effort and product development of AI proposals (time to market)

This KPI assesses the amount of work needed to standardize and develop the different AI proposals. This effort translates to additional standardisation time and thus increase the time-to-market for a new feature, a new scenario or a new service. The amount of effort can be measured approximated by the number of features / protocol layers that can be reused by the multiple AI proposals.

Ability to integrate new AI proposals with LTE-A

This KPI assesses the ability of a proposal to integrate with LTE-A, using the KPIs explained above. There is a consensus in METIS-II that the new 5G AI should not be constrained to be backwards compatible with LTE-A. However, some benefits exist in harmonizing at least some 5G AI aspects with the LTE design, such as the possibility to reuse HW and SW components and perform HW load balancing (see previous subsection), as well as a potential reduction in the standardization effort. Within METIS-II, there is a consensus that LTE and 5G AIVs would likely be integrated on Packet Data Convergence Protocol (PDCP) / Radio Resource Control (RRC) level.

Forward compatibility

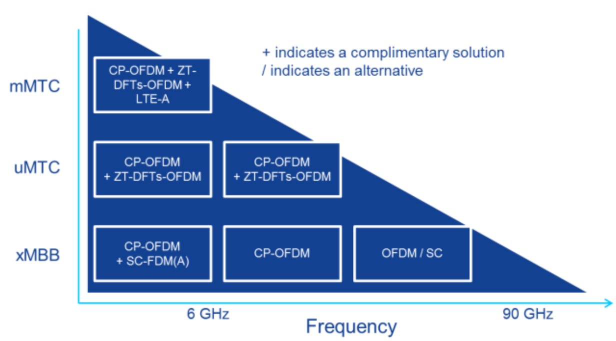
This KPI assesses the ability to efficiently introduce new features and services in the future without the need for an AI re-design. Beyond harmonization, METIS-II also investigated to which extent UP instances related to different bands can be logically aggregated and on which layer(s), and beyond which layer there would be a single CP instance. Different AI design proposals may offer different support of CP features, which needs to be considered.

4.2.3 5G AI Proposals considered in METIS-II

METIS-II elaborated on different proposals for the 5G AI. A significant difference between those proposals is in the use of different waveform concepts. In particular, two approaches have been followed:

1. 5G AI based on Orthogonal Frequency Division Multiplex (OFDM) waveform with variations (in particular with additional windowing and filtering), tailored to meet different 5G service requirements and bands.
2. 5G AI based on multiple waveforms (in particular OFDM and Filterbank Multi-Carrier (FBMC)), providing the system more degrees of freedom to adapt to the requirements of the different services and to enable additional performance gains.

An example for each of these approaches, following the illustration of the 5G landscape as introduced in Figure 4-3, is presented in Figure 4-4 (for further details, refer to [MII-D22]). The two approaches have been evaluated separately by applying the above evaluation criteria; the details of these evaluations can be found in Deliverable D4.1 [MII-D41].



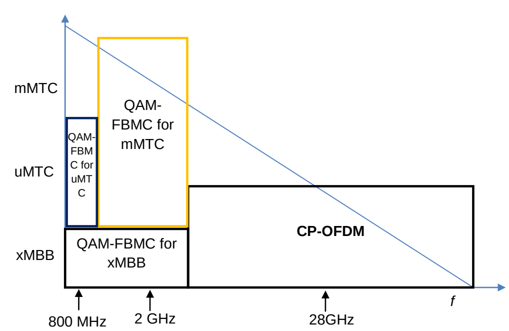


Figure 4-4 Examples for an OFDM based (left) and a multiple waveform based (right) AI framework. CP stands for cyclic prefix in this figure.

Harmonization framework for the implementation of multiple waveforms

In addition to the previous evaluation of the two proposals for the 5G AI waveform, METIS-II has proposed a harmonization framework to implement any of these proposals (for details see [MII-D42]). In particular, six waveforms have been analyzed, including classical Cyclic Prefix-OFDM (CP-OFDM), Windowed-OFDM (W-OFDM), Pulse-shaped-OFDM (P-OFDM), Single Carrier Frequency Division Multiple Access (SC-FDMA) or Zero Tail DFT-spread-OFDM (ZT-DFTs-OFDM), FBMC-Quadrature Amplitude Modulation (FBMC-QAM), and FBMC-Offset QAM (FBMC-OQAM). In this analysis, common blocks in the transmission and reception chains of the waveforms have been found, which allows an implementation that reuses HW in such a way that more than 50% of chip space for the multi-waveform implementation can be saved. This framework will facilitate the definition of a harmonized 5G AI that supports all services in all frequency bands.

This framework allows different implementation possibilities, realizing different trade-offs between chip space and clock speed. More specifically, the same multi-waveform AI can be implemented using a lot of chip space and a low clock frequency, or little chip space and a high clock frequency. For any possible implementation, it is important to take into account that an implementation that requires less chip space, requires higher clock frequency.

5 Overall System Architecture

The overall system architecture is typically standardized in order to enable interoperability among the equipment from multiple manufacturers. The most fundamental part, which is also the focus of this chapter, is the mobile network architecture, which comprises both CN and RAN domains, the definition of NFs, standardized interfaces and protocols running over these interfaces. More precisely, this chapter covers CN considerations, CN / RAN functional split, network interfaces and the protocol architecture for the AIs. The implementation or deployment of the logical architecture on a physical architecture comprises aspects such as x-haul, constraints in terms of HW and SW platforms, placement of the functions in the mobile network sites (access sites, aggregation sites, etc.), usage of cloud environments, centralization and distribution, etc. In addition to these aspects, management and orchestration has gained a lot of attention in the past few years in the context of the 5G architecture. This has happened due to the expectations that at least a subset of the 5G NFs (e.g. CN-specific NFs) would be based on cloud platforms.

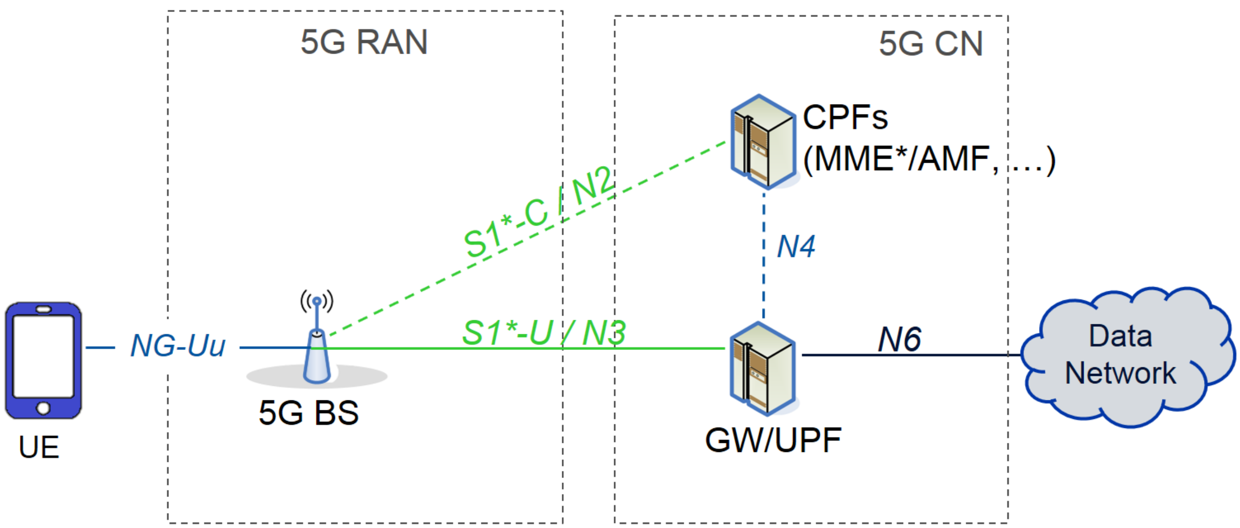


Figure 5-1: System Architecture for the 5G System (5GS)

The 5G System as shown in Figure 5-1 consists of the 5G RAN, the 5G CN, and the UE. In this high level view, the 5G RAN includes the 5G BS, which supports novel 5G AIVs (e.g. New Radio (NR) as known from 3GPP terminology, then the BS would be a 3GPP gNB) and/or the evolved LTE-A AIV (then the BS would be a eNB). The BS can be split in different parts, e.g. in a Distributed Unit (DU) and a Centralized Unit (CU) as shown later in this section. The functions in the 5G RAN can be further split into control plane functions (CPF) and user plane functions (UPF) as already done in the 5G CN [3GPP-23501]. The 5G Mobility Management Entity (MME*), also more recently noted in 3GPP as Access and Mobility Management Function (AMF), is shown in this figure while, for simplicity, other CPFs like the Session Management Function (SMF), the Authentication Server Function (AUSF), and the Policy Control Function (PCF) are not explicitly shown.

Please note that 3GPP is currently introducing the 5G terminology, which is partially different from the terminology used in the METIS-II project. In order to show both terminologies, we use here the notion of METIS-II terminology/new 3GPP terminology. Taking the interface between the 5G BS and the UPF as example, S1*-U/N3 means that in METIS-II, we use the term S1*-U for this interface while latest 3GPP drafts (where the terminology still may change) use the term N3.

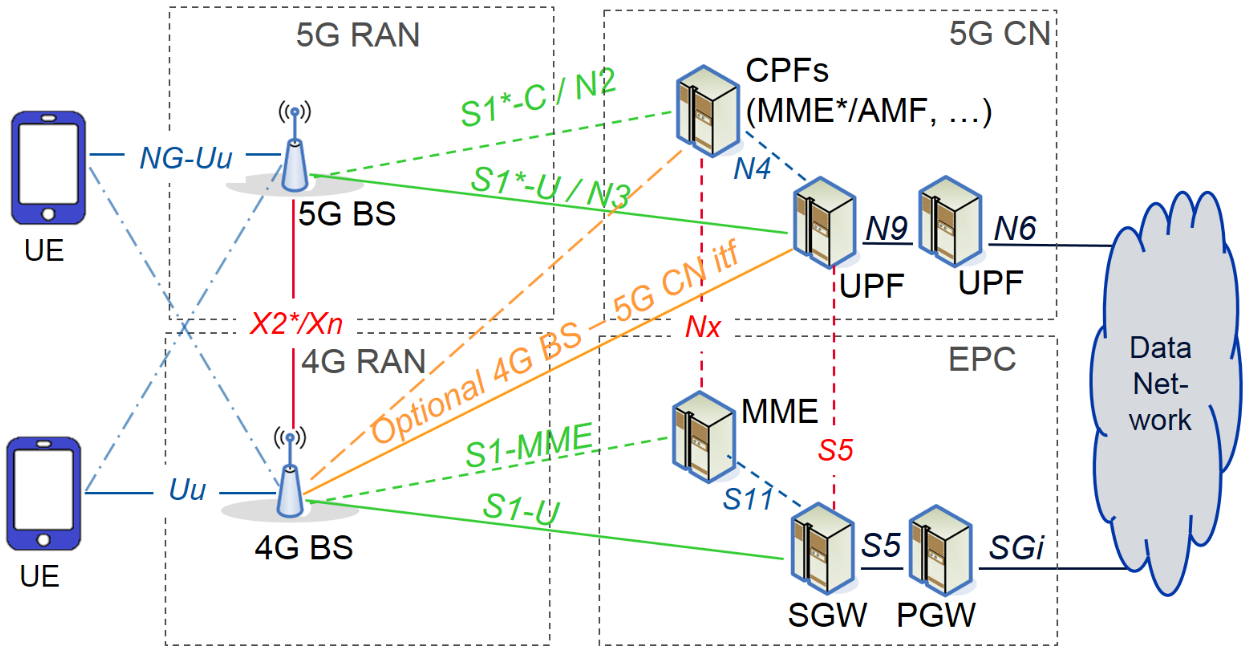


Figure 5-2: System Architecture for the interworking between 4G and 5G

Figure 5-2 shows the system architecture for the interworking between 4G and 5G. The interworking architecture enables users to move between the 4G and the 5G system. Further on, the system architecture must also enable the system to serve a user with 4G and 5G simultaneously e.g. via MC, which is also a key point in the following sections. As a further option as shown in orange colour, an evolved 4G BS/eNB, providing an S1* interface, can also be directly connected to the 5G CN. In such a case, the 4G CN, aka Evolved Packet Core (EPC), can be removed. A further option, which for simplicity is not shown in the figure, but which is especially of interest in 3GPP standardization for initial 5G deployments, would be to connect a 5G BS towards the EPC (via 4G BS for UP/CP or with separate S1-U for UP only). For more details on such options see [3GPP-23799] [3GPP-38801].

5.1 5G QoS Model

The 5G QoS architecture shall allow the detection and differentiation of sub-service flows in order to provide good quality of experience (QoE). The 4G bearer concept fails to cover this requirement. Therefore, a refined QoS model has to be used for 5G where the “QoS flow” is introduced. The QoS flow is the finest granularity for QoS forwarding treatment in the 5G System [3GPP-23501]. All traffic mapped to the same 5G QoS flow receives the same forwarding treatment. Providing different QoS forwarding treatment requires the use of different 5G QoS flows.

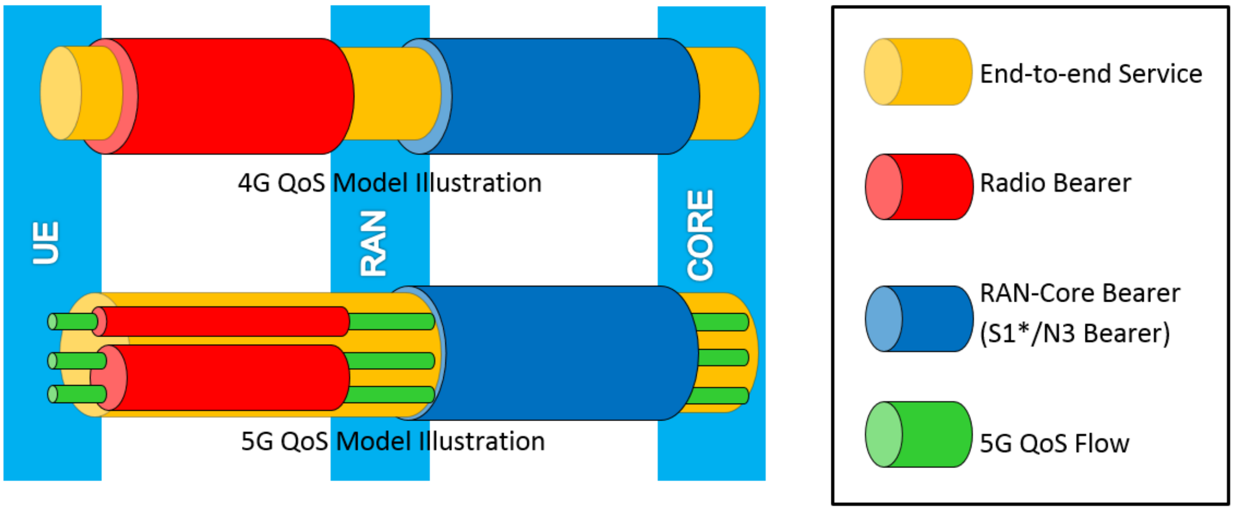


Figure 5-3: QoS model as used in 4G and 5G

An illustration and comparison of the 4G and 5G QoS model is given in Figure 5-3. It shows that the 5G concept allows a flexible mapping of the 5G QoS flows to radio bearers, e.g. the first 5G QoS flow is transported over the first 5G radio bearer while the second and third 5G QoS flows are transported together in the second 5G radio bearer.

For the support of the 5G QoS flows, either existing protocols (e.g. PDCP) need to be enhanced or a new protocol like the Service Data Adaptation Protocol (SDAP) proposed in [3GPP-38300] needs to be used.

Note that the QoS framework is covered in [MII-D62], the dynamic QoS for traffic steering is analysed in [MII-D52] and QoS from a UP aspect is covered in [MII-D42].

With one end-to-end service, a user can connect to one network slice. If the user wants to use multiple slices in parallel, the user must establish at least one end-to-end service to each slice.

5.2 CN/RAN Interface

Different UP protocols can be used on the S1*-U/N3 interface located between the 5G RAN and the 5G CN. Figure 5-4 shows the generic UP protocol stack for 5G where the generically illustrated S1*-U/N3 protocol stack is shown in blue colour. The end-to-end Protocol Data Unit PDU layer shall support different services types, e.g. IP, Ethernet or even unstructured data for non-IP data delivery (NIDD).

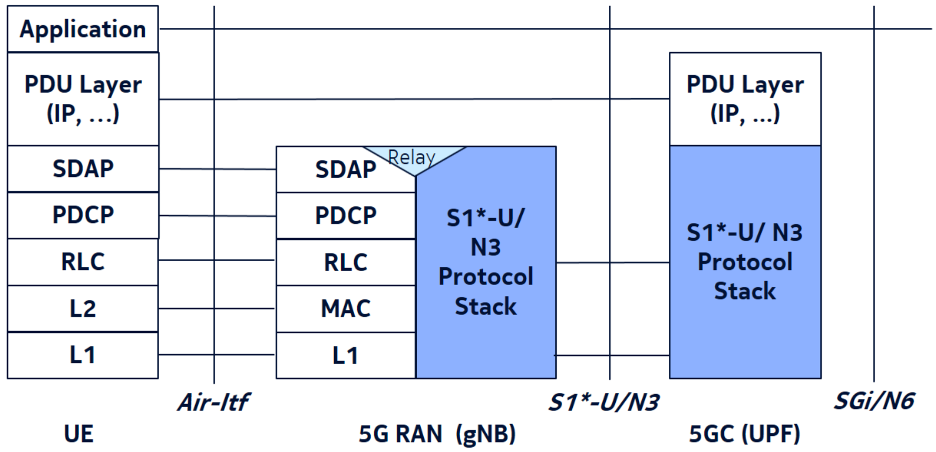


Figure 5-4: User plane protocol stack

This section presents 5 protocol options for this S1*-U/N3 protocol stack. The 5 options (GTP-U, GRE, EoGRE, L2 datagram switching, L3 packet forwarding) are shown in Figure 5-5. Please note that these protocol options can also be applied to other interfaces than the S1*U-interface, especially to the UP interfaces inside the 5G CN.

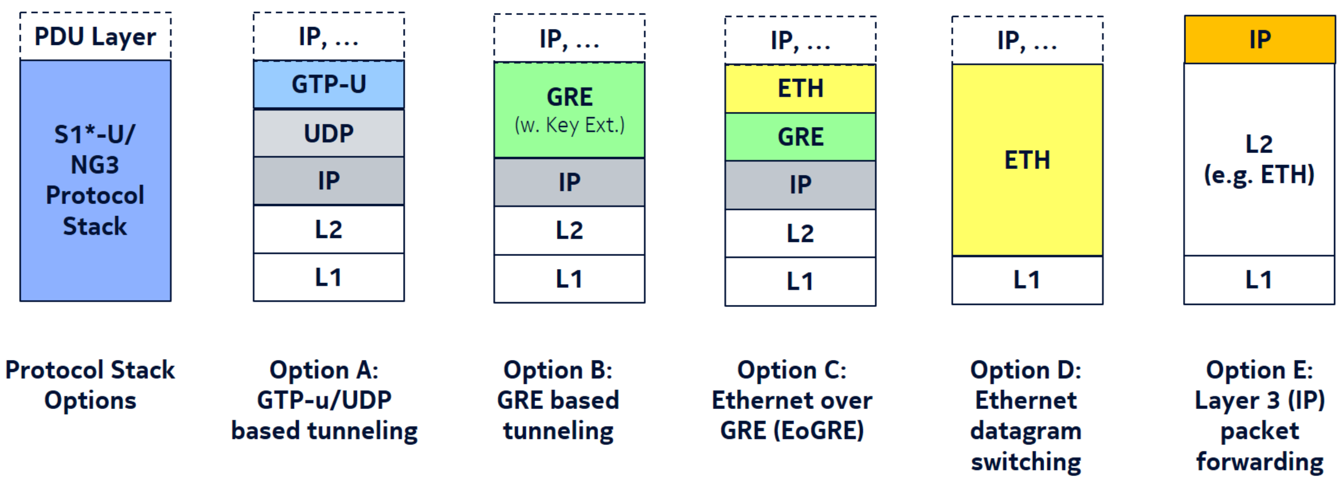


Figure 5-5: User plane protocol options for the S1*-U/N3 interface

The five options can be characterized as follows:

- **Option A "GTP-U"** uses the same protocol stack as in the S1-U interface used in 4G. The GPRS Tunnelling Protocol User Plane (GTP-U) [3GPP-29281] is transported over UDP/IP. The GTP tunnels must be setup each time when a UE enters the active mode or starts a session with new service requirements with a CP protocol like GTP-C. This may be inefficient especially when a UE transmits small amounts of data only sporadically which is the case e.g. for mMTC.

- **Option B "GRE"** uses the Generic Routing Encapsulation (GRE) protocol [IETF-RFC2784]. Using GRE with the "key and sequence numbering extension" [IETF-RFC2890] is similar to Option A.

Both Option A and B use one tunnel per user per service type and traffic direction. Because these tunnels transport only the traffic of a single bearer, they can be denoted as "thin pipes".

- **Option C "EoGRE"** aggregates the traffic of multiple UEs but with similar service characteristics in one GRE tunnel. These tunnels carrying the traffic of multiple users are called "fat pipes". The advantage of the fat pipes concept is a strong reduction of the required CP signalling [GZ16].
- **Option D "Ethernet datagram switching"** simplifies the previous option by using the Ethernet layer without any additional tunnels. As in the previous option, locally administered IEEE MAC addresses are used to identify the 3GPP network interface of a UE. For scalability reasons of the backhaul transport network, methods and protocols like Transport Interconnection of Lots of Links (TRILL) or Shortest Path Bridging (SPB) should be used.
- **Option E "IP packet forwarding"** uses IP forwarding techniques. Typically, the forwarding tables are updated with Software Defined Networking (SDN) methods, e.g. when a new UE attaches or in case of mobility.

While 3GPP currently has only specified Option A "GTP-U", multiple options can be used in 5G, especially in the context of network slicing. An xMBB slice for example may use Option A "GTP-U" as already defined in 3GPP while a slice with a large amount of mMTC traffic may use Option C "EoGRE" in order to reduce the signalling traffic. Option D "Ethernet datagram switching" is well suited for uMTC/URLLC traffic. The Options B, C, D and E are also well suited for access agnostic scenarios as they do not rely on 3GPP specific protocols.

The control plane shall be able to support all these options, especially the control plane shall provide information on which user plane option to use. Further information on CP procedures on the CN/RAN interface can be found in D6.2 [MII-D62].

5.3 Protocol Stack Architecture for the 5G AI

5.3.1 Protocol Functions for 5G

A key question in METIS-II is to which extent protocol functions of the new AIVs may have to be substantially modified to meet the 5G requirements. In this subsection, we will hence explore the different protocol stack layers, list their current functions as in LTE-A, and elaborate on any potential changes in 5G.

In LTE-A, PDCP is responsible for compression and decompression, transfer of UP and CP data, security (i.e. encryption), maintenance of sequence numbers etc. An overview on all PDCP functions and possible changes in 5G is provided in Table 5-1.

For the Radio Link Control (RLC) layer, the main function is Automatic Repeat Request (ARQ) and data segmentation/concatenation, based on which mode (acknowledged or unacknowledged mode) is configured. Possible changes to this layer for 5G are listed in Table 5-2,

The design of MAC in LTE-A has focused on low complexity while maintaining efficient and fast operation. It is envisioned that the current functions of the MAC layer will also be needed in 5G, but there is a need for a more elaborate design as described in Table 5-3.

Table 5-1: PDCP functionalities in LTE-A and possible changes in 5G.

Functionality in LTE-A (Release 13)	Considerations for novel AIVs in 5G
Maintenance of PDCP Sequence Numbers (SNs), duplicate detection/elimination and discarding, and timer-based discard.	No changes foreseen.
Routing and reordering of PDCP PDUs in the case of split bearers (RLC Acknowledged Mode, AM).	No changes foreseen. This functionality is seen as particularly important for the widespread usage of MC in 5G.
Reliability using MC	For URLLC (uMTC) services, packet duplication is supported for both UP and CP in PDCP for reliability purposes. Retransmission coordination with multiple RLC entities may be required. [MII-D42].
Data-recovery procedure for split bearers in Dual-Connectivity (DC) mode (for RLC AM), for instance needed when part of the data transmitted over one radio leg is lost due to bad radio conditions.	No changes foreseen, though in 5G the data-recovery procedure will need to be defined for both MC among LTE-A evolution and novel 5G radio, as well as among multiple novel AIVs.
Retransmission of PDCP Service Data Units (SDUs) at handover: The handover case is very similar to the use case for the data-recovery procedure.	No changes foreseen.
(De-)Ciphering and Robust Header Compression (ROHC)	Due to significant contribution to latency [MII-D42], these functionalities may need to be reviewed especially for URLLC services.

Table 5-2: RLC functionalities in LTE-A and possible changes in 5G.

RLC functionality in LTE-A	Considerations for novel AIVs in 5G
Transfer of upper layer PDUs	No change foreseen.
Error correction through ARQ (only for AM data transfer). By configuring AM RLC, ARQ is supported with an extra layer of retransmission reliability.	For novel 5G AIVs, the combination of ARQ and HARQ should be further studied. Since it may be possible to improve the reliability of MAC HARQ, the ARQ may in some use cases potentially be omitted. ARQ performed on any numerologies / TTI lengths that the logical channel is mapped to. Retransmission coordination with PDCP may be required between RLC entities in multi connectivity scenarios [MII-D42].
Concatenation, segmentation and reassembly of RLC SDUs (only for unacknowledged mode, UM, and AM data transfer), for the purpose of generating RLC PDUs of appropriate size from the incoming RLC SDUs.	Since concatenation and segmentation require the knowledge on the MAC transport block sizes, this RLC functionality is tightly tied to the MAC and hence has to happen on synchronous time scale. Therefore, 3GPP has agreed to move the concatenation to the MAC for NR already [3GPP-38300]. It is still under discussion if the segmentation will be moved to MAC too. RLC still keeps individual queues per RLC entity to avoid head-of-line blocking (i.e. packets in a queue are being held up by the first packet). This way, the remaining RLC functions would be asynchronous, and a function split between RLC and MAC would be a split between asynchronous and synchronous functions. It yet has to be clarified to which extent this would touch standardization, or be a matter of implementation.
Re-segmentation of RLC data PDUs (only for AM data transfer), in the case that these do not fit to the actual transport blocks.	In novel 5G AIVs, the usage of this function may be extended to new scenarios, for example, the usage of unlicensed spectrum, where the transmission may be blocked by channel acquisition. Then, the RLC PDU could be re-segmented to fit the next transmission.
Reordering of RLC data PDUs, duplicate detection and RLC SDU discard (only for UM and AM data transfer), **RLC re-establishment, and protocol error detection** (only for AM data transfer)	RLC does not support re-ordering in 5GNR. However, a T-reordering like functionality is supported to determine the content of the RLC status report. [3GPP-38322].

Table 5-3: MAC functionalities in LTE-A and possible changes in 5G.

MAC functionality in LTE-A	Considerations for novel AIVs in 5G
Error correction through **HARQ**.	Inclusion of HARQ modes for increased reliability. To be able to meet 5G requirements, the HARQ protocol may need to be faster, with lower overhead and operate on a flexible timing base. HARQ parameters

	may need to be configured differently for different services [MII-D42].
Initial Access using the Random Access Channel (RACH) for requesting UpLink (UL) resources.	Add new modes supporting prioritization of initial access to support high reliability services.
Transport format selection	Inclusion of and support for new formats for high data rates.
Priority handling between UEs by means of dynamic scheduling	- Introduce scheduling algorithms that provide increased multi-user gains making use of e.g. the beam-centric design - Differentiation according to the 5G service parameters.

5.3.2 Network entities / possible Function Splits and related Intra-RAN Interfaces

Figure 5-6 provides a high level view on the main changes in RAN architecture design for 5G in comparison to 4G considering a two-dimensional separation of RAN NFs.

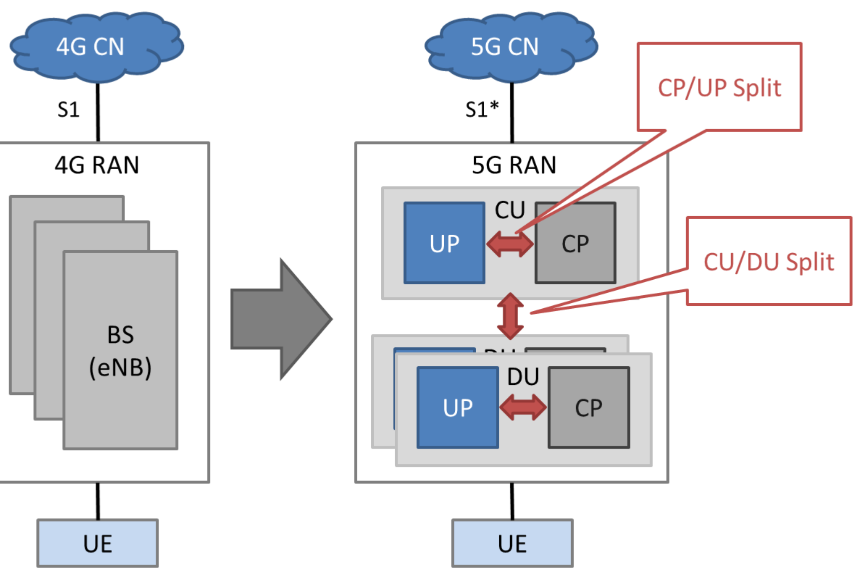


Figure 5-6: High level view on architectural evolution from 4G to 5G RAN considering two-dimensional split in control/user plane (CP/UP) and central/distributed units (CU/DUs)

The first separation step is the differentiation between CPFs and UPFs (aka vertical split), enabling the introduction of SDN principles also in the RAN [TGV+14] [RBB+16] [YHZ+16] [ABB+17] (see also Deliverable D4.2 [MII-D42] for more details). The anticipated benefits of a vertical split are:

- In multivendor networks, a standardized interface to the CP enables a consistent control over network entities and NFs from different vendors, e.g. in terms of interference management for Ultra-Dense Networks (UDNs) [MII-D52].
- Due to the tight coupling of CPFs and UPFs in today's networks, the replacement or upgrade of a CPF often requires also the replacement of UPFs. Avoiding this might offer significant cost savings.

Besides, there are also disadvantages:

- CPFs and UPFs are often tightly coupled, especially in the lower radio protocol stack layers. It might be challenging and could affect the performance, especially if the processing is not collocated (see also Annex 0).
- Standardization is required in case the interfaces between CP and UP have to be extended to introduce new features which might slow down this process. Integrating additional interfaces in a proprietary manner in combination with standardized ones is not a suitable solution, as it would destroy the benefits of a CP/UP split. For example, a flexible change of CPFs in logical network entities would not be possible any more if only selected UPFs support certain proprietary interfaces.
- Additional effort in terms of testing is required to ensure the interoperability of CPFs and UPFs from different vendors (shifting the effort to system integrators supporting the operators instead of doing this work by a single supplier).

The second separation step is related to a horizontal split in the radio protocol stack allowing to concentrate some typically higher layer processing functions in a physical entity called CU, whereas lower radio layer NFs will be placed together with the radio units (RUs) at several so-called DUs near the antenna sites. The main intention of the horizontal split is to enable gains from centralization of RAN NFs in a CU, e.g. through common RM and flow control as anticipated in Cloud-based RAN networks (C-RAN), but it also allows NFs to be placed in CU and DUs according to performance criteria like latency as well as to adapt the placement to the characteristics of the x-haul transport network between CU and DUs [5GC] [5GX]. Centralization of lower layer NFs generally increases the x-haul requirements in terms of bandwidth and latency as known from today's CPRI PHY layer interface implementation [CPRI15]. With 5G, those requirements may be further tightened because of e.g. shortened TTIs, wider frequency bands and strongly increased number of antenna ports with Full Dimension (FD) or Massive MIMO (especially for frequency bands above 6 GHz). In Annex 0, a detailed presentation of the radio protocol stack (both UPFs and CPFs) is given, which also includes different horizontal split options M1 – M8[2] introduced in METIS-II Deliverable D2.2 [MII-D22] and relevant interfaces required for the vertical split. For more information about the impact of horizontal split on the UP (e.g., data rate to be transferred via the corresponding interfaces), please refer to D4.2 [MII-D42].

[2] Please note there is no one-to-one mapping between the split options defined by METIS-II in [MII-D22] and those of 3GPP defined within the study on 5G New Radio (NR) [3GPP-38801].

To fulfil the diverging 5G service requirements via flexible and adaptable RAN NFs there are many feasible combinations of vertical and horizontal splits corresponding to many options for mapping CPFs and UPFs to CU and DU. From a practical perspective however, only a limited number makes sense. Otherwise, it would be hard to handle such a large number of intra-RAN interfaces between logical and physical entities with varying capabilities, taking into account both the operational as well as the standardization effort required.

The most important deployment options considering vertical and horizontal splits and the mapping of related RAN NFs to physical network entities will be discussed in the following (see also [MII-D62]). To allow a simplified presentation, CPFs and UPFs have been structured into three parts with respect to their position in the radio protocol stack [MII-D22]. The meaning of the different layers is given in Table 5-4.

Table 5-4: Mapping of NFs to different CP/UP layers

NF layer (CP/UP)	NFs related to CP/UP layer
CP-H: High layer CPFs	High-level inter-site/AIV resource coordination like ICIC, AIV agnostic Slice Enabler (AaSE) [MII-D52]
CP-M: Medium layer CPFs	User and network specific NFs (e.g. RRC, RAN mobility, admission control)
CP-H: Low layer CPFs	Cell configuration, short-term scheduling, PHY layer control
UP-H: High layer UPFs	QoS/Slice enforcement, PDCP[3]
UP-M: Medium layer UPFs	RLC (asynchronous/synchronous), MAC, Higher PHY
UP-H: Low layer UPFs	Lower PHY

Figure 5-7 denotes the two extreme cases. The first one corresponds to a standard deployment used in 4G systems assuming a flat hierarchy of network entities (here noted as BS) characterized by fully decentralized NFs (aka Distributed RAN (D-RAN)). The second deployment scenario represents the full centralization of all CPFs and UPFs for a certain number of RUs at a CU. The interface between the CU and the DUs related to split option M1 will carry digital baseband data in time domain for each antenna port plus additional control and management information according to CPRI or ORI [ETSI14-ORI] specifications. Hybrid BF approaches as intended especially for Massive MIMO usage in mmW bands require adjusted phase values for the analog RF precoding stage in the RU (see Figure A-1 in the Annex), which necessitates extensions of the interface specifications. Due to tight latency and high bandwidth requirements for M1 such deployment is in contrast to the D-RAN approach only feasible with ideal fronthaul, i.e., via fiber access.

[3] UP-H may also contain asynchronous RLC functions (horizontal split M7; see Figure A-1 in the Annex), so only synchronous RLC functions will remain in UP-M.

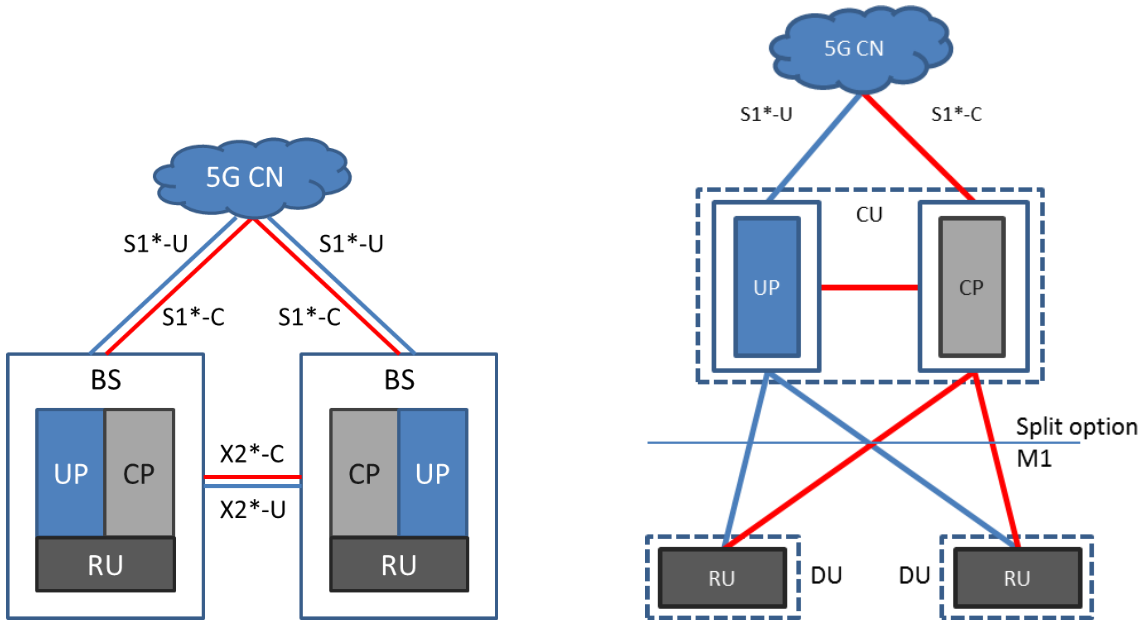


Figure 5-7: Fully decentralized deployment with full RAN NF functionality in each BS (left) vs. fully centralized CP/UP in the CU (right)

In Figure 5-8 two scenarios with partially distributed NFs are shown. For the scenario on the left side the UP is partially centralized, whereas the CP is fully centralized. Two different split options are suited for this scenario. The M2 split requires digital baseband data per antenna port to be carried by the interface. However, the difference to M1 is that the data is in frequency domain which is less bandwidth demanding. Split M3 carries user data after performing Forward Error Correction (FEC) coding before following steps of scrambling, modulation and layer mapping/precoding resulting in further reduction of x-haul bandwidth requirements (see [MII-D42] for more details. A further difference between the split option M3 compared to M2 (and M1) is the fact, that it also involves additional CP/UP interfaces according to Figure A-1 which have to be carried via the x-haul link which again might pose additional requirements, especially in terms of latency.

In the scenario on the right side of Figure 5-8 also the CPFs are distributed between CU and DUs. In that case, synchronous CP/UP NFs are typically deployed at the DUs and the asynchronous ones at the CU. It is worth noting METIS-II also envisions fast-scale operation of various traditionally slow functions, which may also be deployed at the CU, see Section 6.2.1. With respect to horizontal split, options M7 and M8 as shown in Figure A-1 in the Annex fit that approach. With split option M8 the whole RLC NFs are placed at the DUs, with M7 only the synchronous RLC part (asynchronous RLC at the CU). Regarding the CP all asynchronous CPFs stay in the CU, only short-term scheduling (CP-L) will be placed at DUs. The advantage of this deployment is that all CP/UP interfaces with strict timing requirements can be handled DU-internally, which also relaxes the requirements on the x-haul interface. The M8-based approach is especially interesting with respect to MC combining 5G AIVs like NR with LTE-A Pro as it is already applied for the LTE DC feature [3GPP-36300] (see also Section 5.3.4 and Section 6.2.1). As bandwidth and latency requirements for the x-haul in case of higher layer split options M7 and M8 are weakened compared to the lower layer options, deployments based on them are also feasible with non-ideal x-haul (i.e., inclusive of wireless x-haul links).

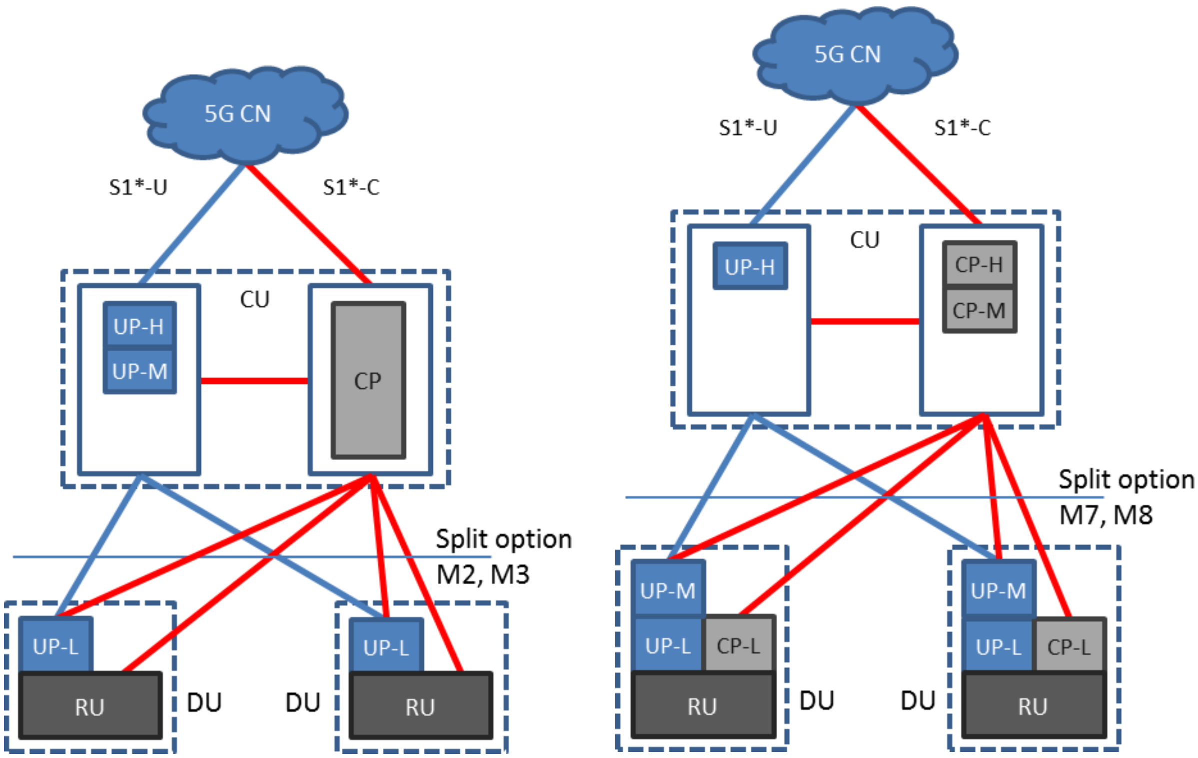


Figure 5-8: Deployments scenarios with different partially distributed splits

For scalability reasons it normally does not make sense to implement a country-wide RAN via a single CU, but to implement several CUs each controlling the radio processing for a certain number of antenna sites (domain) [MII-D42] [ABB+17]. Typically, the NFs running in the CU are implemented as virtual functions (VNFs) on server platforms based on network function virtualization (NFV) principles [ETSI-NFV]. Suitable locations for CUs are e.g. the central offices of fixed or integrated network operators [5GPP16]. To support especially low latency applications, Mobile Edge Computing (MEC) facilities [RBB16] can be also integrated into the CU. The CU approach has a big advantage with respect to mobility handling. If a UE is moving within the range of antenna sites belonging to a single CU, mobility is handled CU-internally only. This can happen through fast UP switching [MII-D52] resulting in low handover interruption time (ideally zero) because of low latency between involved components. In that case, no signalling traffic is required between RAN and CN. This is beneficial especially for UDN deployments (using e.g. mmW bands) with a high number of mobility events.

5.3.3 Overview on CP design

METIS-II assumes a CN/RAN split, which enables an independent evolution of CN and RAN functionalities and allows multi-vendor deployments. It is further assumed that there will be a common CN and CN/RAN interface (denoted S1*) for both the new AIVs and the evolution of LTE-A, see Figure 5-2. This enables a tighter interworking between the new AIVs and LTE-A evolution (see Section 5.3.4), improves the mobility, robustness, and resource usage and minimizes the signalling to the CN. Similar enhancements are also envisioned for the evolution of the X2 interface (denoted X2* herein), which jointly with S1* become interfaces addressing multiple AIVs. METIS-II also proposes a specific-protocol architecture for the 5G RAN, illustrated in Figure 5-9 where two AIVs are exemplarily illustrated. Therein, AIV-overarching mechanisms are located at the Access Network – Outer (AN-O) layer while AIV-specific mechanisms are located at the Access Network – Inner (AN-I) layer. It is worth noting that in this implementation AN-O corresponds to a CU and AN-I corresponds to a DU. Considering the goal of tight interworking between 5G and legacy AIVs, the functional split option is preferred to be at PDCP level not to influence the 5G specification with legacy AIV constraints. This protocol architecture implementation takes into account both synchronous CPFs [MII-D52] and asynchronous CPFs [MII-D62]. The key elements of the common CP and the associated concepts can be outlined as:

- **RAN Moderation** determines the optimum number of active access nodes such so that network energy efficiency can be substantially improved while target service requirements can be fulfilled (see Section 6.1.5 and [MII-D52]),
- **AIV agnostic Slice Enabler (AaSE)** enables performance guaranteeing multi-slice RM with real-time SLA monitoring (see Section 6.2.2 and [MII-D52]),
- **Multi-AIV Coordination** enables AIV adaptation to the semi-dynamically changing network conditions, e.g., during stadium events, AIV tailored for mMTC devices can have extended bandwidth (see [MII-D52]),
- **Multi-AIV Resource Mapping** capitalizes on the interfaces with the aforementioned concepts and proactive link establishment strategies, and provides the interface to AN-I to enable fast routing of data flows to the appropriate AIV(s) comprising both novel 5G AIVs and legacy AIVs (see Section 6.2.1 and [MII-D52]),
- **Interference Management (IM)** deals with different types of interference caused in 5G networks, both conventional interference already existing in legacy systems (such as cell-edge interference or D2D communications) as well as novel interference patterns caused by new network features such as a dynamic topology with NNs (see Section 6.2.3 and [MII-D52]),
- **Real-time Resource Mapping** is a collection of mechanisms that deal with the following functionalities: i) flexible multi-service scheduling where different parameters related to the communication using a certain AIV can be adjusted in real time, ii) resource allocation and mapping for D2D group communications, and iii) UL power control such that the users

transmit in a cooperative way without causing much interference to each other (see [MII-D52]),

- **AIV-specific Context Management** handles all context parameters that are bound to one specific AIV, where both measurements and configurations aspects are contained (see Section 6.2.4 and [MII-D52]).
- **Initial Access:** handles the initial access procedures related to Random Access taking into account the requirements of each service (see Section 6.3 and [MII-D62]).
- **Cell Config:** handles the transmission of the System Information Blocks (SIBs) and Master Information Blocks (MIBs) as well as the Cell-specific Reference Signals (CRSs) and PDCCH resources for every cell (see Section 6.1.4)
- **RRC:** includes the RRC state machine handling and the mobility management functions that should be moved to the RAN to optimize Tracking Area Updates (TAU). Additionally, the way that the UE is configured to perform the measurements for the various AIVs (see Section 6.3 and [MII-D62]).

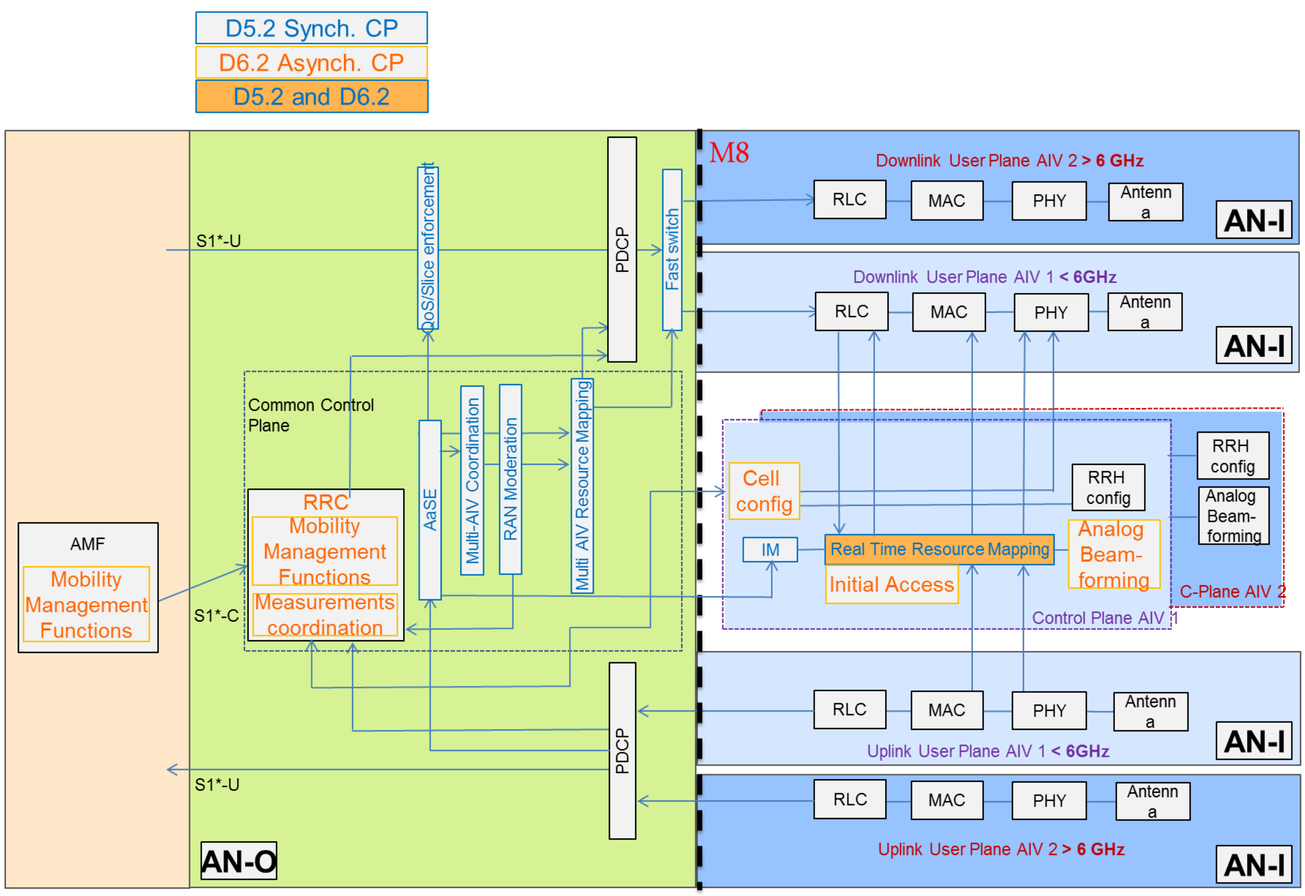


Figure 5-9: Protocol Architecture of METIS-II Control Plane.

5.3.4 CP / UP Architecture for the Interworking of AIVs

In the development of 5G at 3GPP the DC concept from LTE Rel. 12 is being used as a basis for a tighter integration between LTE evolved and 5G AI. It will enable the UE to be connected to LTE and 5G AI (UP and CP) at the same time. DC can increase the UE throughput due to UP aggregation (receiving data from both AIVs at the same time) and make the connection more reliable. The increased reliability also comes from the case when the UE needs to switch (handover) to another Secondary eNB (SeNB), since in this case the UE can still be connected to the Master eNB (MeNB) and reliably receive RRC signalling from the MeNB. The most typical architecture for the LTE-DC is the so-called bearer split "3C" options from LTE DC. In LTE DC MeNB is responsible for splitting (or aggregating in UL) the user plane data over the links. i.e. split /aggregate the data at the PDCP layer (see Figure 5-8 for the M8 split). This split option was assumed at an early stage in METIS-II [MII-D51], [MII-D61] and is also agreed on to use for LTE evolved and 5G AI tight integration in 3GPP [3GPP-38804].

The data is sent from a MeNB to the SeNB via the X2 interface. For LTE DC, only the MeNB CP (RRC) is connected to the CN via the MME. This solution was also adopted at an early stage for the LTE evolved and 5G AI [MII-D51], [MII-D61] and is also the current assumption in 3GPP for 5G, i.e., a common evolved CN/RAN interface for both LTE and 5G will be used [3GPP-38804]. This implies that no extra CN/RAN signalling is needed to add or remove a secondary node. For LTE DC all RRC messages are transmitted via the MeNB. SeNB RRC messages are sent to the MeNB over the X2 interface, and the MeNB makes the final decision of whether to transmit the RRC message to the UE. This has the advantage that there is no need for coordination, since the MeNB always makes the final decision. The disadvantage is that there is no RRC diversity and RRC messages from the SeNB take longer time since they are always routed via the MeNB. Even though it is hard to predict how the RRC for the LTE-NR tight integration will be standardized by 3GPP, it is likely that some disadvantages of the LTE DC will be addressed. Meanwhile 3GPP has agreed for NR that there may be duplication of RRC packets and that the SeNB can send some RRC messages directly to UE.

5.4 Architectural Enablers for Network Slicing

It can be foreseen that a limited number of different RAN configurations will be sufficient to serve the UCs described in Figure 2-1. This can be justified by the UC requirements and the respective grouping, as well as the more restricted ability of virtualization in the RAN. Each combination may be called RAN Configuration Mode (RCM) which is a composition of RAN NFs, specific function settings and associated resources (HW/SW, and network resources). An RCM can be statically defined or fully flexible, and this is up to the implementation and the requirements for flexibility and future-proofness (i.e., in case a totally new UC arises with new unforeseen requirements).

The generic considerations for the RCMs have been presented in details in D6.2 [MII-D62] and are exemplarily captured in Figure 5-10 w.r.t. the 3 main 5G service types. In brief, it can be foreseen that:

- The different RCMs share an RRM function for ensuring the sharing of the common radio resources; also, this function can facilitate that, in the case of the RCMs sharing the lower layer functions the slice isolation can be guaranteed at least using QoS classes. However, each slice anyway can apply its own RRM strategies according to the slice specific characteristics.
- At least a common RRC part for all slices will be present, as it is seen there is a shared part, which enables the slice selection. Each slice can have its own RRC functions and configurations as well so as to tackle the special UC requirements when it comes to particular functions (e.g., discontinuous reception/transmission (DRX/DTX), measurements reporting, TAU periodicity, cell selection strategies, etc.) when particular optimizations can be achieved. One alternative implementation of the common part of the RRC could be a common slice which will provide information for the slice selection
- For PDCP and the RLC, depending on the message size, or the delay requirements certain functions can be either omitted (e.g., header compression, ciphering) or modified (e.g., segmentation, re-ordering, ciphering).
- The RCMs that share the lower layers (PHY, MAC, etc.) should have a joint “Unified Scheduler” for enabling them to share the resources more dynamically.

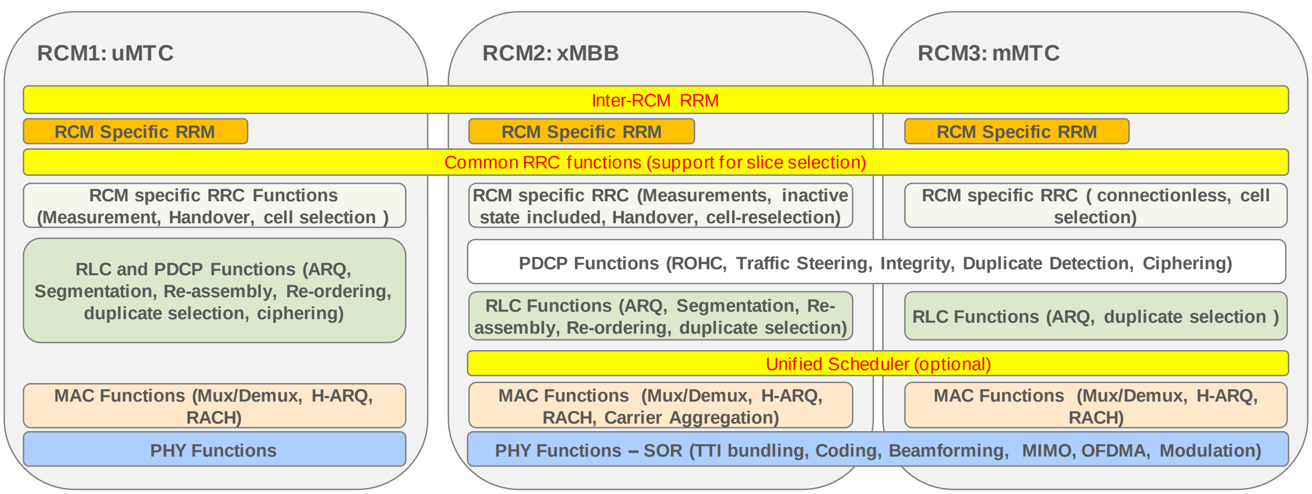


Figure 5-10: Example of RCMs with shared and independent functions

5.5 Flexible Spectrum Management within the Network Management & Orchestration Framework

5.5.1 An introduction to Network Management & Orchestration in 5G[4]

The 5G RAN design of METIS-II is intended to fulfil NGMN's vision for the overall 5G architecture [NGMN15]. This native Software Defined Radio (SDR)/SDN/NFV-based architecture is set up on different layers covering aspects ranging from the devices and the physical infrastructure, NFs to be implemented on it and supporting the network slicing concept, value-enabling capabilities etc., up to all the management functions needed to manage and orchestrate the 5G system (E2E MANO) [ETSI14-NFV-MAN] [3GPP-28801] (see Figure 5-11 for a high level overview). APIs between the different layers are used to flexibly support 5G use cases and business models (Anything as a Service, aka XaaS). This approach is also generally considered in the architectural description provided by the 5G PPP Working Group (WG) "Architecture" [5GPPP16]. Design principles developed by METIS-II on 5G RAN are incorporated into that architecture (see also [MII-D22]).

[4] The MANO framework is not a main research topic in METIS-II, therefore only issues with relevance to RAN design are noted in the following.

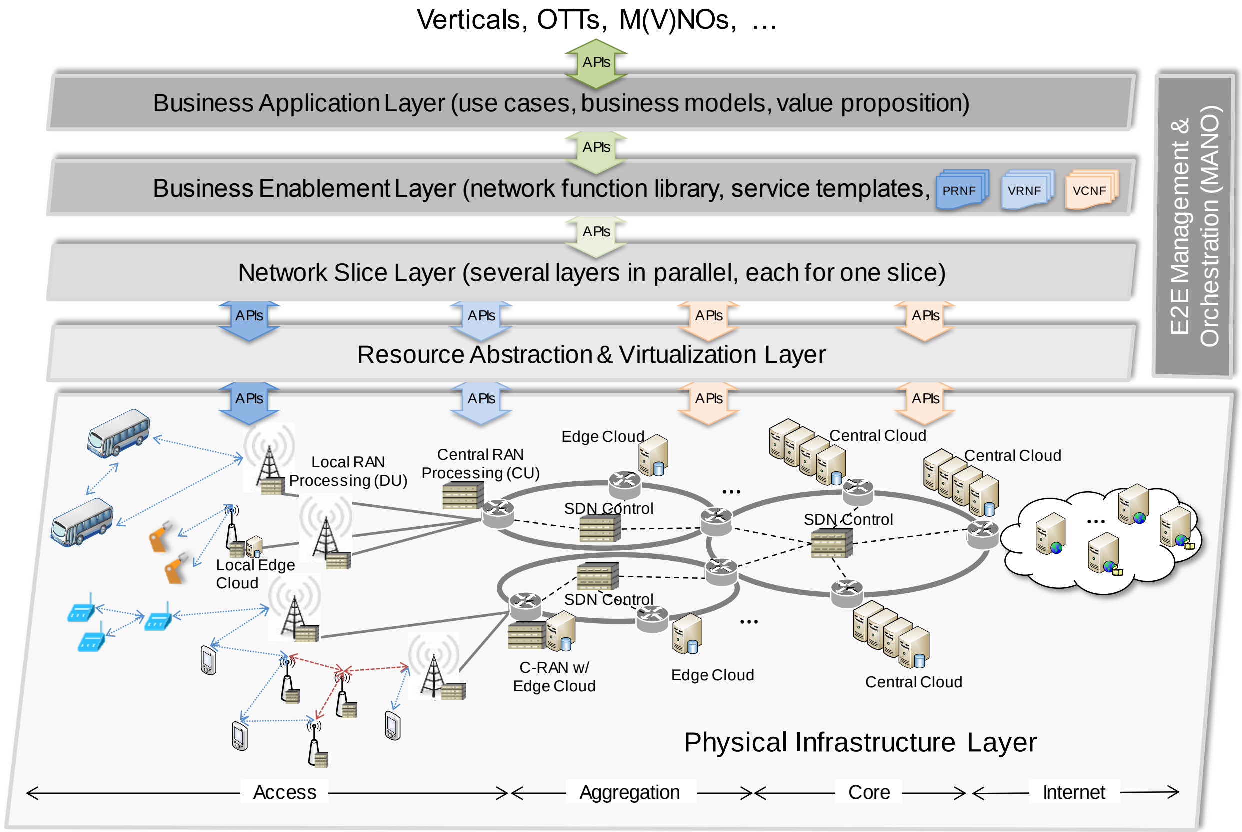


Figure 5-11: Layered 5G system architecture.

The E2E MANO is responsible for the translation of 5G use cases into concrete services and network slices. Depending on defined SLAs, it determines for each slice instance and corresponding service flows, respectively, all relevant NFs[5], AIVs, and parameter configurations, and realizes the geographical mapping onto the available physical network infrastructure consisting of all HW and SW parts of access, transport and core network nodes inclusive of computing and storage resources. The MANO framework takes care of the infrastructure resource sharing among multiple slices (inter-slice coordination) and it provides efficient lifecycle management mechanisms for slice instances (i.e., deployment, operation, monitoring, and termination), both within single and across multiple domains with different administrative owners (operators, infrastructure providers, etc.). It further manages scaling of the capacity of individual NFs and their geographic distribution, as well as Operations/Business Support Systems (OSS/BSS), Domain/Element Management (DM/EM), and SON (Self-Organizing Networks) procedures. The MANO framework will also cover relevant aspects inside the operator domains

[5] Except of the classification into CPFs and UPFs, there exist also a differentiation between Physical NFs (PNFs), tightly coupled with the underlying HW, and Virtual NFs (VNFs) which may run on General Purpose Processors (GPPs) used e.g. in cloud servers. CN-related NFs are usually implemented as VNFs (VCNFs), NFs in the RAN may happen in both variants (PRNFs and VRNFs, respectively).

to realize the innovative spectrum management concept as introduced in Section 4.1.3. The architectural approach needed for that will be described in the following.

5.5.2 Spectrum management architecture

In order to enable the spectrum management concept, a number of technical requirements have to be fulfilled which can be differentiated between requirements inside and outside the MNO domain.

Regulator domain

Requirements outside the MNO domain are basically in the regulator domain, in particular, a "Spectrum Management System" (SMS) entity that would perform the respective spectrum resource request and protection evaluations and decisions based on regulatory terms and rules is needed. The SMS architecture for METIS-II is based on the LSA architecture reference model defined in [ETSI15-103235] which is extended to support several additional sharing methods (like limited spectrum pool, etc.), and to manage spectrum resource user authorization more flexibly to support the limited spectrum pool and mutual renting options. The extensions would allow to introduce further spectrum sharing/usage methods by adding new spectrum resource databases and setting respective spectrum resource protection rules. More information on the SMS is available in [MII-D32].

Operator domain

Main challenges of spectrum management within the MNO domain of a future 5G network are to integrate numerous frequency bands within a wide spectrum range with possibly differing spectrum access schemes, and to cope with the versatile spectrum requirements from different user groups. These challenges are proposed to be addressed by holistic spectrum management architecture, comprising a central "Spectrum Assignment Coordination" (SAC) entity which takes the final assignment decision. The SAC is supported by a "Service-specific Spectrum Requirements" entity and a "Spectrum Resource Storage" entity for providing information on service specific requirements and spectrum availability, a "Spectrum Usage Rules" entity encompassing "Spectrum Access Modes" and "Network Deployment Scenarios" based on operator spectrum policy, and a spectrum usage tools "Spectrum Sharing Enablers" entity and a "Inter-operator Coordination Functions" entity. Interfaces between the SAC and the "Regulatory Spectrum Coordination" (RSC) entity in the regulator domain (Operator-Regulator Interface), and between SAC of different operators (Inter-Operator Interface), are required in order to facilitate cross-operator operation and data exchange on spectrum requests and assignments as well as regulatory requirements. The SAC is further connected with the operator's RRM. A graphical illustration of this holistic spectrum management architecture is shown in Figure 5-12

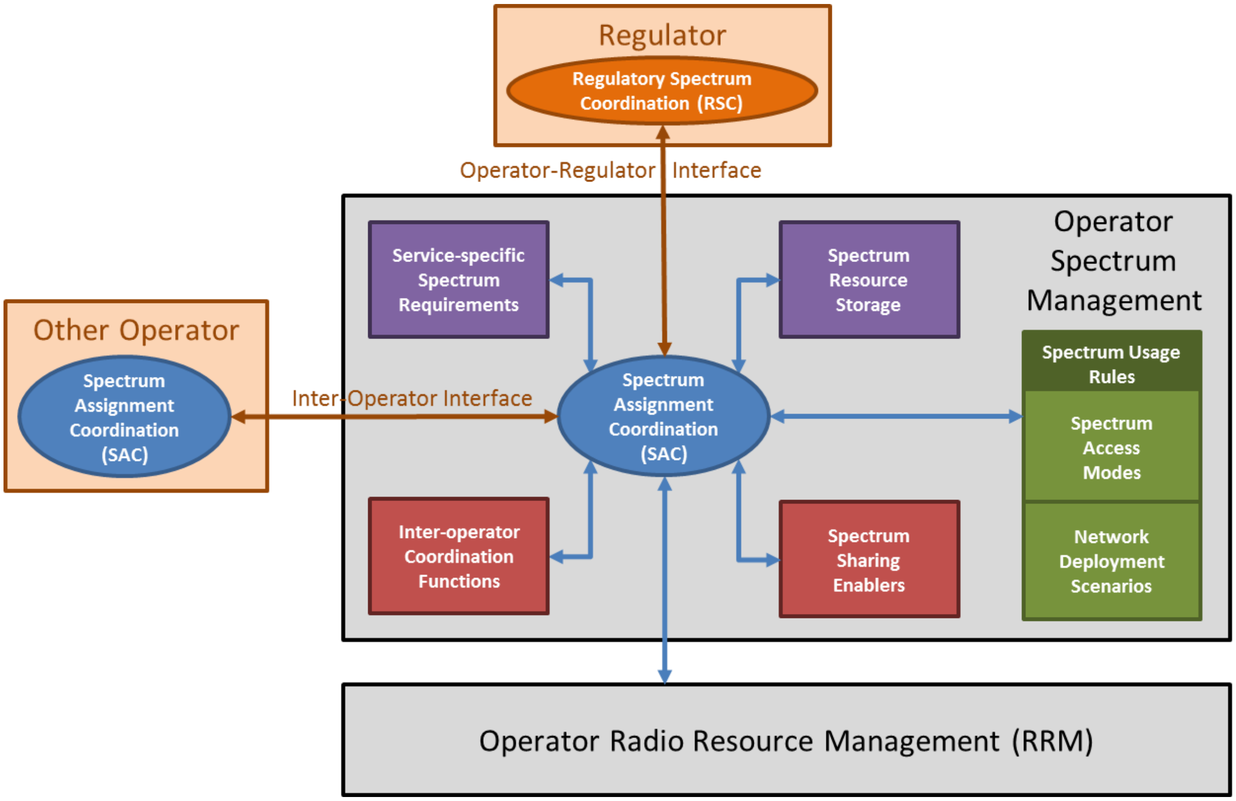


Figure 5-12: Holistic Spectrum Management architecture: Operator Spectrum Management comprising a central "Spectrum Assignment Coordination" (SAC) entity with interfaces to internal and external functional entities.

The different functional entities within the "Operator Spectrum Management" as well as interfaces to external and internal functional entities are described in [MII-D32]. Furthermore, the processing of relevant information in the central SAC in interaction with the functional entities within the "Operator Spectrum Management" as well as the interworking between the SAC and the functional entities outside the "Operator Spectrum Management" are illustrated for dedicated use cases and applications.

5.5.3 Implementation options for the SAC into the MANO framework

In the following, options for implementing the functional spectrum management architecture are briefly considered, by focusing on the implementation of the SAC entity, as the other spectrum management entities may be either connected directly to the SAC or already part of the OSS. More details can be found in [MII-D32]. It has to be noted that current specifications of 3GPP are more related to "traditional" Operation, Administration, & Maintenance (OAM) approaches, whereas with the introduction of 5G there will be a change to an intensified virtualization environment as described in Section 5.5.1.

Implementation of the SAC into the 3GPP OAM system

In this implementation option, the SAC may interact with the OAM system in the MNO domain at the Network Management (NM) level. The main advantage of the SAC being implemented at the NM level is that it can be connected to NM level applications (including the existing network planning and administration tools). Thus, the activities to use spectrum resources are part of the existing processes of managing the RAN.

The functionalities fulfilled by the SAC and the OAM are as follows:

- The SAC determines and provides constraints on cell parameters (e.g., maximum transmission power) upon receiving information on spectrum resource availability from the RSC or another SAC, and upon receiving a notification from the NM describing a change in network deployment.
- The OAM utilizes the constraints on cell parameters received from the SAC in its normal operation (e.g., SON and Configuration Management (CM) functions) and provides the information for utilizing the spectrum resources (e.g., applied transmission power) to the SAC.

Implementation of the SAC into a SON architecture

SON technology enables the autonomic organization of network elements and functions, respectively, as well as optimization of network performance by supporting the implementation of complex solutions in a flexible manner. In a SON environment, the SAC entity may be considered as a "Self-Configuration Spectrum Assignment Function" (SC_SACF). The respective functional architecture, i.e., the implementation of the SAC into the 3GPP SON concept according to the Self-Configuration Reference Model [3GPP-32501], is presented in B.1.

Implementation of the SAC into virtualized networks

In a mixed 3GPP and NFV-MANO architectural network framework [3GPP-32842], both VNFs and PNFs are managed. Assuming the SAC being implemented at the NM level or even being part of the NM, no specific modification of the standardized interactions between the NM and other entities are considered to be required. The constraints on cell parameters received by the NM from the SAC may be further processed through CM capabilities. More details can be found in B.2.

5.5.4 Intra-operator spectrum management with "Open-SON"

In [MII-D22], the concept of an AIV reconfiguration management functionality to manage and control the reconfiguration of the nodes inside the NW was introduced. In the context of the "Open-SON" architecture and the Control-Management Plane (CMP) described in [MII-D62], the AIV reconfiguration management functionality could be introduced as a CMP entity supported by a specific communication protocol (CMP protocol) being able to interact with the AIV specific CP and user UP, by decoupling the logical SON framework from the AIVs and from the respective network architectures. This entity would be in charge of the radio reconfiguration (e.g. spectrum) management, with the goal of self-adapting towards an optimal mix of supported AIVs and frequency bands [MII-D22]. An example of such a CMP functional architecture is given in Figure 5-13 where the supervisor functionalities of a Self-Configuration Supervisor (SCS) and a Reconfiguration Management Functionality (RMF) are in charge to manage and control the spectrum reconfiguration. The communication between the supervisor functionalities and the autonomic functions/agents is achieved through a CMP protocol that can use either the UP (yellow lines in Figure 5-13) or the CP (green lines in Figure 5-13) of the different supported AIVs. In relation to the configuration actions, the CMP protocol foresees to manage the peer-to-peer relations between the SCS and the RMF.

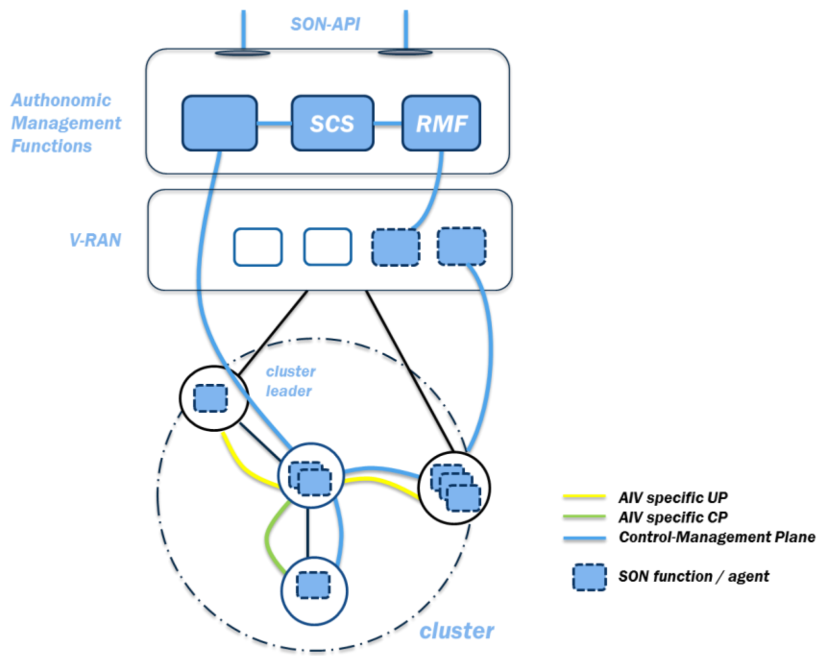


Figure 5-13: Control-Management Plane functional architecture

In relation to current 3GPP activities on 5G, this generic concept of spectrum reconfiguration can be applied in the context of co-existence and interworking between NR and legacy AIVs in an intra-operator domain (see Annex B.3). In such a context, the MANO framework presented in [MII-D22] may find an application in the NR/LTE co-existence management in which the reconfiguration of radio resources (e.g. change of channel bandwidth, activation of a novel AIV in a different frequency band, etc.) of NR and/or LTE could be performed by the AIV RMF on a slower time scale (e.g. on the order of hours), while the allocation of the specific channel resources (i.e. Physical Resource Blocks (PRBs)) to NR and/or LTE could be handled by the agile RM framework presented in section 5.3.4 and [MII-D22]. The presence of an orchestrator entity managing the interaction between these two levels of RM would add more efficiency as well as flexibility to the overall network control and management.

6 Functional Design Considerations

A summary of various functional design considerations of METIS-II is provided in this chapter. General design considerations have a strong implication on the overall 5G RAN design and are listed in Section 6.1. Section 6.2 presents more detailed considerations on the functional design for traffic steering and RM in 5G, and Section 6.3 provides details on design considerations related to initial access and mobility. The key paradigm changes imposed by the functional design recommendations are summarized at the end of this chapter.

6.1 General Considerations

6.1.1 Overall CP functions

As mentioned in Section 3, 5G is expected to fulfil a wide variety of requirements, be able to operate in a wide range of frequencies and consists of several slightly different AIVs (including evolved LTE). The aim with the CP functions/procedures is to be as **common** as possible and support all different AIV variants as well as different frequencies (including the mmW bands). There may be some necessary exceptions, e.g. see Section 5.4 about network slicing and RAN configuration modes.

Figure 6-1 depicts some of the more vital CPFs, treated in more detail in Section 6.3:

1. **Idle mode**: When the UE is not in the active state, the CP must support the cell selection/reselection functions/procedures.
2. **Initial access**: In order for the UE to be able to connect to the system and enter active state (via UE **state handling** functions/ procedures), it must listen to the System Information (SI) as well as the **paging** channel (if connection is network initiated).
3. When the UE enters active state (transmitting data), the CP must support security, **mobility** and radio bearer establishment. A key functionality of NR is the MC ability (for higher reliability) and to handle advanced **BF techniques**, including BF mobility (see Section 6.1.3).
4. In addition to this, 5G AI will support a **tight integration** with the (evolved) LTE.
5. **D2D** will be an integral part of 5G and natively supported in the protocol stacks of 5G, see Section 6.1.6).

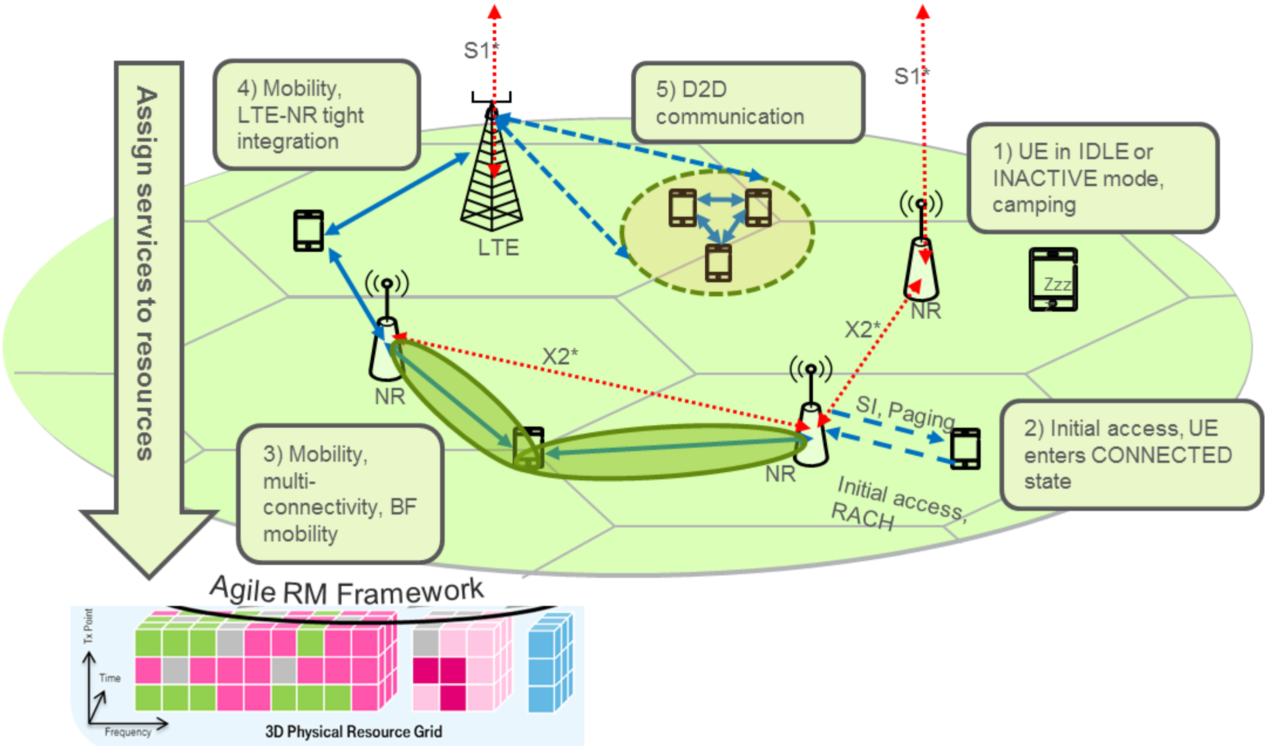


Figure 6-1: Overview of the RAN common control plane functions. Boxes 1-5 shows the asynchronous CPFs treated in this report. The Agile Resource Management is part of the synchronous CPFs treated in [MII-D52].

Figure 6-1 also shows a box of the Agile RM, one the so-called synchronous (fast) CPFs treated in [MII-D52]. A high-level conceptual illustration for the agile RM framework is shown in Figure 6-2. The framework operates over the 5G landscape, consisting of different and novel deployment options, novel communication modes, and new duplexing schemes (e.g., dynamic Time Division Duplex (TDD) in UDN). Accordingly, the agile RM framework aims to dynamically and efficiently assign services to the most suitable resources capitalizing on the available context information obtained through different AIVs. Given the latency-critical services to be enabled by 5G networks, the efficiency of RM mechanisms shall be clearly improved to be agile enough to react sufficiently to service needs. In particular, the framework of agile RM comprises

- synchronous CPFs that ensure the fulfilment of service requirements,
- paradigm changes for efficient operation and improved performance of typical *synchronous* CPFs (e.g., IM mechanisms to adapt to new dynamic radio topologies), and for fast operation of typical *asynchronous* CPFs (e.g., dynamic traffic steering applied on a synchronous level rather than legacy hard handovers as described in Section 6.2.1),
- multi-slice RM that supports one or more services with their associated Service Level Agreements (SLAs),
- AIV-overarching RM, where real-time context is collected from novel AIV(s) and the legacy ones to determine whether a service flow can be mapped onto a given AIV,

- intra-AIV RM based on AIV-specific RM functionalities that are tailored to the AIV characteristics, e.g. PHY frame structure,
- design recommendations to enable the envisioned agile RM framework along with their RAN design implications and analyses.

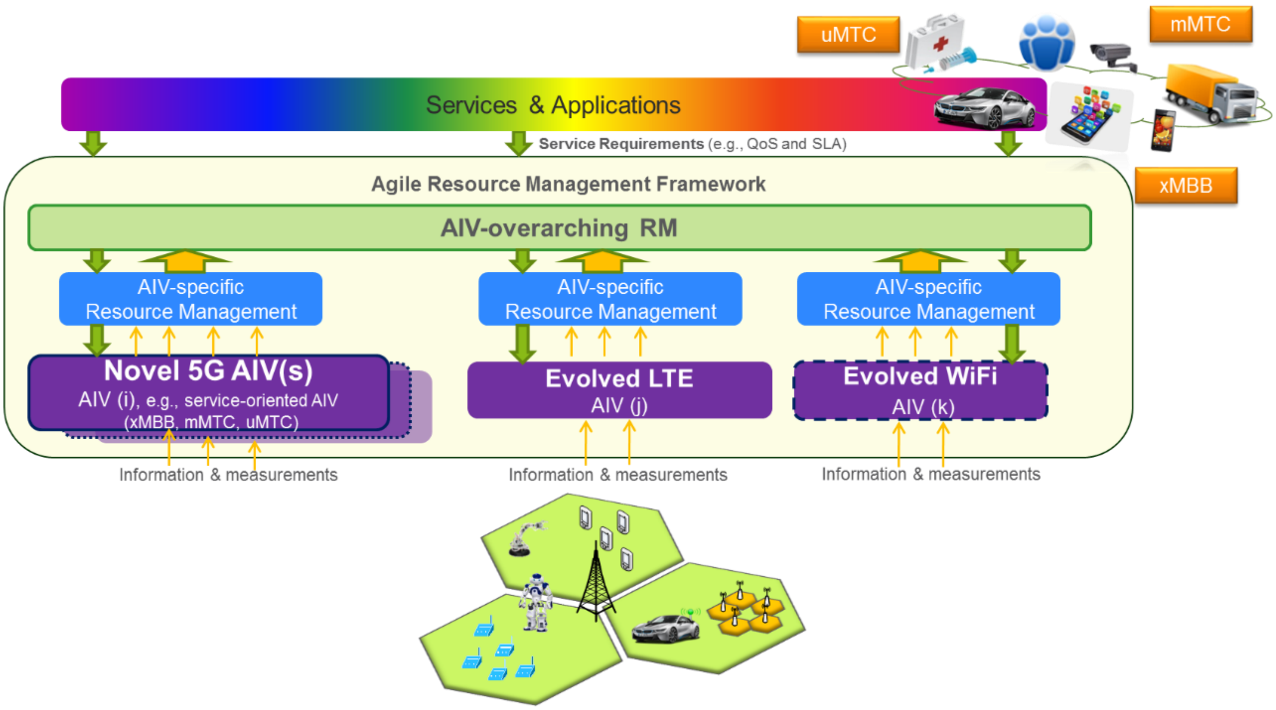


Figure 6-2: High-level illustration of the agile RM framework.

6.1.2 Service-Tailored Network Functions in 5G

To support the wide range of 5G services, it is expected that the NFs in radio protocol stack layers such as RRC, RLC, PDCP etc., must be service tailored. This can for example be that RLC ARQ is used for some service types while for others it may be inactivated. However, METIS-II envisions that the overall 5G AI should ideally be characterized by a large extent of UP protocol harmonization across the AIVs used for different bands, services and cell types (see Sections 4.2.2 and 5.3.4). Also, the aim with the CP is to be as **common** as possible and support all different AIV variants. Table C-1 shows some examples of NFs that could be tailored to specific service needs in 5G [M16II-D22]. In general, there is the common understanding that specific services will likely reuse the same functionalities as other services for a large portion of the protocol stack, differing only for a smaller number of functionalities.

6.1.3 Beam-centric Design

The reliability and interruption delay requirements of 5G are more stringent compared to 4G. In addition to this, 5G is expected to operate in a wider range of frequencies (from below 1 GHz up to 100 GHz) than 4G. This also means that BF techniques may be needed to compensate for the higher propagation loss at high frequencies. Using antenna arrays at radio access nodes with the number of elements in the hundreds, fairly regular grid-of-beams coverage patterns with tens or hundreds of candidate beams per node may be created. The coverage area of an individual beam from such array may be small, down to the order of some tens of meters in width. As a consequence, channel quality degradation outside the current serving beam area probably due to (small) objects that shadow the beam, is quicker than in the case of wide area coverage. Therefore, the BF mobility design should support a fast switching/tracking of the communication beam to combat rapid changes in link quality. Also, the design should be able to exploit MC, i.e. the availability of multiple overlapping beams that can be used for the communication with a single UE. Further on, the beam mobility should have a minimum impact to the RRC layer. One solution to fulfil these requirements is the idea of cluster based mobility. The cluster is a set of nodes that the UE can detect and which are prepared in advance for a fast re-routing of the signalling and user data, see [MII-D61] for more details.

In addition, to efficiently support BF mobility, combined UL and DownLink (DL) measurements should be utilized for 5G, once again, see [MII-D61] for more details.

6.1.4 Lean and Future-Proof Design

When the traffic demand grows, more radio nodes are required to densify the network. One of the drawbacks with LTE was the rather low possibility for the cell to enter a so-called micro sleep (the cell DTX) in this situation, leading to an increased power consumption [ERI11]. METIS-II has at an early stage proposed a so-called lean design of the broadcasted information for 5G (i.e. the system information). There are several signals and channels that need to be addressed to make 5G more lean design than LTE, listed below.

Reference Signal in general

In 5G, the Reference Symbols (RS) necessary for channel estimation should only be transmitted in the same subframe as the data transmission, over the same bandwidth, and in the same beam as the corresponding data. This is different from LTE which can also have the CRS in previous subframes to aid channel estimation, see an example in Figure 6-3 (where the red slots are RS). How this will be done exactly for NR is now up for discussion in 3GPP [3GPP-38912], [3GPP-38804].

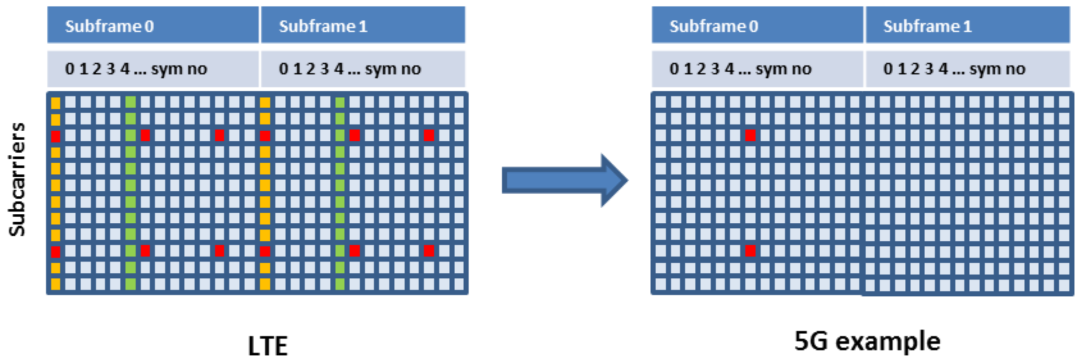


Figure 6-3: Example of 5G lean design compared to LTE for one PRB.

System information transmission using user on-demand approach

In the 3GPP NR discussions, the system information is divided into minimum SI and other SI. Minimum SI is periodically broadcast (as in LTE today). The minimum SI comprises basic information required for initial access to a cell and information for acquiring any other SI broadcast periodically (as in LTE) or provisioned via on-demand basis (new compared to LTE). The other SI encompasses everything not broadcast in the minimum SI. The other SI may either be broadcast, or provisioned in a dedicated manner, either triggered by the network or upon request from the UE [3GPP-R2168858].

PDCCH

In LTE, PDCCH is transmitted across the full system bandwidth i.e. at least one PDCCH symbol is used for all PRBs. This is not especially resource and energy efficient. For 5G, we foresee a more efficient PDCCH transmission, the goal is to be more limited to the resource used by the user data.

Synchronization signals

LTE uses a periodicity of 5 ms. However, if the periods between the synchronization signals can be increased, the BS sleep efficiency can be increased [DDL15]. The reason is that it takes some time to deactivate and reactivate certain components, and given this the longer the sleep duration, the more components can be put to sleep and the lower the sleep power usage becomes.

Figure 6-4: shows the relative power consumption per cell for 5G compared to LTE for a fixed number of users in the area (5 or 50 users per square km) and when the network is densified, i.e. smaller and smaller cell radius. This means that for small cell radiuses there is higher probability for no active users which may enable the cell to enter the Cell DTX. The major difference between 5G and LTE is the ability to utilize the Cell DTX. In Figure 6-4: NR has 4 and 6 times higher probability to enter Cell DTX (if the cell is empty).

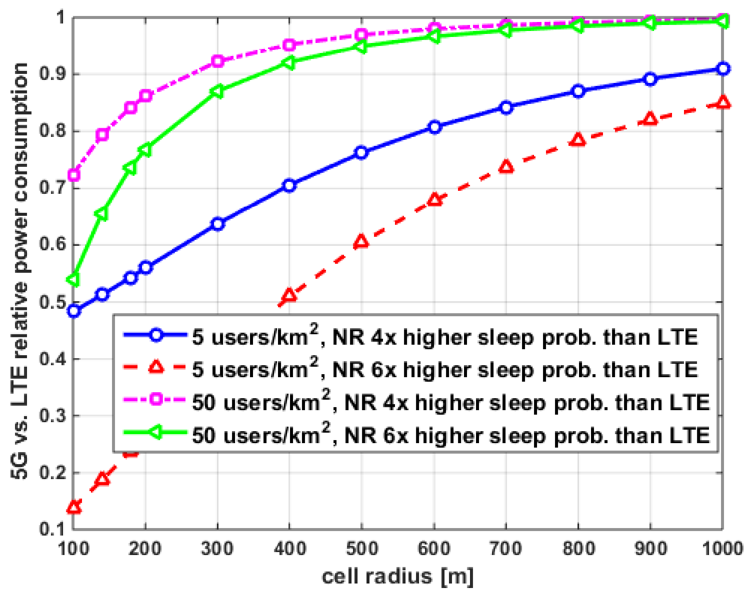


Figure 6-4: Relative 5G power consumption vs. LTE for different NR Cell DTX (sleep) probabilities when the cell has no active users. Note that there are 5 or 50 users per km^2 regardless of cell radius which means that the probability for zero users increases with *decreased* cell radius.

These results show that if the 5G RAN is designed so that it allows better Cell DTX sleep probabilities than LTE, the power consumption can be decreased substantially, also see Section 7 and [MII-D23].

6.1.5 RAN moderation for energy efficient network operation

Energy efficiency is one of the most important system design requirements of 5G, especially taking into account the increased deployment density and operational capacity requirements of such networks. The work done in [MII-D52] investigates whether this could be enabled by defining traditionally asynchronous functions such as RAN moderation on a synchronous, short-term time scale. An overview of the considerations related to this aspect using a deployment architecture diagram is shown in Figure 6-5. Here, we consider the availability of additional link-layer channel quality measurements, along with the BS traffic measurements, transported using newly defined information elements, to assist in the enhanced RAN moderation process. The channel quality measurements are used to derive the communication element for channel quality indication. The BS traffic measurements enable the calculation of real-time traffic demand as well as the signalling of DRX configuration request using RRC signalling. The signalling is used to coordinate the wireless self-BackHauling (sBH) activation / deactivation, while it is finally decided using an energy-aware backhaul / Access Point (AP) controller. Self-backhaul nodes are those BSs that have wired connectivity with the core network, providing wireless connectivity, using the 5G radio protocol stack, for those BSs without such connectivity.

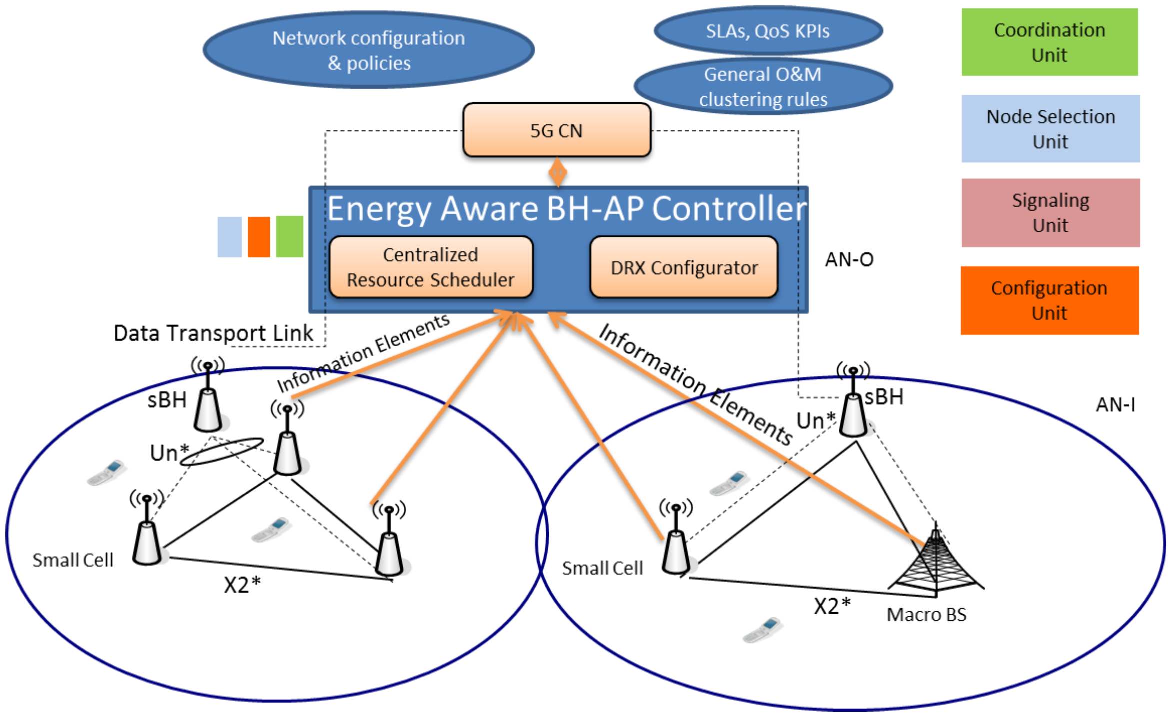


Figure 6-5: Deployment architecture of energy efficient 5G RAN moderation [MII-D52]

Various further inputs are considered for the RAN moderation mechanism, such as the user context and QoS information, in order to enable efficiency RAN moderation decisions. Since the mechanism inherently assumes lean system design concept, where the BSs are active only when there is active data to be sent to the end user, the output from the function could be radio resource allocation, explicit switch on/off command for the BS or DRX configuration for the sBH nodes.

In addition, we present a key technology enabler for RAN moderation in 5G which uses centralized resource scheduling or distributed active-mode coordination for maximizing energy efficiency. In a system supporting lean design paradigm, centralized resource scheduling has significant potential in enabling energy efficient design, mainly due to the fact that optimizing resource scheduling could help the system to operate only the optimal number of access points to transport the traffic to the end user. Distributed coordination in this context further allows simplifying the requirements for the interfaces that link various entities within the network. In this design recommendation, we mainly focus on how the concept of joint RAN – BH coordinated operation can enable such energy efficiency maximization.

An overview of the coordinated operation is as shown in Figure 6-6, where the RAN nodes indicate over the backhaul link the resource scheduling paradigms in terms of active and inactive durations. Here the BSs in the RAN are assumed to be in sleep mode when there is no active data scheduled for transmission. This enables the centralized traffic aggregation node (AN-O), which could be considered as a BH node using the sBH paradigm, to coordinate the distributed resource allocation paradigms using a centralized scheduling mechanism. Here, the interface between the sBH nodes and aggregation nodes is assumed to be an enhanced Un interface [3GPP-36300], called Un*. From the aggregation node perspective, the sBH nodes appear as UEs, due to which explicit DRX configurations are made in a centralized manner to coordinate the resource scheduling. This enables the sBH nodes to enter inactive / sleep state in a coordinated manner that allows the traffic aggregation node to enter sleep mode during a subset of the time as well. From the Figure 6-6 (c), also presented in [MII-D52], we can observe that such a coordinated scheduling operation would enable significant increase to the inactive time of the traffic aggregation node during low-load conditions, leading to higher network energy efficiency. The savings are observed in comparison to legacy LTE where RAN and BH nodes operate independently, and with 5G having dedicated BH or self-BH. The detailed parameters used can be found in [MII-D52], and the BS sleep modes based power consumption model is based on [DDL15]. In particular, Sleep Mode-1 (sleep duration of 0.071 ms) provides around 64% power savings for 5G BS with a fixed fiber access BH link, and around 90 % improvement with sBH links. Sleep Mode-4 (sleep duration of 1 s) shows around 68 % power savings for 5G BS with a fixed fiber access BH link, and around 98 % improvement with sBH links.

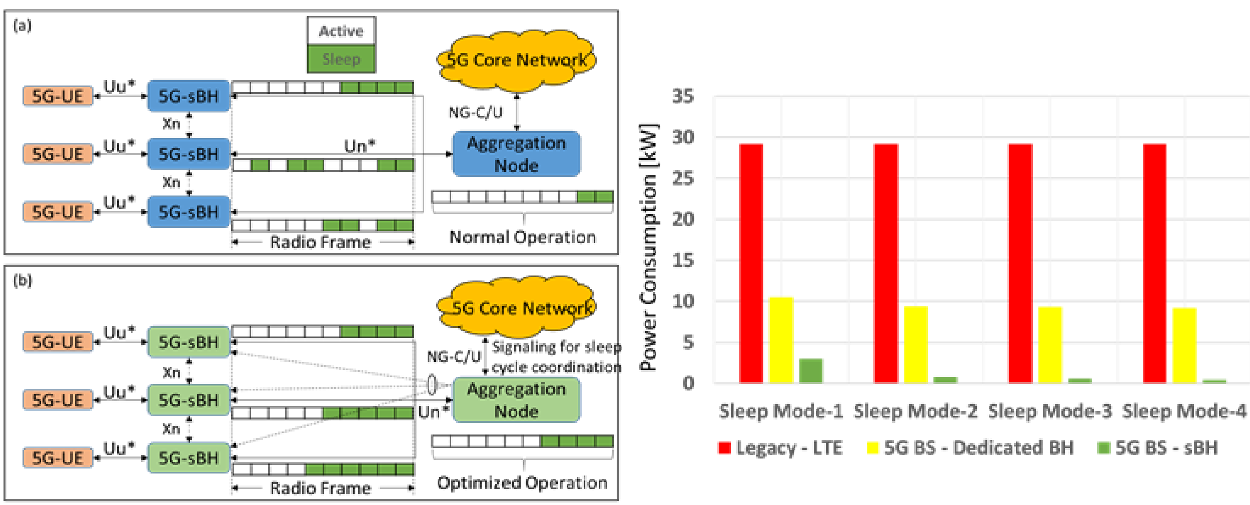


Figure 6-6: Coordinated RAN-BH sleep mechanism, (a) Normal operation, (b) Optimized Operation, (c) Energy Saving gains using joint RAN-BH operation [PUM17] [MII-D52].

6.1.6 Native Relaying, Self-backhauling and D2D Support in 5G

A key design requirement of the 5G system is the native support of relaying, self-backhauling and D2D, as opposed to legacy systems like LTE-A, where these features are either introduced as an extension to the original design or have not yet been introduced. Such add-on approach in many cases naturally involves compromises w.r.t. a potentially better design. METIS-II has been exploring communication scenarios and solution design related to relaying, self-backhauling and D2D. These includes: grouping of devices in proximity with similar communication needs, deep coverage extension for mMTC services, D2D communication in the context of mobility, cooperative communication and wireless self-backhauling in very dense 5G deployments. These design details, solutions as well as evaluation outcomes are discussed in details in [MII-D61] and [MII-D62].

Self-Backhauling: Self-backhauling is seen as a very important enabler for facilitating future 5G deployments. Using self-backhauling, a 5G base station can provide a coverage extension solution in the absence of a fixed wired backhaul link. Limited fragmented bandwidth availability in frequency bands below 6 GHz implies that 5G capacity and throughput targets, especially in dense urban centres will have to be fulfilled using cmWave and mmW frequency bands where large chunks of unused bandwidth still are available. Due to the inherent environmental sensitivity in terms of high path and penetration losses in the mmW range, it is imperative that the average inter-site distance for such cells will be drastically reduced compared to today's deployments. Even in highly developed urban centres, current fibre access is not sufficient to accommodate such a dense deployment. Self-backhauling provides a cost-effective way to overcome these challenges as it does not rely on fibre availability to extend cell coverage.

In the absence of fibre, a normal 5G base station can act as a self-backhauling base station by providing backhaul connectivity wirelessly to its users via another fibre fed 5G base station. In comparison to other wireless backhaul solutions such as microwave based point to point transport links, 5G self-backhauling uses the same access technology for backhaul and access while retaining most of the hardware functionality of a standard base station. Backhaul and access links can either be in-band or out-of-band depending on the available carriers, however the use of same access technology and form factor drastically reduces the cost compared to dedicated transport solutions which rely on well-directed deployments. Antenna panels for backhaul and access can be kept at a reasonable size due to expected beamforming gains at mmW bands, which makes it also easier to combat self-interference thanks to highly directed beams. From standardization perspective, self-backhauling is expected to have minimal impact on the RAN design with various architectural options under consideration and the common goal of sharing maximum possible functionality with other base stations. From radio access perspective, optimal performance can be ensured due to dynamic scheduling of resources between users and self-backhauling base stations. The early support of self-backhauling in the standardization process also promises forward compatibility and overcomes many of the performance bottlenecks which led to limited success seen by LTE relaying.

D2D Relay: To overcome the propagation constraints and bottlenecks in signalling channels (such as the RACH, Non-Access Stratum (NAS)) in mMTC communication and related power consumption challenge at device side, METIS-II studied the exploitation of context-aware D2D communication for mMTC [MII-D61]. In such cases, certain UEs are selected by the network to act as relay UEs for mMTC devices located in cell boarder or in deep indoor. In order to optimize the system performance in terms of service availability and device power consumption, context information is collected and exploited by the network to efficiently set up D2D pairs. The signalling diagrams introduced in [MII-D61] illustrate the radio link enablers for the proposed scheme, and correspondingly, the numerical results given by [MII-D62] show that this scheme improves the service availability from 85% to 99%. In addition to that, the percentage of UEs, who can meet the ten years battery life requirement, is improved from 75% to 90%.

D2D Relaying mechanisms introduced by METIS-II in [MII-D61] and [MII-D62] manage to reduce the signalling cost using group communications for TAU and reduce the collision rate in the RACH. In the former case, the group head undertakes the TAU process for all the members of the group whereas in the latter the group head performs the group RACH request. The details as well as the evaluation results for the two schemes can be found in [MII-D61] and [MII-D62].

D2D in Underlay Mode: Without any additional infrastructure deployment and spectrum demand, the network controlled D2D is an efficient approach to offload some cellular traffic to the local information exchange procedure, in order to improve the overall system capacity. In this scheme [MII-D62], D2D links transmit on the same time and frequency resource as the cellular UL transmission. However, in this approach, the network controls the introduced mutual interference between the cellular and D2D links in a smart way. The signalling schemes proposed in [MII-D62] enable a context-aware RRM algorithm. In addition, the given system performance [MII-D62] shows the improvement w.r.t. the overall system capacity. Depending on the transmission power setting, the proposed scheme can provide a system capacity higher than the legacy cellular network. For example, using the simulation settings in Section 7.3 of [MII-D62] where we assume that 50% of the users can use D2D links for communication with another UE, the system capacity can be increased up to 60% with an undelay mode D2D. .

SL Mobility Management: Mobility management of the interface between two UEs over which direct communication is ongoing, i.e. PC5 interface according to 3GPP, is an important design requirement of several 5G services including V2X. In a general scenario, two UEs participating in D2D communication are considered. However, the number of D2D and/or V2V devices participating or being part of a particular group can be different depending on the particular application scenario. For example, a platoon of vehicles might consist of a leader and several followers, whereas a group of wearable devices or IoT devices might have significantly more communicating devices than a platoon. The group mobility issue arises when UEs due to mobility reach the cell edge and all group members may or may not fully satisfy the handover condition simultaneously. In the absence of a reliable scheme to handover a D2D and/or V2V group, the established D2D and/or V2V link within the group or between a pair of D2D devices would be interrupted, leading to packet loss. Moreover, each UE in the D2D and/or V2V device group are likely to be handed over to the target cell in an individual fashion, which leads to extra signalling overhead.

To address these mobility aspects, suitable mobility management schemes targeting moving D2D and V2V devices have been designed. The approach is based on four general handover steps focused on SL mobility problems: signal quality measurement, coordination between the source and target BSs, resource allocation of the target BS and packet switch from the source BS to the target BS. Each of these four steps specific to D2D scenarios are studied in details in [MII-D62].

Figure 6-7 compares the D2D mobility management scheme and the LTE handover scheme. It shows the D2D communication reliability with respect to D2D devices mobility. The handover delay of each individual UE is kept at either 2 ms or 200 ms during the simulation.

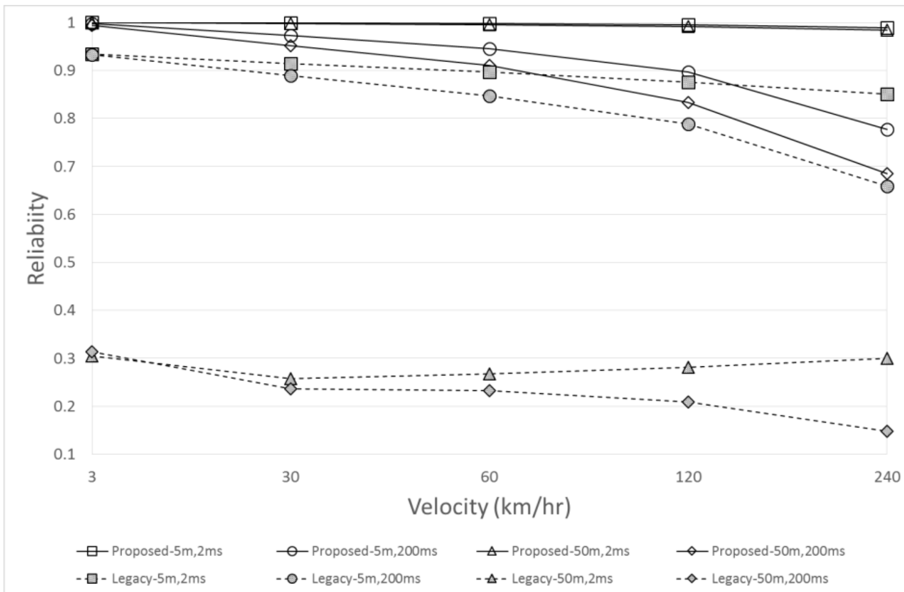


Figure 6-7: Effect of mobility on D2D communication reliability

<u>Cooperative D2D</u>: Cooperative D2D communications where D2D pairs implement relay functionalities to facilitate transmission between a cellular user and its BS is a way to improve spectrum efficiency. In such scenarios there is unicast D2D communication and/or one-to-many/all D2D communication among pairs of devices over the PC5* interface. PC5* is the enhanced reference point between ProSe-enabled UEs used for control and user plane for ProSe Direct Discovery, ProSe Direct Communication and ProSe UE-to-Network Relay. So, in this case one of these devices can be source (D2D transmitter) while the other devices are the destination (D2D receiver). The cooperative communication scheme enables 5G RAN to dynamically allow cooperative D2D mode selection and communication, at the same time ensure interference mitigation e.g. in case of simultaneous D2D communication and cellular user to BS communication over the shared radio resources, etc. To enable cooperative D2D communications, among others, approaches for cooperative mode selection, relay selection, cooperative transmission and resource allocation are discussed in METIS-II Deliverable D6.1 [MII-D61].

In D2D communication, interference management is one of the key issues to ensure high SE. Various techniques involving MIMO signal processing, power control, and transmission mode selection have been proposed to reduce the interference between the D2D pair and the cellular user or BS, especially when multiple D2D pairs are allowed to share the same channel. Some mechanisms need to be designed to further mitigate the interference both among D2D pairs and between the D2D transmitters and the cellular system. By allowing cooperation among D2D transmitters, more D2D pairs can be allowed to transmit simultaneously in the same and limited spectrum resource, increasing the spatial spectrum utilization of the system. This approach is studied in details in [MII-D62].

Performance comparisons between proposed cooperative D2D transmission method and the non-cooperative method at different D2D pair number shows that the average sum rate increases with the number of D2D pairs regardless of whether cooperation exists. This is because more D2D pairs are considered for data transmission. However, improvement gets saturated when the number of D2D pairs is large enough. Besides, the average sum rate in proposed cooperative D2D transmission method is better than that in non-cooperative method because resource balancing and fairness (rate-gain constraint) are considered.

Another cooperative communication concept discussed in [MII-D52] is Group Transmission which is a way to implement joint transmission to increase the coverage and user bit rate compared to legacy single UE transmission.

6.2 Functions related to Agile Traffic Steering and Resource Management

The agile RM framework is described in Section 2.3 and Section 5.3.4 from conceptual and protocol perspectives. A simplified version of the functional architecture of that innovation pillar [MII-D52] is depicted in Figure 6-8. The agile RM framework is designed to take into account the key 5G RAN design requirements outlined in Section 3. In particular, the overall functional architecture is formed by functionality frameworks of AIV-overarching RM and Intra-AIV RM. AIV-overarching RM comprises functionalities, which are operating over multiple AIVs to map the data flows to appropriate AIVs based on the context received. The functionalities that need to be tailored to each AIV construct the Intra-AIV RM. In Figure 6-8, a hierarchical CP design is illustrated, where the AIV-overarching functionalities are mainly located with the AN-O layer, which comprises of mechanisms which are essentially not limited by how an AIV is defined. The overarching functionalities could be applied to different AIVs simultaneously, as well. Intra-AIV functionalities are constrained by the AIV design and, hence, are assumed to be located within the AN-I layer. The communication between overarching and Intra-AIV functionalities are assumed to be based on quantized or abstracted values, which could be applied to any AIV. Thus, e.g., the load measurements reported by the AN-I layer to AN-O layer would be quantized in such a way that similar measurement values would be reported by multiple AIVs encountering the same load condition. Recall that, in this implementation, AN-O corresponds to a CU and AN-I corresponds to a DU

Accordingly, the functional architecture of the agile RM framework is designed to provide flexibility in order to cope with future needs. For example, whenever a new AIV is added, the framework functionality can be extended by modifying the relevant functionalities, which are referred to as Logical Entities (LEs) and LE Groups (LEGs). In the following, various essential elements of the agile RM framework are presented briefly with main conceptual descriptions, key design recommendations, and achievable gains.

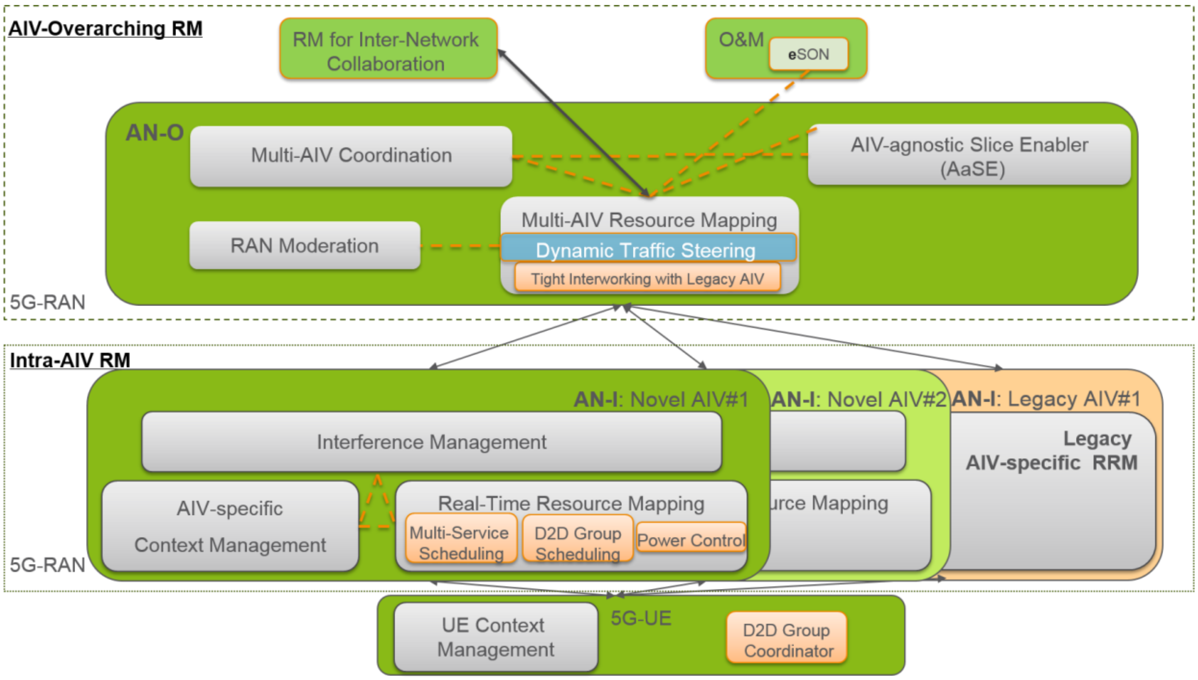


Figure 6-8: Functional architecture perspective of agile RM framework as one of the key innovation pillars developed by METIS-II.

6.2.1 Multi-AIV Resource Mapping

For multi-AIV resource mapping, we combine the elements from related building blocks presented in [MII-D52], such as tight integration with evolved legacy AIVs and Dynamic Traffic Steering (DTS), in order to provide a unified resource mapping function. The deployment architecture Multi-AIV resource mapping is as shown in Figure 6-9. In order to enable enhanced feedback configuration and mode selection between DC and Fast Switching (FS) for tight integration with legacy AIVs, enhanced measurement and configuration signalling functions with such information transported using newly defined information elements over the transport link are considered, which communicates with a coordination and configuration unit, in order to make final decisions.

Similar considerations are made for the DTS function as well, with additional focus on dynamic QoS / application detection functionality present in the RAN, in order to enable dynamic modification of service flows to serve the end users efficiently. Aspects related to the beam-based system design has also been incorporated into the resource mapping paradigm, with physical resources within directional beams considered to be one possible way of mapping the resources for the end users. The Pre-emptive Geometrical Interference Analysis (PGIA) and Resource Sharing Cluster (RSC) management function primarily relies on mmW node localization to generate the DTS and scheduler signalling to enable interference-aware DTS and multi-AIV resource mapping.

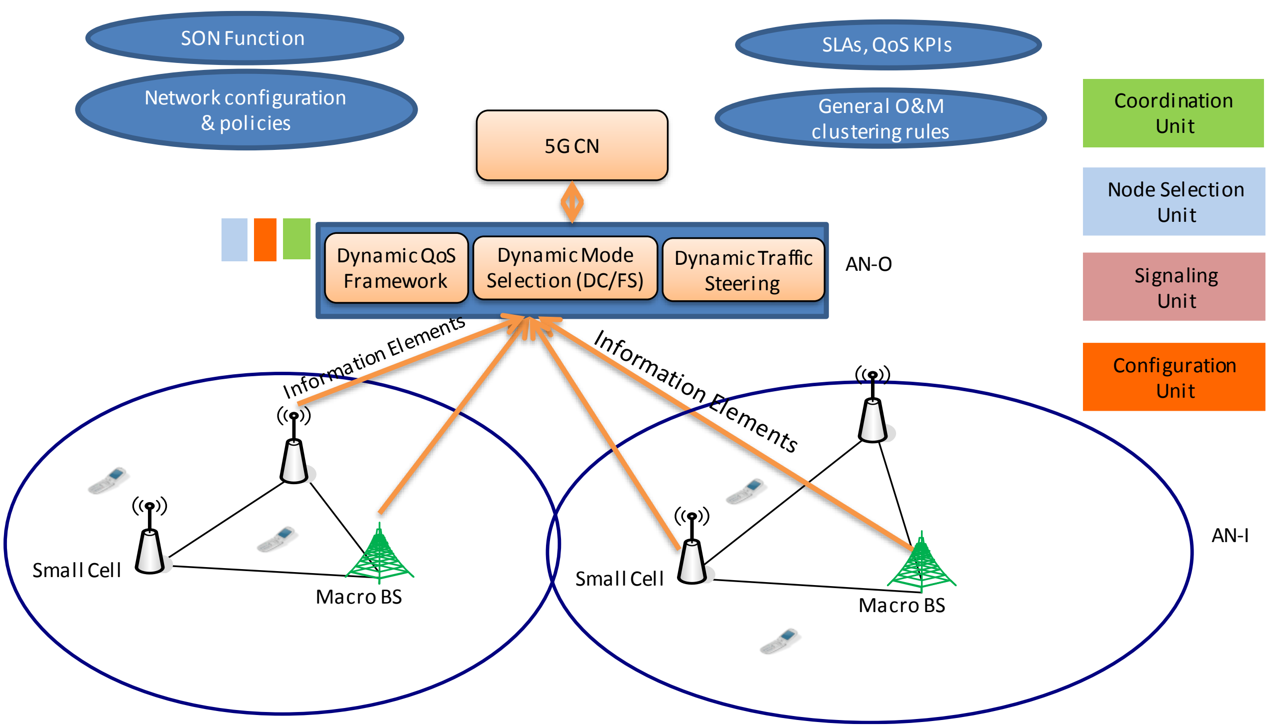


Figure 6-9: Deployment architecture for Multi-AIV resource mapping [MII-D52]

In addition, we present the following technology enablers for multi-AIV resource mapping in 5G along with the quantitative results being given in Table 6-1, where detailed analyses and system models are provided in [MII-D52]:

Enable PDCP level FS between novel and legacy AIVs: This design recommendation proposes to enable switching between novel / 5G and legacy AIVs on a fast / synchronous timescale in order to achieve better traffic aggregation and higher data rates for the end users. One of the KPIs to be considered while making the resource mapping over the set of available AIVs is the interference conditions. Here it is considered that the UE could be scheduled over the link with lower total interference, in order to improve the reliability and throughput experienced by the user. In terms of resultant RAN implications, it is expected that a common S1* interface needs to be defined between CN and RAN, with the interface terminating at the MeNB, with the SeNB link paradigms controlled by the SeNB. Additional measurement information exchange is also expected for such enhancements, with the information being exchanged in a faster timescale.

Enable Faster Traffic Steering: While FS mechanism focuses on enhancing the currently defined DC paradigm to enable better service provisioning and improved link reliability in 5G, the main focus of this enabler is to leverage the flexible protocol split (not only limited to PDCP layer) and dynamic QoS considerations in 5G. The goal is to enhance the RAN architecture to support the divergent use cases and services currently envisioned for the system. Here we consider that the QoS definition would be still done at the CN, while QoS enforcement functions would be defined in the AN-O layer of the RAN, where the higher layer RAN logical functions are located. Such enhancements would enable fast link selection, QoS modification depending on dynamic link conditions and packet duplication to provision eMBB and uMTC type of traffic. The key RAN design implications include the definition of dynamic QoS enforcement functions in the RAN, new information elements for transporting additional measurements for enabling fast traffic routing.

Enable Pro-active DTS rather than being reactive: This mechanism proposes an interference aware traffic steering and beam management mechanism called pro-active PGIA which limits the probability of transmission collisions over multiple links, which limits the achievable rates in the system. The key assumption for the mechanism is that the network is aware of the geometric position of all the users and mmW nodes in the system, which enables the logical elements such as mmW node localization function and location signalling function. These LEs enable the final decision making PGIA and RSC management functions which does the scheduler and DTS signalling.

Increase Environmental Awareness for RM Mechanisms: Increased RAN environmental awareness is an essential RM requirement in 5G, especially for mmW type of deployments. In this enabler, we propose the increased awareness in terms not only the direct beam transmissions in an ultra-dense mmW deployment scenario, but also of the strong reflected beams that a UE receives. Through additional RRC configuration, the UE is configured to report the strongest direct and reflected beams, so that the network can build a reflected environment maps (RefMaps) in order to reach the UE, in case there is a link blockage over direct beams. This enables the UE to remain connected to the network even if the direct beam is blocked due to the dynamic variations in the network environment, thereby improving the reliability of the system. The key RAN implications could include defining new information elements to transport the RefMaps information between 5G-RAN and SON or network management entities, as well as the possible creation and maintenance of RefMaps database in the system.

Table 6-1: Overview of Technical Enablers for Multi-AIV Resource Mapping

Technical Enabler	Gains and Results
Enable PDCP level FS between novel and legacy AIVs	**Low-Load** (0.1 users per cell): Dual Connectivity provides 25 % higher throughput than Fast Switching **High-Load** (1.6 users per cell): Fast Switching provides 33 % higher throughput than Dual Connectivity
Enable Faster DTS	Reliability / Signal-to-Interference and Noise Ratio (SINR) improvements at approximately 50-th percentile: ~1.5 dB and 3 dB for two and three links cooperating as compared to LTE baseline
Enable Pro-active DTS rather than being reactive	**Low-density** (10 links per sq. km): Approximately 5 % for PGIA based mechanisms, as compared to without PGIA **High-density** (200 links per sq. km): Approximately 95 % for original PGIA, and with clusters & sum, ~85 % for PGIA with clusters.
Increase Environmental Awareness for RM Mechanisms	Number of active connected mode users: 20 % active users for the enhanced mechanism with reflection environment maps with a 5 dB offset (RefMaps-5dB case)

6.2.2 Resource Management for Network Slices

Network slicing enables end-to-end service chain optimization for different services. While the CN optimizes the placement of VNFs, the RAN needs to handle slice specific configuration rules [5GN-D32] in addition, such as advanced KPI requirements of a single service as well as business driven SLAs when it comes to radio resource allocation among slices.

A Key Design Recommendation in this regard is to **enable AIV-Agnostic Network Slicing Support by using SLA based QoS adaptation and slice-adaptive RRM placement**, as detailed in the following.

Figure 6-10 shows an overview of the functional architecture proposed to enable network slicing. A new functional entity, the AaSE (see also Section 5.3.3) consists of three elements: an AIV overarching monitoring entity, a logical entity to control and dimension RAN slicing as well as an element to adapt QoS specific functionality, such as Admission Control (AC), Allocation and Retention Priority (ARP), and Quality of Service Class Identifier (QCI) according to the slice requirements.

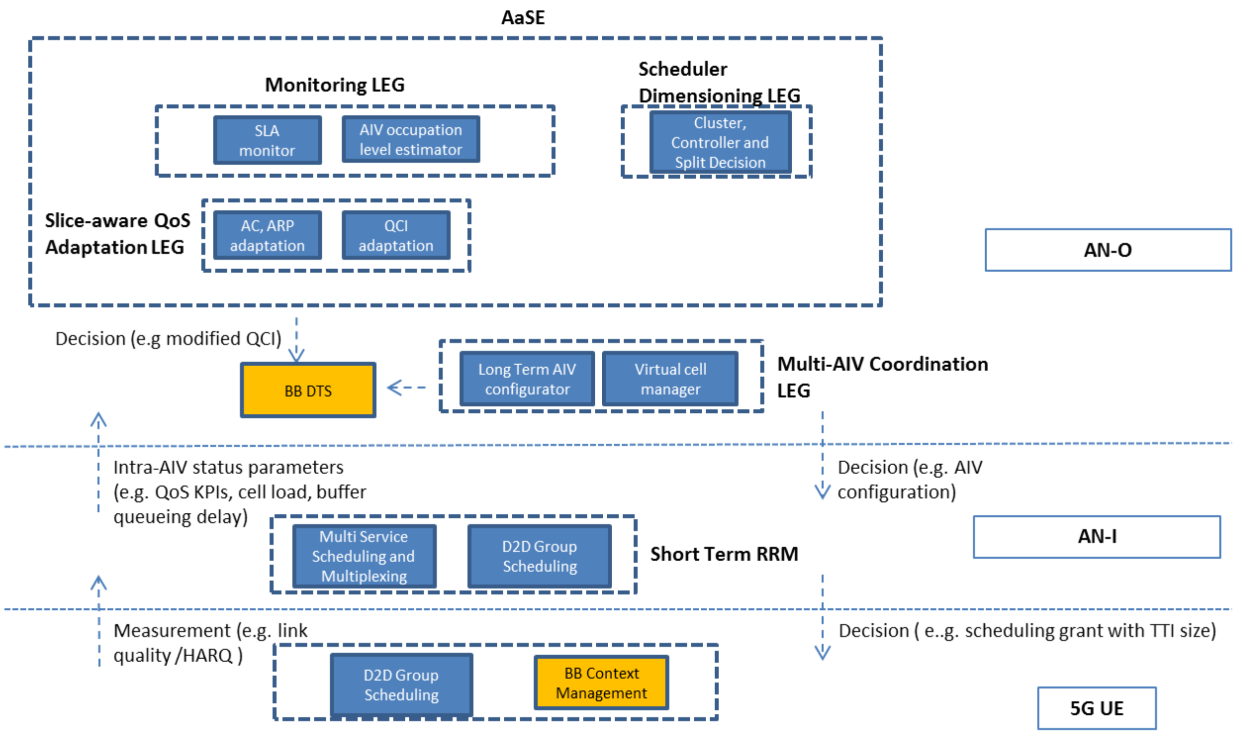


Figure 6-10: Functional Decomposition for Multi-slice and Multi-service Holistic RM

With respect to the RAN, an efficient sharing of scarce radio resources among the network slices is the key challenge, which is achieved by slice awareness. It is realized with the help of the AaSE which is responsible for monitoring and enforcing SLAs for individual slices by mapping the abstract slice specific SLA definition to the QoS policies. It monitors the status of the SLAs and adapts QoS parameters accordingly. It could, for example, in case of a network slice with a latency guarantee, assign a certain QoS class to all corresponding data flows that are part of it. Using ARP, the importance of individual data streams can be configured. It is then a task of the multi-AIV resource mapping, interference management, and real-time resource mapping to realize the corresponding QoS. More details on the proposed solution as well as simulation results can be found in [MII-D52].

Furthermore, a key functionality of AaSE can be the adaptive placement of intra-slice RRM functionalities to the RAN nodes, assuming that schedulers can coordinate clusters of APs. By taking into account the slice requirements, the backhaul/access channel conditions and the traffic load, AaSE can assign schedulers to BSs for pre-defined clusters of nodes, as well as RRM functionalities with different levels of centralization in order to meet the per slice SLAs (in terms of throughput, reliability, latency).

The multi-slice RM functionality is demonstrated by the following simulation example (for detailed assumptions see Annex A.10 of [MII-D52]): In a today's scenario, 2 dedicated RANs (subnetworks) may be operated in parallel for independent businesses, each covering a channel bandwidth of 10 MHz. Subnetwork 1, representing a special purpose network which is overprovisioned to guarantee a high quality of service, serves 100 users with low data demand resulting in a low network load. In contrast, subnetwork 2 represents a low-cost best effort network that serves 710 users causing a fully loaded system with lower performance per user. Figure 6-11 shows the probabilities for the achievable user throughput in both subnetworks and the total one (see the red curves).

With the slicing concept both subnetwork types (now called slice 1 and 2) may run as logical networks onto a common RAN infrastructure which also allows sharing their frequency resources (resulting total bandwidth of 20 MHz). For slice 1 an SLA is assumed to guarantee the same overall network capacity as it was the case with the dedicated subnetwork 1, whereas users of slice 2 are still served via best effort. For this setup, it is not expected that the slices will achieve the same user-specific throughput performance as the dedicated subnetworks for two reasons: Different RRM approaches (subnetwork-specific scheduling vs. joint scheduling with prioritization of slice 1) and different interference conditions (especially the low interference in subnetwork 1 vs. the fully loaded shared network).

The results are depicted in Figure 6-11 via the blue curves. For slice 1, the user throughput distribution has changed because of the higher interference occurring when both slices are served on the shared band. The scheduler compensated the users of slice 1 with lower channel quality by allocating more resources to the users with higher quality to keep the target SLA of slice 1. Users in slice 2, even when served via best effort only, profit from increased resource space. In this way, also the overall performance is strongly improved by that concept compared to the scenario with two dedicated networks (see solid lines). As demonstrated, multi-slice RM can achieve performance gains due to resource pooling while protecting the performance of individual slices. For simplicity reasons the example was only related to network capacity as KPI, but the concept also allows guaranteeing a mix of different KPIs like throughput, latency, and/or reliability.

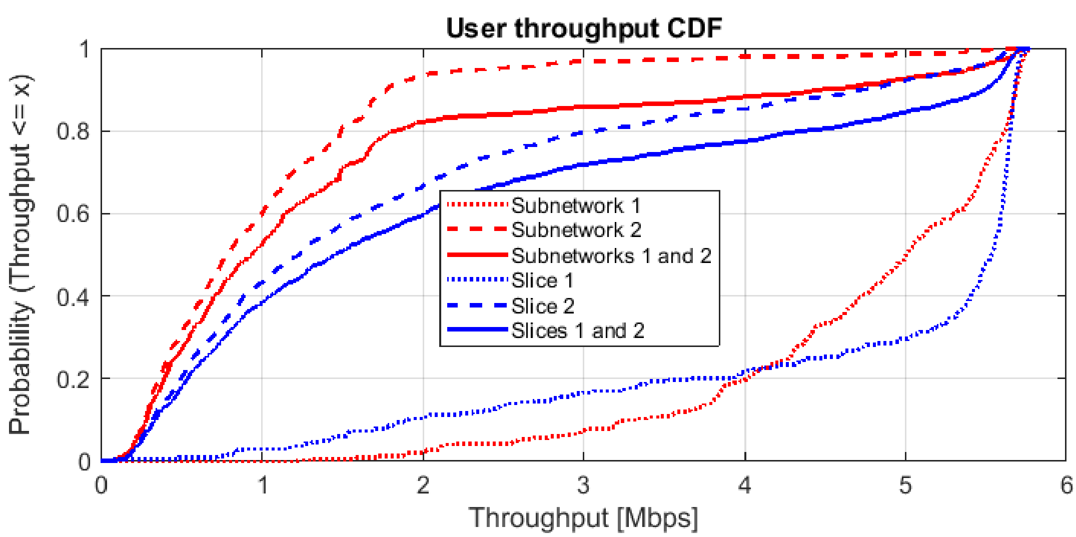


Figure 6-11: **Simulation results of Multi-slice RM**

One additional evaluation study is shown for the scheduler dimensioning and placement of RRM functionalities. For different slices, we may have different requirements for SE and RRM centralization. For the example shown in a practical scenario (see Annex A.11 of [MII-D52]), for uMTC (URLLC) more than 1 bps/Hz is an acceptable level, while for eMBB more than 2.5 bps/Hz SE is required. Thus, we select the level of centralization considering these requirements and the interference levels (e.g., for cell edge users we might need centralization to benefit from MC at cell edges). The per-AP SE for this particular simulation setup can be seen in Figure 6-12.

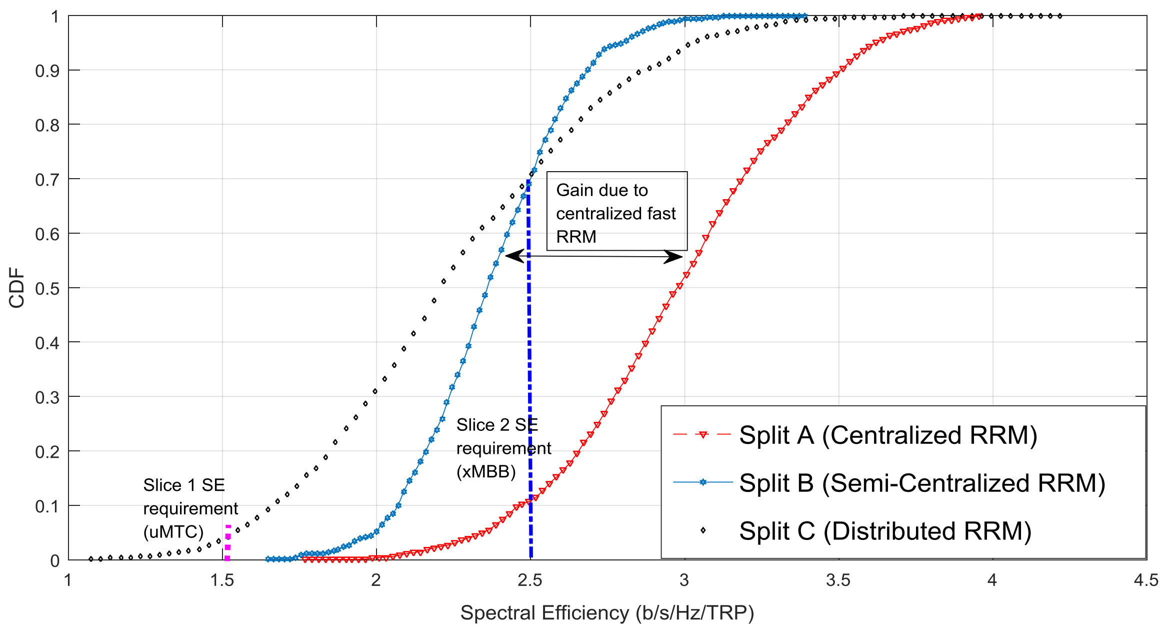


Figure 6-12: **CDF of Spectral Efficiency – Comparison of different splits**

As we can observe from the Cumulative Distribution Function (CDF) of SE, for the uMTC slice we do not need to centralize RRM, unless the users are near the cell edge (e.g., 5 percentile), since the SE KPI is fulfilled. On the other hand, for eMBB the higher the centralization the higher gain we can achieve.

6.2.3 RAN Enablers for Interference Management

Overcoming interference is essential to ensure high capacity and wide coverage, as well as robust and efficient communication. In METIS-II, we have proposed an overarching building block as part of the overall RRM architecture and a set of design recommendations for IM enablers in 5G. The reader is referred to [MII-D52] for all the details.

The internal functionalities of this building block, referred to as logical entities and mapped to the deployment architecture, are provided in Figure 6-13. Details of each entity are as follows:

- *Signaling*: Performs long-term and short-term measurements on the (self-) backhaul (for the case of a dynamic topology comprised of NNs), access (i.e., regular UE-BS channel), and on the channel between APs (as needed for cross-link IM in dynamic TDD scenarios). It also handles all the signaling coming from different network elements as explained above.
- *Configuration*: Handles all the information messages needed to configure all network elements involved in any IM scheme, i.e., it takes care of the network coordination part.
- *Node Selection*: Carries out the actual decision making on the nodes affected by the scheme. As an example, it determines whether UEs belong to the group of users that should be scheduled with an interference-resistive modulation such as Frequency Shift Keying and Quadrature Amplitude Modulation (FQAM). This LE also provides an interface to the RAN moderation building block as the set of serving NNs needs to be selected from the available set of candidate NNs.
- *Coordination & Scheduling*: This LE contains the intelligence related to all network coordination algorithms and schemes necessary to apply the IM schemes. Nodes involved in this LE include both static and dynamic APs (i.e., BSs and NNs) as well as BS clusters.

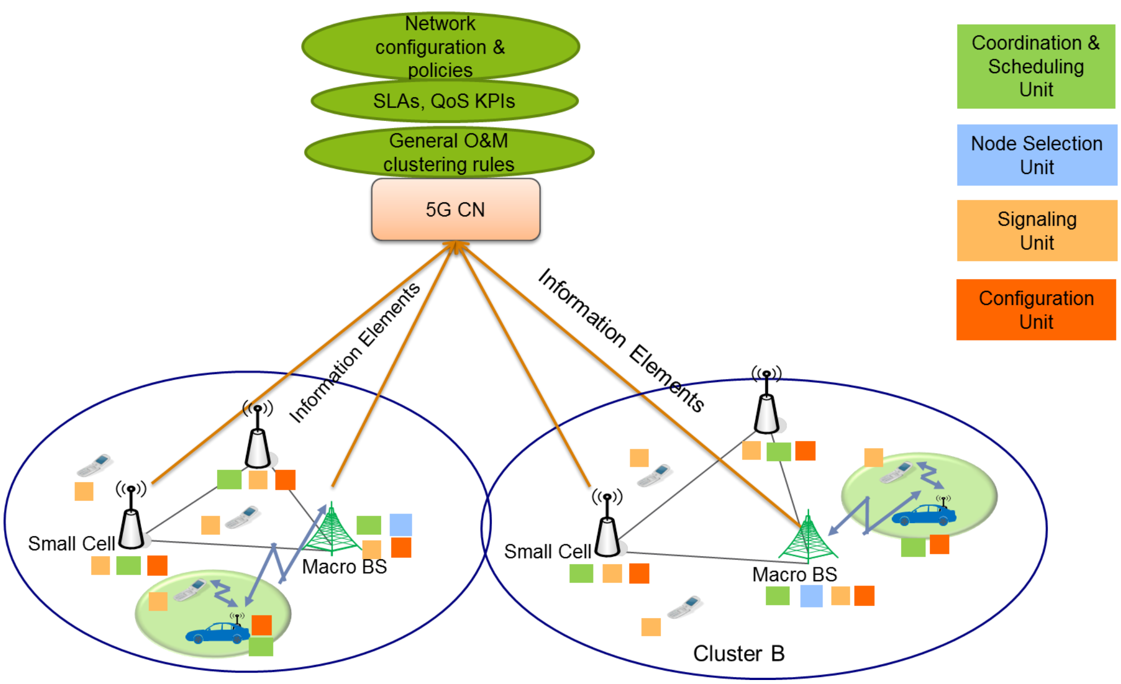


Figure 6-13: Deployment architecture for interference management

In addition, we present the following key IM enablers for 5G along with the key findings in Table 6-2, where detailed analyses and system models are provided in [MII-D52]:

Enable adaptive interference mitigation to cope with dynamic radio topologies: A key aspect of the IM building block is to provide UE-centric IM in heterogeneous UDNs by means of selecting overlays of access nodes that can serve users individually, given their diverse service requirements. On top of that, coordinated resource allocation and JT will be applied adaptively based on the BH conditions, the load constraints and the service type. In [MII-D52], we provide a case study for a hotspot area and a 5G RAN consisting of a number of NNs under a macro-cell umbrella. In particular, we consider a dynamic network topology comprising such non-static access nodes, which emerges as a promising notion enabling flexible network deployment and new services. Up to 45% and 52% higher user throughput can be achieved at the 90th (cell center) and 10th percentile (cell edge) of the CDF in case of activating more NNs with interference management, respectively

Enable adaptive interference mitigation exploiting interference-resistive design via advanced modulation and coding techniques: One key requirement for the 5G system is the enhancement of cell edge user performance to ensure that every user is supported with consistent experience anywhere in the network. This can be achieved by employing advanced modulation and coding schemes embedded in advanced transceivers. An active interference design to improve anywhere performance, particularly in the low SINR regime, can be achieved by applying a recently proposed new type of modulation scheme FQAM, which could change the distribution of interference and therefore improve channel capacity. In this design recommendation, a resource partitioning scheme to support FQAM in interference intensive scenarios was proposed. The proposed scheme partitions radio resources into orthogonal parts for QAM and FQAM, respectively, along two different resource dimensions, namely, space and frequency. This can be achieved by incorporating advanced BF algorithms, e.g., full-dimension (FD)-MIMO), and performing a frequency-based split of FQAM resources to effectively improve the data rate of the edge users experiencing heavy interference. Results show that cell edge UE throughput can potentially be significantly improved as much as by a factor of 5, for the scenario under study, by applying FQAM to those UEs experiencing high level of interference, boosting average and the 5% rates of the user rate CDF curve. This benefit would have to be balanced against potential drawbacks (reduced spectral efficiency in the baseline).

Employ transmit precoding to mitigate same- and other-entity interference for dynamic TDD in UDN: This design recommendation proposes a novel way to mitigate both BS-to-UE and BS-to-BS interferences by means of network-wide JT where single-antenna BSs cooperate to construct one large spatially distributed antenna array in the DL. JT is facilitated using zero forcing transmit precoding in order to cancel BS-to-UE interference. UEs equipped with single antennas are however unable to perform transmit precoding in the same way and therefore transmit independently. To deal with BS-to-BS interference, it is proposed that UL BSs in terms of their complex-valued BS-to-BS channels be included in the precoder design. Since DL BSs are not aware of which symbols UL UEs will transmit beforehand, dummy symbols are transmitted virtually with zero power. The proposed scheme is denoted as JT with Dummy Symbols (JT-DS). Results show that at low and medium utilization, both UL and DL performance can be significantly improved with proposed scheme (JT-DS) by 10-20%. At high utilization, ill-conditioning limits the received signal power in the DL.

Support dynamic selection of transmission path (DL or UL) for pilots in dynamic TDD systems: This scheme shows that interference can be avoided in a dynamic TDD system with at least one massive MIMO BS as follows: i) selecting the right transmission paths for the pilot signals in BSs without overhead constraints and ii) selecting the right order in the data slots of all cells. Hence, selecting a TDD configuration at small cells that avoids beamformed interference is dependent on both the communication path selected for the small cell pilot signals and the configuration of the data slots (i.e., UL or DL). Different algorithmic solutions could be designed to mitigate cross-link interference. However, for any specific method designed to have a significant impact on performance, it is crucial that communication paths for pilot signals can be dynamically selected at each subframe while the number of slots are determined by the load distribution. Specifications should therefore provide support for that additional level of flexibility if massive MIMO arrays are enabled in at least some BSs. Results show that UL and DL spectral efficiencies can be boosted by an average of 1bps/Hz in the whole CDF distribution.

Enable Interference Avoidance in high SINR scenarios: A design recommendation is needed for 5G usage scenarios in which it is required to increase coverage of nodes with high SINR, and the interference can be generated by a multitude of sources, as in multi-layered high density deployments. This TeC aims to configure a procedure for orthogonalizing neighbor BSs transmissions, by means of CP based IM information (spreading & scrambling codes) between BSs grouped in a BS cluster. In order to keep frequency band usage limited, the approach is based on time spreading of the complex baseband symbols transmitted in the 5G time-frequency grid. The spreading codes, also known by the UE will allow the recovery of the complex symbols sent with increased level over orthogonalized signals from other BSs and even over the uncorrelated noise. Results for a cluster of 4 BSs offer around 6 dB gain taking into account the lower value at which a few kbps connection is feasible, and a cluster of 8 BSs further improve the gain by 3 dB.

Table 6-2: Overview of gains and results for interference management enablers

Technical Enabler	Gains and Results
Interference mitigation in dynamic radio topologies	Up to 45% and 52% higher user throughput can be achieved at the 90th (cell center) and 10th percentile (cell edge) of the CDF in case of activating more NNs with interference management, respectively
Advanced modulation and coding techniques	Cell edge UE throughput can be improved as much as by a factor of 5 by applying FQAM to those UEs experiencing high level of interference, boosting average and the 5% rates of the user rate CDF curve.

Precoding for dynamic TDD in UDN	At low and medium utilization, both UL and DL performance can be significantly improved with proposed scheme (JT-DS) by 10-20%. At high utilization, ill-conditioning limits the received signal power in the DL.
Dynamic selection of pilot transmission path in dynamic TDD	Managing the interference caused by the pilot contamination effect in a HetNet by means of the TDD configuration introduces large gains on the attainable user rates: UL and DL spectral efficiencies can be boosted by an average of 1bps/Hz in the whole CDF distribution
Interference avoidance in high interference scenarios	Orthogonalizing neighbor BSs transmissions via BS clustering and selection of modulation and coding schemes introduces large gains for low SNR/SINR users. Results for a cluster of 4 BSs offer around 6 dB gain taking into account the lower value at which a few kbps connection is feasible, and a cluster of 8 BSs further improve the gain by 3 dB,

6.2.4 Novel UE Context Management in 5G

Context awareness is defined as delivering real time context information of the network, devices, applications, the user and his environment to application and network layers in the context of IMT-2020 [ITU-R14]. The context data are gathered by UE and BS, and then they are sent to specific databases in the network and exploited by extended and new radio management algorithms; see also [MII-D62]. While designing the UE context in 5G networks, the amount of data to be gathered and the complexity of RM algorithms need to be considered carefully between the network performance enhancements they make available and the load they impose on both the BS and the UE in terms of data gathering, signaling, processing and storage.

To address the above challenge, we have proposed an adaptive framework for context management, which has the following essential LEs:

- Measurement Functions, in which the UE and the BS perform measurements,
- Measurement Communication Function, which sends the UE measurements to the BS and vice versa,
- Configuration Function, which selects the most suitable UE measurement configuration profile.

In what follows, we describe how the different LEs interact with each other.

Firstly, a set of the so-called Measurement Configuration Profiles (MCPs) is defined and stored at both the BS and the UE. Each MCP contains a predefined set of UE measurement configurations (e.g. UE measurement intervals, measurement sampling rate, maximum number of measured cells, etc.). The framework allows the UE and the BS to select the best suitable MCP according to variety of parameters. Those parameters can be categorized into different groups as following:

- **UE-Calculated Parameters:** This group contains all the parameters calculated by the UE (e.g., UE mobility state, UE power state, UE capability, etc.) and then reported to the BS.
- **Infrequently-Changing BS-Calculated Parameters:** This group contains all the infrequently changing parameters calculated by the BS (e.g., number of neighbor cells, BS served cell size, BS capabilities, etc.), and then sent to the UE (either through dedicated or broadcasted signaling).
- **Frequently-Changing BS-Calculated Parameters:** This group contains all the frequently changing parameters calculated by the BS (e.g., current active radio bearers, load of neighbor BSs, etc.).

The algorithm defining the interaction among the LEs consists of 3 main steps:

- **STEP-1:** UE selects the best suitable MCP according to the “UE-Calculated Parameters” and the “Infrequently-Changing BS-Calculated Parameters”. Subsequently, the UE shall adopt the RRM scheme indicated by the selected UE-MCP (e.g., adjust measurement intervals according to the selected profile).
- **STEP-2:** BS reselects (fine-tunes) the “active UE-MCP”. When UE establishes a connection with BS, it shall transmit to the BS the "UE-Calculated Parameters". Therefore, the BS may reselect the UE-MCP taking into consideration the “UE-Calculated Parameters”, “Infrequently-Changing BS-Calculated Parameters”, as well as the “Frequently-Changing BS-Calculated Parameters”. As a result of this reselection (fine-tuning) of the suitable UE-MCP, the BS may command the UE to adjust the current active UE-MCP.
- **STEP-3:** UE and BS both update each other with latest calculated parameters. Whenever the UE detects that the “UE-Calculated Parameters” are different from the values transmitted to the BS, it shall inform the BS with the updated parameter set. Similarly, the BS shall inform the UE when the BS detects that the values in the last calculated parameters set differ from the ones which have been provided by the BS to the UE. Consequently, the best suitable UE-MCP shall be reselected accordingly.
- The introduced framework should provide BS with flexibility to extend the defined MCPs by adding new MCPs. The BS shall send the new MCPs to the UE (either through dedicated or broadcasted signaling).

The introduced framework should provide BS with flexibility to extend the defined MCPs by adding new MCPs. The BS shall send the new MCPs to the UE (either through dedicated or broadcasted signaling).

6.3 Functions for Initial Access and Mobility

Initial access refers to a set of CP functions across multiple layers of the RAN protocol stack (e.g. PHY, MAC and RRC) and, at some extent, the CN / RAN interface as in the case of paging and state handling. In LTE, some of these functions are synchronization (time and frequency, UL/DL), Cell Search, System information distribution and acquisition, Random access and Paging [3GPP-36300]. This section presents particular enhancements and/or changes required in 5G. Specifically:

- Random Access procedures addressing diverse access latency requirements and for a wide frequency range.
- Paging optimizations for RRC Connected (Inactive) UEs.

Apart from the mechanisms presented in this chapter, the System information distribution and acquisition have been optimized in METIS-II. The particular procedures have been presented briefly in Section 6.1.4 and in details in [MII-D62].

The rest of the presented functions and procedures in this subsection is related to Mobility Management control plane functions and to the RRC State Management. The former incorporates a set of Mobility Management enhancements that relate to the introduction of certain innovative functions and the RRM State Management relates to how the RRC protocol is affected by the introduction of the new RRC Connected Inactive state.

6.3.1 Random Access Channel Solutions

As the number of the MTC devices increases the cellular network will require for innovative solutions [3GPP-36300] [3GPP-38300] to be able to handle this increase. To efficiently support MTC, it is required to design new schemes that will lead to the reduction of signaling messages both in DL and in UL communication and avoid potential communication bottlenecks for a 5G operator in channels such as the random access. In LTE, for accessing the network, the UE follows the contention-based random access procedure, which occurs in every Random Access Opportunity. However, such network designs are unlikely to be able to handle the MTC applications, where a large number of machines will attempt to transmit simultaneously small amounts of data.

Up to now several schemes have been proposed in the literature for handling the RACH procedure. These schemes may be classified into two large groups, namely pull-based and push-based [CKS+15]. These schemes however, are designed mainly for prioritizing access based on the transmission requirements and are not, on the one hand, targeting the solution of the collision rate problem, and on the other hand, are not focusing on 5G use cases. Even in the cases where the solutions are applied for MTC scenarios, the number of the considered devices is rather small, thus making their applicability in scenarios where big number of devices is considered.

For the random access of a vast amount of devices, a solution based on the grouping of the devices seems to be appropriate. Instead of having all the group members to proceed in random access using one of the 64 preambles when they have to transmit we could aggregate the transmission requests and only one device (the group head) will perform the RACH request. This will result in significant reduction in the collision rate in the RACH. According to the proposed solution the devices are being grouped by the network based on their mobility and their communication characteristics. The network then schedules the cluster heads' transmission opportunities based on their transmission requirements. As shown in Figure 6-14 the group based system access reduces the collision rate significantly. This is related to the reduction of the number of the devices that compete for the RACH resources (only the group heads) which reduces the collisions and the consequent delays.

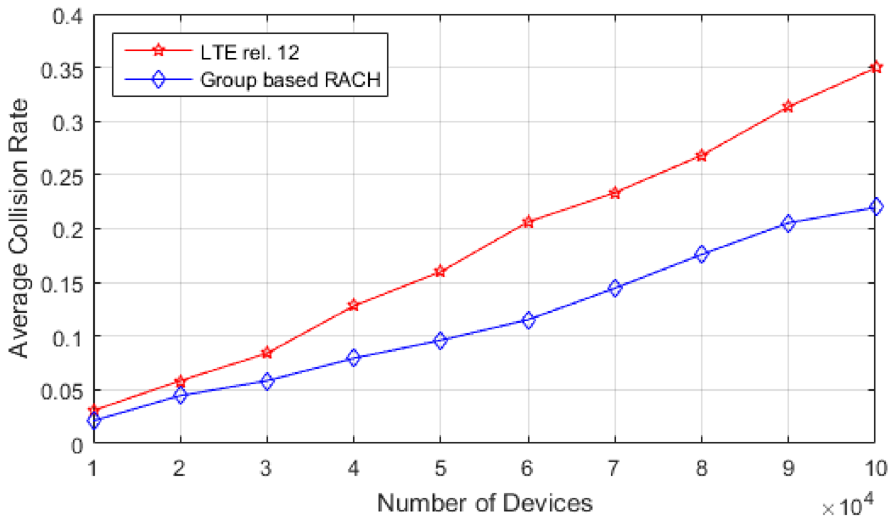


Figure 6-14: Number of collisions for the Group Based System Access compared with LTE-A [MII-D62]

For the devices with strict latency requirements instead of reserving a set of dedicated preambles for the use of devices with high priority random access requests associated with delay sensitive services could be configured to apply a combination of preamble signatures at a given random access time slot. The aforementioned approach would enable requests with stricter delay requirements to have higher priority, since combinations of preambles can always be identified by the receiver. This way, requests with higher priority are significantly less prone for collisions and the retransmissions.

6.3.2 RRC State Management

For the users in RRC Connected state, the mobility procedures of the active connections can be maintained for all mobility profiles, even when the user is moving at very high speeds. For users without RRC connection to the network, the users are said to be in RRC Idle state. In 5G, a new RRC Connected Inactive state has been proposed where the UE is always connected form 5G CN perspective, also during the low activity periods in RAN.

The main characteristic of RRC Connected state is the active RRC connection between the UE and the network and allocation of logical dedicated unicast resources for the transfer of CP signaling or UP data in UL or DL. The UE has Access Stratum (AS) context in RAN and RAN knows the cell where the UE is located. The RRC in RAN controls the mobility by performing handovers and cell changes and the UE location is known at the cell level.

The new proposed RRC Connected Inactive state was introduced in the beginning of METIS-II and documented in [MII-D61]. RRC Connected Inactive state was identified as the primary low activity state for the 5G access in [SMS+16]. This new state will maintain the UE AS context in UE and RAN allowing low system access latency from power saving state to ready to transmit/receive data. The RRC Connected Inactive supports a wide diversity of services with different requirements in terms of power consumption and access delays, thus the RRC procedures are proposed to be configurable. The mobility is UE controlled based on cell reselections within the RAN defined area and UE can be reached by paging from RAN. RRC Connected Inactive can allow multi-AIV camping where also the Evolved-LTE nodes are connected to 5G CN and therefore the LTE evolution is tightly integrated to the 5G RAN.

The RRC Idle state is the power saving state where the UE context is not stored in RAN. The UE will be paged from CN and the CN maintains the Tracking Area where to reach the UE. The RRC Idle state in 5G mobile systems is needed for initial registration procedure, initial Public Land Mobile Network (PLMN) selection and for fault recovery mechanisms. RRC Idle will be used also for core network based location tracking and paging.

The proposed RRC state model consists of three states: “RRC Idle”, “RRC Connected” and “RRC Connected Inactive”, according to Figure 6-15.

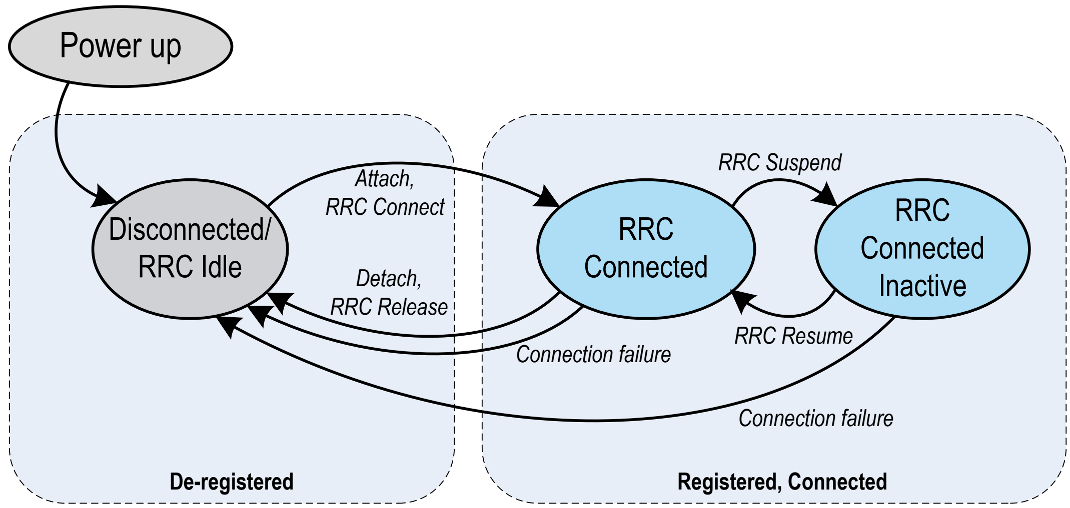


Figure 6-15: 5G UE RRC state transitions

Mobility during RRC Connected Inactive state may cause frequent path switching. Therefore, the UE context transfers due to cell reselections can be reduced by keeping the UE context and the data path(s) terminated in the Last Serving 5G BS where UE was in RRC Connected state. Now the Last Serving 5G BS takes the role of a mobility anchor, which allows keeping the C-plane and the U-plane RAN connections unmodified towards the 5G CN.

The Figure 6-16 illustrates the signalling flow of state transition from RRC Connected state to RRC Connected Inactive state and back to RRC Connected state. In this case, UE resumes connection to its Last Serving 5G BS, i.e. it does not move while being in the Connected Inactive state. When network commands the UE to Connected Inactive state, the Last Serving 5G BS sends an RRC Connection Suspend message to the UE. The message that contains (at least) Resume ID (in this case the Last 5G BS ID), Connected Inactive state related timing Information (e.g. Registration period), up-to-date TA List in which UE is allowed to move without TAU and Security Information for UE identification while re-connecting to the network.

Continuing with the example in Figure 6-16, connectivity is needed again when an application needs to send data. The UE is already connected to the network so it reconnects via the selected cell and sends RRC Connection Resume Request message to the 5G BS including (at least) UE ID, Resume ID, RRC Connected Inactive state related timing Information (e.g. time spent in inactive state), and Security Information to verify the UE context. The 5G BS responds to the UE with the RRC Connection Resume Complete message and UE is back to RRC Connected state.

Some use cases, such as low latency applications, might require switching the S1* connection to the optimal 5G BS location as soon as possible. For example, the S1* connection might need to be immediately switched when the UE moves to a new cell which is not located in the 5G BS currently terminating the S1* connection.

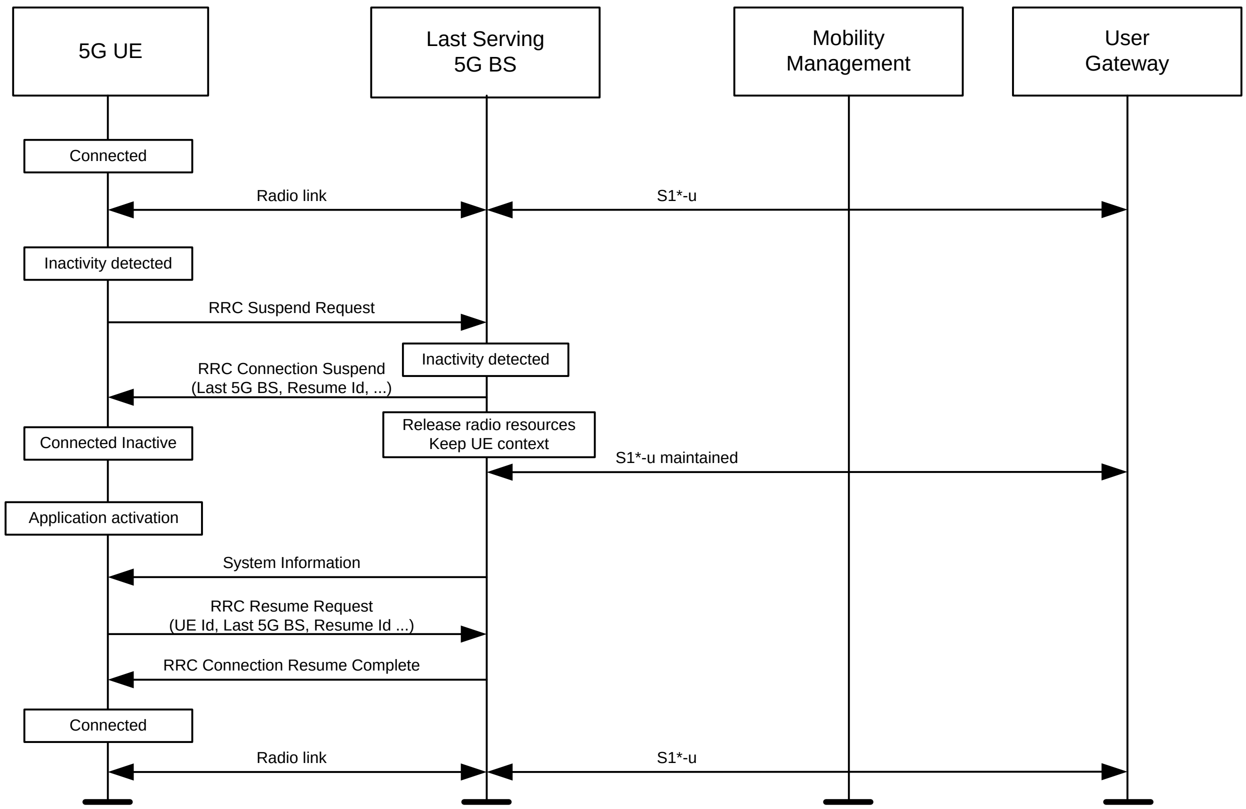


Figure 6-16: Signalling procedure of mobility during RRC Connected Inactive and RRC activation/inactivation

The characteristics of the 5G RRC state model are summarized in the Table 6-3.

Table 6-3: RRC states in 5G

5G State	Mobility procedure	Monitoring Dedicated Physical Channels	Allowed Mode for DL Channel Monitoring	UE Location Known on	UL Activity Allowed	Storage of RAN Context Information
RRC Idle	Cell selection & reselection	No	Discontinuous with DRX	Tracking Area list level	No	No
RRC Connected Inactive	Cell selection & reselection	Configurable, yes/no	Discontinuous with DRX	RAN Tracking Area level	Configurable, Contention based UL data	Yes
RRC Connected	Network controlled handover	Yes	Both continuous and discontinuous with DRX	Cell level	Yes	Yes

6.3.3 RAN-based Paging

The RRC Connected Inactive state assumes that the connection between RAN and CN is maintained during the low activity periods. Therefore, in RAN the anchor 5G BS can control the UE location tracking and paging and store the UE context. Anchor 5G BS also terminates the 5G CN connection for the UE. The RAN is partitioned into group of 5G BSs and cells which are called RAN Tracking Areas (RTA), where every cell broadcasts its RTA Identity (RTA ID). The anchor 5G BS provides the UE with the list of RTA IDs that the UE may move without updating its location. If the UE moves out of its list of RTA IDs, it sends a location update to the RAN which may trigger an anchor 5G BS relocation.

Figure 6-17 describes the procedure where the anchor 5G BS receives an MT (Mobile Terminated) data and triggers paging in RAN to reach the UE. The UE is paged through all the cells in its list of RTA IDs. In case the list of RTA IDs of the UE includes multiple 5G BSs, a horizontal Paging inter-5G BSs interface is necessary. This requires anchor 5G BS to maintain the inter-5G BS relationships with all 5G BSs of any RTA which it has given to the UEs. In addition, the anchor 5G BS needs to buffer and forward the UE MT data until the anchor 5G BS is relocated. Upon receiving the paging message, the UE responds to the paging and is ready to receive DL UP or CP data with existing RRC configuration.

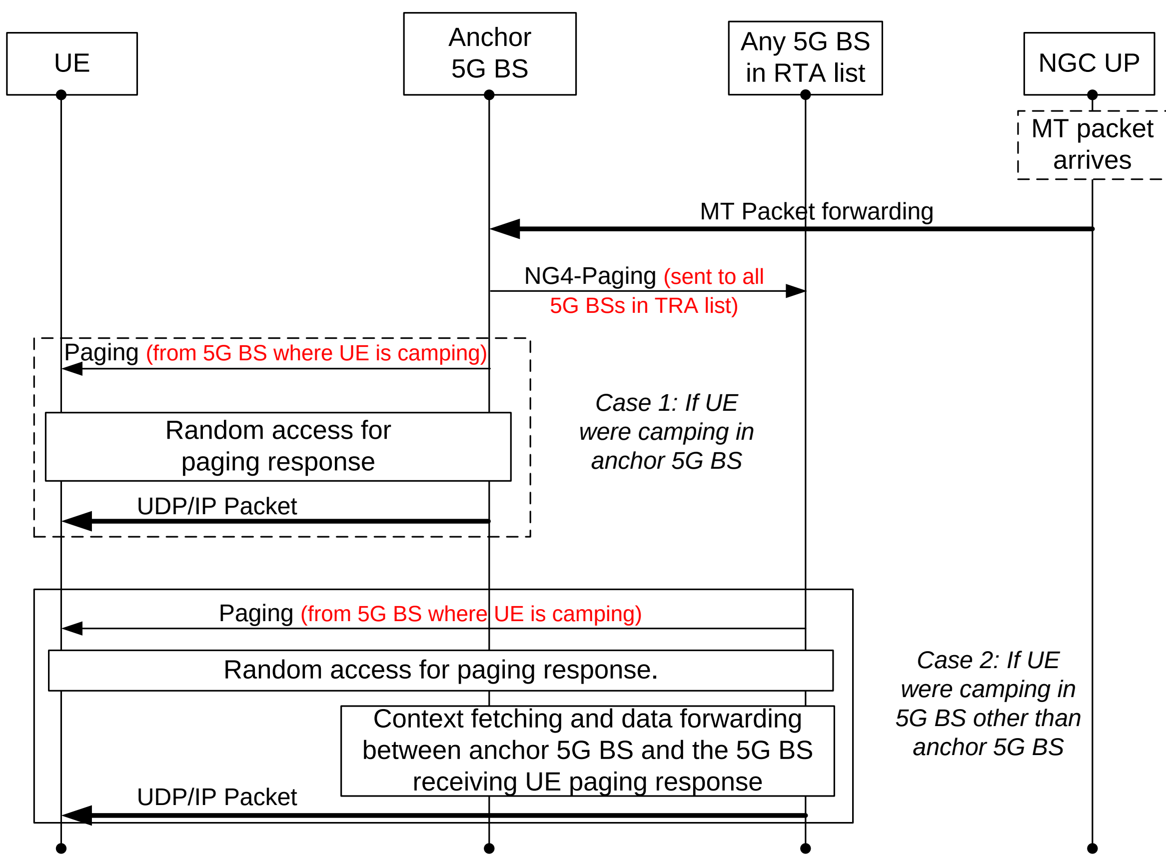


Figure 6-17: RAN initiated paging in the RRC Connected Inactive

For the performance comparison, both approaches of CN and RAN based paging can be evaluated by counting the number of messages for signalling overhead assuming 5G architecture in Figure 5-1, e.g. [3GPP-23799]. The procedure used for evaluating the CN based paging follows the paging principle of LTE and the paging is initiated from Mobility Management Control Function located in 5G CN.

Performance is analysed using a macro-cellular deployment scenario with hexagonal cells, where each 5G BS consists of three cells. The considered traffic model is characterized by Poisson distribution with average arrival rate of 1 packet per 60 seconds. UE mobility is assumed to follow a trajectory over a straight line. The considered UE speed values are {3, 30, 60, 90, 120} km/h. CN initiated paging are taken as a baseline for comparison.

Figure 6-18 shows the total number of paging and location update signalling messages per hour of a UE with FTP traffic with average packet arrival rate of 60 packets per hour. This illustrates the total paging and location tracking signalling messages of a UE assuming an inactivity timer of 10 s. The FTP traffic model leads to a relatively higher number of paging events such that the paging signalling is significantly dominant overhead over the signalling from location updates. Thus, the RAN initiated paging overall has significantly lower signalling overhead than the CN initiated paging due to its smaller paging area.

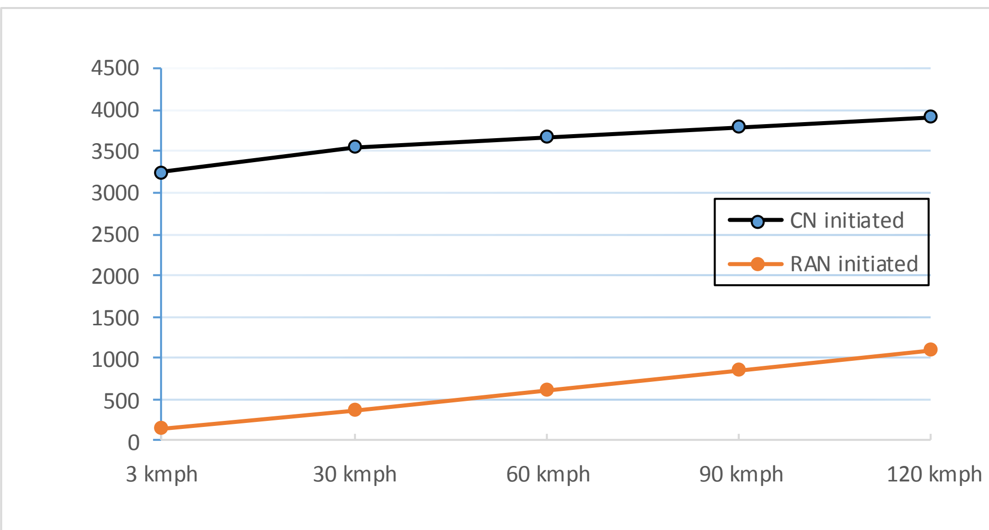


Figure 6-18: Total number of paging and location update signaling messages [/h] of a UE with FTP traffic.

6.3.4 Mobility Management

Mobility in the 5G framework needs to cover use cases with active users in RRC Connected state and low activity users in RRC Connected Inactive state and in RRC Idle state. In addition, 5G must support the tight interworking between LTE and 5G AIVs for mobile users (see Section 5.3.4). The 5G mobility framework consisting of several new methods including UE autonomous mobility, make-before-break handover and mobility concepts for URLLC.

Mobility and multi-connectivity in C-RAN

One possible realization of C-RAN comprises a multi-layer RAN where the RAN functions are functionally divided between a CU and DUs that are connected by a non-ideal x-haul interface, as shown in Figure 5-8. Such architecture provides various opportunities for optimizing mobility and MC, stemming from:

- No context fetch is needed when the UE moves between DUs.
- No re-location of the RAN-CN interface is needed when the UE moves between DUs.
- CU has a global control and visibility over multiple DUs for mobility and MC.

These enablers can be exploited by various INACTIVE and ACTIVE state procedures to reduce signaling overhead, decrease handover interruption time, provide faster activation of MC, and reduce the UE power consumption, as described in more detail in [MII-D62].

UE mobility

The Deliverable D6.2 [MII-D62] addresses impacts of 5G mobility, in particular RRM measurements and capability signaling from a UE perspective.

The extended frequency range above 6 GHz in 5G requires support of much larger UE channel bandwidth and an evolved Carrier Aggregation (CA). BF operation offers many options how the UE determines the best cell in particular taking antenna arrays and beams into account. Therefore, 5G numerologies require support of flexible slot types and structures and duplexing modes. But the design of RRM measurements for 5G must avoid unnecessarily complex or restricted operation of different functions and services. So one of the key issues on 5G mobility design from an UE mobility point of view is whether to use the common RRM measurement for both IDLE and CONNECTED mode. Options are to use either the same RS resources (5G Primary Synchronization Signal (5G-PSS) or 5G Secondary Synchronization Signal (5G-SSS), Multi-port multi-beam Reference Signal (MRS), 5G-SSS and DeModulation Reference Signal (DMRS) for Physical Broadcast Channel (PBCH)) or not the same RS resources in IDLE and CONNECTED mode.

We propose to evolve channel state information (CSI) with a dedicated designed CSI-RS. Then one optional set of RS are 5G-SSS in IDLE and MRS in CONNECTED; 5G-PSS and/or 5G-SSS in IDLE; 5G-PSS and/or 5G-SSS and Channel State Information Reference Signal (CSI-RS) in CONNECTED [3GPP-R11700334]. The proposed evolved CSI for UE related mobility is sufficient for the Reference Signal Received Power (RSRP) accuracy measured with SSS in some 5G scenarios. CSI-RS is only required in certain scenarios as common RRM measurement for both IDLE and CONNECTED mode.

Another aspect of UE mobility is the UE capability signalling. In LTE-5G tight interworking scenarios, both LTE and 5G systems must provide configuration information to the UE of their RRC and protocol stack. However, some UE capabilities can be shared between LTE and 5G systems, and these capabilities were part of the research. We propose an evolution of UE capability signaling with container splitting for LTE-5G tight-interworking scenarios [3GPP-R2168507]. The container split can also be used for cell change and simplified for no configuration changes. If there are no configurations that need to be verified by LTE eNB, 5G BS only sends one container that is to be forwarded to the UE and no specific action is needed by LTE eNB. It is also possible to send the UE configuration directly from 5G BS to UE over 5G radio. When LTE eNB needs to reconfigure the UE with parameters that need coordination, it provides a container with those parameters to 5G BS.

As part of the coordination container checks, LTE eNB and 5G BS ensure that they can comprehend all of the fields included in the container. If at least one of the fields cannot be comprehended, the container data shall get rejected for becoming part of configuration and it indicates there is a mismatch in the LTE and 5G capabilities for the parameters that need further coordination. One further solution to avoid that a LTE eNB must implement the 5G RRC is to use LTE RRC to provide an “equivalent” configuration as the 5G configuration enabling it to check for UE capability violation. It assumes that there is the capability to mapping a 5G configuration onto a LTE RRC configuration. Further study will be needed to evaluate such a solution.

6.4 Summary

Key functional design considerations developed by METIS-II are briefly provided in the previous sections, while further details and evaluations are captured by WP4 [MII-D42], WP5 [MII-D52], and WP6 [MII-D62]. Table 6-4 summarizes these functional design considerations for 5G, and highlights their key benefits, the differences to LTE-A, and the main implications on the overall 5G RAN design.

The benefits of the key design recommendations include higher coverage and capacity (i.e., beam-centric design, relaying, D2D communications, self-backhauling, and IM), increased energy efficiency (i.e., lean design, energy efficient RAN moderation, optimized UE context measurement), increased flexibility and reliability (i.e., multi-AIV interworking, AIV configuration, and DTS), as well as network slicing enablers (i.e., multi-slice RM). The functional considerations factor in both in fixed and dynamic topologies. Additionally, signaling overhead is reduced using efficient and optimized mobility management, a new RRC state model, and optimized initial access.

Major differences compared to LTE-A include service-oriented designs (e.g., service-oriented AIVs, optimized initial access for service prioritization, a service-tailored RRC state transition handling), network slicing in 5G, and legacy interworking at the RAN level. The novel network design, contrary to that of LTE-A, enables the system information distribution and the reference signals transmission only when needed. Furthermore, optimized agile RM techniques provide faster operation of conventionally slow functions like traffic steering to avoid hard handovers and, thus, to reduce latency. IM schemes are designed to cope with not only dense fixed topologies but also dynamic radio network topologies including non-static access nodes. Certain functionalities, such as, D2D and self-backhauling are natively integrated in the 5G system. Finally, the UE measurements, and mobility management will in 5G focus on the new needs with multiple AIVs available, and an extensive use of BF.

Table 6-4: Summary on key functional design considerations for 5G.

5G Functional Design Paradigm	Key benefits	Key difference to LTE-A evolution	Implication on overall RAN design
Beam-centric Design	Better coverage, capacity and data rates in higher bands	Narrow beams possibly swept instead of omni-directional cells	Major; all control signals beamformed; all mobility and initial access procedures need native beam-centric design
Lean and Future-proof Design	Energy efficiency and future-proofness, potentially also improved C-plane scalability	Reference signals not always on, not full band, not in all subframes	Significantly more configurable reference signals and mobility procedure
RAN moderation for energy efficient network operation	Reduction in overall network energy consumption via optimal active-mode operation	Exploitation of flexible self-backhauling and access node coordination to attain high energy efficiency leveraging on the QoS and channel quality awareness	Additional RRC signalling for coordinating the access node sleep-mode operation and channel measurement coordination for energy-aware controller at aggregation node, i.e., AN-O (CU)
Native Relaying, Self-backhauling and D2D support in 5G	Efficient support of 5G services that can benefit, e.g., from capacity, resource reuse, power consumption and coverage	Native integration since the beginning of 5G system design (e.g., in terms of CP functionalities, frame structures, etc.) rather than an add-on feature	CP and UP functionalities ranging from PHY to higher layers should consider native D2D and self-backhauling support

5G Functional Design Paradigm	Key benefits	Key difference to LTE-A evolution	Implication on overall RAN design
	gains offered by these technologies	on top of an already mature system like LTE	
Multi-AIV Resource Mapping on Fast Time Scale	AIV-agnostic RM; higher reliability; reduced latency; tight interworking with legacy increasing capacity; fast data routing	DTS and tight interworking on PDCP level avoiding hard inter-AIV and inter-cell handovers	New control information elements between AN-O/CU and AN-I/DU (new fronthaul interface), e.g., new signaling for AIV quality metric; fast addition and deletion of a new CP connection in DC to a UE along with lightweight signaling to support ultra-reliability
RM for Network Slicing	Share a common RAN for multiple businesses and services with diverging requirements	Network slicing is a new feature which is not part of LTE-A	New multi-slice RM concepts required to implement slice aware resource assignment; AaSE as new entity that performs multi-slice RM
RAN Enablers for IM	Higher cell-edge use throughput, larger capacity and better coverage	Advanced cooperative IM techniques targeted at dynamic topologies and dense deployments, for instance with flexible UL/DL TDD	RAN impact is mostly characterized by the need for signaling and procedures over the wired or wireless backhaul using X2* interface to support the exchange of information among cooperating BSs
Novel UE Measurement Context in 5G	Reduced overhead, enhanced energy efficiency	Functional extensions and changes in the UE measurement context	New information and configurations in the UE measurement context; option that a UE may maintain multiple measurement contexts
Novel RRC State Model	Reduced UE power dissemination, CP latency and CN/RAN signalling, esp. suitable for	UEs are always connected from a CN perspective; significantly larger possibilities for	Context fetching needs to be specified and supported. Novel mobility

5G Functional Design Paradigm	Key benefits	Key difference to LTE-A evolution	Implication on overall RAN design
	bursty connectivity and massive access	service-spec. configuration	procedures for new state to be defined
Service Prioritization at Initial Access	Service differentiation already at first access; lower latency for mission-critical services	Different levels of service prioritization for diverse sets of delay requirements without reserving resources for certain service classes	New MAC procedures required for RACH to enable service prioritization; signalling to higher layers
Mobility Management	Mobility with very low interruption delays and efficient BF mobility	Support for extreme low interruption handover and functions to handle massive BF	Major; BF mobility requires new set of measurements and signalling; new mobility procedures to handle handovers with low interruption delay
RAN-based Paging	Reduced CN/RAN signalling; reduced CP latency	In LTE paging is a CN function, which is now moved into the RAN	Entire re-design of paging functionality, signalling etc.; change of usage of CN/RAN interface

7 Evaluation Results

One of the METIS-II goals was to enable and perform system-level simulations that feed the process of design of the future 5G system. This task started with the establishment of a framework and methodology for system-level simulations. [MII-D21] provided simulation guidelines to align assumptions, methodology and simulation use cases in order to allow for a direct comparison of the different technology components. This was to address the need of guaranteeing valid simulation results for the evaluation of the METIS-II concept at the last phase of the project. In order to ensure consistency of results, a procedure for calibration, guidelines for simulation and a mechanism to support and control the validity for the simulations performed in the technical work within the project was set up.

Partners involved in the technical research have extensively used these guidelines in their performance evaluations, resulting in already-valid results ready for the benchmarking process. This allowed for the more accurate decision making process in the identification of promising techniques and the final system design. The level of use of the agreed simulation assumptions and guidelines defined in METIS-II is a great indicator of the huge collaboration between partners during this evaluation process.

Evaluation assumptions and results are summarized in [MII-D23]. Following the guidelines described in [MII-D21], an analysis was made to assess the impact of the different TeCs proposed in METIS-II on the 5G requirements established in the beginning of the project for the 5 selected UCs [MII-D11]. Results permitted drawing very interesting conclusions, as for instance that latency requirements can be only achieved with a big reduction of the TTI together with the definition of a lean architecture for the direct communication between peers. Furthermore, in D2.3 we demonstrated that it is not feasible to reach the objectives of reliability for safety conditions unless V2V communications are enhanced with the techniques proposed in METIS-II.

These simulation results, together with the self-evaluation performed by technical WPs 2–6, provided the basis for the final 5G design described in this document. The rest of this section studies to which extent the METIS-II 5G system concept is able to reach 5G KPIs based on the system simulation results in [MII-D23]. The main findings of this huge simulation effort have allowed us to identify and quantify, under certain assumptions, the potential impact of some fundamental technology enablers of the 5G mobile and wireless communication system.

7.1 5G KPI analysis

A set of KPIs was evaluated in [MII-D23], including inspection, analytical and simulation-based indicators.

Concerning the first group, that is, inspection KPIs, we can confirm that all the ones defined in the beginning of the project are fulfilled by the final METIS-II 5G system, including the following concepts:

- **Bandwidth and channel bandwidth scalability**. METIS-II system can operate with different bandwidth allocations [MII-D41] and in bands up to 100 GHz [MII-D32].
- **Coexistence with LTE**. The METIS-II 5G RAN has been designed for coexistence with LTE (cf. e.g., RRM schemes or RAN moderation solutions captured in [MII-D51] and [MII-D52]), and the same spectrum bands can be used by both technologies, which could share resources depending on the specific AI needs. This flexible allocation also contemplates re-farming of spectrum from LTE to the 5G technology.
- **Deployment in IMT bands**. METIS-II has addressed this KPI through work in [MII-D31].
- **Interworking with 3GPP legacy technologies and 802.11 WLAN**. METIS-II 5G RAN has been designed to support interworking with 3GPP legacy technologies, that is, GSM, UMTS and LTE family of standards (cf. Section 2.3.2 in [MII-D61]), and IEEE 802.11 family of WLANs (cf. Section 6.2 in [MII-D61]). This interworking guarantees the seamless connectivity in case of inter-system handover to any of those mentioned technologies.
- **Operations above 6 GHz**. METIS-II addresses this KPI through spectrum-related activities in [MII-D32] (e.g. analysis of coexistence with fixed service links operating on mmW, or feasibility studies for outdoor-to-indoor deployment at higher frequencies) as well as through appropriate UP and CP design [MII-D41] [MII-D52] [MII-D62].
- **Spectrum flexibility and sharing**. The ability to adapt to different DL/UL traffic patterns and capacity for paired and unpaired bands has been addressed by METIS-II through specific UP design concepts [MII-D41] [MII-D42] and system level solutions. METIS-II has also investigated mechanisms to facilitate sharing between 5G and other technologies in licensed or unlicensed spectrum [MII-D32].
- **Support of wide range of services**. This has been addressed by METIS-II through numerous technical solutions in all technical WPs.
- **Low cost requirements**. METIS-II 5G RAN is designed to support low cost devices, as well as low cost operation and maintenance enabled by e.g., mMTC solutions captured in [MII-D23], lean signaling and energy efficiency [MII-D52] [MII-D62], spectrum sharing [MII-D32] and self-organizing networks [MII-D51].

With respect to the analytical evaluation of KPIs, this also concluded the ability of the 5G RAN designed by METIS-II to fulfil the 5G system requirements. Evaluation results indicate that 5G RAN can deliver peak data rates in the order of 21 Gbps in DL and 12 Gbps in UL. Comparing to 4G operations, 5G RAN designed in METIS-II will also enable significant reduction of UP and CP latencies (measured as E2E one-way latencies), down to 0.763 ms and 7.125 ms, respectively. In UP, it is of paramount importance the reduction of the sub-frame length to 0.125 ms. On the other hand, CP latency reduction was enabled by new RRC Connected Inactive state (see Section

6.3.2). It has been also proved that for mMTC operations a single battery life time exceeding 10 years is possible for devices that sporadically upload data to network (see [MII-D23] for more details).

Finally, also the simulation-based evaluation work for the five METIS-II 5G UCs has pointed out that the METIS-II simulation KPI requirements for 5G [MII-D11] have been fulfilled using a subset of the TeCs proposed in the project. In UC1, for dense urban environment and HetNet deployment, users can expect data rates above 300 Mbps and operators can support traffic volumes greater than 750 Gbps/km^2. In this UC, significant energy efficiency gains have been demonstrated as well. In UC2, high frequency bands and massive antenna systems enable Gbps data rates indoors, reaching up 7.85 Gbps (that is above the 5 Gbps target). In UC3, traffic volumes of 700 Mbps and 650 Mbps are supported in DL and UL for the required user data rates of 50 Mbps and 25 Mbps, considering an LTE system at 800 MHz with BF capabilities. With 3.5 GHz, UC3 required data rates can be supported with 10 times higher load. In addition, energy efficiency analysis shows that the system with BF consumes half the energy of the system without BF, when sleeping capabilities are considered. In UC4, it is shown that, depending on the traffic profile, 5G will cater for more than 1 million devices per km^2. For devices transmitting once every 100 s, proposed access scheme support more than 6.9 million devices per km^2. In UC5, although results captured in [MII-D23] were not enough to reach the requirements, latest incorporation of adaptive transmission schemes made METIS-II 5G system proposal also reach the defined target. In this sense, reliability provided in urban scenarios for 5 ms end-to-end latency is close to 99.999% for the required range of 50 m with 40 MHz of available spectrum. However, in the highway scenario, the required coverage range of 1000 m can only be achieved with an allocation of 100 MHz to the V2V communication link.

The next tables summarize the outcome of the METIS-II 5G system concept evaluation, for both analysis and simulation KPIs.

Table 7-1: Analysis KPI evaluation.

KPI	Requirement	METIS-II performance	Key contributor
CP latency	< 10 ms	7.125 ms	RRC Connected Inactive, reduction of processing time in BS and UE
UP latency	< 1ms	0.763 ms	Shortening of TTI, reduction of processing time in BS and UE
mMTC energy efficiency	> 10 years on a single 5 Wh battery	> 10 years on a single 5 Wh battery	Extension of DRX, CP latency reduction, deep sleep energy conservation features

KPI	Requirement	METIS-II performance	Key contributor
Peak data rates	> 20/10 Gbps for DL/UL	21.7/12.4 Gbps for DL/UL	MIMO spatial multiplexing (for lower frequencies), exploitation of mmW bands
Mobility interruption time	0 ms	0 ms	MC + make-before-brake

Table 7-2: Simulation KPI evaluation.

KPI	Requirement	METIS-II performance	Comment
User throughput (UC1, UC2 and UC3)	UC1: 300 Mbps UC2: up to 5 Gbps UC3: 50/25 Mbps for DL/UL	UC1: 1 Gbps+ UC2: up to 7.85 Gbps UC3: 50/25 Mbps for DL/UL	Only DL values for UC1 and UC2 Different methodology applied for UC3 evaluation
mMTC device density (UC4)	> 1 mln/km^2	4 mln/km^2	Depends heavily on the traffic/report periodicity of mMTC devices. 1 upload of 1000 bits every 100 s was used in METIS-II
Reliability (UC5)	99.999% at 50/1000m for urban/highway	99.999% with 40/100 MHz for urban/highway	Evaluation of V2V solutions with dynamic resource allocation techniques. Required channel bandwidth is identified
Network energy efficiency (UC1, UC3)	Should follow (at least) capacity improvement	For the capacity x1000, network energy efficiency improvements of 350-7500 times were reported	Evaluation done only for Dense Urban environment. Savings depend on the load level in LTE-A/5G network

7.2 Service related KPIs

METIS-II has dealt with 5G challenges by developing a set of TeCs. In [MII-D23] selected TeCs were compared to legacy solutions to provide service-related numerical evaluation results that address 5G KPIs as defined in [MII-D11].

TeCs have been grouped according to the 5G generic services, depending on the service that is most related to the concept. This could be xMBB, mMTC and uMTC. In addition, a fourth group of TeCs that enable handling more than one service has been also studied.

From the isolated analysis of TeCs, some key concepts could be highlighted. The tight integration of 5G with LTE-A has proved to be useful in initial deployment phases. The new roles of infrastructure and user devices such as NNs, mobile relays, cluster heads etc., have demonstrated their ability to increase system performance (throughput, energy efficiency, etc.). The dynamic cell switching off is a powerful tool to increase energy efficiency when traffic load is not high. Random access optimizations (based on grouping of accesses, preamble multiplexing, etc.) allow to increase the number of supported mMTC devices and to differentiate services appropriately. The AI flexibility, with regard to e.g. granularity of resources in frequency (bandwidths) or time (subframe durations), improves network and user performance in terms of e.g. data rates or latency, when handling different services at the same time. Traffic steering and network slicing enable tailored QoS support of different services. Harmonization of AIs is needed to facilitate an optimal RRM across different AIs. New waveforms that provide improved spectrum confinement, flexibility and better coverage (operating at lower SINR values for a given BLER) enable active interference design for additional ICI reduction. Finally, the RRC Connected Inactive state provides CP latency reduction and mMTC energy consumption improvements.

8 Key RAN Design Questions Addressed

Table 8-1. Status of the METIS-II work on answering key 5G RAN design questions.

No	Key RAN Design Aspect / Question
1	**What is the general spectrum usage foreseen for 5G?** 5G networks have to integrate numerous of frequency bands within a wide range of spectrum and with differing spectrum authorizations, and to cope with the versatile spectrum requirements from different user groups. Frequency bands for 5G and a concept for spectrum management and sharing are briefly introduced in Sections 4.1.1 and 4.1.3. More details can be found in [MII-D31] and [MII-D32].
2	**Given the various characteristics of different spectrum bands, which band should be used for what type of service, air interface and how much spectrum needs to be made available for mobile communications in the different bands?** A brief summary on which band should be used for what type of service is given in Section 4.1.1, with more details in [MII-D31]. Initial considerations on spectrum needs in different bands are given in [MII-R31]. Two exemplary results of spectrum demand analyses are outlined in Section 4.1.2, and described in detail in [MII-D32]. The question of which bands should be used for which service is also covered in [MII-D41], by determining which overall set(s) of AIVs, e.g. operating in different spectrum bands and / or tailored towards certain services, would be most suitable to address the overall 5G requirements space.
3	**Which air interface variants are expected to be introduced in the context of 5G, and which are to be evolved from existing standards?** An AIV for below 6 GHz is expected to be an evolution of current 4G standards, at least from a UP design point of view. Filtering may be applied for reduced in-band interference. The numerology parameters (subcarrier spacing and symbol duration and slot duration) can be adapted to the use case, e.g. shorter symbols may be chosen for low latency traffic. The case of D2D communications may require special waveforms to counteract the effects of asynchronicity. Above 6 GHz new AIVs with special frame structures may be required to, e.g., manage massive MIMO and channel estimation. Vehicular communications, especially for road safety, may require new AIVs to efficiently deal with multicasting, asynchronicity and reliability.
4	**How many different novel and legacy air interface variants should different devices support? Which forms of concurrent connectivity (e.g. multi-standard and multi-cell connectivity, concurrent device-to-device and device-to-infrastructure connectivity) will be required in 5G?** This will depend on the purpose of each device. The harmonized 5G AI should allow that purpose-specific devices implement only necessary functionalities. For instance, a laptop

No	Key RAN Design Aspect / Question
	thought to be in a static or quasi-static indoor environment should implement AIVs for below and above 6 GHz with massive MIMO support and high-order modulations. On the other hand, an in-car communication unit for V2V communications should implement an AIV with high reliability and multicasting features. Hence, this in-car unit may not require high-order modulations and massive MIMO support for safety-related applications. Further, concurrent connectivity in the form of MC (e.g., multi-AIV connectivity) or concurrent D2D and device-infrastructure connectivity is beneficial for agile RM as enablers, as highlighted in Section 6.2.1 and Section 6.2.2. For example, DTS exploits MC to address diverse service requirements and to increase reliability while grouped D2D communications can improve network coverage. Furthermore, possible implications of concurrent connectivity on the device complexity have been taken into account for the investigation on potential UE context extensions, as described in Section 6.2.4. Further details in this direction are captured in [MII-D52].
5	**How tightly are novel air interface variants expected to be integrated with each other and with legacy technologies (e.g. LTE evolution and WLAN), to which extent should they be harmonized or have common functionality in the protocol stack, and on which level should different transmission forms be aggregated?** It is concluded that the integration among legacy AIV (LTE-A evolution) and novel AIVs, or the integration among multiple novel AIVs, should be possible on RAN level [MII-D52] [MII-D62]. Furthermore, PHY harmonization of novel AIVs in the form of multi-waveform implementation may be a feasible option if necessary, as it decreases the implementation complexity and required chip space versus a single waveform approach. In addition, the coexistence of different numerologies and frame structures may be required to better support different service characteristics [MII-D42]. The harmonization level in the protocol stack must be carefully selected to allow sufficient backward compatibility. For LTE-A evolution and novel AIVs PDCP layer aggregation is seen as feasible, see Section 5.3, where a protocol stack implementation of the common CP is depicted. Among novel AIVs, a large extent of protocol stack function harmonization should be strived for (i.e. at least a harmonized MAC and higher layers) [MII-D41]. The interworking with other access technologies, such as WLAN can be performed, e.g., via radio maps to determine transmit power levels [MII-D52]. Further, the use of unlicensed band in an LAA manner is analysed in dynamic radio topologies [MII-D52]. These considerations are further detailed in D4.2, D5.2, and D6.2.
6	**How can one efficiently handle interference in an ultra dense environment? What kind of information is required, at what time scale and how fast the system must react?** Various mechanisms constituting the Agile RM Framework of METIS-II are targeting IM that respond to this question as summarized in Section 6.2.3 and detailed in [MII-D52]. It is emphasized here that the way of handling interference depends on the operational scenario and use case. To this end, the same-entity interference in dynamic TDD operation shall be tackled, where JT with dummy symbols is found to provide a good trade-off between UL and DL performances. Also, to overcome the pilot contamination in dynamic TDD with massive MIMO, dynamic selection of transmission paths shall be supported.

No	Key RAN Design Aspect / Question
	These schemes are applied on a subframe basis, e.g., few ms. The IM schemes shall be adaptive to cope with the dynamic radio network topologies based on non-static access nodes. The time scale of modifying the interference mitigation scheme depends on the changes of the topology, which can range from minutes to hours depending on the availability of the non-static access nodes in a target service region and day time. Furthermore, interference resistive design can be exploited to mitigate inter-cell interference, where coordination is needed on the X2* interface. The concept of lean design for common signals reduces the amount of interference, which is an important enabler for the 5G system to handle ultra-dense environments, see Section 0.
7	**What will be considered as "resource" in a 5G system? How can we manage these resources effectively in order to achieve the 5G KPIs?** As captured in [MII-D52], it is envisioned that, in 5G, the notion of a resource is extended beyond conventional RRM to attain the optimum mapping of 5G services to any available resources when and where needed within this extended realm of resources. In addition to the licensed radio frequency bands, the extended realm of resources includes the unlicensed bands, whose usage shall be adaptive and be coupled with the changing radio network topology, energy, as well as HW and SW resources. With respect to how this extended notion of resource will be managed efficiently, various considerations have been presented in [MII-D52], including mechanisms pertaining to IM, RAN moderation, DTS and multi-slice RM.
8	**On which time scale should certain 5G radio access network functionality (e.g. radio RM, radio resource control, mobility) operate, and consequently, how should the necessary functionalities be best abstracted, grouped and tackled in standardization and implementation?** Various 5G functionalities are envisioned to be handled on a faster time scale than in legacy systems. For instance, as summarized in Section 6.2.1, METIS-II RAN design enables mobility and MC among LTE-A evolution and novel 5G AIVs on RAN level, inherently allowing for a faster setup of new MC constellations and switching among these. Further, the proposed DTS among different AIVs, which was so far done via hard handover, is performed on lower protocol stack layers and consequently on a much faster time scale. The envisioned agile RM framework groups RM mechanisms under intra-AIV and AIV-overarching RM functionality framework, see Section 6.2 and [MII-D52]. As described in Section 5.3.4, the mechanisms pertaining to AIV-overarching RM are envisioned to be implemented in a CU (AN-O) while intra-AIV RM schemes are envisioned to be implemented in DUs (AN-I).
9	**How will the concepts from dynamic spectrum management interwork with the control plane architecture (new network elements and interfaces for this purpose and/or some level of integration to the control plane design)?** The METIS-II architecture concept embraces the regulator domain covered by a "Spectrum Management System" (SMS), and the operator domain which consists of a central Spectrum Assignment Coordination (SAC) entity supported by a number of further

No	Key RAN Design Aspect / Question
	functional blocks (see Section 5.5.2). The SAC is going to be integrated into the 5G Network MANO framework as briefly outlined in Section 5.5.3. More details can be found in [MII-D32].
10	**What will be the network elements and interfaces in the 5G system architecture and, assuming these, how would these interfaces look like, i.e. which functionalities will they have, which programmability level will be adopted, what level of openness, what level of abstraction, etc.?** In 5G the RAN NFs may be distributed across different network elements in a centralized or a distributed way (CU vs. DU) according to the service demand to be supported. In each unit the NFs can be split into a CP part (CPFs) and a UP part (UPFs). On the interface between RAN and CN, the UP may not be transported over a single protocol as in 4G but each service or slice may use the protocol (e.g. GTP, GRE, EoGRE, ETH) best suited for the service. Regarding intra-RAN interfaces, it is assumed that an evolved X2* interface between access nodes is required. It is expected that this interface will also be crucial for agile IM in 5G, as listed in Section 6.2.3 and described in further detail in [MII-D52]. Furthermore, a hierarchical CP design is envisioned, where AIV-agnostic control schemes are implemented at a CU, while AIV-specific control schemes are located at the DUs. Accordingly, a new x-haul interface between the CU and DUs is crucial to attain the promising gains of the developed mechanisms. New signaling schemes are then required for multi-AIV resource mapping, as summarized in Section 6.2.1. The AIV-specific radio link feedbacks depend on the AIV (e.g., carrier frequency) and UE context (e.g., speed) [MII-D52].
11	**What type of control and user plane functionalities should be centralized or distributed depending on the 5G use cases associated to them? Out of these functionalities, what are the most promising candidates to be implemented as virtual network functions?** The degree of centralization and the associated benefits also constitute an important aspect for the RM functionalities within the agile RM framework, which considers both centralized and distributed control functions. As highlighted in Section 5.3.4 and Section 6.2, the CPFs regarding AIV-overarching RM are envisioned to be implemented in a CU. These centralized functions enable efficient mapping of the service flows to the appropriate AIVs. Also, multi-slice RM requires AaSE functionality to be centralized so that SLAs can be fulfilled despite changing network conditions. The corresponding UPFs should be centralized accordingly (e.g., centralized PDCP processing is required in case of the aforementioned AIV-overarching RM). AIV-specific CPFs, such as, dynamic scheduling and IM, can be implemented in a distributed way. Nevertheless, RM mechanisms can also be implemented in a centralized way provided that the physical deployment allows such a centralization, e.g., C-RAN deployments. Besides the above-mentioned strategies for function placement, a flexible network architecture is of importance. With the help of different options for centralization (a lower /

No	Key RAN Design Aspect / Question
	higher degree of centralization, as described in Section 5.3.2), the network can be adapted to fully exploit the underlying transport network.

9 Summary and Outlook

In this deliverable, we have presented the 5G RAN design worked out by the METIS-II project.

This starts with the 5G AI that may be composed of multiple AIVs – one of them can be evolved legacy LTE plus others that are tailored to support the different services and frequency bands. Furthermore, we have described the 5G architecture on which the project has reached a high level of consensus: the view on the different horizontal and vertical splits and the interface between CN and RAN, the ways to integrate different AIVs including LTE-A, the mapping of different network functions to a physical architecture or the architectural enablers for network slicing.

Moreover, we have given design recommendations to achieve better coverage and higher capacity, increased network energy efficiency, increased flexibility and reliability of the network.

We have evaluated the proposed technologies and given (or referred to) the results along with the design in the technical chapters, thus motivating the design decisions taken. The evaluation was done using the common simulation guidelines established within the project. In addition, we have added a dedicated Chapter on the overall KPI analysis of the system which confirmed that all inspection KPIs that were defined in the beginning of the project can be fulfilled by the final METIS-II system.

Standardization of 5G NR in 3GPP System Architecture (SA) and RAN WGs has started during the project period of METIS-II, and we have seen a number of the METIS-II concepts and results being discussed in 3GPP. To ease this process, we have therefore taken care to keep this document easily readable thus maximizing the use of project's results in that standardization process.

Since 3GPP works in a phased approach, the adaptation of the concepts described in this document may span a longer period corresponding to several 3GPP releases. To keep the adaptation path open, much attention will have to be paid to keeping the evolving standard forward compatible.

References

[3GPP-22261] 3GPP TS 22.261, "Service requirements for next generation new services and markets; Stage 1 (Release 15)", V15.0.0, March 2017.

[3GPP-22862] 3GPP TR 22.862, "Feasibility study on new services and markets technology enablers for critical communications; Stage 1 (Release 14)", V14.1.0, October 2016.

[3GPP-23501] 3GPP TS 23.501, "System Architecture for the 5G System; Stage 2 (Release 15)", V0.4.0, April 2017.

[3GPP-23799] 3GPP TR 23.799, "Study on Architecture for Next Generation System (Release 14)", V14.0.0, December 2016.

[3GPP-28801] 3GPP TS 28.801, "Study on management and orchestration of network slicing for next generation network (Release 14)", V1.1.0, March 2017.

[3GPP-29281] 3GPP TS 29.281, "General Packet Radio System (GPRS) Tunnelling Protocol User Plane (GTPv1-U) (Release 13)", V13.3.0, June 2016.

[3GPP-32501] 3GPP TS 32.501, "Telecommunication management; Self-configuration of network elements; Concepts and requirements (Release 14)", V14.0.0, April 2017.

[3GPP-32842] 3GPP TR 32.842, "Telecommunication management; Study on network management of virtualized networks (Release 13)", V13.1.0, December 2015.

[3GPP-38300] 3GPP TS 38.300, "NG Radio Access Network; Overall Description; Stage 2 (Release 15)", V0.2.0, May 2017.

[3GPP-38322] 3GPP TS 38.322, "NR; Radio Link Control (RLC) protocol specification (Release 15)", V0.2.0, June 2017.

[3GPP-38801] 3GPP TR 38.801, "Study on New Radio Access Technology; Radio Access Architecture and Interfaces (Release 14)", V14.0.0, March 2017.

[3GPP-38804] 3GPP TR 38.804, "Study on New Radio Access Technology; Radio Interface Protocol Aspects (Release 14)", V14.0.0, March 2017.

[3GPP-38912] 3GPP TR 38.912, "Study on New Radio (NR) Access Technology (Release 14)", V14.0.0, March 2017.

[3GPP-R2168858]. 3GPP Tdoc R2-168858, "Text Proposal to TR 38.804 on on-demand SI provisioning for NR", NTT DOCOMO, November 2016.

[3GPP S1172416], 3GPP Tdoc S1-172416 "Feasibility Study on LAN Support in 5G".

[3GPP-SP170169] 3GPP Tdoc SP-170169 "5G Communication for Automation in Vertical Domains", March 2017.

[3GPP-R11700031] 3GPP Tdoc R1-1700031, "Coexistence of NR DL and LTE", Huawei, HiSilicon, January 2017.

[3GPP-R11700334]. 3GPP Tdoc R1-1700334, "Considerations for NR mobility", Intel, January 2017.

[3GPP-R11700841] 3GPP Tdoc R1-1700841, "NR LTE Coexistence", Qualcomm, January 2017.

[3GPP-R2168507] 3GPPP Tdoc R2-168507, "LTE and NR interaction for configuration coordination", Intel, November 2016.

[5GARCH16-WP] 5G PPP, Architecture White Paper, May 2016.

[5GC] 5G PPP project 5G-Crosshaul, http://5g-crosshaul.eu/.

[5GNORMA] 5G PPP project 5GNORMA, https://5gnorma.5g-ppp.eu/.

[5GN-D32] ICT671584 5G NORMA Deliverable D3.2, "5G NORMA network architecture - Intermediate report", January 2017.

[5GNOW] FP7 project 5GNOW, http://www.5gnow.eu/.

[5GPPP16] 5G PPP Architecture Working Group: White Paper "View on 5G Architecture", Version 1.0, July 2016.

[5GX] 5G PPP project 5G-XHaul, http://www.5g-xhaul-project.eu/.

[ABB+17] P. Arnold, N. Bayer, J. Belschner, G. Zimmermann, "5G Radio Access Network Architecture based on Flexible Functional Control / User Plane Splits", EuCNC, June 2017.

[CKS+15] Konstantinos Chatzikokolakis, Alexandros Kaloxylos, Panagiotis Spapis, et al., On the Way to Massive Access in 5G: Challenges and Solutions for Massive Machine Communications, CrownCom 2015.

[CPRI15] "Common Public Radio Interface (CPRI); Interface Specification", V7.0, October 2015.

[DDL15] B. Debaillie, C. Desset and F. Louagie, "A Flexible and Future-Proof Power Model for Cellular Base Stations," IEEE 81st Vehicular Technology Conference (VTC Spring), Glasgow, 2015, pp. 1-7.

[DDM13] Uwe Dötsch, Mark Doll, Hans-Peter Mayer, Frank Schaich, Jonathan Segel, and Philippe Sehier, "Quantitative Analysis of Split Base Station Processing and Determination of Advantageous Architectures for LTE", Bell Labs Technical Journal 18(1), pp. 105-128.

[ERI11] M. Ericson, "Total Network Base Station Energy Cost vs. Deployment", ," IEEE Vehicular Technology Conference (VTC-Spring), Budapest, Hungary, 2011.

[ETSI14-NFV-MAN] ETSI GS NFV-MAN 001 V1.1.1, „Network Functions Virtualisation (NFV); Management and Orchestration", December 2014.

[ETSI14-ORI] ETSI GS ORI 002-1/2, "Open Radio equipment Interface (ORI); ORI interface Specification; Part 1: Low Layers / Part 2: Control and Management", V4.1.1, October 2014.

[ETSI15-103235] ETSI TS 103 235 V1.1.1, "Reconfigurable Radio Systems (RRS); System architecture and high level procedures for operation of Licensed Shared Access (LSA) in the 2 300 MHz - 2 400 MHz band", October 2015.

[ETSI-NFV] European Telecommunications Standards Institute (ETSI): Industry Specification Group "Network Function Virtualization" (ISG NFV), http://www.etsi.org/technologies-clusters/technologies/nf.

[EU16-COM588] COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS, COM(2016) 588, "5G for Europe: An Action Plan", September 2016.

[FANTASTIC-5G] 5G PPP project FANTASTIC-5G, http://fantastic5g.eu/

[GZ16] J. Gebert, D. Zeller, „Fat pipes for user plane tunnelling in 5G", IEEE CSCN 2015, Berlin, October 2014.

[IETF-RFC2784] IETF RFC 2784, Generic Routing Encapsulation (GRE).

[IETF-RFC2890] IETF RFC 2890, Key and Sequence Number Extensions to GRE.

[IRT] https://www.irt.de/en/activities/radio-systems/mobile-broadcast/

[ITU-R14] ITU-R Report M.2320-0, "Future technology trends of terrestrial IMT systems," Nov. 2014.

[ITU15] International Telecommunication Union – Radio (ITU-R) Working Party WP 5D: Draft New Recommendation "IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond", Doc. R12-SG05-C-0199 (approved by Study Group SG5), June 2015.

[KOI] Koi Project, http://www.koi-projekt.de/.

[METIS-II] METIS-II web page, https://metis-ii.5g-ppp.eu/.

[MII-R31] ICT-671680 METIS-II, Report R3.1 , "Preliminary spectrum scenarios and justification for WRC Agenda Item for 5G bands above 6 GHz", October 2015.

[MII-D11] ICT-671680 METIS-II, Deliverable D1.1, "Refined scenarios and requirements, consolidated use cases, and qualitative techno-economic feasibility assessment", February 2016.

[MII-D12] ICT-671680 METIS-II, Deliverable D1.2, "Quantitative techno-economic feasibility assessment", June 2017.

[MII-D21] ICT-671680 METIS-II, Deliverable D2.1, "Performance evaluation framework", February 2016.

[MII-D22] ICT-671680 METIS-II, Deliverable D2.2, "Draft Overall 5G RAN Design", June 2016.

[MII-D31] ICT-671680 METIS-II, Deliverable D3.1, "5G spectrum scenarios, requirements and technical aspects for bands above 6 GHz", May 2016.

[MII-D41] ICT-671680 METIS-II, Deliverable D4.1, "Draft air interface harmonization and user plane design", May 2016.

[MII-WP] ICT-671680 METIS-II, White Paper, "Preliminary Views and Initial Considerations on 5G RAN Architecture and Functional Design", March 2016.

[MII-D23] ICT-671680 METIS-II, Deliverable D2.3, “Performance evaluation results”, March 2017.

[MII-D32] ICT-671680 METIS-II, Deliverable D3.2, “Enablers to secure sufficient access to adequate spectrum for 5G”, June 2017.

[MII-D42] ICT-671680 METIS-II, Deliverable D4.2, “Final air interface harmonization and user plane design”, April 2017.

[MII-D52] ICT-671680 METIS-II, Deliverable D5.2, “Final Considerations on Synchronous Control Functions and Agile Resource Management for 5G”, March 2017.

[MII-D61] ICT-671680 METIS-II, Deliverable D6.1, “Draft Asynchronous Control Functions and Overall Control Plane Design”, June 2016.

[MII-D62] ICT-671680 METIS-II, Deliverable D6.2, “5G Asynchronous Control Functions and Overall Control Plane Design”, April 2017.

[METIS] FP7 project METIS, https://www.metis2020.com/.

[MET-D54] ICT-317669 METIS, Deliverable D5.4 Version 1, “Future spectrum system concept”, April 2015.

[mmMagic] 5G PPP project mmMagic, https://5g-mmmagic.eu/.

[NGMN15] NGMN, “5G White Paper”, March 2015.

[PUM17] A. Prasad, M.A. Uusitalo, A. Maeder, "Energy Efficient Coordinated Self-Backhauling for Ultra-Dense 5G Networks," 85th IEEE Vehicular Technology Conference, Sydney, June 2017.

[RBB+16] P. Rost, A. Banchs, I. Berberana, et al., “Mobile Network Architecture Evolution toward 5G”, IEEE Communications Magazine, May 2016.

[RSPG16-032] RADIO SPECTRUM POLICY GROUP Opinion on spectrum related aspects for next-generation wireless systems (5G), “STRATEGIC ROADMAP TOWARDS 5G FOR EUROPE”, November 2016.

[SMS+16] I. Da Silva, G. Mildh, M. Säily, S. Hailu, “A Novel State Model for 5G Radio Access Networks”, IEEE ICC Workshop, 2016.

[TGV+14] R. Trivisonno, R. Guerzoni, I. Vaishnavi, D. Soldani: “SDN-based 5G mobile networks: architecture, functions, procedures and backward compatibility”, Transactions on Emerging Telecommunications Technologies (ETT), November 2014.

[YHZ+16] F. Yang, M. Huang, S. Zhang, S. Sirotkin, “Radio Access Network Re-architecture to meet 5G Requirements – A SDN-based Paradigm Shift for Cellular Networks”, ICC 2016, 5G RAN Design WS, May 2016.

A Functional Split Options within the RAN

The following figure gives an overview about different functional split options within the UP of the radio protocol stack (horizontal split) denoted here as M1 – M8 [MII-D22]. More details with respect to the impact of horizontal splits especially on x-haul bandwidth and latency for different AIV parametrizations can be found in METIS-II Deliverable D4.2 [MII-D42].

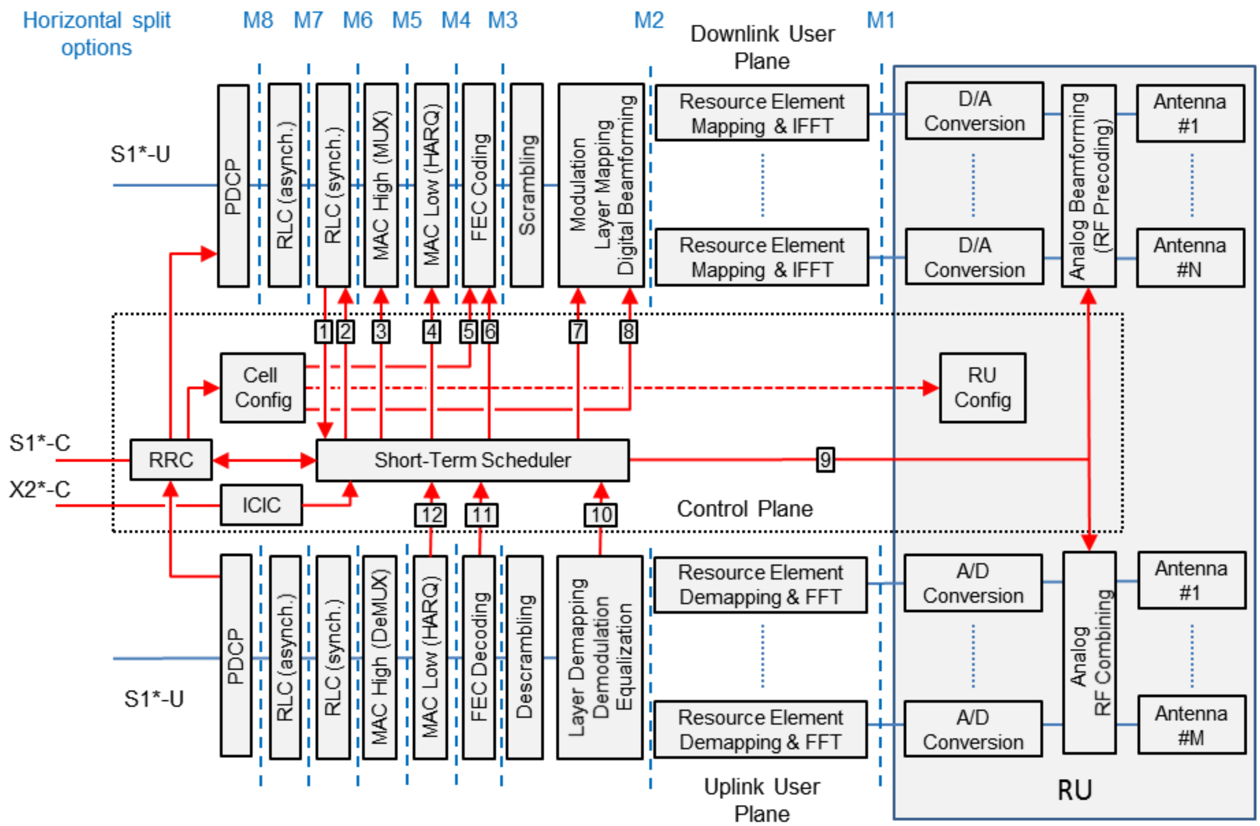


Figure A-1: Control and user plane decomposition and interactions in the radio access network (network infrastructure part only; single radio protocol stack)

In addition, the figure also shows the separation between CPFs and UPFs (vertical split) and the corresponding interfaces in between (see the red arrows marked by (1) – (12)). There meaning is as follows:

(1) DL buffer status

(2) Payload selection

(3) DL resource assignment and generation of UL transmission grants

(4) Retransmission control

(5) Cell related broadcast information settings (cell ID etc.)

(6) FEC coding scheme

(7) Antenna mapping, precoding, modulation scheme

(8) Reference symbol settings

(9) Antenna weights in case of analog BF (e.g. for Massive MIMO)

(10) CSI from UL sounding

(11) CSI from reporting, UL scheduling request

(12) HARQ status.

Figure A-1 visualizes the tight coupling of CPFs and UPFs in the RAN, which also shows the huge effort in terms of standardization that is required to achieve a fully flexible separation.

B Examples for Spectrum Management within the MANO Framework

B.1 Example for the implementation of the SAC into a SON architecture

The functional architecture, i.e., the implementation of the SAC into the 3GPP SON concept according to the Self-Configuration Reference Model [3GPP-32501], is illustrated in Figure B-1, showing the three levels of the management model, namely NM, DM/EM and Network Element (NE).

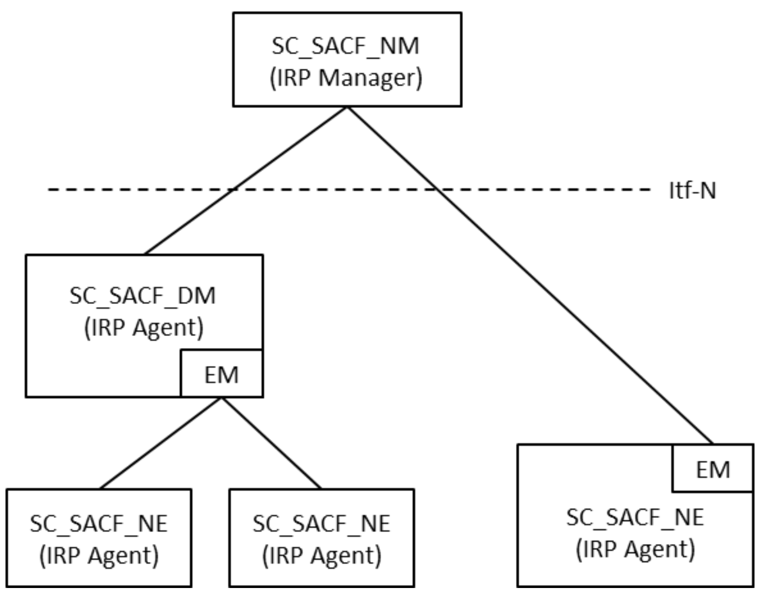


Figure B-1: Implementation of the SAC into the 3GPP SON concept according to the Self-Configuration Reference Model.

The SC_SACF_NM functional block represents the NM portion of SC_SACF (i.e., policy, control, and monitor functions), as well as the related Integration Reference Point (IRP) Manager functionality. It takes the spectrum assignment decision for a dedicated use case, service or application, after processing of relevant information from other functional entities. This assignment decision is communicated to the SC_DM/(EM) and the SC_SACF_NE functional blocks. These blocks are representing the portion of the SC_SACF operating below the Itf-N interface, as well as related IRP Agent functionality. The SC_SACF_NE functional block (where the SON algorithms are located) is responsible for the respective configurations of cell parameters according to the decision made by the SC_SACF.

B.2 Example for the implementation of the SAC into virtualized networks

Figure B-2 illustrates the 3GPP management architecture which manages both, virtualized and non-virtualized network functions, and also clarifies the relationship between 3GPP management framework and NFV-MANO framework [ETSI14-NFV-MAN].

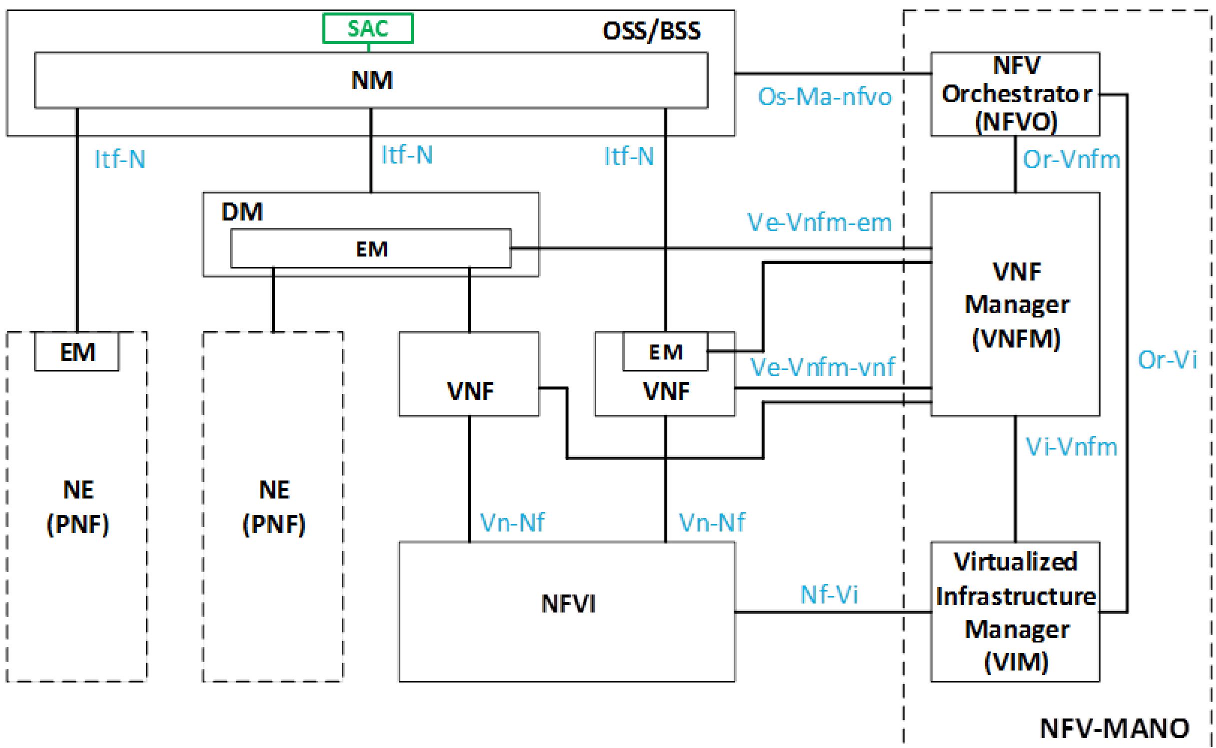


Figure B-2: Mixed network management mapping relationship between 3GPP and NFV-MANO architectural framework [3GPP-32842], SAC added as green box.

The 3GPP CM has the system modification functions and system monitoring functions in order to support the operations of NE. The NFV CMCMCM includes the configuration of VNF application specific parameters and the configuration of VNF deployment specific parameters. If VNF deployment specific parameters or VNF application specific parameters are changed, this parameter change should be notified to the NM by using CM capabilities.

The constraints on cell parameters received by the NM from the SAC may be further processed similar to the use case “NFV configuration management” described in [3GPP-32842]:

1. NM sends the configuration request to NFVO over the interface “Os-Ma-nfvo”. NFVO interprets it into a specific VNF lifecycle management request and sends the request to VNFM over the interface “Or-Vnfm”.
2. VNFM receives this request and implements the corresponding operation to the VNF and completes configuration of VNF deployment specific parameters over the interface “Ve-Vnfm-vnf”.
3. After the VNF deployment specific parameters are configured by VNFM, VNF is created, terminated or updated and the resource of the VNF is changed.
4. After the completion of the configuration, VNFM returns the success response to EM with necessary VNF change notification over the interface “Ve-Vnfm-vnf”.
5. EM performs post-operation activities (e.g. adjusting neighbour nodes of the affected VNF, configuring the VNF with application specific parameters).
6. EM notifies NM of all needed VNF resource changes through CM capabilities over the interface “Itf-N”.

From the functional areas defined in [3GPP-32842] and [ETSI14-NFV-MAN], the “Virtualized Resource Management” (VRM) and the “Policy Administration” are considered as most relevant for spectrum management. For example, the NM may need to trigger certain VRM functions over the interface “Os-Ma-nfvo”, allowing the agile resource management of 5G RAN to apply dedicated resources (i.e. spectrum, infrastructure, processing power, etc.) for flexible spectrum usage [MII-D52]. Furthermore, VRM data may need to be correlated with VNF application specific data over the interface “Itf-N”. Concerning policy administration, the NM may need to configure VRM policies over the interface “Os-Ma-nfvo” which may be forwarded over the interfaces “Or-Vnfm” and “Or-Vi”.

B.3 Co-existence and interworking between the NR (New Radio) and legacy AIVs

In 3GPP, activities on co-existence and interworking between NR and legacy AIVs are currently part of a study item work [3GPP-38913]. In an intra-operator domain, it is under discussion that NR should be able to support flexible allocation of resources (e.g. time, frequency) between NR and the legacy AIVs (e.g. LTE) operating in the same block of spectrum (with possible bandwidth overlap). Resource allocation granularity in the time/frequency domain, as well as the potential guards between NR and LTE resources are still to be determined. NR should be able to use these resources at least for DL, UL and eventually SL. This resource allocation should work regardless whether legacy AIVs are supported by the same BS as NR, or the two AIVs are supported by different BSs. On such basis, a flexible allocation of resources may also enable a smooth introduction of NR in the same band used by a legacy AIV, i.e. the band allocated to a legacy AIV can be progressively reduced (by steps of 5 MHz) in order to make spectrum available for the allocation to NR. Focusing on LTE as a legacy AIV, the coexistence of NR and LTE can be categorized into two main categories [3GPP-R11700031]:

- FDM (Frequency-Division Multiplexing): in this case, NR and LTE have no bandwidth overlap and to fulfill adjacent channel coexistence requirements. Guard bands between NR and LTE are needed. Additionally, bandwidth adaptation or cell (de)activation mechanisms can be used to balance the traffic loads of NR and LTE.
- TDM (Time-Division Multiplexing): NR and LTE have bandwidth overlap. Because of such tight co-channel coexistence, special mechanisms for IM based on dynamic sub-frame allocation are needed for both NR and LTE.

In addition to the NR-LTE coexistence in both DL and UL carriers, NR and LTE can only share the same UL carriers but need to have separate DL carriers. The approaches described above can be further categorized in static, semi-static and dynamic [3GPP-R11700841]:

- *Static FDM*: The spectrum partitioning between NR and LTE (e.g. 5/15, 10/10, or 15/5 MHz partitioning assuming a 20 MHz bandwidth) can be adjusted based on UE penetration.
- *Semi-static FDM with CA*: LTE has a static bandwidth allocation as anchor for LTE and NR (e.g. 5 MHz), and the remaining bandwidth is allocated to LTE SCell (Secondary Cell) and/or NR as needed. Another possibility would be that LTE and NR has static PCell (Primary Cell) bandwidth allocation (e.g. 5 MHz). The remaining bandwidth can be allocated to LTE SCells and/or NR SCells as needed.
- *Semi-static TDM*: In this case one would utilize LTE DL MBSFN sub-frames and resources with unused UL sub-frames to schedule NR.
- *Dynamic Resource Sharing*: NR utilizes unused LTE resources dynamically at PRB level in frequency and sub-frame level in time.

C Service-specific Network Functions

Table C-1: Service-specific flavors of network functions, input partially from the 5G Architecture WG White Paper [5GARCH16-WP].

Type of network function		Possible service-specific flavor
General connectivity	Connectivity model	E.g., bearer-based (for high throughput services), or connection-less (for internet of things, IoT).
	Multi-Connectivity	MC at different network layers (micro/macro), technologies (WLAN/LTE), spectrum (sub-6 GHz/mmW), user plane layers (MAC/RLC/PDCP) depending on service, deployment and AIV (see, e.g., Section 6.2.1 and [MII-D52] [MII-D62]).
Spectrum Access		Service-dependent operation in licensed, unlicensed, or license-assisted spectrum, or time-frequency multiplexed in common spectrum (see, e.g., the extended notion of resources and specific considerations in [MII-D52]).
Advanced SON schemes		Support of the dynamic densification through agile RAN schemes, e.g., Nomadic Nodes (see, e.g., interference management based on dynamic radio topology in [MII-D51]).
RRC related	Mobility	No (metering), local (enterprises), in groups (trains), very high speed (cars/trains/aircraft), on demand/forward (tracking sensors) or always/backward (pedestrian broadband) handover.
	Cell discovery	Sub-6 GHz MIMO (broadcast), massive MIMO mmW (sub-6 GHz assisted), small cells in ultra-dense networks (via macro coverage layer) cell discovery.
PDCP		Potential service-specific omitting of header compression and ciphering even for user plane traffic.

Type of network function		**Possible service-specific flavor**
RLC		Potential service specific unacknowledged mode only (e.g. sensor) or acknowledged mode only (e.g. mission-critical services), or transparent mode.
MAC / PHY	Carrier Aggregation	CA may not be needed in each scenario as it also impacts battery consumption; it could further include very distinct spectrum.
	Multi-Cell Cooperation	Service, load, deployment and channel-dependent tight cooperation (symbol-synchronized operation, RNTIs/scrambling/CSI-RS/scheduling/precoding coordination up to joint Tx/Rx CoMP) or loose cooperation (ICIC) (for specific considerations see, e.g., Section 6.2 and [MII-D52]).
	Scheduling	Service specific scheduling schemes, as for instance semi-persistent scheduling on sidelinks using geo-location information to improve V2X communication performance.
	RACH	Service specific RACH schemes where priorities can be introduced (please note a specific proposal for RACH service prioritization, which is described further in [MII-D61]). Also, grant free schemes can be considered for services to minimize the establishment of signaling channels or the transmission of emergency data.
	HARQ	Optimized for spectral efficiency (massive broadband) coverage (sensor, IoT), reliability (mission critical services) or latency (tactile Internet).
	Coding	For 3GPP NR, flexible LDPC coding is currently considered for all block sizes of eMBB data. Polar coding will be used for signaling information (except for very small block lengths where repetition/block coding may be preferred) [3GPP-38300].